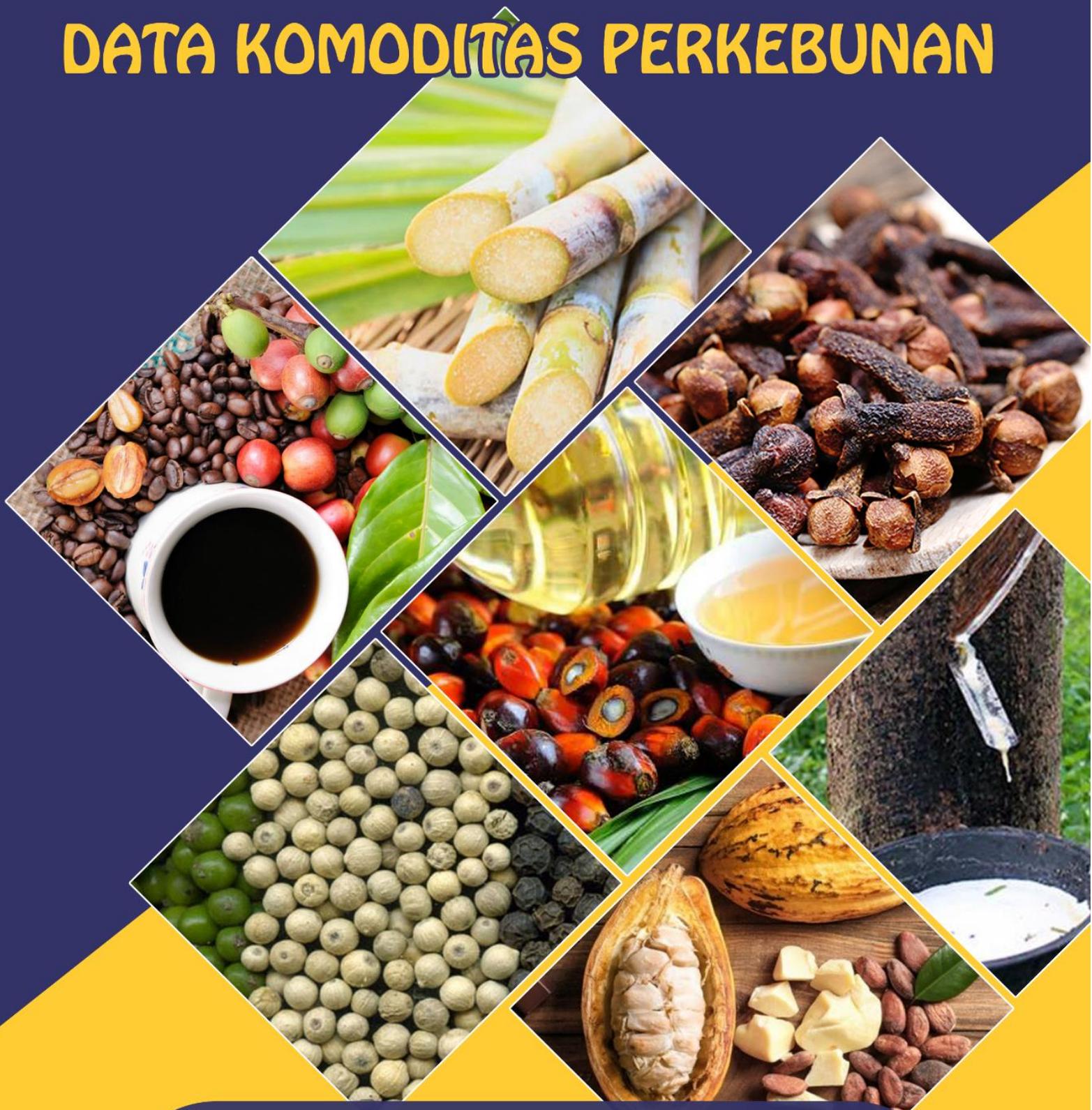


# KAJIAN METODE ESTIMASI DATA KOMODITAS PERKEBUNAN



**Pusat Data dan Sistem Informasi Pertanian  
Sekretariat Jenderal - Kementerian Pertanian  
Tahun 2020**



# **KAJIAN METODE ESTIMASI DATA KOMODITAS PERKEBUNAN**

**Pusat Data dan Sistem Informasi Pertanian  
Sekretariat Jenderal Kementerian Pertanian**

**2020**



# **KAJIAN METODE ESTIMASI DATA KOMODITAS PERKEBUNAN**

## **Pengarah :**

Dr. Akhmad Musyafak, SP, MP.

## **Penanggung Jawab :**

Dr. Ir. Anna Astrid Susanti, M.Si.

Rhendy Kencana Putra W, S.Si., M.AppStat.

## **Ketua Tim :**

Ir. Mohammad Chafid, M.Si.

## **Anggota Tim :**

1. Ir. Efi Respati, M.Si. - Pusdatin
2. Ir. Roch Widaningsih, M.Si. - Pusdatin
3. Diah Indarti, SE. MM - Pusdatin
4. Yuliawati Rohmah, SP., M.SE. - Pusdatin
5. Roydatul Zikria, SSi, M.SE - Pusdatin
6. Ir. Vera Junita S - Pusdatin
7. Lasmiyati, SST, SAB M.Stat – BPS
8. Diah Mekita Sari S.Tr.Stat. – BPS

## **Desain Sampul :**

Suyati, S.Kom

## **Diterbitkan oleh :**

Kementerian Pertanian  
Pusat Data dan Sistem Informasi Pertanian  
2020



## KATA PENGANTAR

Puji dan syukur kami panjatkan kehadiran Allah SWT atas berkat rahmat dan hidayah-Nya sehingga Buku Kajian Metode Estimasi Data Komoditas Perkebunan dapat diselesaikan. Buku Kajian Metode Estimasi Data Komoditas Perkebunan disusun dalam rangka melakukan pengembangan metode estimasi data perkebunan, sehingga diharapkan akan membantu penyusunan Angka Estimasi Data Perkebunan. Buku ini berisi kumpulan hasil analisis estimasi komoditas utama perkebunan dengan menggunakan model-model statistik.

Buku Kajian Metode Estimasi Data Komoditas Perkebunan ini merupakan hasil kerjasama Pusat Data dan Sistem Informasi Pertanian – Kementerian Pertanian (Pusdatin), Direktorat Jenderal Perkebunan – Kementerian Pertanian (Ditjenbun), dan Direktorat Statistik Tanaman Pangan Hortikultura dan Perkebunan - Badan Pusat Statistik (BPS) dengan narasumber utama adalah akademisi di Jurusan Statistika – Institut Pertanian Bogor. Buku ini mencakup hasil kajian estimasi produksi dan luas areal beberapa komoditas perkebunan strategis dengan pendekatan Model Arima, Model Regresi, Model Fungsi Transfer, dan Model VAR (*Vector Auto Regression*).

Keberhasilan dalam menyusun angka estimasi data perkebunan yang lebih akurat sangat ditentukan oleh kesungguhan dan kesadaran akan pentingnya akurasi data estimasi yang dihasilkan. Dalam hal ini, pemahaman akan metode analisis statistika yang tepat dan sesuai dengan data yang tersedia adalah salah satu komponen utama. Sehingga, dengan terbitnya buku ini, diharapkan dapat menjadi acuan dan referensi dalam menyusun angka estimasi data perkebunan yang lebih baik.

Kami mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penyusunan Buku Kajian Metode Estimasi Data Komoditas Perkebunan ini. Kami menyadari bahwa buku ini masih perlu disempurnakan lagi, sehingga saran dan masukan untuk perbaikan buku ini akan kami terima.

**Jakarta, Desember 2020**

**Kepala Pusat Data dan  
Sistem Informasi Pertanian**



**Dr. Akhmad Musyafak, SP, MP.**  
**NIP. 197304051999031001**



# DAFTAR ISI

Halaman

<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>Kajian Model Estimasi Produksi Karet Nasional Pendekatan Model Regresi ARIMA, Fungsi Transfer dan VAR (Vector Auto Regressive)</b>	
Mohammad Chafid .....	1-48
<b>Kajian Model Peramalan Produksi Sawit (<i>Crude Palm Oil/CPO</i>) di Indonesia Pendekatan Model ARIMA, Regresi, Fungsi Transfer dan VAR</b>	
Lasmiyati.....	49-64
<b>Kajian Metode Estimasi Produksi Gula Indonesia</b>	
Efi Respati.....	65-83
<b>Kajian Pengembangan Metode Estimasi Produksi Kakao</b>	
Yuliawati Rohmah .....	85-100
<b>Kajian Model Peramalan Produksi Kopi di Indonesia Pendekatan Model ARIMA, Regresi, Fungsi Transfer dan VAR</b>	
Roch Widaningsih.....	101-123
<b>Kajian Metode Estimasi Produksi Teh Indonesia</b>	
Diah Indarti .....	125-147
<b>Estimasi Produksi Lada di Indonesia</b>	
Roydatul Zikria .....	149-173
<b>Pengembangan Metodologi Estimasi Data Produksi Cengkeh 2020</b>	
Vera Junita Siagian .....	175-200
<b>Kajian Model Estimasi Luas Areal Karet Nasional Pendekatan Model Regresi ARIMA, Fungsi Transfer dan VAR (Vector Auto Regressive)</b>	
Mohammad Chafid .....	201-247
<b>Kajian Model Peramalan Luas Areal Kelapa Sawit (<i>Crude Palm Oil/CPO</i>) Di Indonesia Pendekatan Model ARIMA, Regresi, Fungsi Transfer dan VAR</b>	
Lasmiyati.....	249-264
<b>Kajian Metode Estimasi Luas Panen Tebu Indonesia</b>	
Efi Respati.....	265-283

<b>Kajian Pengembangan Metode Estimasi Luas Areal Kakao</b>	
Yuliawati Rohmah .....	285-299
<b>Kajian Model Peramalan Luas Areal Kopi di Indonesia</b>	
<b>Pendekatan Model ARIMA, Regresi, Fungsi Transfer dan VAR</b>	
Diah Mekita Sari .....	301-320
<b>Kajian Metode Estimasi Luas Areal Teh Indonesia</b>	
Diah Indarti .....	321-341
<b>Estimasi Luas Areal Lada di Indonesia</b>	
Roydatul Zikria .....	343-367
<b>Pengembangan Metodologi Estimasi Data Produksi Cengkeh 2020</b>	
Vera Junita Siagian .....	369-394

# **KAJIAN MODEL ESTIMASI PRODUKSI KARET NASIONAL PENDEKATAN MODEL REGRESI, ARIMA, FUNGSI TRANSFER DAN VAR (*Vector Auto Regressive*)**

**Mohammad Chafid**

*Center for Agricultural Data and Information System-Ministry of Agriculture  
Jalan Harsono RM No. 3 Gd D Lt 4, Ragunan-Jakarta Selatan, Indonesia  
E-mail: mohammad.hafidz1@gmail.com*

## **ABSTRAK**

Komoditas karet merupakan salah satu komoditas strategis penghasil devisa negara karena sebagian besar produksi karet nasional untuk diekspor. Status Angka statistik perkebunan terdiri dari Angka Sementara dan Angka Estimasi. Tujuan penulisan ini adalah mencari model alternatif lain untuk menyusun angka estimasi produksi karet sehingga akurasi menjadi lebih baik yang ditandai dengan semakin kecilnya MAPE baik untuk data training maupun testing.

Model yang digunakan untuk menyusun angka estimasi produksi karet meliputi, Metode Arima, Metode Regresi linier dan kuadratis dengan variabel bebas tahun, Metode Fungsi Transfer dengan peubah input harga karet dunia, Metode VAR (*Vector Autoregressive*) dengan variabel produksi, luas areal, harga karet dalam negeri, harga karet dunia, volume ekspor karet dan volume impor karet. Sumber data yang digunakan untuk variabel produksi karet (1967 – 2019), luas areal karet (1970 – 2019), harga karet nasional (1971-2019), volume ekspor dan impor karet ( 1970 – 2019) berasal dari Ditjen. Perkebunan. Untuk variabel harga karet dunia (1971 - 2019) berasal dari World Bank. Run model menggunakan software RStudio.

Untuk analisis ini data dibagi menjadi 2 kelompok, yaitu data training tahun 1967 – 2013, dan data testing tahun 2014 – 2019. Data training untuk penyusunan model, sedangkan data testing untuk uji coba model dalam melakukan estimasi 6 tahun kedepan. Untuk estimasi produksi karet alternatif model pertama adalah Model ARIMA. Model Arima terbaik adalah ARIMA (0,2,1), menghasilkan MAPE untuk data training 4,31%, dan MAPE data testing 6,34%. Model regresi linier menghasilkan MAPE data training sebesar 15,49% dan MAPE data testing 16,88%. Model regresi kuadratik menghasilkan MAPE data training sebesar 6,61%, MAPE data testing 5,05%. Model Fungsi Transfer dengan variabel input harga karet dunia, menghasilkan MAPE untuk data training sebesar 3,80% dan MAPE data testing 5,03%. Untuk model VAR(2) dengan pengaruh trend dan konstanta, menghasilkan MAPE data training 3,90% dan data MAPE data testing 4,92%.

Berdasarkan perbandingan besarnya MAPE baik data testing maupun data training dan hasil estimasi produksi 5 tahun kedepan, maka model terbaik yang terpilih adalah model VAR(2) karena menghasilkan akurasi yang cukup tinggi sehingga bias rata-rata sebesar 4,92%. Hasil estimasi produksi karet nasional dalam bentuk karet kering untuk model VAR (2) untuk tahun 2020 sebesar 3.479.390 Ton, tahun 2021 sebesar 3.589.243 Ton, tahun 2022 sebesar 3.617.694 Ton, tahun 2023 sebesar 3.713.331 Ton, dan tahun 2024 sebesar 3.808.592 Ton. Laju pertumbuhan estimasi produksi karet nasional selama 5 tahun kedepan rata-rata 2,29% per tahun.

**Kata Kunci : *Produksi Karet, Regresi, Arima, Fungsi Transfer, VAR (Vector Autoregressive)***

## PENDAHULUAN

Karet (*Hevea brasiliensis*) termasuk dalam genus *Hevea* dari familia *Euphorbiaceae*, yang merupakan pohon kayu tropis yang berasal dari hutan Amazon. Di dunia, setidaknya 2.500 spesies tanaman diakui dapat memproduksi lateks, tetapi *Hevea brasiliensis* saat ini merupakan satu-satunya sumber komersial produksi karet alam. Karet alam mewakili hampir separuh dari total produksi karet dunia karena sifat unik mekanik, seperti ketahanan sobek, dibandingkan dengan karet sintetis.

Data statistik perkebunan yang disajikan merupakan hasil sinkronisasi dan validasi yang dilaksanakan secara berjenjang dari tingkat kabupaten sampai tingkat nasional dengan mengacu pada Pedoman Pengelolaan Data Komoditas Perkebunan yang diterbitkan oleh Direktorat Jenderal Perkebunan. Secara periodik, data perkebunan disajikan dalam 3 status angka yaitu Angka Tetap (ATAP), Angka Sementara (ASEM) dan Angka Estimasi (AESTI). Angka Tetap merupakan angka hasil rekapitulasi dari pelaporan yang sudah tetap, sehingga tidak dilakukan estimasi, sedangkan untuk penentuan Angka Sementara dan Estimasi perlu dilakukan estimasi dengan metode estimasi yang paling relevan dan tepat (PDKP, 2013). Data Angka Tetap (ATAP) merupakan data 2 tahun yang lalu ( $n-2$ ), Angka Sementara (ASEM) merupakan data tahun lalu ( $n-1$ ), dan Angka Estimasi (AESTI) merupakan data tahun yang berjalan ( $n$ ).

Metode estimasi yang digunakan adalah Metode *Exponential Smoothing* (Peramalan Pemulusan Eksponensial) yang merupakan salah satu kategori metode time series yang menggunakan pembobotan data masa lalu secara eksponensial. Dalam kategori ini terdapat dua metode yang umum dipakai yaitu metode Pemulusan Eksponensial Tunggal (*Single Exponential Smoothing*) dan metode Pemulusan Eksponensial Ganda (*Double Exponential Smoothing*). Pemilihan model *Single Exponential Smoothing* atau *Double Exponential Smoothing* harus mempertimbangkan nilai *Mean Absolute Percentage Error (MAPE)* serta kerealistisan hasil estimasi bila dibandingkan dengan series data sebelumnya. MAPE adalah pengukur tingkat akurasi (ketepatan) nilai dugaan yang dihasilkan oleh model dalam bentuk presentase. Model yang mempunyai nilai MAPE lebih kecil dianggap sebagai model yang lebih baik. Keunggulan dari metode estimasi ini adalah dapat digunakan untuk meramalkan data yang berisi trend atau pola musiman. Namun metode estimasi ini juga memerlukan keahlian khusus dalam menginterpretasikan hasil estimasi yang diperoleh (PDKP, 2013).

Metode untuk menghasilkan angka estimasi (AESTI) yang diliris oleh Ditjen. Perkebunan pada waktu tahun berjalan ( $n$ ) perlu dikaji kembali, agar didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik dari metode sebelumnya, sehingga tingkat kesalahannya lebih kecil. Guna merumuskan kebijakan produksi, pasokan dan distribusi komoditas karet sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Metode ramalan produksi karet dapat digunakan dengan menggunakan beberapa metode statistik yang biasa digunakan untuk melakukan peramalan.

Pada analisis ini akan dikaji metode ARIMA, Regresi, Fungsi Transfer dan VAR dalam melakukan pemodelan dan peramalan produksi karet nasional.

Oleh karenanya, tujuan dari disusunnya analisis ini adalah:

- a. Melakukan analisis dan peramalan data produksi karet nasional menggunakan model ARIMA, Regresi, Fungsi Transfer dan VAR
- b. Menentukan metode terbaik dalam meramal data produksi karet nasional

## METODOLOGI

Data yang digunakan dalam melakukan pemodelan produksi karet nasional adalah data series produksi karet nasional tahun 1967 - 2019. Untuk analisis ini data dibagi menjadi 2 kelompok, yaitu data training tahun 1967 – 2013, dan data testing tahun 2014 – 2019. Data training untuk penyusunan model, sedangkan data testing untuk uji coba model dalam melakukan estimasi 6 tahun kedepan.

Peubah yang diasumsikan mempengaruhi besaran produksi karet untuk Fungsi Transfer adalah harga karet dunia. Harga karet dunia diasumsikan akan berpengaruh terhadap produksi, karena jika terjadi kenaikan harga karet dunia maka produksi karet cenderung akan terus ditingkatkan, kondisi sebaliknya maka produksi karet cenderung stagnan, pertumbuhan produksi rendah, atau cenderung turun.

Peubah yang berpengaruh terhadap produksi untuk pemodelan VAR adalah luas areal, harga karet domestik, harga karet dunia, volume ekspor dan impor karet. Harga karet dunia dan harga karet domestik diduga kuat berpengaruh pada tingkat produksi. Untuk volume ekspor dan impor juga berpengaruh terhadap produksi, dimana semakin tinggi permintaan ekspor karet maka produksi cenderung akan semakin tinggi.

Peramalan data produksi karet, dilakukan uji coba beberapa metode yakni mengkaji metode ARIMA, Regresi, Fungsi Transfer dan VAR (Vector Autoregression), menggunakan software *R Studio*.

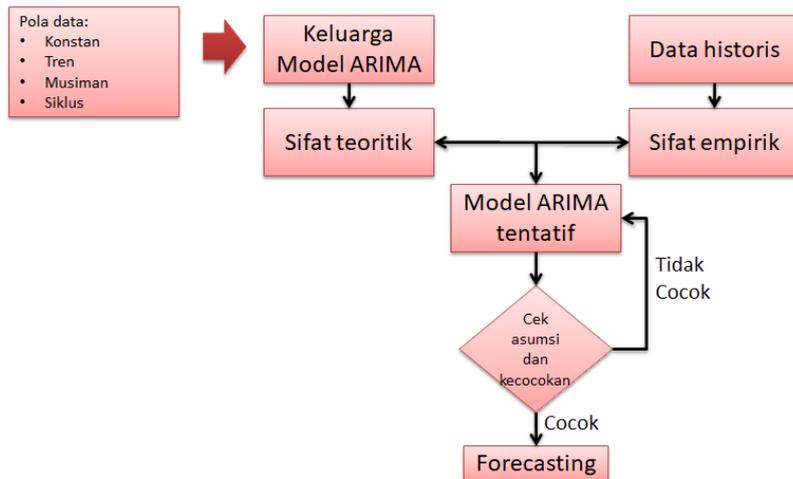
### a. ARIMA (*Autoregressive Integrated Moving Average*)

Metode Autoregressive Integrated Moving Average (ARIMA) atau biasa disebut juga sebagai metode Box-Jenkins merupakan metode yang secara intensif dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1970 (Iriawan, 2006).

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung flat (mendatar/konstan) untuk periode yang cukup panjang.

Model *Autoregresif Integrated Moving Average* (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (time series) secara statistik berhubungan satu sama lain (dependent).

Model ARIMA terdiri dari tiga langkah dasar, yaitu tahap identifikasi, tahap penaksiran dan pengujian, dan pemeriksaan diagnostik. Selanjutnya model ARIMA dapat digunakan untuk melakukan peramalan jika model yang diperoleh memadai.



Gambar 1. Prosedur Peramalan Model Arima (Box- Jenkins)

Hal yang perlu diperhatikan adalah bahwa kebanyakan deret berkala bersifat stasioner. Stasioner berarti tidak terdapat pertumbuhan atau penurunan pada data. Data secara kasarnya harus horizontal sepanjang sumbu waktu. Dengan kata lain, fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan varians dari fluktuasi tersebut pada pokoknya tetap konstan setiap waktu. Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan differencing. Yang dimaksud dengan differencing adalah menghitung perubahan atau selisih nilai observasi. Nilai selisih yang diperoleh dicek lagi apakah stasioner atau tidak. Jika belum stasioner maka dilakukan differencing lagi. Jika varians tidak stasioner, maka dilakukan transformasi logaritma.

Model Box-Jenkins (ARIMA) dibagi kedalam 3 kelompok, yaitu: model autoregressive (AR), moving average (MA), dan model campuran ARIMA (autoregressive moving average) yang mempunyai karakteristik dari dua model pertama. ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampaunya. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $Y_{t-p}$  = data *time series* pada kurun waktu ke (t-p)
- $\mu$  = suatu konstanta
- $\theta_1, \theta_q, \phi_1, \phi_n$  = parameter-parameter model
- $\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

Model dikatakan baik jika nilai error bersifat random, artinya sudah tidak mempunyai pola tertentu lagi. Dengan kata lain model yang diperoleh dapat menangkap dengan baik pola data yang ada. Untuk melihat kerandoman nilai error dilakukan pengujian terhadap nilai koefisien autokorelasi dari error, dengan menggunakan salah satu dari dua statistik berikut, yaitu Uji Q-Box and Pierce dan uji Ljung-Box.

## b. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih peubah. Pada analisis regresi, dibedakan menjadi dua, yaitu peubah respon atau biasa juga disebut peubah bergantung (*dependent variable*) dan peubah *explanatory* atau biasa disebut penduga (*predictor variable*) atau disebut juga peubah bebas (*independent peubah*). Model regresi digunakan untuk meramalkan atau memprediksi nilai masa depan dari peubah respon berdasarkan nilai-nilai tertentu yang diberikan oleh peubah prediktor.

Asumsi yang mendasari penggunaan analisis regresi linier adalah:

1. Sisaan mengikuti fungsi distribusi normal;
2. Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas);
3. Tidak terdapat autokorelasi antara sisaan untuk setiap data pengamatan; dan
4. Tidak terdapat multikolinearitas antara peubah respon.

Model regresi linier sederhana melibatkan satu peubah prediktor dan diformulasikan dalam model matematis sebagai berikut:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

dimana:

- $y$  = peubah respon/dependen/terikat
- $x$  = peubah prediktor/independen/bebas
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi peubah bebas
- $\varepsilon$  = error/residu/sisaan

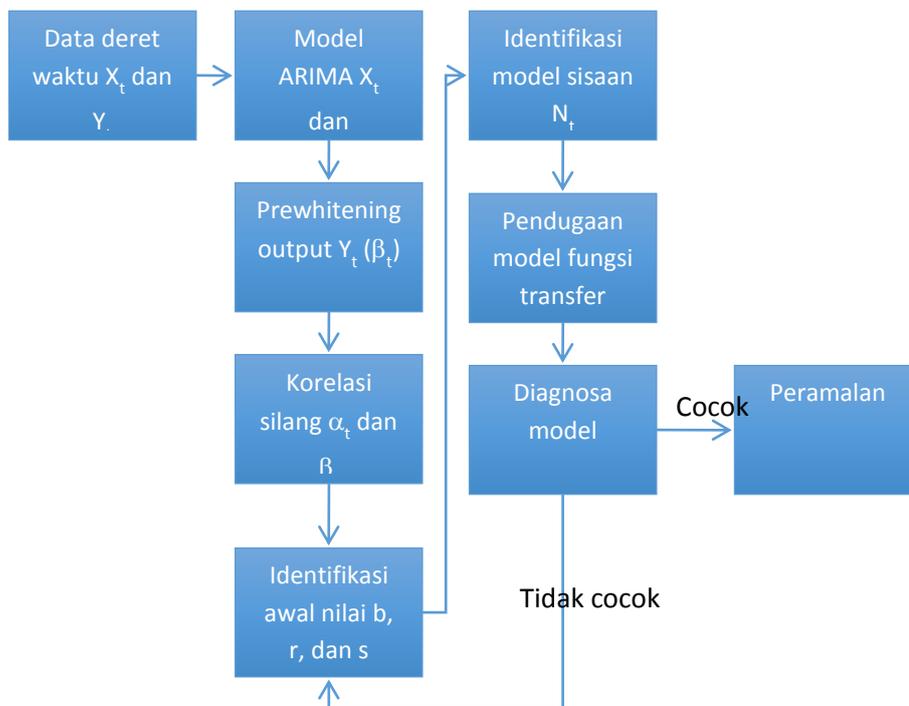
Parameter model atau koefisien regresi yaitu  $\beta_0$  dan  $\beta_1$  masing-masing memiliki interpretasi sebagai intercept dan slope dari suatu garis lurus.  $\beta_1$  mengukur perubahan rata-rata dari peubah respon  $y$  untuk setiap perubahan dari peubah prediktor  $x$ . Parameter ini biasanya tidak diketahui dan harus diprediksi dari data sampel.  $\varepsilon$  merupakan kesalahan statistik, didefinisikan sebagai sisa atau perbedaan hasil antara nilai data pengamatan peubah respon terhadap nilai peubah respon hasil prediksi.

## C. Model Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau  $Y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $Y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $X_t$ ) dengan deret output tersebut. Model fungsi transfer merupakan fungsi dinamis yang pengaruhnya tidak hanya pada hubungan linier antara deret input dengan deret output pada waktu ke- $t$ , tetapi juga pada waktu

$t+1, t+2, \dots, t+k$ . Hubungan seperti ini pada fungsi transfer dapat menimbulkan delay (waktu senjang) antara peubah *input* dan peubah *output*.

Tujuan pemodelan fungsi transfer adalah untuk menetapkan model yang sederhana, yang menghubungkan deret output ( $Y_i$ ) dengan deret input ( $X_i$ ) dan gangguan/noise( $n_i$ ). Wei (1994) juga menjelaskan bahwa di dalam fungsi transfer terdapat rangkaian output yang mungkin dipengaruhi oleh rangkaian multiple input. Pada kasus single input peubah, dapat menggunakan metode korelasi silang yang dianjurkan oleh Box and Jenkins (1976). Teknik ini juga dapat digunakan ketika terdapat single input peubah yang lebih dari satu selama antar variable input tidak berkorelasi silang. Jika beberapa atau semua peubah input berkorelasi silang maka teknik prewhitening atau metode korelasi silang tidak dapat digunakan secara langsung. Alasan utama bagi perlunya suatu perencanaan atau peramalan adalah adanya tenggang waktu pengambilan keputusan yang dapat berkisar dari beberapa hari atau sampai beberapa tahun. Pada analisis fungsi transfer untuk peramalan deret berkala univariate, terdapat deret berkala output yang diperkirakan dipengaruhi oleh deret berkala input dan input-input lain yang digabungkan dalam satu kelompok yang disebut gangguan (noise). Deret input mempengaruhi deret output melalui sebuah fungsi transfer yang mendistribusikan pengaruhnya secara dinamis melalui beberapa periode waktu yang akan datang dengan persentase tertentu yang disebut sebagai bobot respons impuls atau bobot fungsi transfer.



Gambar 2. Langkah-langkah melakukan pemodelan Fungsi Transfer

Model umum Fungsi Transfer:

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\varphi_p(B)} \varepsilon_t \quad \dots\dots\dots(7)$$

Dimana:

- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$
- $q \rightarrow$  ordo MA bagi noise  $N_t$

**c. Model Vector Autoregression (VAR)**

*Vector Autoregression* (VAR) dikemukakan pertama kali oleh Christopher Sims (1980). Sims mengembangkan model ekonometri dengan mengabaikan pengujian asumsi secara apriori. VAR dikembangkan oleh Sims sebagai kritik atas metode simultan. Jumlah peubah yang besar dan klasifikasi endogen dan eksogen pada metode simultan merupakan dasar dari kritik tersebut. Penggunaan pendekatan struktural atas pemodelan persamaan simultan biasanya menerapkan teori ekonomi di dalam usahanya untuk mendeskripsikan hubungan antar peubah yang ingin diuji. Akan tetapi sering ditemukan bahwa teori ekonomi saja ternyata tidak cukup kaya di dalam menyediakan spesifikasi yang ketat dan tepat atas hubungan dinamis antar peubah (Gujarati, 2010). Model VAR merupakan jalan keluar atas permasalahan ini melalui pendekatan non-struktural atau tidak mendasarkan pada teori ekonomi tertentu (*atheoretical*). Metode VAR memperlakukan seluruh peubah secara simetris tanpa memperlakukan peubah dependen dan independen atau dengan kata lain model ini memperlakukan seluruh peubah sebagai peubah endogen., karena pada kenyataannya suatu peubah dapat bertindak sebagai endogen atau dependen dari suatu peubah yang lainnya (Gujarati, 2010).

Kelebihan dalam penggunaan metode VAR (Gujarati, 2010):

- a. Kemudahan dalam penggunaan, tidak perlu mengkhawatirkan tentang penentuan peubah endogen dan peubah eksogen.
- b. Kemudahan dalam estimasi, metode *Ordinary Least Square* (OLS) dapat diaplikasikan pada tiap persamaan secara terpisah.
- c. *Forecast* atau peramalan yang dihasilkan pada beberapa kasus ditemukan lebih baik daripada yang dihasilkan oleh model persamaan simultan yang kompleks.
- d. Perangkat estimasi yang digunakan adalah *Impulse Respon Function* (IRF) untuk melacak respon dari peubah dependen dalam sistem VAR terhadap *shock* dari *error term* dan *Variance Decomposition* yang memberikan informasi mengenai pentingnya masing-masing *error term* dalam mempengaruhi peubah-peubah dalam VAR.

Di sisi lain, terdapat beberapa kritik terhadap model VAR menyangkut permasalahan berikut (Gujarati, 2010) :

- 1) Model VAR merupakan model yang *atheoritic* atau tidak berdasarkan teori, hal ini tidak seperti pada persamaan simultan. Pada persamaan simultan, pemilihan peubah yang akan dimasukkan dalam persamaan memegang peranan penting dalam mengidentifikasi model.
- 2) Pada model VAR penekanannya terletak pada *forecasting* atau peramalan sehingga model ini kurang cocok digunakan dalam menganalisis kebijakan.
- 3) Permasalahan yang besar dalam model VAR adalah pada pemilihan *lag length* atau panjang lag yang tepat. Karena semakin panjang lag, maka akan menambah jumlah parameter yang akan bermasalah pada *degrees of freedom*.
- 4) Peubah yang tergabung pada model VAR harus stasioner. Apabila tidak stasioner, perlu dilakukan transformasi bentuk data, misalnya melalui *first difference*.
- 5) Sering ditemui kesulitan dalam menginterpretasi tiap koefisien pada estimasi model VAR, sehingga sebagian besar peneliti melakukan interpretasi pada estimasi fungsi *impulse respon*.

#### e. Kriteria Pemilihan Model Terbaik

Untuk menguji kebaikan suatu model ada beberapa kriteria yang digunakan. Pada buku pedoman teknis ini kriteria yang digunakan adalah MAPE. Model time series/arima, model regresi, model fungsi transfer, dan model VAR dapat digunakan untuk melakukan estimasi variabel untuk beberapa tahun ke depan. Untuk model *time series* baik analisis ARIMA, model regresi, model fungsi transfer, dan model VAR, ukuran kelayakan model berdasarkan nilai kesalahan dengan menggunakan statistik MAPE (*mean absolute percentage error*) atau kesalahan persentase absolut rata-rata yang diformulasikan sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right| \cdot 100$$

Dimana :  $X_t$  adalah data aktual

$F_t$  adalah nilai ramalan.

Semakin kecil nilai MAPE maka model yang diperoleh semakin baik, karena makin mendekati nilai aktual.

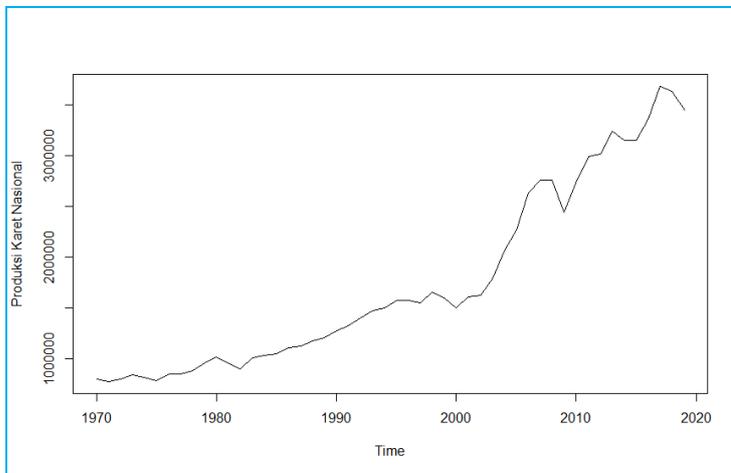
#### f. Sumber Data

Data yang digunakan untuk analisis ini bersumber dari Ditjen. Perkebunan, Kementerian Pertanian, meliputi data : produksi karet dalam bentuk karet kering satuan ton tahun 1967 – 2019, luas areal karet satuan hektar tahun 1970 – 2019, harga riil karet sheet satuan Rp/kg tahun 1971 – 2019, volume ekspor dan impor karet nasional satuan ton tahun 1969 – 2019. Disamping itu juga ada data harga karet dunia yang bersumber dari World Bank, tahun 1970 – 2019

## HASIL DAN PEMBAHASAN

### A. Model ARIMA

Eksplorasi data produksi karet nasional dalam bentuk karet kering berupa data tahunan dari tahun 1970 sampai 2019, seperti yang terlihat pada Gambar 1. Pada gambar 1 terlihat produksi karet pada tahun 1970 sebesar 800,15 ribu ton, produksi karet terus naik secara landai sehingga pada tahun 2000 produksi karet mencapai 1,50 juta ton. Pertumbuhan produksi karet tahun 1980 – 2000 rata-rata sebesar 2,06 %/tahun. Pada tahun 2001 sampai 2019 produksi karet nasional naik signifikan, jika tahun 2001 produksi karet sebesar 1,61 juta ton, maka tahun 2019 produksi karet sudah mencapai 3,45 juta ton, rata-rata pertumbuhan lebih tinggi pada periode sebelumnya yaitu mencapai 4,62%/tahun. Berdasarkan Gambar 1 juga bisa terlihat bahwa data belum stasioner karena masih mengalami perubahan seiring perubahan waktu.



Gambar 3. Perkembangan Produksi Karet Nasional Tahun 1970 - 2019

Dalam melakukan pemodelan produksi karet menggunakan model Autoregressive Integrated Average (ARIMA), data yang digunakan adalah periode tahun 1970 sampai 2019. Periode data tersebut kemudian dipisahkan menjadi data set training dan testing. Perlunya pemisahan data training dan testing adalah untuk menguji tingkat akurasi dalam melakukan peramalan. Panjang series data pada data set training adalah tahun 1970 sampai 2013, sementara dataset testing adalah periode 2014 sampai 2019. Dataset training digunakan untuk melakukan penyusunan model, sementara dataset testing digunakan untuk validasi model.

Uji kestasioneran data seperti yang disyaratkan apabila melakukan pemodelan ARIMA dilakukan secara visual menggunakan hasil plot data maupun uji formal statistik. Gambar 3 menunjukkan tidak ada fluktuasi yang muncul secara regular setiap bulannya, sehingga dapat disimpulkan bahwa data belum stasioner atau nilai rata-rata dan variansi dari data time series karet mengalami perubahan secara stokastik sepanjang waktu atau sebagian ahli menyatakan rata-rata dan variannya belum konstan (Narchrowi dan Haridus usman, 2006).

Tabel 1. Hasil Uji Augmunted Dickey-Fuller Produksi Karet

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-351518 -55201  -3404   64774 264455

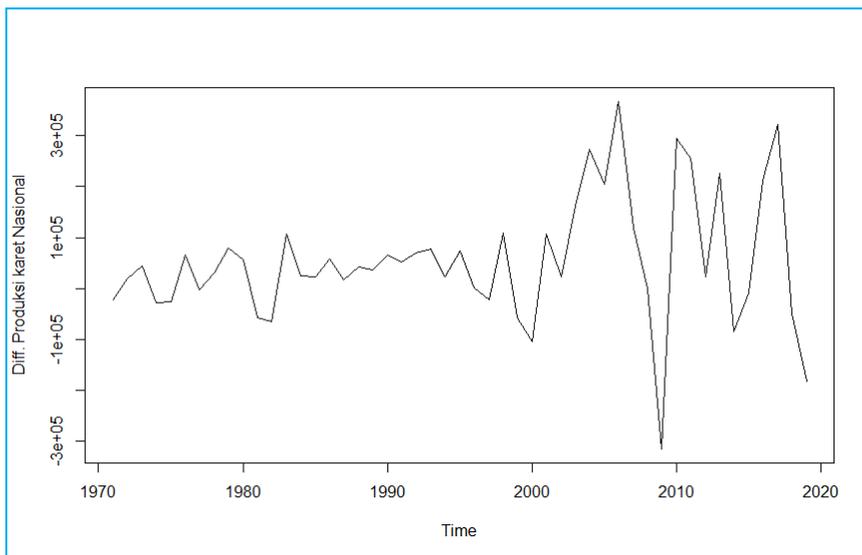
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.786e+04  3.916e+04   0.967  0.3390
z.lag.1      -1.237e-01  6.174e-02  -2.004  0.0512 *
tt           8.734e+03  3.839e+03   2.275  0.0278 *
z.diff.lag   1.757e-01  1.562e-01   1.125  0.2666
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 121300 on 44 degrees of freedom
Multiple R-squared:  0.1303, Adjusted R-squared:  0.07101
F-statistic: 2.198 on 3 and 44 DF, p-value: 0.1018

Value of test-statistic is: -2.0042 3.6204 2.6596

Critical values for test statistics:
      1pct  5pct 10pct
tau3  -4.15 -3.50 -3.18
phi2   7.02  5.13  4.31
```

Hal ini diperkuat oleh hasil uji formal statistik yaitu dengan uji Augmunted Dickey-Fuller yang mengindikasikan bahwa data produksi karet adalah belum stasioner, terlihat dari hasil uji tes statistik sebesar  $-2,0042$  sementara nilai kritis pada tingkat kepercayaan  $5\% = -3,50$  atau lebih kecil dari nilai uji statistik sehingga sehingga  $H_0$  tidak ditolak, atau data produksi karet belum stationer. Oleh karena itu, selanjutnya data produksi karet dilakukan pembedaan (differencing) satu kali. Hasil plot setelah dilakukan differencing satu kali seperti pada Gambar 4.



Gambar 4. Perkembangan Produksi Karet Nasional Setelah Difference 1

Hasil plot produksi karet setelah dilakukan differencing 1, menunjukkan bahwa data sudah terlihat stasioner untuk rata-rata. Hal ini juga didukung dengan uji Augmunted Dickey-Fuller yang mengindikasikan bahwa data produksi karet setelah differencing 1 sudah stasioner, terlihat dari hasil uji tes statistik sebesar  $-4,1214$  sementara nilai kritis pada tingkat kepercayaan  $95\% = -1,95$  dan tingkat kepercayaan  $99\% = -2,62$  atau lebih besar dari

nilai uji statistik sehingga sehingga  $H_0$  ditolak, atau data produksi karet setelah diifrencing 1 sudah stationer.

Tabel 2. Hasil Uji Augmunted Dickey-Fuller Produksi Karet Difference 1

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-305065 -21530  33083  95147 400703

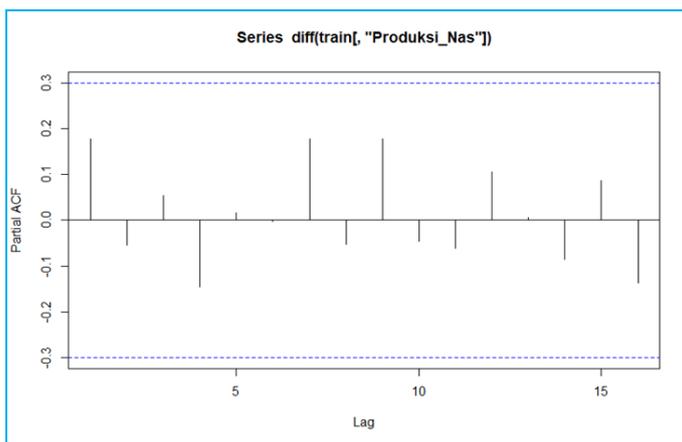
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.7351    0.1784  -4.121  0.00016 ***
z.diff.lag  0.0735    0.1514   0.485  0.62973
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

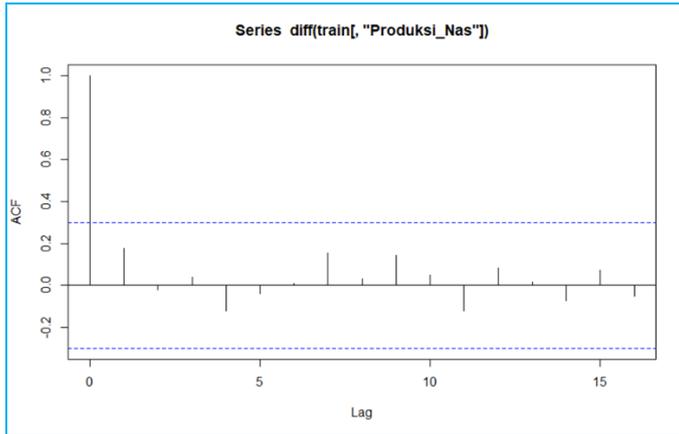
Residual standard error: 133500 on 45 degrees of freedom
Multiple R-squared:  0.3363, Adjusted R-squared:  0.3068
F-statistic: 11.4 on 2 and 45 DF, p-value: 9.885e-05

Value of test-statistic is: -4.1214

Critical values for test statistics:
    1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Identifikasi ordo AR dan MA untuk data produksi karet sebelum ada difference menunjukkan pola ACF *tail off* sementara pola PACF tidak ada yang signifikan. Hal ini menunjukkan bahwa benar data karet belum signifikan, sehingga perlu dilakukan differencing 1 agar pola ACF dan PACF lebih jelas terlihat.





Gambar 5. Plot ACF dan PACF Produksi Karet Difference 1

Pada Gambar 5 menunjukkan bahwa baik plot ACF maupun plot ACF setelah *Difference* 1 keduanya sudah *cut off*. Identifikasi ordo AR dan MA seperti tersaji pada Gambar 5 menunjukkan pola ACF *cut off* setelah lag 0, sementara pola PACF tidak ada yang menunjukkan signifikan, sehingga model tentatif belum dapat ditentukan.

Tabel 3. Model Arima Tentatif Berdasarkan Automodel

```
Series: train[, "Produksi_Nas"]
ARIMA(0,2,1)

Coefficients:
      ma1
      -0.8889
s.e.      0.0751

sigma^2 estimated as 1.377e+10: log likelihood=-550.14
AIC=1104.28   AICc=1104.58   BIC=1107.75

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 20862.56 113292.8 75846.47 1.118655 4.497759 0.8522773 0.07893533
```

Pengamatan secara visual pada plot ACF dan PACF sulit menentukan orde ARIMA, setelah dilakukan run model dengan menggunakan *auto arima* maka orde ARIMA yang disarankan adalah ARIMA (0,2,1), artinya model ARIMA tentative terbaik untuk melakukan estimasi produksi karet nasional adalah untuk orde AR nilai  $p=0$ , untuk orde MA nilai  $q=1$ , dan difference  $d=2$ . Berdasarkan Tabel 3 dengan menggunakan ARIMA (0,2,1) maka untuk data training, akan menghasilkan MAPE = 4,49% artinya data berdasarkan model arima akan menyimpang sekitar -4,5% sampai +4,5% dari data aktual.

Disamping metode pemilihan model Arima berdasarakan automodel, digunakan juga metode lain untuk mendapatkan orde ARIMA terbaik, yaitu dengan metode *Arima selection*. Pada metode ini akan dikeluarkan beberapa model terbaik. Setelah dilakukan pemilihan model metode terbaik tetap pada *differencing* 2. Model terbaik pertama menurut metode ini adalah ARIMA (2,2,0) menghasilkan nilai sbc paling kecil yaitu sebesar 999,69. Model terbaik kedua adalah ARIMA (1,2,0) dengan nilai sbc = 998,48. Model terbaik berdasarkan auto arima berada pada urutan ketujuh, yaitu ARIMA (0,2,1) dengan nilai sbc=1006,23.

Tabel 4. Model Arima Tentatif Berdasarkan Arima Selection

	p	q	sbc
[1,]	2	0	996.6099
[2,]	1	0	998.4833
[3,]	0	0	999.6473
[4,]	3	0	1000.4565
[5,]	4	0	1003.5162
[6,]	5	0	1006.3272
[7,]	0	1	1006.5376
[8,]	3	3	1007.3866
[9,]	2	4	1007.5090
[10,]	0	2	1008.1330

Untuk menentukan model ARIMA terbaik dari ketiga model tentative tersebut dilakukan pengujian berdasarkan data training dan data testing. Model terbaik adalah model yang menghasilkan MAPE terkecil. Pada analisis ini sudah dilakukan pembagian data training dan data testing. Data training merupakan data produksi karet dari tahun 1967 – 2014, sementara data testing diambil 6 data terakhir, yaitu produksi karet tahun 2014 – 2019. Untuk menguji performa model ARIMA terbaik, dilakukan pengujian dengan data testing. Hasil pengujian data training dan testing terlihat pada tabel 5.

Tabel 5. Hasil Pengujian MAPE untuk beberapa model Arima Tentative Terbaik

Model Arima	MAPE Training	MAPE Testing	Kesimpulan
ARIMA (2,2,0)	4,3762	13,8570	-
ARIMA (1,2,0)	4,9578	12,3090	-
ARIMA (0,2,1)	4,3131	6,3442	Terbaik

Selanjutnya dilakukan pengujian model ARIMA (0,2,1) apakah koefisien sudah signifikan. Untuk model ARIMA (0,2,1) koefisien  $ma_1$  sebesar -0,8924 dan koefisien ini signifikan pada taraf 1%. Sehingga model ARIMA (0,2,1) layak digunakan.

Tabel 6. Uji Koefisien Model Arima (0,2,1)

```
Call:
arima(x = train[, "Produksi_Nas"], order = c(0, 2, 1))

Coefficients:
      ma1
-0.8924
s.e.    0.0710

sigma^2 estimated as 1.256e+10:  log likelihood = -587.85,  aic = 1179.71

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1 -0.892422   0.071025 -12.565 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Salah satu syarat kebaikan model ARIMA adalah sebaran sisaan LJung-Box. Hasil pengujian nilai p-value lag 5 sampai dengan lag 30 tidak ada yang signifikan, sehingga dapat disimpulkan bahwa sisaan bersifat random dan tidak ada autorelasi. Hal ini menunjukkan bahwa ARIMA (0,2,1) layak digunakan.

Tabel 7. Uji Sisaan LJung Box Model Arima (0,2,1)

lags	statistic	df	p-value
5	4.545437	5	0.4738272
10	6.452847	10	0.7758913
15	9.287861	15	0.8619915
20	10.086629	20	0.9665738
25	10.555918	25	0.9948639
30	16.415244	30	0.9789470

Selanjutnya dilakukan pengepasan model untuk seluruh data. Untuk Model ARIMA (0,2,1) koefisien ma sebesar -0,9418. Jika melakukan run model ARIMA (0,2,1) untuk seluruh data yaitu dari tahun 1967 – 2019 maka akan dihasilkan MAPE sebesar 4,39%. Hal ini menunjukkan bahwa antara data estimasi dengan data actual akan berbeda rata-rata berkisar anantara -4,39% sampai +4,39%. Untuk metode stimasi dengan bias masih dibawah 5% dianggap masih cukup baik dan akurat.

Tabel 8. Model Arima (0,2,1) untuk Seluruh Data

```
Series: karet[, "Produksi_Nas"]
ARIMA(0,2,1)
Coefficients:
      ma1
    -0.9418
s.e.    0.0611

sigma^2 estimated as 1.542e+10:  log likelihood=-671.15
AIC=1346.31  AICC=1346.56  BIC=1350.17

Training set error measures:
Training set  ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
13796.38  120614.5  83572.25  0.5687763  4.393883  0.909479  0.1075829
```

Dengan menggunakan model ARIMA (0,2,1) menghasilkan angka estimasi produksi karet untuk 5 tahun ke depan. Hasil Estimasi dengan model ARIMA ini pada tahun 2020 produksi karet sebesar 3,51 juta ton. Pada tahun 2021 produksi karet diestimasi akan naik sebesar 1,91% menjadi 3,58 juta ton. Pada tahun 2022 sampai 2024 juga menunjukkan kenaikan. Rata-rata pertumbuhan produksi karet tahun 2020 – 2024 rata-rata sebesar 1,85%.

Tabel 9. Ouput Permalan Model Arima (0,2,1) untuk Produksi Karet

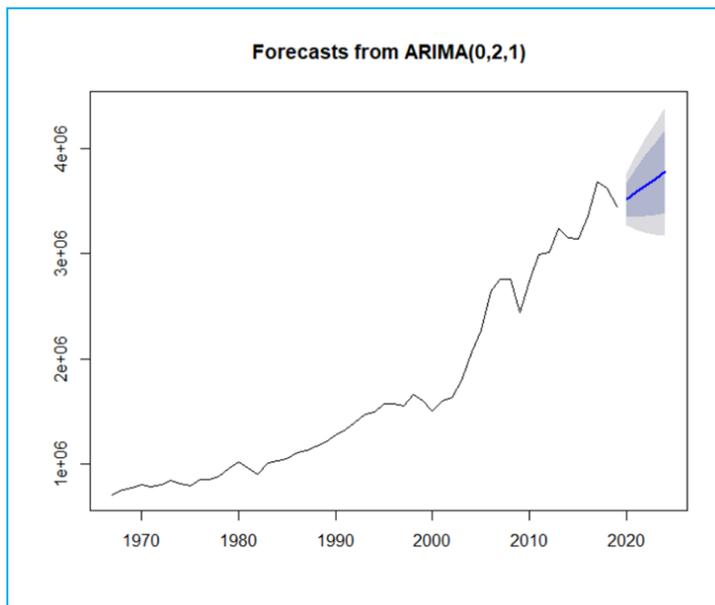
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	3515956	3356795	3675117	3272540	3759371
2021	3583129	3351375	3814884	3228692	3937567
2022	3650303	3358233	3942373	3203620	4096986
2023	3717477	3370650	4064303	3187051	4247902
2024	3784650	3386112	4183189	3175138	4394163

Tabel 10. Hasil Estmasi Produksi Karet dengan Model ARIMA (0,2,1)

Tahun	Estimasi Produksi Karet	Pertumbuhan
	(Ton)	(%)
2019	3 448 782	
2020	3 515 956	1,948
2021	3 583 129	1.911
2022	3 650 303	1.875
2023	3 717 477	1.840
2024	3 784 650	1.807
	Rata - rata	1.858

Tahun 2019 : Angka Sementara (Ditjen. Perkebunan)

Tahun 2020 – 2024 : Angka Estimasi Model Arima (0,2,1)



Gambar 6. Hasil Estimasi Produksi Karet Tahun 2021 – 2024 Model Arima (0,2,1)

## B. Model Regresi

Model kedua untuk meramalkan produksi karet adalah model regresi. Model regresi yang diuji adalah model regresi linier sederhana dan model regresi kuadratis. Untuk regresi linier sederhana maupun regresi kuadratis sebagai variabel tak bebas bebas adalah produksi, sementara untuk variabel bebas adalah waktu.

Sebelum dilakukan analisis data dibagi dulu menjadi data training dan data testing. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk validasi model. Dari 50 observasi data produksi karet dilakukan pembagian data training dan testing, dimana periode 1970-2013 menjadi data training sedangkan 2014-2019 menjadi data testing (enam tahun terakhir).

## Regresi Linier Sederhana

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dengan produksi karet. Data input yang digunakan untuk pendugaan model adalah data training produksi karet dari tahun 1970-2013.

Tabel 11. Ouput Model Regresi Linier

```

Call:
lm(formula = Prod_Total ~ Tahun, data = train)
Residuals:
    Min       1Q   Median       3Q      Max
-468538 -177119  -97976  203809  596285
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -101288829   6320861  -16.02  <2e-16 ***
Tahun         51629       3174    16.27  <2e-16 ***

---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 267300 on 42 degrees of freedom
Multiple R-squared:  0.863,    Adjusted R-squared:  0.8598
F-statistic: 264.6 on 1 and 42 DF,  p-value: < 2.2e-16
    
```

Dari hasil pendugaan model diatas dijelaskan bahwa tahun signifikan terhadap produksi karet. Model ini layak dengan nilai F hitung = 264,6 dan nilai p-Value < 0.05 artinya model signifikan pada tingkat kepercayaan 95%. Koefisien regresi untuk variabel bebas tahun, juga menunjukkan nilai yang signifikan pada alpha = 0.1%. Model ini cukup layak dimana R<sup>2</sup> Adjusted sebesar 85,98%. Dari model diatas dapat dibuat model produksinya yaitu :

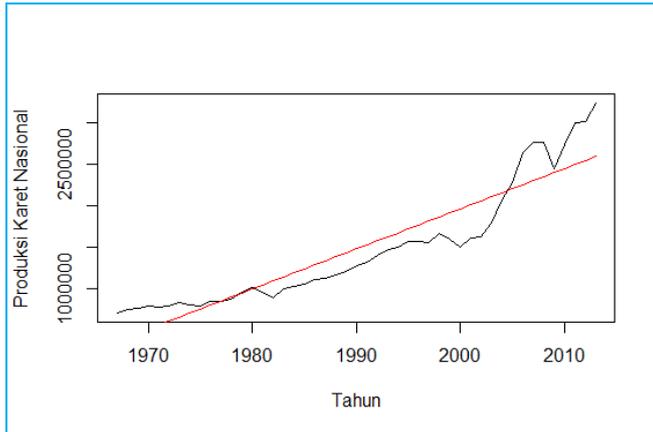
$$\text{Produksi} = -101288829 + (51629 * \text{Tahun})$$

Untuk menguji kelayakan model dalam melakukan estimasi maka dilakukan pengujian perbandingan MAPE untuk data training dan data testing. Hasil pengujian model menunjukkan untuk regresi sederhana ini menghasilkan MAPE untuk data training rata-rata sebesar 15,49%, sedangkan untuk MAPE data testing lebih besar lagi yaitu 16,88%.

Tabel 12. MAPE Data Training dan Testing Model Regresi Linier

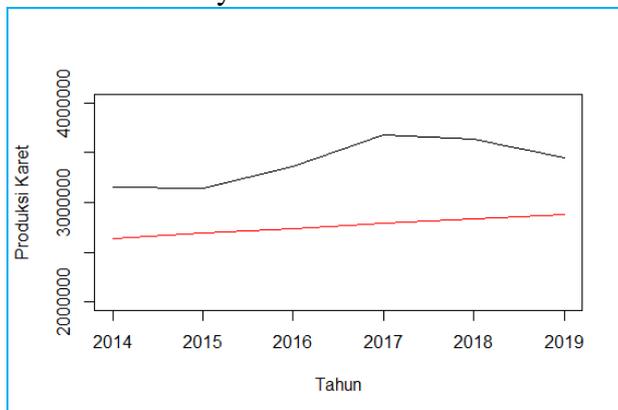
MAPE data training : 13,77%						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.2342	9.0730	13.7713	15.4914	17.2847	47.5053	
MAPE data testing : 15,67%						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
12.75	14.48	15.67	16.88	19.29	22.63	

Untuk melihat tingkat akurasi dari hasil pemodelan regresi ini, selanjutnya dilakukan plot data training dengan data actual. Data training yang digunakan adalah produksi karet dari tahun 1970-2013. Hasil plot pada Gambar 13 menunjukkan hasil plot regresi cenderung lurus, sementara data actual berflutuasi bisa di atas atau di bawah garis regresi linier. MAPE untuk data training ini cukup besar yaitu 15,49%.



Gambar 7. Plot Data Training Regresi Linier dan Data Aktual Produksi Karet 1970-2019

Hal yang sama juga dilakukan untuk plot data testing dan data aktual produksi karet tahun 2014 – 2019. Jika model yang dihasilkan dari pola data data tahun 1970 – 2013, digunakan untuk mengestimasi produksi tahun 2014 – 2019, menunjukkan akan dihasilkan MAPE data testing sebesar 16,88%, artinya rata-rata perbedaan data aktual dan data estimasi sebesar 16,88%. Hal ini ditunjukkan pada Gambar 14, data hasil estimasi model selalu di bawah dari data aktualnya.



Gambar 8. Plot Data Testing Regresi Linier dan Data Aktual Produksi Karet 2014-2019

Hasil estimasi menggunakan regresi sederhana menunjukkan hasil estimasi yang cenderung under estimate. Hasil estimasi selama 5 tahun kedepan produksi karet nasional berkisar antara 3,00 juta ton sampai 3,21 juta ton. Produksi ini masih lebih rendah dari produksi Angka Sementara tahun 2019 yaitu sebesar 3,45 juta ton. Hal ini juga didukung nilai MAPE data testing yang terlampau besar 16,88%, sehingga akurasi menjadi rendah.

Tabel 13. Hasil Estimasi Model Regresi Linier Produksi Karet Tahun 2020-2024

Tahun	Produksi Karet (Ton)	Pertumbuhan (%)
2019	3 448 782	
2020	3.002.554	-12,939
2021	3.054.183	1,720
2022	3.105.813	1,690
2023	3.157.442	1,662
2024	3.209.072	1,635
	Rata-rata	1,677

Tahun 2019 : Angka Sementara (Ditjen. Perkebunan)

Tahun 2020 – 2024 : Angka Estimasi Model Regresi Linier

### Regresi Kuadratik

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dan tahun kuadrat dengan produksi karet. Data input yang digunakan untuk pendugaan model adalah data training produksi karet dari tahun 1970-2013. Hasil output run model untuk regresi kuadratik seperti pada Tabel 14.

Tabel 14. Model Regresi Kuadratik untuk Produksi Karet

```
Call:
lm(formula = Produksi_Nas ~ t + t2, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-362806  -56937   25423   67027  308529

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  892831     67184  13.289 < 2e-16 ***
t            -16462     6737  -2.444  0.0188 *
t2             1509       142  10.628 1.76e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 143600 on 42 degrees of freedom
Multiple R-squared:  0.9646,    Adjusted R-squared:  0.9629
F-statistic: 572.7 on 2 and 42 DF,  p-value: < 2.2e-16
```

Persamaan regresi kuadratik yang dihasilkan adalah :

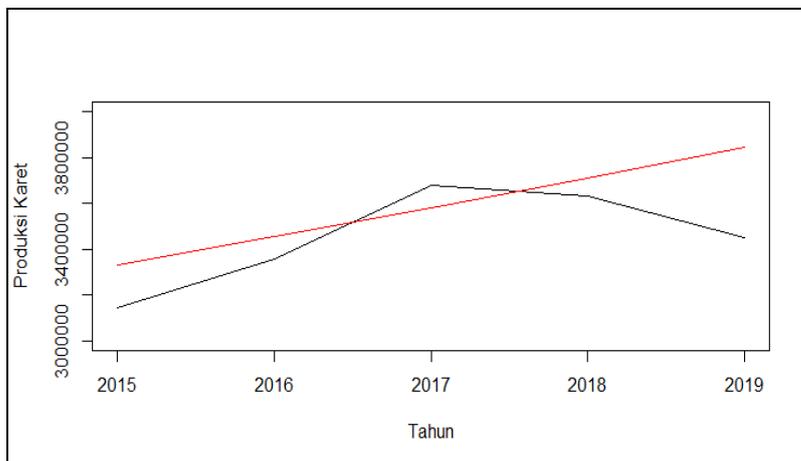
$$\text{Produksi} = 892831 + (-16462 \cdot t) + (1509 \cdot t^2)$$

Model regresi kuadratik menghasilkan model yang layak dengan nilai F-hitung= 572.7 dan nilai p-value < 0.05 . Nilai R<sup>2</sup>-Adjusted lebih tinggi dari regresi linier yaitu sebesar 96,92%. Untuk melihat apakah koefisien regresi signifikan atau tidak digunakan uji t, hasil uji t menunjukkan intersep dan koefisien t<sup>2</sup> signifikan pada tingkat kepercayaan 99%, koefisien t signifikan pada tingkat kepercayaan 95%.

Untuk menguji kelayakan model regresi kuadratik dalam melakukan estimasi maka dilakukan pengujian perbandingan MAPE untuk data training dan data testing. Hasil pengujian model menunjukkan untuk regresi sederhana ini menghasilkan MAPE untuk data training rata-rata sebesar 6,61%, sedangkan untuk MAPE data testing lebih baik lagi yaitu 5,01%. Jika dibandingkan dengan regresi linier biasa, maka MAPE regresi kuadratik lebih baik, sehingga selisih antara data estimasi dan aktual lebih kecil. Untuk menggambarkan plot data hasil estimasi dan data actual pada Gambar 9, menunjukkan, garis warna merah merupakan hasil estimasi sedangkan warna hitam data actual, kedua garis ini relative lebih berdekatan dibandingkan dengan regresi linier. Hal ini menunjukkan bahwa regresi kuadratis menghasilkan model dan estimasi yang lebih baik dibandingkan regresi linier.

Tabel 15. Perbandingan MAPE Data Training dan Data Testing Regresi Kuadratik

MAPE data training					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4081	3.8529	5.0921	6.6129	8.8691	22.3774
MAPE data testing					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.192	2.729	2.833	5.005	5.843	11.427



Gambar 9. Plot Data Testing Regresi Kuadratis dan Data Aktual Produksi Karet 1970-2019

Hasil analisis estimasi menunjukkan pada tahun 2020 produksi karet mencapai 3,97 juta ton, atau meningkat sebesar 15,37%. Hasil estimasi ini diperkirakan terjadi over estimate, karena terjadi lonjakan produksi yang sangat tinggi. Angka pertumbuhan hasil estimasi rata-rata sebesar 3,42% per tahun, dan rata-rata produksi karet setiap tahun meningkat sekitar 100 ribu ton. Meskipun model regresi kuadratik menghasilkan MAPE yang data testing yang lebih kecil 5,05% tetapi hasil estimasi diduga terlampaui tinggi (Over Estimate).

Tabel 16. Hasil Estimasi Model Regresi Kuadratik Produksi Karet Tahun 2020-2024

Tahun	Produksi Karet (Ton)	Pertumbuhan (%)
2019	3.448.782	
2020	3.978.853	15,370
2021	4.117.846	3,493
2022	4.259.856	3,449
2023	4.404.885	3,405
2024	4.552.933	3,361
	Rata-rata	3,427

Tahun 2019 : Angka Sementara (Ditjen. Perkebunan)

Tahun 2020 – 2024 : Angka Estimasi Model Regresi Kuadratik

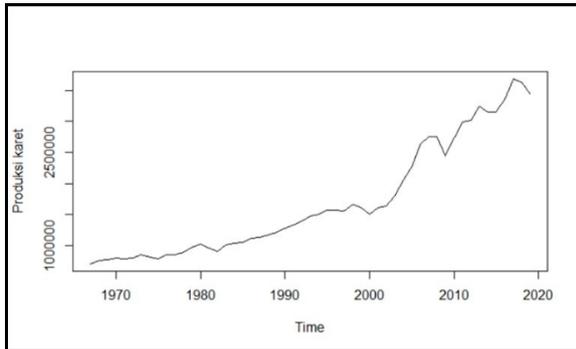
### C. Model Fungsi Transfer

Model fungsi transfer merupakan gabungan dari karakteristik analisis regresi berganda dengan karakteristik deret berkala ARIMA (*Autoregressive Integrated Moving Average*). Dalam model ini, selain mengandung keterkaitan dengan kejadian pada waktu-waktu sebelumnya, juga keterkaitan yang sifatnya kausal dengan variabel lain sebagaimana pada model regresi. Model fungsi transfer merupakan suatu model yang menggambarkan nilai prediksi dari suatu time series (deret output atau  $Y_t$ ) berdasarkan nilai-nilai dari deret itu sendiri ( $Y_t$ ) dan berdasarkan pula pada data time series yang mempunyai hubungan (deret input atau  $X_t$ ) dengan deret output. Model fungsi transfer yang bersifat dinamis berpengaruh tidak hanya pada hubungan linier antara waktu ke- $t$  input  $X_t$  dan waktu ke- $t$  output  $Y_t$ , tetapi juga saat input  $X_t$  dengan saat  $t, t+1, \dots, t+k$  pada output  $Y_t$  (Fitriani et al., 2013). Dalam melakukan estimasi dengan fungsi transfer diperlukan dua variabel, yaitu output  $Y_t$  dan input  $X_t$ . Input  $X_t$  merupakan variabel yang diduga paling berpengaruh terhadap output  $Y_t$ . Pada analisis ini dicontohkan estimasi untuk produksi karet (output  $Y_t$ ) yang diduga dipengaruhi oleh harga karet dunia.

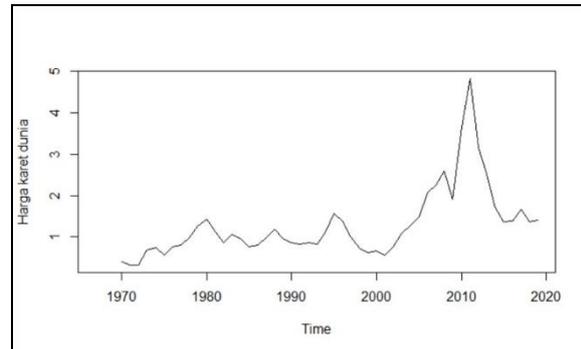
Dalam bahasan ini, model fungsi transfer variabel output produksi karet akan melibatkan variabel harga karet dunia sebagai variabel input. Langkah awal dalam mencari model fungsi transfer menggunakan software RStudio adalah penyiapan data, baik output  $Y_t$  (produksi karet) maupun input  $X_t$  (harga karet dunia). Produksi karet dalam satuan ton bersumber dari Direktorat Jenderal Perkebunan, Kementerian Pertanian, sedangkan harga karet dunia dalam satuan US\$/Kg bersumber dari World Bank. Banyaknya series data yang digunakan baik untuk produksi maupun harga karet dunia harus seragam. Pada bahasan ini, series data yang digunakan yaitu 1970-2019 atau sebanyak 50 observasi.

Eksplorasi data dilakukan dengan menampilkan plot data produksi maupun harga karet dunia. Berdasarkan plot data dapat diketahui pola data series 50 tahun yang akan digunakan untuk pemodelan. Berdasarkan Gambar 10 dan Gambar 11, terlihat bahwa terdapat data produksi karet nasional memiliki tren meningkat dari tahun ke tahun sedangkan harga karet

dunia cenderung berfluktuasi. Harga karet dunia cenderung turun pada sekitar 10 tahun terakhir, akibat persaingan harga dengan produk karet sintetis. Produksi karet nasional maupun harga karet dunia terindikasi tidak stasioner berdasarkan plotnya.



Gambar 10. Plot Data Produksi Karet, 1970-2019



Gambar 11. Plot Data Harga Karet Dunia, 1970-2019

Tahapan penyusunan model Fungsi Transfer produksi karet dengan variabel input harga karet dunia adalah sebagai berikut :

- a. Pembagian series data awal menjadi series data *training* dan *testing*
- b. Pemeriksaan kestasioneran
- c. Pencarian model tentatif untuk variabel input
- d. *Prewhitening* dan korelasi silang
- e. Identifikasi model noise
- f. Pengepasan model
- g. Peramalan berbasis fungsi transfer

Data produksi dan harga karet tahun 1970-2019 sebanyak 50 series akan dibagi menjadi series data *training* untuk periode 1970-2013 dan series data *testing* untuk periode 2014-2019. Selanjutnya dilakukan uji kestasioneran data untuk data input  $X_t$  yaitu harga karet dunia menggunakan uji Augmented Dickey-Fuller (ADF).

Hipotesis pada uji ADF ini adalah:

$H_0$  : data tidak stasioner

$H_1$  : data stasioner

Nilai test-statistic yang lebih besar dari critical values (nilai tau3), baik untuk taraf 1%, 5% maupun 10% menunjukkan bahwa  $H_0$  gagal ditolak, atau series data harga karet dunia belum stasioner. Oleh karena itu akan dilakukan pembedaan/differencing satu kali dan kemudian dilakukan uji ADF. Output hasil uji Augmented Dickey-Fuller sebelum dilakukan *differencing* menunjukkan belum stasioner. Setelah dilakukan Differencing 1 hasil ini menunjukkan sudah stasioner seperti pada Tabel 15.

Tabel 17. Hasil uji Kestationeran untuk Harga Karet Dunia Setelah Differencing 1

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-0.67299 -0.12130  0.03059  0.29409  1.95838

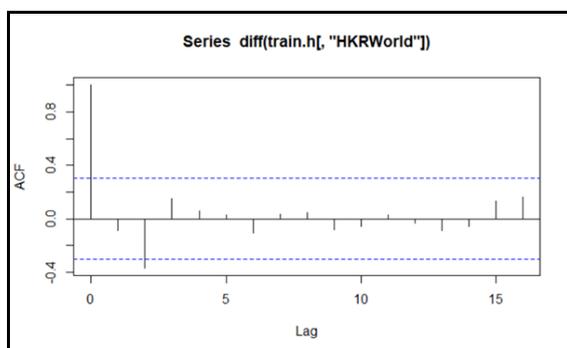
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1      -1.5588     0.2327  -6.698 6.33e-08 ***
z.diff.lag   0.6042     0.1976   3.059 0.00406 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4493 on 38 degrees of freedom
Multiple R-squared:  0.5618, Adjusted R-squared:  0.5388
F-statistic: 24.36 on 2 and 38 DF, p-value: 1.553e-07

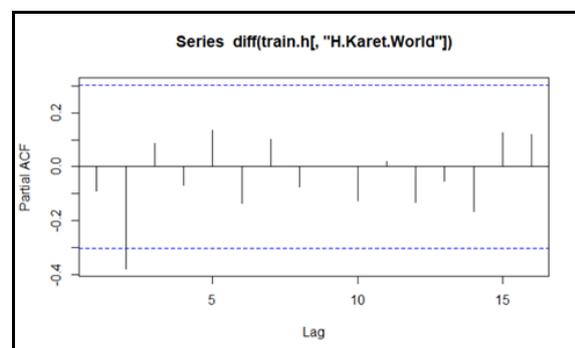
Value of test-statistic is: -6.6976
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Uji ADF pada data yang telah dilakukan *differencing* satu kali menunjukkan bahwa nilai *test-statistic* yaitu -6.6976 lebih kecil dari *critical values* menunjukkan bahwa  $H_0$  ditolak yang berarti data harga karet dunia telah stasioner.

Stasioneritas data dapat juga dilihat dari plot ACF dan PACF. Jika lag pada ACF maupun PACF menunjukkan *tail off* maka data tidak stasioner. Berikut adalah script untuk menunjukkan plot ACF dan PACF harga karet dunia setelah differencing 1.



Gambar 12. Plot ACF Harga Karet Dunia dengan *Differencing* satu kali



Gambar 13. Plot PACF Harga Karet Dunia dengan *Differencing* satu kali

Plot ACF cenderung cut off dimana hanya lag 0 yang keluar dari *confidence interval*. Plot PACF menunjukkan cut off pada lag 2, ditunjukkan berada di luar garis *confidence interval*. Hal tersebut memperkuat hasil uji ADF yaitu data harga karet dunia telah stasioner.

Selain itu, plot ACF dan PACF juga dapat digunakan untuk mengidentifikasi ordo AR dan MA, dengan cara melihat lag yang nyata atau keluar dari garis *confidence interval*. Namun demikian, pada plot ACF diatas, terlihat nyata pada lag=0 yang sering terjadi pada data riil, sehingga sulit untuk mengidentifikasi ordo AR dan MA melalui plot ACF dan PACF.

Pencarian model tentatif variabel input harga karet dunia dilakukan melalui penelusuran menggunakan model ARIMA. Model terbaik dapat dipilih menggunakan script *auto.arima* yang tersedia pada RStudio. Data yang digunakan untuk memilih model terbaik adalah series data *training*.

Tabel 18. Model Auto Arima Harga Karet Dunia

```
Series: train.h[, "H.karet.world"]
ARIMA(0,1,3)
Coefficients:
      ma1      ma2      ma3
s.e.  0.1518  0.1283  0.1665

sigma^2 estimated as 0.1864:  log likelihood=-23.24
AIC=54.47  AICC=55.55  BIC=61.42

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.08375659 0.4111303 0.269561 2.14057 20.07916 0.9016135 -0.01900243
```

Berdasarkan pemilihan orde ARIMA menggunakan *automodel* menyarankan bahwa model terbaik untuk harga karet dunia adalah ARIMA (0,1,3) dengan MAPE 20,07%. Model ARIMA (0,1,3) untuk data harga karet dunia tersebut menghasilkan nilai AIC = 54,47 sebagaimana output yang disajikan pada Tabel 18.

Tabel 19. Pengujian Koefisien Model Arima (0,1,3) Harga Karet Dunia

```
Call:
arima(x = train.h[, "H.karet.world"], order = c(0, 1, 3))
Coefficients:
      ma1      ma2      ma3
s.e.  0.1518  0.1283  0.1665

sigma^2 estimated as 0.1731:  log likelihood = -23.24,  aic = 54.47
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1 -0.03861    0.15179  -0.2544  0.79921
ma2 -0.38616    0.12828  -3.0103  0.00261 **
ma3  0.38526    0.16651   2.3136  0.02069 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Berdasarkan output di atas terlihat bahwa komponen MA1 tidak signifikan, MA2 signifikan pada taraf 1% dan MA3 signifikan pada taraf 5%. Selain menggunakan script *auto.arima* model tentatif dapat juga dipilih dengan *arima selectioan*. Pada Tabel 18, output yang dihasilkan untuk memilih model tentative terbaik untuk factor input  $X_t$  yaitu harga karet dunia.

Tabel 20. Model Arima Tentatif untuk Harga Karet Dunia

	p	q	sbc
[1,]	2	0	-64.13388
[2,]	3	0	-61.77537
[3,]	0	0	-61.72972
[4,]	4	0	-57.77817
[5,]	1	0	-57.54780
[6,]	5	0	-52.93032
[7,]	3	1	-47.56537
[8,]	1	3	-45.24218
[9,]	0	3	-44.32915
[10,]	2	1	-43.94674

Hasil output R-Studio akan menunjukkan sepuluh model tentatif dimana idealnya model terbaik adalah model yang memiliki nilai SBC terkecil. Model ARIMA yang direkomendasikan ditunjukkan dari nilai p,d,q. Sebagai contoh model pertama dengan nilai  $p=2$  dan  $q=0$ . Karena data harga karet dunia telah dilakukan *differencing* satu kali berarti  $d=1$ , artinya model yang direkomendasikan adalah ARIMA (2,1,0). Untuk mengetahui apakah model ARIMA (2,1,0) lebih baik dari hasil *auto.arima* yaitu ARIMA (0,1,3) maka dapat dibandingkan dari nilai AIC.

Tabel 21. Pengujian Koefisien Model Arima (2,1,0) Harga Karet Dunia

```
Call:
arima(x = train.h[, "H.karet.world"], order = c(2, 1, 0))

Coefficients:
      ar1      ar2
  0.0337  -0.5489
s.e.  0.1667   0.1730

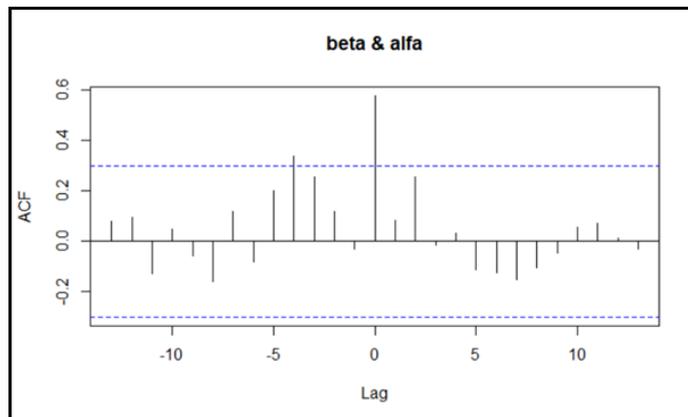
sigma^2 estimated as 0.1831:  log likelihood = -24.31, aic = 54.61

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.033724   0.166692  0.2023 0.839674
ar2 -0.548946   0.172999 -3.1731 0.001508 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model ARIMA (2,1,0) menghasilkan nilai AIC = 54,61 tidak berbeda jauh dari model ARIMA (0,1,3) dengan nilai AIC = 54,47. Pada model ARIMA (2,1,0) ditemukan bahwa AR1 tidak signifikan sedangkan AR2 signifikan pada taraf 1%. Oleh karena ARIMA (2,1,0) memiliki SBC yang paling kecil maka selanjutnya yang digunakan untuk peramalan harga karet dunia menggunakan ARIMA (2,1,0).

Tahap selanjutnya untuk penyusunan model fungsi transfer ini adalah *prewhitening* dan korelasi silang. Korelasi silang menggambarkan struktur hubungan antara  $X_t$  dengan  $Y_t$ . Untuk mengidentifikasi pengaruh  $X_t$  terhadap  $Y_t$  maka deret  $X_t$  harus stasioner atau sudah distasionerkan. Dalam konteks pemodelan  $X_t$  terhadap  $Y_t$ , untuk membuat  $X_t$  stasioner tidak dengan pembedaan (*differencing*) namun dengan mengambil komponen *white noise* dari  $X_t$  (*prewhitening*). *Prewhitening* dilakukan terhadap deret input  $X_t$  yang didefinisikan sebagai alfa serta deret input  $Y_t$  yang didefinisikan sebagai beta. Hasil output untuk *prewhitening* dan korelasi silang berupa grafik ACF untuk beta dan alfa.



Gambar 14. Plot korelasi silang Produksi Karet dengan Harga Karet Dunia

Hasil plot korelasi silang digunakan untuk mengidentifikasi ordo  $r, s$ , dan  $b$ . Ordo  $r$  adalah panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$ , ordo  $s$  adalah panjang lag  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$ , dan ordo  $b$  adalah panjang jeda pengaruh  $X_t$  terhadap  $Y_t$ . Identifikasi ordo  $r, s$  dan  $b$  hanya dilihat pada lag yang positif.

Plot korelasi silang diatas menunjukkan bahwa hanya lag 0 yang keluar dari garis signifikansi, maka nilai  $b=0$ . Kemudian, tidak ada tambahan lagi lag yang signifikan maka nilai  $s=0$ . Mengingat data produksi karet dan harga karet dunia merupakan data tahunan yang tidak mengandung musiman maka diasumsikan nilai  $r=0$ . Nilai  $b=0$  menunjukkan tidak ada jeda pengaruh antara harga karet dunia pada waktu  $t$  terhadap produksi karet pada waktu  $t$ . Nilai  $s=0$  berarti ada korelasi antara produksi dan harga karet dunia pada tahun yang sama. Dengan kata lain, dampak dari harga karet dunia terhadap produksi dirasakan pada waktu yang sama ( $t$ ).

Tahap selanjutnya dilakukan pengepasan model, untuk nilai  $r, s$  dan  $b$ . Hasil pengujian fungsi transfer dengan nilai  $r=0$ ,  $s=0$ , dan  $b=0$  menghasilkan nilai MAPE yang cukup besar yaitu 24,71%.

Tabel 22. Pengujian Orde r, s, dan b ARIMA (0,0,0) Produksi Karet

```
Series: train.h[, "Produksi_nas"]
Regression with ARIMA(0,0,0) errors

Coefficients:
      intercept          xreg
      730043.77    615109.08
s.e.    99838.49    65638.68

sigma^2 estimated as 1.518e+11:  log likelihood=-613.53
AIC=1233.05  AICC=1233.67  BIC=1238.33

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -4.570655e-07 380444.2 333676.2 -7.693225 24.70841 3.891259 0.7916605
```

Untuk menghasilkan orde yang paling tepat untuk menentukan orde Arima fungsi transfer dengan melakukan identifikasi model noise. Untuk menghasilkan model terbaik dengan menggunakan *auto-arima* pada R Studio, model maka noise yang disarankan adalah Arima (0,1,0). Model ini ternyata masih kurang tepat, karena menghasilkan MAPE yang cukup besar yaitu 94,68%.

Tabel 23. Model Residual Tentatif ARIMA (0,1,0) Produksi Karet

```
Series: res
ARIMA(0,1,0)

sigma^2 estimated as 5.822e+10:  log likelihood=-580.13
AIC=1162.27  AICC=1162.37  BIC=1164

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 12069.65 238468.6 141687.9 20.70138 94.68429 0.9767728 -0.04015689
```

Tabel 24. Model Residual Tentatif ARIMA Untuk Produksi Karet

```
p q      sbc
[1,] 2 0 1038.811
[2,] 0 0 1041.977
[3,] 3 0 1042.194
[4,] 4 0 1042.836
[5,] 1 0 1046.627
[6,] 5 0 1047.641
[7,] 0 2 1055.458
[8,] 4 1 1059.043
[9,] 2 1 1059.123
[10,] 1 2 1059.191
```

Oleh karena model *autoarima* disarankan differencing tingkat 1, maka solusinya akan dicari model alternative. Model alternative yang diberikan untuk *model noise* seperti Tabel 18. Model tentatif pertama yang disarankan untuk data residual adalah ARIMA (2,1,0). Selanjutnya model tersebut didefinisikan sebagai *modelres* dan dilihat signifikansi AR. Model noise untuk residual dengan Arima (2,1,0) menghasilkan komponene ar1 yang tidak signifikan, tetapi ar2 signifikan pada tingkat kepercayaan 99%.

Tabel 25. Model Residual Tentatif ARIMA (2,1,0) Untuk Produksi Karet

```
Series: res
ARIMA(2,1,0)

Coefficients:
      ar1      ar2
      0.1131 -0.6778
s.e.  0.1848  0.1733

sigma^2 estimated as 4.547e+10:  log likelihood=-574.54
AIC=1155.07  AICC=1155.7  BIC=1160.29
> coeftest(modelres)

z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1  0.11313    0.18477  0.6123   0.5404
ar2 -0.67779    0.17329 -3.9113  9.181e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tahap berikutnya adalah dilakukan pengepasan model dengan noise. Pengepasan model dengan noise dilakukan dengan menggabungkan nilai r,s,b dan hasil identifikasi model noise yaitu ARIMA(2,1,0). Selanjutnya diperoleh output model fungsi transfer ARIMA (2,1,0) (Tabel 19).

Tabel 26. Model Fungsi Transfer ARIMA (2,1,0) Untuk Produksi Karet

```
Series: train.h[, "Produksi"]
Regression with ARIMA(2,1,0) errors

Coefficients:
      ar1      ar2      xreg
      0.5465 -0.0629 152566.47
s.e.  0.1704  0.1728 26510.46

sigma^2 estimated as 8.206e+09:  log likelihood=-537.59
AIC=1083.17  AICC=1084.25  BIC=1090.12

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 24543.2 86268.56 62607.59 1.371311 3.798703 0.7301161 -0.06647602
```

Model fungsi transfer produksi karet dengan ARIMA(2,1,0) menghasilkan MAPE 3,79%. Untuk mengetahui *performance* dari model fungsi transfer tersebut dilakukan pengujian *coefstest* untuk mengetahui tingkat signifikansi dari masing- masing koefisien, seperti terlihat pada Tabel 27.

Tabel 27. Pengujian Koefisien Model Fungsi Transfer ARIMA (2,1,0) Produksi Karet

```
z test of coefficients:

      Estimate Std. Error z value  Pr(>|z|)
ar1  5.4649e-01 1.7041e-01 3.2069 0.001342 **
ar2 -6.2903e-02 1.7277e-01 -0.3641 0.715802
xreg 1.5257e+05 2.6510e+04 5.7550 8.667e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Berdasarkan hasil uji Z untuk koefisien AR ditemukan bahwa AR1 dan komponen fungsi transfer xreg yaitu variabel harga karet dunia signifikan. Namun komponen AR2 tidak signifikan. Oleh karena itu akan dicobakan model fungsi transfer dengan model tentatif komponen noise ARIMA (1,1,0). Hasil output untuk fungsi transfer dengan model noise ARIMA adalah sebagai berikut:

Tabel 28. Model Tentatif Fungsi Transfer ARIMA (1,1,0) Untuk Produksi Karet

```
Series: train.h[, "Produksi"]
Regression with ARIMA(1,1,0) errors

Coefficients:
      ar1      xreg
    0.5149 150167.10
s.e. 0.1459 25338.84

sigma^2 estimated as 8.027e+09: log likelihood=-537.65
AIC=1081.31 AICC=1081.94 BIC=1086.52

Training set error measures:
Training set  ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
23588.77 86413.09 62916.21 1.316241 3.803698 0.7337151 -0.02274797
```

MAPE yang dihasilkan dari fungsi transfer ARIMA (1,1,0) adalah 3,80% tidak berbeda jauh dari MAPE fungsi transfer ARIMA(2,1,0). Selanjutnya dilakukan pengujian *coefstest* untuk fungsi transfer ARIMA(1,1,0) untuk melihat signifikansi dari komponene AR.

Tabel 29. Pengujian Koefisien Fungsi Transfer ARIMA (1,1,0) Untuk Produksi Karet

```
z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1  5.1489e-01 1.4586e-01  3.5301 0.0004154 ***
xreg 1.5017e+05 2.5339e+04  5.9264 3.097e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dari hasil pengujian di atas, ditemukan bahwa komponen AR1 signifikan dan komponen fungsi transfer xreg juga signifikan. Berdasarkan hasil penelusuran model dan uji yang telah dilakukan maka disimpulkan bahwa model fungsi transfer produksi karet dengan *noise* ARIMA (1,1,0) lebih baik dibandingkan model ARIMA fungsi transfer (2,1,0).

### Peramalan berbasis fungsi transfer

Berdasarkan model fungsi transfer dengan model noise ARIMA (1, 1, 0), dilakukan peramalan berbasis nilai aktual dimana produksi karet diestimasi menggunakan data aktual harga karet dunia periode 2014-2019. Meskipun data aktual produksi karet periode 2014-2019 telah ada, dilakukan peramalan produksi karet untuk mengecek *performance* model fungsi transfer.

Tabel 30. Pengujian MAPE Model Fungsi Transfer ARIMA (1,1,0) Dengan Faktor Input Harga Karet Dunia Aktual tahun 2014 - 2019

```
Series: test.h[, "Produksi"]
Regression with ARIMA(1,1,0) errors

Coefficients:
      ar1      xreg
 0.5149 150167.1
s.e. 0.0000      0.0

sigma^2 estimated as 8.027e+09: log likelihood=-66.24
AIC=134.47  AICC=135.8  BIC=134.08

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 44886.52 121324.5 94586.18 1.336987 2.705989 0.6984282 -0.1100538
```

Peramalan produksi karet periode 2014-2019 menggunakan fungsi transfer ARIMA (1,1,0) dengan input harga karet dunia **nilai aktual** menghasilkan MAPE 2,70%.

Tujuan melakukan pemodelan fungsi transfer adalah untuk mendapatkan nilai ramalan periode ke depan, yakni produksi karet tahun 2020-2024. Karena data series input harga karet dunia tersedia hingga tahun 2019, maka perlu dilakukan peramalan harga karet dunia terlebih dahulu atau dengan kata lain peramalan produksi dilakukan berbasis nilai ramalan harga karet.

Oleh karenanya, terlebih dahulu dilakukan estimasi harga karet dunia periode 2020-2024 menggunakan model ARIMA (2,1,0) sebagaimana yang telah diperoleh dari tahap pencarian model tentatif untuk variabel input, sebagai variabel input harga karet dunia. Pemilihan variabel input harga karet dunia karena sangat berpengaruh pada harga karet nasional, dan secara tidak langsung berpengaruh juga pada produksi karet nasional. Selanjutnya dilakukan peramalan produksi karet dengan fungsi transfer ARIMA (1, 1, 0) sebagai model terbaik berdasarkan tahapan pengepasan model dengan noise. Peramalan produksi dengan fungsi transfer ARIMA (1,1,0) menggunakan **nilai ramalan** harga karet dunia yang telah diestimasi dengan ARIMA (2,1,0).

Tabel 31. Pengujian MAPE Model Fungsi Transfer ARIMA (1,1,0) Dengan Faktor Input Harga Karet Dunia Ramalan tahun 2014 - 2019

```
Series: test.h[, "Produksi"]
Regression with ARIMA(1,1,0) errors

Coefficients:
      ar1      xreg
 0.5149 150167.1
s.e. 0.0000      0.0

sigma^2 estimated as 8.027e+09: log likelihood=-68.92
AIC=139.85  AICC=141.18  BIC=139.46

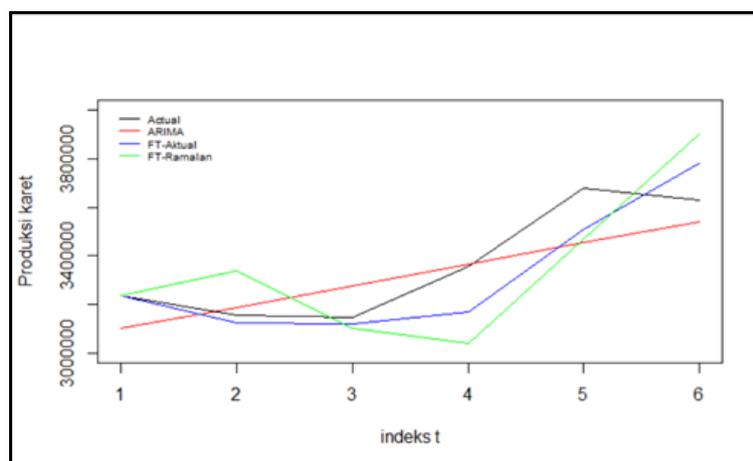
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 19691.33 207698.6 172974.7 0.5500169 5.032109 1.277252 0.0276403
```

Estimasi produksi karet berbasis fungsi transfer dengan model noise ARIMA (1,1,0) selama 6 tahun terakhir (2014-2019) menggunakan input harga karet dunia hasil **angka ramalan** ARIMA (2,1,0) menghasilkan MAPE 5,03%.

Setelah dilakukan peramalan produksi karet baik menggunakan input (harga karet dunia) data aktual maupun ramalan, tahapan berikutnya adalah pengepasan model arima output. Pengepasan model ARIMA output dimaksudkan untuk membandingkan hasil ramalan produksi baik berdasarkan data training (1970-2013) maupun data testing (2014-2019). Model ARIMA biasa, tanpa fungsi transfer, terbaik untuk variabel produksi yang dipilih adalah ARIMA (0,2,1). Model ARIMA (0,2,1) tersebut merupakan ARIMA biasa dan bukan hasil dari fungsi transfer. Hasil ramalan model ARIMA tanpa fungsi transfer dan model ARIMA dengan fungsi transfer selanjutnya akan dibandingkan.

Hasil peramalan produksi karet dengan ARIMA biasa atau ARIMA tanpa fungsi transfer menunjukkan bahwa hasil ramalan jika menggunakan data training menghasilkan MAPE = 4,31%. Di sisi lain, MAPE yang dihasilkan dari data testing yaitu 6,34%. Artinya jika model yang digunakan hanya ARIMA tanpa fungsi transfer maka MAPE data testing lebih besar dibandingkan data training.

Untuk membandingkan ketepatan model estimasi, dilakukan perbandingan hasil estimasi terhadap data aktual produksi karet pada tahun 2014-2019 (data testing). Hasil ramalan yang dibandingkan yaitu ramalan dengan model ARIMA(0,2,1), model fungsi transfer ARIMA(1,1,0) dimana input harga karet dunia yang digunakan adalah data aktual maupun ramalan. Selanjutnya hasil ramalan dibandingkan dalam sebuah grafik seperti Gambar 15.



Gambar 15. Perbandingan Hasil Ramalan Produksi Karet Tahun 2014-2019

Dari grafik di atas terlihat jika dibandingkan dengan data aktual produksi karet 2014-2019 (warna hitam), maka hasil peramalan dengan model ARIMA tanpa fungsi transfer (warna merah) tidak mengikuti pola data aktualnya. Di sisi lain, peramalan dengan fungsi transfer khususnya jika input harga karet dunia yang digunakan adalah data aktual maka hasil

ramalan produksinya (warna biru) sangat menyerupai pola data produksi aktual selama 6 tahun terakhir. Jika input harga karet dunia yang digunakan adalah hasil ramalan, maka estimasi produksinya (warna hijau) hampir menyerupai pola data asli, meskipun tidak terlalu berimpit seperti estimasi menggunakan data aktual.

Hasil identifikasi model untuk meramalkan produksi karet dengan ke-3 cara diatas menghasilkan nilai MAPE masing-masing sebagai berikut:

Tabel 32. Perbandingan MAPE untuk Arima biasa dan Arima Fungsi Transfer

No	Model	MAPE (%)
1	ARIMA biasa (tanpa Fungsi transfer)	4,31 (data training) dan 6,34 (data testing)
2	Fungsi Transfer dengan ARIMA input data aktual	2,70
3	Fungsi Transfer dengan ARIMA input data ramalan	5,30

Berdasarkan nilai MAPE yang dihasilkan maka model fungsi transfer ARIMA dengan input data aktual lebih bagus jika dibandingkan dengan model fungsi transfer ARIMA dengan input data ramalan maupun model ARIMA tanpa fungsi transfer. Namun demikian, karena tujuan penyusunan model adalah untuk mendapatkan angka ramalan tahun 2020 -2024, maka digunakan model Fungsi Transfer berbasis input data ramalan, yakni dengan terlebih dahulu dilakukan estimasi harga karet dunia lima tahun ke depan menggunakan model ARIMA (2,1,0). Hasil estimasi harga karet dunia tersebut kemudian menjadi input untuk meramalkan produksi karet lima tahun ke depan menggunakan model fungsi transfer ARIMA (1,1,0) dan sebagai variabel input harga karet dunia. Hasil ramalan produksi karet disajikan pada Tabel 26 dan Tabel 27.

Selain mencari model terbaik untuk meramalkan produksi karet, akan diestimasi juga produksi karet lima tahun ke depan (2020-2024) menggunakan fungsi transfer ARIMA (1,1,0) dengan menggunakan seluruh data. Berikut adalah output hasil ramalan lima tahun ke depan:

Tabel 33. Hasil Peramalan Model Fungsi Transfer Arima (1,1,0) Tahun 2020 - 2024

```
Time Series:
Start = 50
End = 54
Frequency = 1
[1] 3611873 3621989 3628211 3626798 3625179
```

Tabel 34. Hasil Peramalan Produksi Karet Tahun 2020-2024

Tahun	Produksi Karet (Ton)	Pertumbuhan (%)
2019	3.448.782	
2020	3.611.873	4,729
2021	3.621.989	0,280
2022	3.628.211	0,172
2023	3.626.798	-0,039
2024	3.625.179	-0,045
Rata-rata pertumbuhan (%)		0,092

Tahun 2019 : Angka Sementara Ditjenbun

Tahun 2020 – 2024 : Angka Estimasi Berdasarkan Model Fungsi Transfer

Pada tahun 2019 Angka Sementara untuk produksi karet sebesar 3.448.782 ton (Ditjen Perkebunan, 2019). Hasil peramalan dengan fungsi transfer ARIMA (1,1,0) menunjukkan bahwa selama lima tahun ke depan produksi karet akan meningkat 0,09%. Tahun 2020 produksi karet diestimasi sebesar 3.611.873 ton. Hasil estimasi tersebut lebih tinggi dibandingkan Angka Estimasi Ditjenbun untuk produksi karet tahun 2020 adalah sebesar 3.545.693 ton. Produksi karet diramalkan terus meningkat hingga tahun 2022, kemudian mengalami penurunan mulai tahun 2023 hingga pada tahun 2024 produksinya diramalkan sebesar 3.625.179 ton. Pertumbuhan produksi karet tahun 2020 sampai 2024 rata-rata meningkat 0,092% per tahun.

#### D. Model VAR (*Vector Auto Regressive*)

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen. Model VAR berlaku pada saat nilai setiap variabel dalam sebuah system tidak hanya bergantung pada lag-nya sendiri, namun juga pada nilai lag variabel lain.

Tahapan yang dilakukan dalam penyusunan model VAR adalah sebagai berikut: persiapan data, pembagian data training dan testing, pemilihan lag dan type, pengajuan asumsi, ramalan data training, testing, penghitungan mape, dan plot, pemilihan model terbaik, dan pengepasan model untuk seluruh data dan peramalannya. Disamping itu dibahas Interpretasi Impulse Response Function (IRF) dan Variance Decomposition.

Variabel yang digunakan untuk estimasi model VAR adalah produksi karet dalam satuan Ton, luas areal (areal) dalam satuan Hektar, harga karet sheet (Hsheet) dalam satuan Rp/kg, harga karet dunia (H.Karetworld) dalam satuan US\$/kg, volume ekspor karet (volekkaret) dalam satuan Ton, dan volume impor (volimkaret) dalam satuan Ton. Data produksi karet, luas areal, volume ekspor karet, volume impor karet, dan harga karet sheet diperoleh dari publikasi Ditjen Perkebunan, sementara data variabel harga karet dunia diperoleh dari World Bank. Series masing-masing variabel berbeda karena

keterbatasan ketersediaan data. Series data produksi karet adalah dari tahun 1967-2019, series luas areal adalah dari tahun 1970-2019, series data harga sheet adalah dari tahun 1971-2019, series data harga karet dunia adalah dari tahun 1970-2019, sementara series data volume ekspor dan impor karet adalah dari tahun 1970-2019. Format data yang digunakan bisa dalam bentuk excell, CSV, ataupun SPSS.

### **Pembagian Data Training dan Testing**

Series data yang digunakan adalah series tahun 1971 – 2019 akan dibagi menjadi 2 set data yakni set data training (tahun 1971-2013) atau 43 titik dan set data testing (2014-2019) atau 6 titik.

### **Pemilihan Lag (p) dan Type**

Dalam permodelan VAR karet ini digunakan enam variabel, yaitu produksi karet (produksi), luas areal (areal), harga sheet (HRsheet), harga karet dunia (HKaretworld), volume ekspor karet (volekkaret), dan volume impor (volimkaret). Selain komposisi variabel tersebut, komponen konstanta dan trend juga perlu dilakukan pengujian untuk mengetahui signifikan atau tidak keberadaannya. Jika keduanya signifikan, maka komponen tersebut harus dimasukkan ke dalam model VAR dengan type “both”. Jika hanya konstanta yang signifikan, maka trend perlu dikeluarkan dari model VAR dengan model VAR type “const”. Jika hanya trend yang signifikan maka konstanta dikeluarkan dari model menggunakan model VAR type “trend”, dan jika keduanya tidak signifikan, maka type yang digunakan model VAR adalah “none”.

Keberadaan konstanta dan trend dapat dideteksi dari plot data awal, namun terkadang hal tersebut sulit dilakukan. Untuk mengantisipasi hal tersebut, sebaiknya dilakukan uji *coba/trial and error* menggunakan model VAR dari lag  $p=1$  s.d 6 dengan type “both” untuk mengetahui signifikan atau tidak keberadaannya. Untuk panjang lag maksimum bisa dilakukan trial error sampai tidak memungkinkan untuk dilakukan permodelan.

Untuk data karet ini, setelah dilakukan running model VAR dengan lag  $p=1$  type “both”, diperoleh informasi komponen konstanta dan trend signifikan, sehingga pada pengujian berikutnya dilakukan running model VAR dengan menggunakan lag  $(p) = 1$  s.d lag  $(p)=4$  dengan menyertakan komponen konstanta dan trend. Pemilihan lag  $p$  ditentukan dengan melihat banyaknya variabel yang signifikan dalam lag tersebut sekaligus memastikan harus ada variabel/peubah yang signifikan pada lag terpilih dimaksud. Setelah dilakukan run model ternyata ada 2 kandidat model VAR terbaik yaitu Var (1) type “both” dan Var (2) type “both”. Type “both” artinya persamaan Var itu mengandung constanta dan faktor trend.

Untuk model VAR yang pertama adalah Model VAR (p=1) dengan konstanta dan trend (type=both). Hasil model Var(1) type both, untuk mengestimasi Produksi (t) karet menunjukkan bahwa koefisien constanta dan trend menunjukkan hasil yang signifikan pada tingkat kepercayaan 95%. Untuk variabel lain yang signifikan untuk mengestimasi produksi (t) antara lain Produksi, luas areal, dan harga karet sheet pada lag-1. Ketiga variabel tersebut signifikan dengan tingkat kepercayaan 95%.

Model Var (1) type both ini menghasilkan nilai Adjusted R Square = 97,95%, artinya keragaman produksi dipengaruhi oleh variabel variabel penjelasnya sebesar 97,95%. Nilai p-value untuk model produksi ini sangat kecil atau jauh lebih kecil dari 0,05, sehingga model layak untuk digunakan. Jumlah variabel yang signifikan ada 3 variabel dari total 6 variabel. banyak sehingga model ini layak digunakan dan menjadi kandidat terbaik untuk model VAR.

Tabel 35. Output Model VAR(1) type=both

```

VAR Estimation Results:
=====
Endogenous variables: Produksi, Areal, HRSheet, HKaretworld, volekkaret, volimkaret
Deterministic variables: both
Sample size: 41
Log Likelihood: -2287.319
Roots of the characteristic polynomial:
0.9919 0.677 0.677 0.6147 0.1906 0.05295
Call:
VAR(y = karet[6:47, c(2, 3, 7, 8, 11, 12)], p = 1, type = "both")

Estimation results for equation Produksi:
=====
Produksi = Produksi.l1 + Areal.l1 + HRSheet.l1 + HKaretworld.l1 + volekkaret.l1 +
volimkaret.l1 + const + trend

      Estimate Std. Error t value Pr(>|t|)
Produksi.l1  1.082e+00  2.482e-01  4.358  0.00012 ***
Areal.l1     -3.117e-01  9.740e-02 -3.200  0.00303 **
HRSheet.l1   2.897e+01  1.258e+01  2.303  0.02770 *
HKaretworld.l1 -1.573e+04  3.826e+04 -0.411  0.68365
Volekkaret.l1 -3.529e-01  2.939e-01 -1.201  0.23841
Volimkaret.l1 -7.640e-01  3.110e+00 -0.246  0.80749
const        7.326e+05  2.395e+05  3.059  0.00438 **
trend        2.584e+04  7.518e+03  3.437  0.00161 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 101600 on 33 degrees of freedom
Multiple R-Squared:  0.9831, Adjusted R-squared:  0.9795
F-statistic: 274.4 on 7 and 33 DF, p-value: < 2.2e-16

```

Tabel 36. Output Model VAR(2) type=both

VAR Estimation Results:

```

=====
Endogenous variables: Produksi, Areal, HRSheet, HKaretworld, volekkaret, volimkaret
Deterministic variables: both
Sample size: 40
Log Likelihood: -2173.99
Roots of the characteristic polynomial:
0.9639 0.7709 0.7709 0.7148 0.7148 0.6591 0.6591 0.6507 0.6507 0.4151 0.4151 0.1585
Call:
VAR(y = karet[6:47, c(2, 3, 7, 8, 11, 12)], p = 2, type = "both")

Estimation results for equation Produksi:
=====
Produksi = Produksi.l1 + Areal.l1 + HRSheet.l1 + HKaretworld.l1 + volekkaret.l1 +
volimkaret.l1 + Produksi.l2 + Areal.l2 + HRSheet.l2 + HKaretworld.l2 + volekkaret.l2
+ volimkaret.l2 + const + trend

```

	Estimate	Std. Error	t value	Pr(> t )
Produksi.11	1.254e+00	3.857e-01	3.251	0.003171 **
Area1.11	-7.218e-01	3.736e-01	-1.932	0.064338 .
HRSheet.11	2.399e+01	1.241e+01	1.933	0.064204 .
HKaretworld.11	-9.209e+04	5.389e+04	-1.709	0.099404 .
volekkaret.11	-2.164e-01	3.193e-01	-0.678	0.503915 .
Volimkaret.11	5.876e-01	4.194e+00	0.140	0.889663 .
Produksi.12	7.398e-02	3.363e-01	0.220	0.827613 .
Area1.12	3.482e-01	3.915e-01	0.889	0.381935 .
HRSheet.12	-5.872e-01	1.325e+01	-0.044	0.964990 .
HKaretworld.12	8.398e+04	5.131e+04	1.637	0.113744 .
volekkaret.12	-5.580e-01	3.280e-01	-1.701	0.100830 .
Volimkaret.12	-6.981e+00	4.066e+00	-1.717	0.097900 .
const	9.444e+05	2.648e+05	3.566	0.001432 **
trend	3.504e+04	9.338e+03	3.752	0.000889 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96620 on 26 degrees of freedom  
Multiple R-Squared: 0.9876, Adjusted R-squared: 0.9814  
F-statistic: 159.6 on 13 and 26 DF, p-value: < 2.2e-16

Hasil model Var(2) type both, untuk mengestimasi Produksi (t) karet menunjukkan bahwa koefisien konstanta dan trend menunjukkan hasil yang signifikan pada tingkat kepercayaan 99%. Untuk variabel lain yang signifikan untuk mengestimasi produksi (t) antara lain Produksi (t-1), luas areal (t-1), harga karet sheet (t-1), dan harga karet dunia (t-1), serta volume impor karet (t-1) dengan tingkat kepercayaan 90% - 95%. Jumlah variabel yang signifikan sebanyak 5 variabel dari total 12 variabel. Model VAR(2) type=both ini juga layak perlu diuji MAPE untuk melihat apakah lebih baik dari model VAR(1) type=both.

Selanjutnya dilakukan pengujian model VAR, dengan nilai p=3, p=4 type both, dan model VAR p=3, p=4 type konstan. Hasil pengujian untuk Var(3) type 'both' menghasilkan koefisien konstanta dan trend signifikan, serta ada 6 variabel yang signifikan. Jika dibandingkan dengan p=2, untuk p=3 penambahan variabel yang signifikan relatif kecil. Selanjutnya dilakukan pengujian untuk nilai p=4, ternyata konstanta dan trend sudah tidak signifikan, sehingga berhenti pada p=4. Selanjutnya dilakukan pengujian untuk nilai p=3 dan p=4 tetapi type 'constant', menunjukkan hasil koefisien konstanta tidak signifikan, sehingga berhenti untuk dilakukan pengujian lebih lanjut. Berdasarkan pengujian ini yang terbaik adalah untuk nilai p=1 atau p=2 dengan type 'both'.

Tabel 37. Hasil Pengujian Model VAR pada Beberapa Tingkat p dan Type

Lag (p)	Type	Sig. Type	Var. Syg	Jumlah Var	Adj - R <sup>2</sup>
p=1	both	Const dan Trend Syg	3	6	0,9795
p=2	both	Const dan Trend Syg	5	12	0,9814
p=3	both	Const dan Trend Syg	6	18	0,9800
p=4	both	Const dan Trend tidak Syg	6	24	0,9908
p=3	constant	Const tidak Syg	5	18	0,9756
p=4	constant	Const tidak Syg	6	24	0,9905

VAR dengan lag  $p$  untuk penulisan ringkasnya dapat dituliskan VAR ( $p$ ). Uji coba running model VAR dapat dilakukan sampai lag  $p$  di mana pada lag tersebut jumlah variabel yang signifikan sudah berkurang dibandingkan dengan keseluruhan variabel yang digunakan dalam model atau bahkan sudah tidak ada lagi variabel yang signifikan. Model juga tidak perlu terlalu kompleks karena pada akhirnya dapat menyebabkan overfitting. Pada contoh model VAR produksi karet ini, uji coba model VAR dilakukan sampai lag  $p=4$  type “both” karena pada lag tersebut banyaknya variabel yang signifikan sudah berkurang termasuk tidak signifikannya konstanta dan trend. Dua model terpilih sebagai kandidat model terbaik adalah model VAR (1) dan VAR (2) dengan type “both” karena banyaknya variabel yang signifikan dibandingkan dengan keseluruhan variabel yang digunakan dalam model tersebut. Sekaligus model tersebut tidak terlalu kompleks jika nantinya digunakan dalam permodelan ramalan. Dilihat dari Adjusted R-squared, keempat model di atas sudah memenuhi syarat karena nilainya cukup besar di atas 98% artinya kemampuan variabel bebas menjelaskan produksi karet di atas 98%.

### Pengujian Asumsi

Selanjutnya dilakukan pengujian asumsi non autokorelasi, normalitas, dan homoskedastisitas pada sisaan model VAR kandidat terbaik. Untuk data karet akan dilakukan pengujian sisaan pada dua model kandidat terbaik VAR (1) dan VAR (2) type “both”.

### Pengujian Asumsi VAR (1) type ‘both’.

Pemeriksaan autokorelasi residual model menggunakan fungsi “serial.test” yang di dalamnya dilakukan pengujian *Portmanteau-and Breusch-Godfrey test*. Nilai p-value lebih besar dari 0,05 maka sisaan saling bebas atau asumsi non autokorelasi terpenuhi. Pengujian Jarque-Bera tests untuk menguji kenormalan, hasil pengujian menunjukkan Nilai p-value lebih kecil dari 0,05, namun karena jumlah data yang digunakan cukup banyak, maka series tersebut dapat dianggap normal. Pemeriksaan heteroskedastisitas model menggunakan fungsi “arch.test” yang di dalamnya dilakukan pengujian *ARCH-LM tests*. Nilai p-value lebih besar dari 0,05 maka ragam sisaan homogen atau asumsi homoskedastisitas terpenuhi.

Tabel 38. Ouput Pengujian Asumsi VAR(1) type=both

```
Portmanteau Test (asymptotic)
data: Residuals of VAR object varhsheet.b1
Chi-squared = 527.32, df = 540, p-value = 0.6438

$JB
JB-Test (multivariate)
data: Residuals of VAR object varhsheet.b1
Chi-squared = 347.12, df = 12, p-value < 2.2e-16
$Skewness
Skewness only (multivariate)
data: Residuals of VAR object varhsheet.b1
Chi-squared = 80.695, df = 6, p-value = 2.554e-15
$Kurtosis
kurtosis only (multivariate)
```

```

data: Residuals of VAR object varhsheet.b1
Chi-squared = 266.42, df = 6, p-value < 2.2e-16

ARCH (multivariate)
data: Residuals of VAR object varhsheet.b1
Chi-squared = 756, df = 2205, p-value = 1

```

### **Pengujian Asumsi VAR (2) type “both”**

Kesimpulan yang sama dengan sisaan model VAR (1) type “both” diperoleh untuk sisaan model VAR dengan lag 2 type “both”. Jika sekiranya masih ada model lain yang cukup baik jika dilihat dari signifikansi maupun adjusted R-squared, maka model tersebut dapat menjadi kandidat model terbaik sehingga perlu dilakukan pengujian asumsi untuk melihat layak atau tidaknya model tersebut digunakan dalam peramalan. Jika hasil pengujian asumsi kandidat-kandidat model terbaik tersebut layak, maka dapat dilakukan peramalan, pengeplotan, dan penghitungan MAPE agar dapat dibandingkan hasilnya. Dari ketiga asumsi non autokorelasi, normalitas, dan homoskedastisitas, model yang layak harus memenuhi asumsi non autokorelasi dan homoskedastisitas.

Tabel 39. Output Pengujian VAR(2) type=both

```

Portmanteau Test (asymptotic)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 494.3, df = 504, p-value = 0.6127

$JB
JB-Test (multivariate)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 98.237, df = 12, p-value = 1.221e-15
$Skewness
Skewness only (multivariate)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 26.928, df = 6, p-value = 0.0001494
$Kurtosis
Kurtosis only (multivariate)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 71.31, df = 6, p-value = 2.202e-13

ARCH (multivariate)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 714, df = 2646, p-value = 1

```

### **Ramalan Data Training, Testing, Penghitungan MAPE, dan Plot**

Selanjutnya dilakukan peramalan data, baik untuk data training maupun untuk data testing sekaligus dilakukan penghitungan MAPE. MAPE (Mean Absolute Percentage Error) merupakan rata-rata dari keseluruhan persentase kesalahan (selisih) antara data aktual dengan data hasil peramalan.

### **Model VAR (1) type “both”**

Untuk menguji suatu model lebih baik dengan model yang lain, maka dilakukan pengujian model dengan membandingkan Nilai MAPE baik untuk data training maupun data testing. Data Testing hasil ramalan produksi dengan nilai  $p=1$  type=both menghasilkan MAPE =5,82%. Nilai ini dapat diartikan bahwa rata-rata seluruh

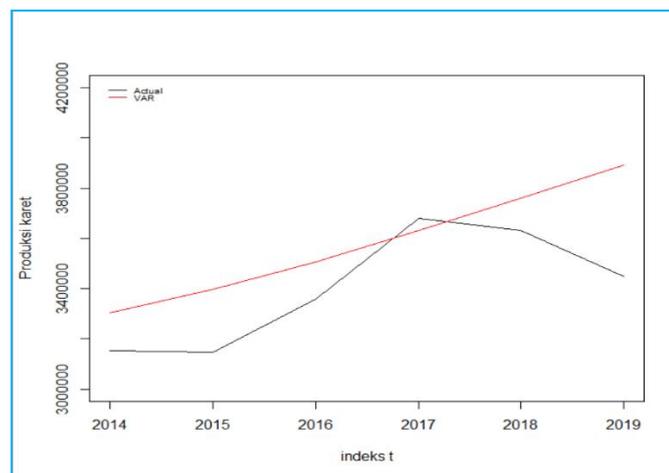
persentase kesalahan antara data aktual dengan data hasil ramalan pada data testing adalah 5,82%.

Tabel 40. Pengujian Nilai MAPE untuk Model VAR(1) type=both

Data Testing					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.354	3.778	4.621	5.821	7.171	12.810
Data Training					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.04291	1.22067	3.06778	4.16305	5.32296	16.76369

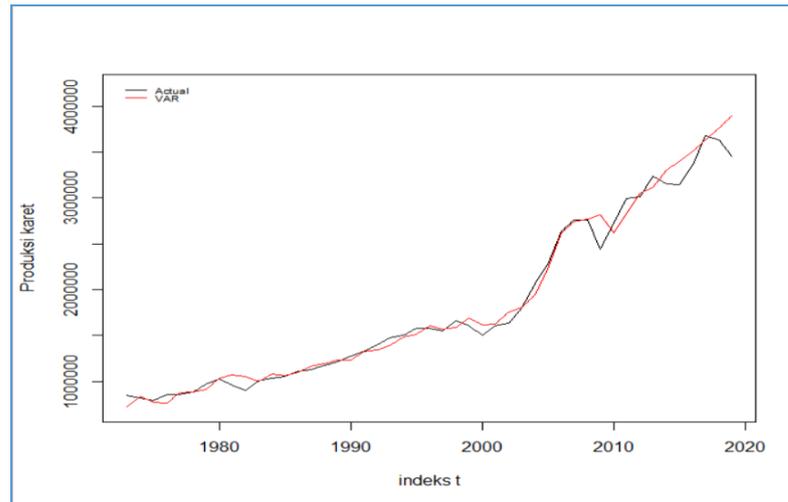
Data Training Hasil Ramalan Produksi dengan nilai  $p=1$  type=both menghasilkan MAPE =4,16%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data training adalah 4,16%

Dari grafik di bawah dapat dilihat bahwa pergerakan ramalan pada data testing tidak mengikuti pergerakan data asli/aktual. Sehingga mungkin model VAR (1) type “both” bukan model yang terbaik. Hasil peramalan dengan model model VAR (1) ini cenderung lebih tinggi dari data actual, kecuali di tahun 2017 data actual lebih tinggi dari data estimasi dengan VAR (1). Untuk data testing ini rata-rata penyimpangannya adalah sebesar 5,82%.



Gambar 16. Plot Ramalan dan Aktual Data Testing VAR (1) Type “Both”

Jika plot antara data testing dan data training digabungkan maka bentuk plotnya seperti Gambar 16. Untuk data tahun 1970 - 2014 atau data training plot sangat baik, karena antara data aktual dan estimasi dengan model VAR (1) hampir selalu berimpit plotnya, sehingga MAPE akan kecil. MAPE hasil pengujian untuk data training adalah sebesar 4,16%, suatu nilai yang cukup kecil karena di bawah 5%, artinya model cukup akurat. Namun plot tahun 2014 – 2019 menunjukkan data aktual dengan data estimasi dengan Model VAR (1) ini mulai agak renggang, dimana hasil estimasi rata-rata lebih tinggi dari data aktual. Hasil MAPE data testing ini, menunjukkan nilai yang lebih besar dari MAPE training yaitu sebesar 5,82%.



Gambar 17. Plot Ramalan dan Aktual Data Training dan Testing VAR (1) Type “Both”

### Model VAR (2) type “both” p=2

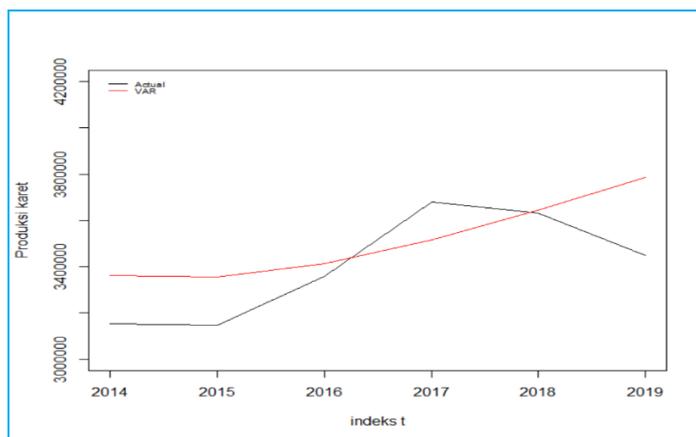
Untuk menguji model VAR(2) lebih baik dengan model yang lain, maka dilakukan pengujian model dengan membandingkan Nilai MAPE baik untuk data training maupun data testing. Data Testing hasil ramalan produksi dengan nilai  $p=2$  type=both menghasilkan MAPE =4,92%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data testing adalah 4,92%.

Tabel 41. Pengujian Nilai MAPE untuk Model VAR(2) type=both

MAPE data Testing					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.3856	2.3173	5.5380	4.9244	6.6549	9.8142
MAPE data Training					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1704	1.1758	2.4612	3.9013	4.7393	14.3177

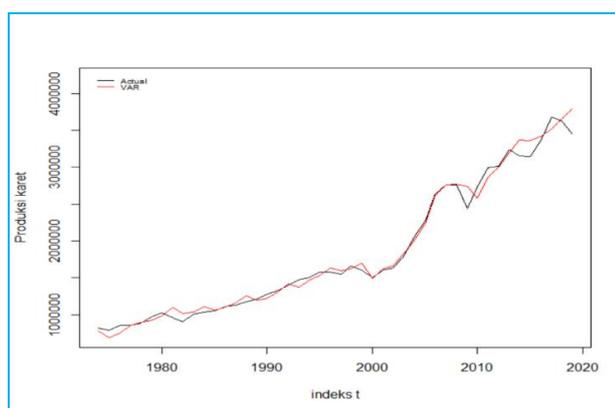
Data Training Hasil Ramalan Produksi dengan nilai  $p=2$  type=both menghasilkan MAPE =3,90%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data training adalah 3,90%

Dari grafik plot di bawah dapat dilihat bahwa pergerakan ramalan pada data testing ada mengikuti pergerakan data aktual. Sehingga mungkin model VAR (2) type “both” merupakan salah satu kandidat model yang terbaik. Hasil peramalan dengan model model VAR (2) ini cenderung lebih tinggi dari data actual, kecuali di tahun 2017 data actual lebih tinggi dari data estimasi dengan VAR (2). Untuk tahun 2018 data hasil permalan/ data testing hampir sama dengan data aktualnya. Untuk data testing ini rata-rata penyimpangannya adalah sebesar 4,92%.



Gambar 18. Plot Ramalan dan Aktual Data Testing VAR (2) Type “Both”

Jika plot antara data testing dan data training digabungkan maka bentuk plotnya seperti Gambar 3.26. Untuk data tahun 1970 - 2014 atau data training plot sangat baik, karena antara data aktual dan estimasi dengan model VAR (2) hampir selalu berimpit plotnya, sehingga MAPE akan kecil. MAPE hasil pengujian untuk data training adalah sebesar 3,90%, suatu nilai yang cukup kecil karena di bawah 5%, artinya model cukup akurat. Namun plot tahun 2014 – 2019 menunjukkan data aktual dengan data estimasi dengan Model VAR (2) ini mulai agak renggang, dimana hasil estimasi rata-rata lebih tinggi dari data aktual, kecuali di beberapa titik agak mendekati. Hasil MAPE data testing ini, menunjukkan nilai yang lebih besar dari MAPE training yaitu sebesar 4,92%.



Gambar 19. Plot Ramalan dan Aktual Data Training dan Testing VAR (2) Type “Both”

### Pemilihan Model VAR Terbaik

Salah satu dasar penentuan model terbaik dapat dilakukan dengan membandingkan nilai MAPE data testing dan training yaitu dengan memilih nilai yang paling kecil, terutama untuk data testing. Untuk mencari model VAR terbaik

dilakukan uji coba pada setiap lag dengan maupun tanpa salah satu atau kedua komponen (konstanta dan trend) untuk kemudian memilih nilai MAPE yang terkecil (khususnya data testing). Selain MAPE yang terkecil, pola pergerakan ramalan juga harus diperhatikan. Pilihlah plot yang paling berhimpit/bersesuaian dengan data asli/aktual atau dengan kata lain performa hasil ramalan seiring dengan data historisnya. Selain itu, Adjusted R-Squared yang besar juga dapat dipertimbangkan dalam memilih model VAR terbaik.

Tabel 42. Ringkasan MAPE

	Training	Testing
VAR (1) type “both”	4,16%	5,82%
VAR (2) type “both”	3,90%	4,92%

Berdasarkan nilai MAPE di atas, dapat dilihat bahwa MAPE data testing maupun training pada model VAR (2) type “both” lebih kecil dibandingkan nilai MAPE pada model VAR (1) type “both”. Sedangkan berdasarkan plot ramalan data training dan testing sebelumnya, baik model VAR (1) type “both” maupun VAR (2) type “both” memberikan pergerakan yang tidak begitu berhimpit untuk ramalan data testing sebaliknya untuk data training terlihat lebih berimpit antara nilai katual dan nilai ramalan. Oleh karena itu, model terbaik yang dipilih adalah model VAR (2) dengan type “both” berdasarkan nilai MAPE data testing dan training yang paling kecil.

### Pengepasan Model untuk Seluruh Data dan Peramalannya

Model VAR terbaik dari hasil perbandingan nilai MAPE adalah model VAR (2) type “both”, sehingga dapat dilanjutkan pengepasan model pada keseluruhan data. Setelah dilakukan pengepasan model, maka dilakukan estimasi produksi karet untuk 5 tahun kedepan yaitu tahun 2020 -2024.

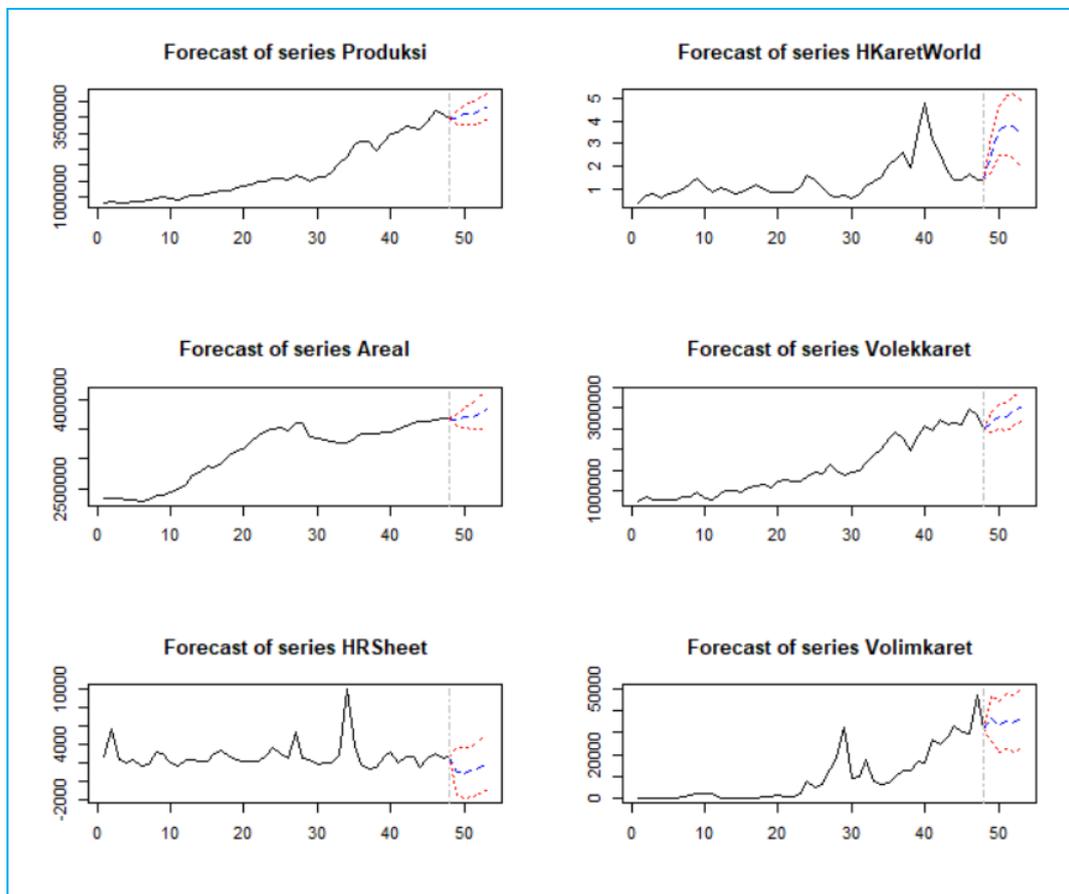
Peramalan dengan model VAR (2) type “both” menggunakan keseluruhan data menunjukkan produksi karet tahun 2020 sebesar 3,47 juta ton, naik kembali pada tahun 2021 menjadi 3,59 juta ton, terus naik, sehingga pada tahun 2024 produksi karet diperkirakan mencapai 3,81 juta ton.

Tabel 43. Hasil Peramalan Produksi Karet Tahun 2020-2024, Model VAR(2)  
Type=both

## Ouput Hasil Peramalan dengan VAR(2) type both ##
[1] 3479390 3589243 3617694 3713331 3808592

Keunggulan model VAR ini pada saat dilakukan permalan maka semua variabel dalam system akan ikut diramalkan. Hasil peramalan bukan hanya target variabel, tetapi semua variabel yang masuk dalam sistem. Hasil permalan 5 tahun ke

depan untuk variabel produksi menunjukkan produksi akan terus meningkat secara linier, begitu juga luas areal karet juga cenderung terus meningkat secara linier. Sebaliknya harga karet sheet dalam negeri cenderung masih terus akan turun, sementara harga karet dunia diperkirakan akan meningkat, tetapi ke depan akan turun kembali. Volume impor karet juga diperkirakan akan meningkat seiring dengan peningkatan produksi, sementara itu untuk volume impor karet diperkirakan akan tetap stagnan.



Gambar 20. Hasil Estimasi untuk Semua Variabel dalam Sistem Model VAR(2) Both

Untuk model VAR (2) type “both”, pergerakan hasil ramalan produksi karet cenderung mengalami peningkatan. Dari plot juga terlihat bahwa data cenderung naik. Jadi berdasarkan nilai MAPE, maka model terbaik yang terpilih untuk peramalan data produksi karet adalah model VAR (2) type “both” karena memberikan nilai MAPE training dan testing yang lebih kecil sekaligus memberikan plot ramalan yang cenderung meningkat sesuai dengan pergerakan data historis sebelumnya. Nilai ramalan 5 tahun ke depan adalah 3.479.390 Ton, 3.589.243 Ton, 3.617.694 Ton, 3.713.331 Ton, dan 3.808.592 Ton. Jika dibandingkan dengan tahun 2019 maka produksi karet hasil estimasi tahun 2020 akan meningkat sebesar 0,89%, sehingga penambahan cukup wajar. Hal ini didukung dari realitas produksi karet yang hanya

sedikit saja mengalami peningkatan, karena harga karet dunia yang masih cenderung turun. Pertumbuhan produksi karet hasil estimasi selama 5 tahun kedepan adalah sekitar 2,29%, sangat mendekati angka historis pertumbuhan 5 tahun ke belakang yaitu sebesar 2,50%.

Tabel 44. Hasil Estimasi Produksi Karet Nasional dengan Model VAR(2), Type= both

Tahun	Produksi Karet (Ton)	Pertumbuhan (%)
2019	3.448.782	
2020	3.479.390	0,888
2021	3.589.243	3,157
2022	3.617.694	0,793
2023	3.713.331	2,644
2024	3.808.592	2,565
Rata-rata pertumbuhan (%)		2,290

Tahun 2019 : Angka Sementara Ditjenbun

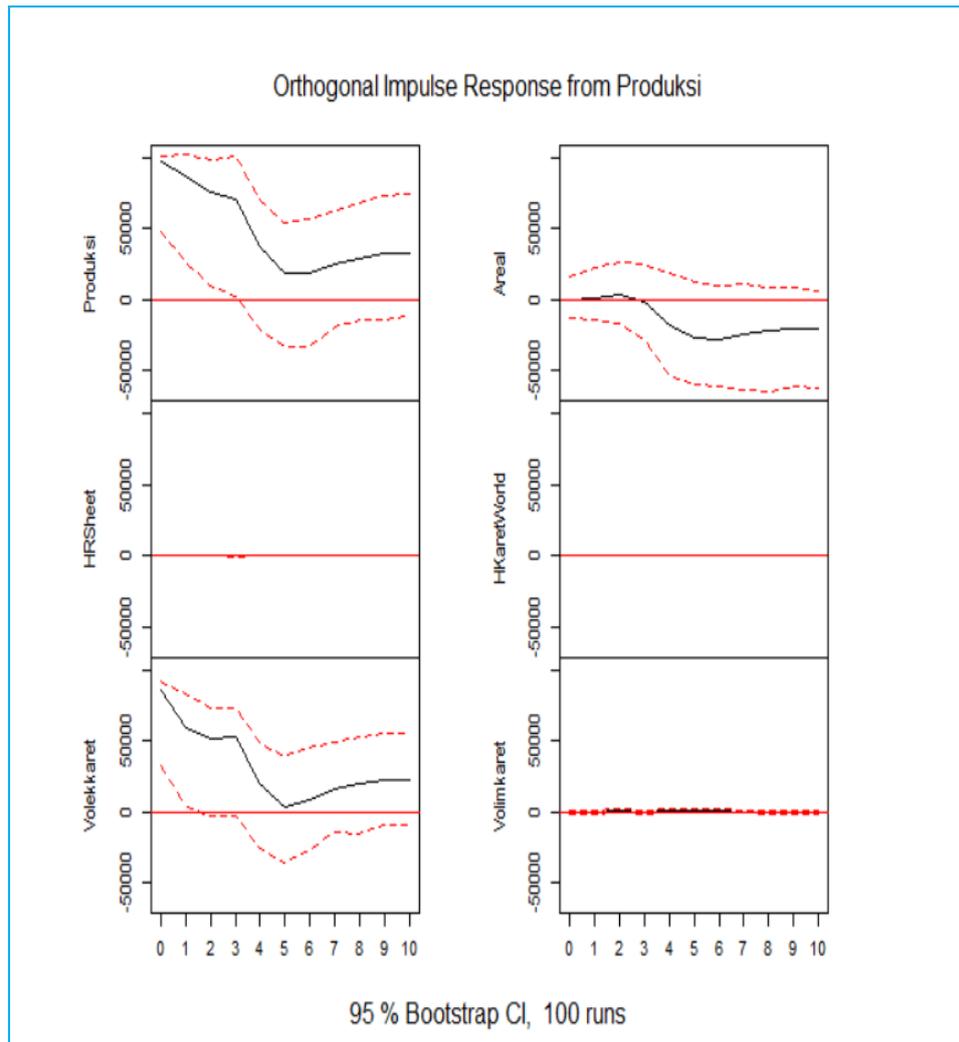
Tahun 2020 – 2024 : Angka Estimasi Berdasarkan Model VAR(2) both

### Interpretasi Impulse Response Function (IRF) dan Variance Decomposition

Setelah diperoleh model terbaik, untuk permodelan VAR dapat dilakukan interpretasi lebih mendalam terkait Impulse Response Function (IRF) dan Variance Decomposition. Berikut interpretasi ketiga hal di atas untuk model terbaik VAR (2) type “both”.

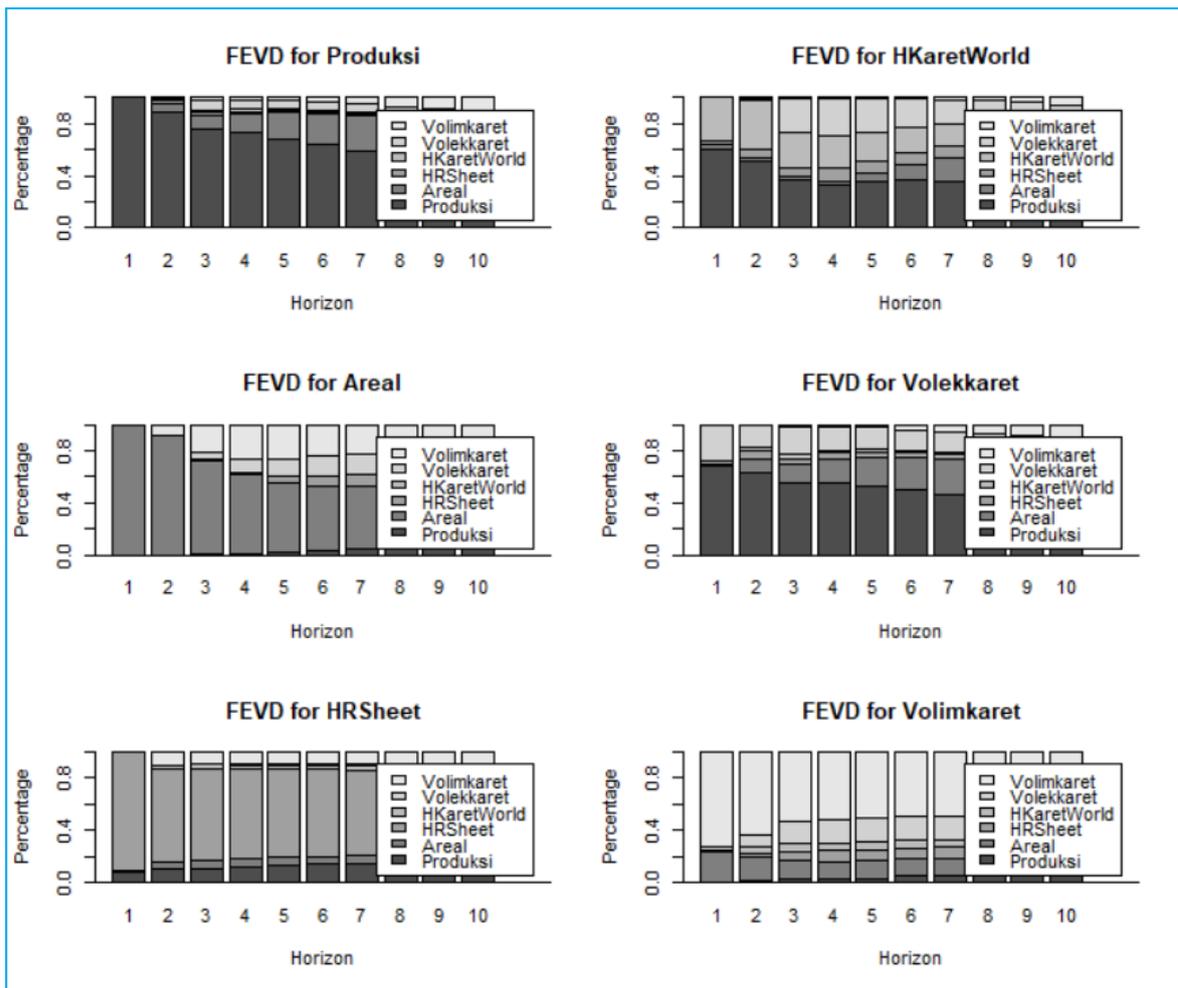
Impulse Response Function akan menjelaskan bagaimana perubahan atau gejala yang terjadi pada suatu variabel/peubah di tahun tertentu akan berdampak pada variabel/peubah lain di tahun tertentu dan tahun-tahun setelahnya.

Dari grafik Impulse Response Function produksi model terbaik VAR (2) type “both”, dapat dilihat bahwa jika terjadi perubahan pada produksi di tahun tertentu maka akan berdampak pada produksi itu sendiri sampai 3 tahun ke depan. Dampak tersebut akan hilang setelah 3 tahun. Sedangkan dampak perubahan produksi tidak berdampak pada harga sheet, luas areal, harga karet dunia, dan volume impor karet, namun masih berdampak pada volume ekspor kedepan. Perubahan produksi berdampak pada volume ekspor sampai dengan dua tahun dan akan hilang dampaknya setelah melewati tahun kedua tersebut.



Gambar 21. Impuls Respon Beberapa Variabel Terhadap Produksi Model VAR(2) Both

**Variance Decomposition** atau dekomposisi keragaman digunakan untuk melihat variabel apa saja yang mempengaruhi komposisi keragaman suatu variabel/peubah. Dari grafik dekomposisi keragaman model terbaik VAR (2) type “both” di atas dapat dilihat bahwa komposisi produksi pada tahun pertama dipengaruhi sepenuhnya oleh produksi itu sendiri. Pada tahun kedua, komposisi produksi 90% dipengaruhi oleh produksi itu sendiri, 5% dipengaruhi oleh harga sheet, dan 5% dipengaruhi oleh luas areal, harga karet dunia, volume ekspor karet, serta volume impor karet. Pada tahun ketiga, komposisi produksi dipengaruhi 85% oleh produksi itu sendiri, 5% dipengaruhi oleh harga sheet, dan 5% dipengaruhi oleh volume impor karet, dan 5% dipengaruhi oleh luas areal, harga karet dunia, serta volume ekspor karet. Semakin bertambahnya tahun, pengaruh produksi karet terhadap keragaman produksi karet sendiri semakin berkurang diikuti makin tingginya pengaruh luas areal, harga sheet, harga karet dunia, volume ekspor karet, dan volume impor karet.



Gambar 22. Dekomposisi Keragaman Beberapa Variabel Untuk Model VAR(2) Both

## E. Pemilihan Model Terbaik Estimasi Karet Nasional

Salah satu dasar penentuan model terbaik dapat dilakukan dengan membandingkan nilai MAPE untuk data testing dan training yaitu dengan memilih nilai MAPE yang paling kecil, terutama untuk data testing. Selain MAPE yang terkecil, pola pergerakan ramalan juga harus diperhatikan. Pilihlah plot yang paling berhimpit/bersesuaian dengan data asli/aktual atau dengan kata lain performa hasil ramalan seiring dengan data historisnya.

Berdasarkan data historis yang ada produksi karet nasional berfluktuasi, produksi tahun 2015 sebesar 3,14 juta ton atau turun 0,25%. Pada tahun 2016 dan 2017 produksi karet nasional meningkat masing-masing sebesar 6,76% dan 9,60%, sehingga produksi karet tahun 2017 sebesar 3,68 juta ton. Pada tahun 2018 produksi karet nasional mulai turun sebesar

1,36%, kemudian pada tahun 2019 kembali turun sebesar 5%. Rata-rata pertumbuhan produksi karet nasional selama 5 tahun terakhir atau tahun 2015 – 2019 sebesar 2,50%.

Tabel 45. Produksi Karet Nasional Tahun 2015 – 2020

Tahun	Produksi (TON)	
	Indonesia	Pertumb. (%)
2015	3.145.398	-0,25
2016	3.357.951	6,76
2017	3.680.428	9,60
2018	3.630.357	-1,36
2019 *)	3.448.782	-5,00
2020 **)	3.545.693	2,81
Rata-rata pertumbuhan (%)		
2015 - 2019		2,50

Sumber : Direktorat Jenderal Perkebunan, diolah Pusdatin

Untuk menyusun angka estimasi produksi karet telah dilakukan uji coba dengan 4 (empat) model. Model yang pertama adalah model time series atau ARIMA, model terbaik untuk ARIMA adalah pada orde ARIMA (0,2,1). Untuk model estimasi produksi karet nasional dengan ARIMA (0,2,1) menghasilkan MAPE untuk data training sebesar 4,31% dan MAPE untuk data testing sebesar 6,34%. Hal ini menunjukkan bahwa kemampuan untuk melakukan estimasi dengan model ARIMA ini rata-rata akan mengalami kesalahan sekitar 6,34% lebih tinggi atau 6,34% lebih rendah.

Model yang kedua adalah menggunakan regresi linier sebagai variabel bebas adalah waktu atau tahun. Untuk model regresi linier sederhana menghasilkan MAPE yang cukup besar yaitu 15,49% untuk data training dan 16,88% untuk data testing. Model regresi linier kurang tepat untuk mengestimasi produksi karet karena menghasilkan bias besar. Untuk model regresi kuadratik dengan variabel bebas waktu (tahun) menghasilkan MAPE yang lebih baik, yaitu sebesar 6,61% untuk data training dan 5,05% untuk data testing. MAPE untuk model regresi kuadratik lebih kecil, namun jika dilihat dari hasil estimasi angka pertumbuhan angka estimasi kurang realistis karena selama 5 tahun ke depan rata-rata pertumbuhan produksi karet 3,57%, dan angka lonjakan produksi cukup signifikan.

Metode estimasi yang ketiga adalah dengan model fungsi transfer, untuk melakukan estimasi produksi karet dengan variabel bebas adalah harga karet dunia. Untuk model fungsi transfer ini menghasilkan MAPE data training 3,80%, sementara untuk MAPE data testing sebesar 5,03%. Model fungsi transfer ini menghasilkan MAPE yang lebih kecil dibandingkan model ARIMA dan model Regresi, sehingga model fungsi transfer lebih akurat dalam melakukan estimasi. Hasil estimasi juga menunjukkan angka yang lebih realistis, dengan angka estimasi tahun 2020 sebesar 3,611 juta ton. Namun untuk estimasi 5 tahun kedepan

angka pertumbuhan sebesar 0,092%/tahun, sementara angka pertumbuhan 5 tahun sebelumnya sebesar 2,5%/tahun, jadi jauh lebih kecil dibandingkan data historisnya.

Tabel 46. Perbandingan MAPE Model Arima, Regresi, Fungsi Transfer dan VAR

No	Urain Model	MAPE		Hasil Estimasi Produksi Karet (Ton)					Pertumbuhan (%) Tahun 2020 - 2025
		Training	Testing	2020	2021	2022	2023	2024	
1	ARIMA (0,2,1)	4,31	6,34	3.515.956	3.583.129	3.650.303	3.717.477	3.784.650	1,858
2	Regresi Linier	15,49	16,88	3.002.554	3.054.183	3.105.813	3.157.442	3.209.072	1,677
	Regresi Kuadratik	6,61	5,05	3.978.853	4.117.846	4.259.436	4.428.930	4.578.484	3,572
3	Arima (1,1,0) xreg=hrkworld	3,80	5,03	3.611.873	3.621.989	3.628.211	3.626.798	3.625.179	0,092
4	Var (2) p=2 type=both	3,90	4,92	3.479.390	3.589.243	3.617.694	3.713.331	3.808.592	2,290
Angka Sementara Tahun 2019 :		3.448.782 ton							
Rata-rata Pertumbuhan 2015 - 2019 :		2,50%							

Untuk model estimasi yang terakhir adalah dengan model VAR (Vector Auto Regressive). Untuk model VAR ini menggunakan 6 variabel yaitu produksi, luas areal, harga karet Sheet, harga karet dunia, volume ekspor dan volume impor karet. Model yang terbaik untuk Model VAR adalah nilai  $p=2$  dan  $type="both"$ ,  $p=2$  artinya menggunakan variabel bebas sampai lag-2 dan ada konstanta dan factor trend. Estimasi produksi karet dengan menggunakan model VAR ini menghasilkan ketelitian yang cukup tinggi yaitu MAPE untuk data training 3,90% dan MAPE untuk data testing 4,92%. MAPE untuk data testing ini paling kecil dibandingkan dengan model-model lainnya, sehingga model VAR ini disimpulkan paling kecil kesalahan estimasinya atau model VAR model terbaik untuk estimasi produksi karet nasional. Jika dibandingkan angka pertumbuhan produksi karet antara hasil estimasi 5 tahun kedepan dengan rerata pertumbuhan 2,29%, sangat mendekati angka pertumbuhan 5 tahun terakhir yaitu sebesar 2,50%. Angka hasil estimasi untuk produksi karet nasional tahun 2020 sebesar 3,47 juta ton, sementara untuk angka sementara tahun 2019 sebesar 3,44 juta ton atau meningkat sebesar 0,89%.

## KESIMPULAN

Untuk meningkatkan akurasi dalam penyusunan angka estimasi, maka dilakukan pengembangan metode estimasi produksi karet. Metode estimasi data perkebunan selama ini menggunakan model *Single Smoothing Exponential (SSE)* atau menggunakan *Double Smoothing Exponential (DSE)*. Meskipun dua metode tersebut dapat menghasilkan angka estimasi yang cukup baik, namun masih perlu melakukan pengembangan model alternatif yang diharapkan lebih akurat. Setelah dilakukan uji coba metode estimasi dengan metode lain, ternyata beberapa metode menghasilkan angka MAPE yang lebih kecil dibandingkan dengan estimasi dengan metode sebelumnya.

Untuk analisis ini data dibagi menjadi 2 kelompok, yaitu data training tahun 1967 – 2013, dan data testing tahun 2014 – 2019. Data training untuk penyusunan model, sedangkan data testing untuk uji coba model dalam melakukan estimasi 6 tahun kedepan. Untuk estimasi produksi karet alternatif model pertama adalah Model ARIMA. Model ARIMA terbaik adalah ARIMA (0,2,1), menghasilkan MAPE untuk data training 4,31%, dan MAPE data testing 6,34%. Model regresi linier menghasilkan MAPE data training sebesar 15,49% dan MAPE data testing 16,88%. Model regresi kuadratik menghasilkan MAPE data training sebesar 6,61%, MAPE data testing 5,05%. Untuk model yang ketiga dengan menggunakan Fungsi

Transfer dengan variabel input harga karet dunia, menghasilkan MAPE untuk data training sebesar 3,80% dan MAPE data testing 5,03%. Untuk model yang keempat model VAR(2) type 'both' ada pengaruh trend dan konstanta, menghasilkan MAPE data training 3,90% dan data MAPE data testing 4,92%.

Berdasarkan perbandingan besarnya MAPE baik data testing maupun data training dan hasil estimasi produksi 5 tahun kedepan, maka model terbaik yang terpilih adalah model VAR(2) tipe 'both' karena menghasilkan akurasi yang cukup tinggi sehingga MAPE rata-rata sebesar 4,92%. Hasil estimasi produksi karet nasional dalam bentuk karet kering untuk model VAR (2) tipe 'both' untuk tahun 2020 sebesar 3.479.390 Ton, tahun 2021 sebesar 3.589.243 Ton, tahun 2022 sebesar 3.617.694 Ton, tahun 2023 sebesar 3.713.331 Ton, dan tahun 2024 sebesar 3.808.592 Ton. Laju pertumbuhan estimasi produksi karet nasional selama 5 tahun kedepan rata-rata 2,29%.

## DAFTAR PUSTAKA

- Anonim. 2013. Pedoman Pelaksanaan Pengelolaan Data Komoditas Perkebunan (PDKP). Dirjen Perkebunan – Kementerian Pertanian.
- Enders, W. 2010. Applied Econometric Time Series. USA: University of Alabama. Wiley, Third Edition.
- Fitriani, D.R, Darsyah, M.Y., & Wasono, R. 2013. Peramalan Fungsi Transfer pada Harga Emas Pasar Komoditi. Seminar Nasional Pendidikan Sains dan Teknologi, Fakultas MIPA, Universitas Muhammadiyah Semarang.
- Guha, B and Bandyopadhyay, G. 2016. Gold Price Forecasting Using ARIMA Model. Journal of Advanced Management Science Vol. 4, No. 2, March 2016
- Gujarati, D.N. and D.C. Porter, 2010. Dasar-dasar Ekonometrika. Jakarta: Penerbit Salemba Empat, Buku 2, Edisi 5.
- Kementerian Pertanian. 2019. Statistik Perkebunan Indonesia 2018-2020 (Karet). Jakarta: Direktorat Jenderal Perkebunan.
- M. Firdaus 2020. Aplikasi Ekonometrika dengan E-Views, Stata dan R. Institut Pertanian Bogor (IPB) Press.
- Montgomery DC, Johnson LA & Gardiner JS. 1990. Forecasting and Time Series Analysis. Singapore:Mc-Graw Hill.
- Myers R. 1994. Classical And Modern Regression with Applications. Boston: PWS – KENT Publishing Company.
- Ryan TP. 1997. Modern Regression Methods. New York,USA: John Wiley & Sons, INC

# KAJIAN MODEL PERAMALAN PRODUKSI SAWIT (*CRUDE PALM OIL/CPO*) DI INDONESIA:

## PENDEKATAN MODEL ARIMA, REGRESI, FUNGSI TRANSFER, DAN VAR

Lasmiyati

*Subdirektorat Statistik Tanaman Perkebunan-Badan Pusat Statistik  
Jalan Dr. Sutomo no. 6-8 Gd 4 Lt 4, Pasar Baru-Jakarta Pusat, Indonesia*

### ABSTRAK

*Crude Palm Oil (CPO) atau minyak kelapa sawit merupakan bahan baku dari berbagai barang kebutuhan sehari-hari bagi masyarakat di dunia. Kebutuhan dunia akan CPO yang cukup tinggi sungguh sangat menguntungkan bagi Indonesia sebagai negara beriklim tropis yang cocok untuk pertumbuhan kelapa sawit. Guna merumuskan kebijakan terkait produksi, sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Penelitian ini akan mengkaji empat metode yaitu metode *Auto-Regressive Integrated Moving Average* (ARIMA), regresi, fungsi transfer dan *Vector Auto Regression* (VAR) dengan menggunakan R Studio guna melakukan pemodelan produksi sawit (CPO) di Indonesia. Metode yang dipilih guna melakukan peramalan data produksi sawit (CPO) adalah metode yang terbaik yang dikaji dari nilai *Mean Absolute Percentage Error* (MAPE) terkecil. Berdasarkan dari hasil nilai MAPE disimpulkan bahwa fungsi transfer adalah metode yang terbaik dengan MAPE 4,49% untuk produksi sawit (CPO) Indonesia.*

### ABSTRACT

*Crude Palm Oil (CPO) or palm oil is the raw material for various daily necessities for people in the world. The world's demand for CPO which is quite high is very beneficial for Indonesia as a tropical country that is suitable for the growth of oil palm. In order to formulate policies related to production, the availability of up-to-date data is urgently demanded, and even forecasts for the next few periods. This study will examine four methods, namely the *Auto-Regressive Integrated Moving Average* (ARIMA) method, regression, transfer function and *Vector Auto Regression* (VAR) using R Studio to model oil palm production (CPO) in Indonesia. The method chosen to forecast palm oil production (CPO) data is the best method that is assessed from the smallest *Mean Absolute Percentage Error* (MAPE) value. Based on the results of the MAPE value it is concluded that the transfer function is the best method with a MAPE of 4.49% for Indonesian palm (CPO) production.*

## PENDAHULUAN

*Crude Palm Oil* (CPO) atau minyak kelapa sawit merupakan bahan baku dari berbagai barang kebutuhan sehari-hari bagi masyarakat di dunia. Kebutuhan dunia akan CPO yang cukup tinggi, sungguh sangat menguntungkan bagi Indonesia sebagai negara beriklim tropis yang cocok untuk pertumbuhan kelapa sawit. Penggunaannya sebagai bahan baku pada berbagai produk, misalnya sabun, kosmetik, minyak goreng, mentega, dll menjadikannya terus dibutuhkan ditengah polemiknya sebagai salah satu kandidat perusak lingkungan. Oleh karena itu, menjadi sangat penting untuk mempelajari estimasi data produksi sawit (CPO) sebagai dasar penentu berbagai kebijakan terkait.

Guna merumuskan kebijakan terkait produksi, sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Makalah ini akan mencoba membandingkan beberapa metode statistik yang telah dikembangkan oleh para ahli, diantaranya metode *Auto-Regressive Integrated Moving Average* (ARIMA), Regresi, Fungsi Transfer dan *Vector Auto Regression* (VAR) dalam melakukan pemodelan dan peramalan produksi sawit (CPO) di Indonesia.

Oleh karena itu, tujuan dari disusunnya kegiatan ini adalah:

- a. Melakukan analisis dan peramalan data produksi sawit (CPO) menggunakan model ARIMA, Regresi, Fungsi Transfer dan VAR di Indonesia.
- b. Membandingkan metode tersebut dalam memperoleh ramalan data produksi komoditas sawit (CPO).
- c. Menentukan metode terbaik dalam meramal data produksi komoditas sawit (CPO) di Indonesia.

## BAHAN DAN METODE

Data yang digunakan dalam melakukan pemodelan produksi sawit (CPO) adalah data series produksi sawit (CPO) dalam satuan ton di Indonesia tahun 1967 s.d. 2018, luas areal kelapa sawit dalam satuan hektar di Indonesia tahun 1967-2018, harga CPO dunia dalam satuan US\$ tahun 1967-2018, dan volume ekspor serta volume impor CPO dalam satuan ton di Indonesia tahun 1981-2018. Peubah yang diasumsikan mempengaruhi besaran produksi sawit (CPO) untuk Fungsi Transfer adalah harga CPO dunia. Sedangkan untuk pemodelan VAR produksi sawit (CPO), peubah yang diasumsikan mempengaruhi adalah harga CPO dunia, luas areal kelapa sawit, volume ekspor CPO, dan volume impor CPO. Harga CPO dunia diperkirakan berpengaruh karena kenaikan harga akan mendorong petani meningkatkan produksi sawit (CPO). Sedangkan luas areal akan menentukan banyak tidaknya produksi sawit (CPO) yang dihasilkan dan sebaliknya. Peningkatan volume ekspor dan penurunan impor diperkirakan akan membangkitkan minat petani untuk meningkatkan produksi sawit (CPO).

Peramalan data produksi sawit (CPO) dilakukan melalui pengujian dengan beberapa metode, yakni ARIMA, Regresi, Fungsi Transfer, dan VAR menggunakan R Studio.

Penjelasan metode yang digunakan dalam melakukan peramalan produksi sawit (CPO) di Indonesia adalah sebagai berikut:

d. ARIMA

ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampainya. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

$Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t

$Y_{t-p}$  = data *time series* pada kurun waktu ke (t-P)

$\mu$  = suatu konstanta

$\theta_1 \theta_2 \dots \theta_p$  = parameter-parameter model

$\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

e. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih variabel.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

dengan

$y$  = variabel respon/dependen/terikat

$x_1, x_2, \dots, x_k$  = variabel prediktor/independen/bebas

$\beta_0$  = konstanta/intercept

$\beta_1, \beta_2, \dots, \beta_k$  = koefisien regresi variabel bebas

$\varepsilon$  = error/residu/sisaan

f. Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau  $Y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $Y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $X_t$ ) dengan deret output tersebut.

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\varphi_p(B)} \varepsilon_t$$

Dimana:

- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$

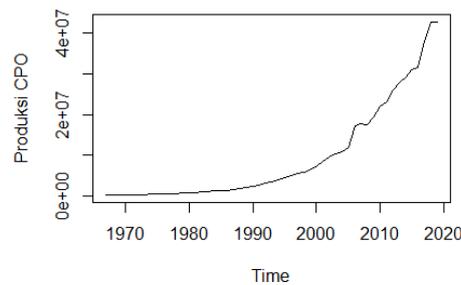
g. VAR

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini

memperlakukan seluruh variabel sebagai variabel endogen, karena pada kenyataannya suatu variabel dapat bertindak sebagai endogen atau dependen dari suatu variabel yang lainnya (Gujarati, Damodar. N dan Porter, Dawn. C. 2009).

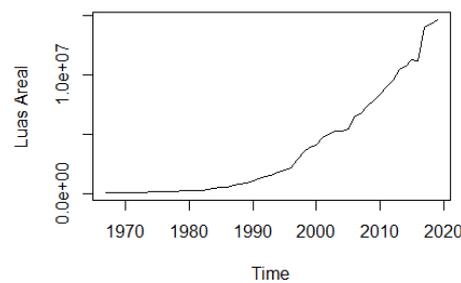
## HASIL DAN PEMBAHASAN

### a. Eksplorasi Data



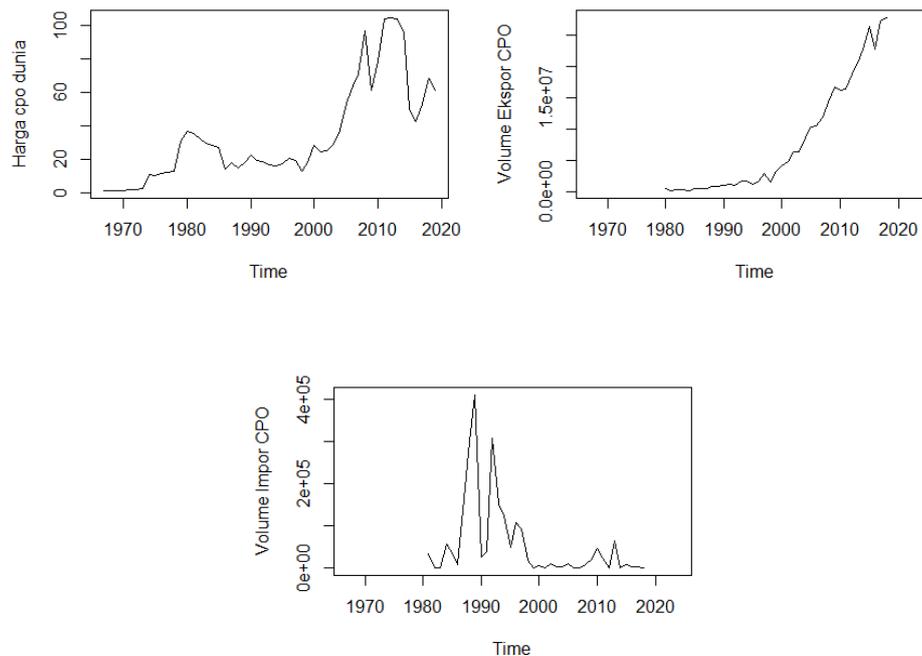
**Gambar 1.** Perkembangan Produksi Sawit (CPO) (Ton), 1967- 2018

Produksi sawit (CPO) di Indonesia selama periode 1967 – 2018 cenderung mengalami trend peningkatan dari tahun ke tahun dengan rata-rata pertumbuhan sebesar 11,49% per tahun (Gambar 1).



**Gambar 2.** Perkembangan Luas Areal Sawit (Ha), 1967-2018

Luas areal sawit di Indonesia selama periode 1967 – 2018 juga cenderung mengalami trend peningkatan dari tahun ke tahun dengan rata-rata pertumbuhan sebesar 10,18% per tahun (Gambar 2). Pertumbuhan tersebut sedikit lebih rendah dibandingkan dengan pertumbuhan produksi sawit (CPO).



**Gambar 3.** Perkembangan Harga CPO dunia (US\$) 1967-2018, Volume Ekspor Karet (Ton) 1980-2018, dan Volume Impor Karet (Ton) 1981,2018

Pergerakan harga CPO dunia selama periode 1967 – 2018 sangat berfluktuasi dari tahun ke tahun dengan kecenderungan meningkat (Gambar 3). Rata-rata pertumbuhan harga CPO dunia sebesar 14,01% per tahun. Pergerakan volume ekspor CPO selama periode 1980 – 2018 cenderung menunjukkan trend peningkatan, sedangkan pergerakan volume impor karet selama periode 1981 – 2018 menunjukkan pergerakan yang cenderung tidak stabil.

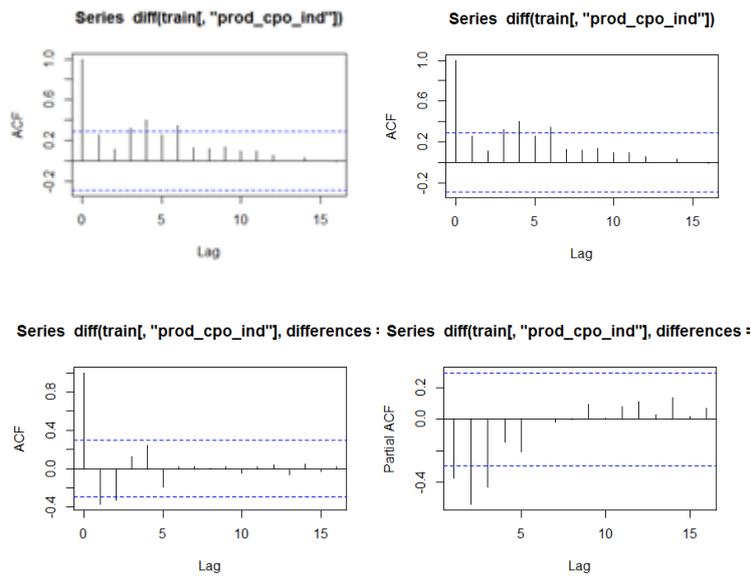
#### **b. ARIMA**

Dalam melakukan pemodelan produksi sawit (CPO) menggunakan ARIMA, data yang digunakan adalah periode 1967-2018. Periode data tersebut kemudian dipisahkan menjadi data *training* dan *testing*. Panjang series data pada data *training* adalah tahun 1967-2012, sementara data *testing* adalah periode tahun 2013-2018. Data *training* digunakan untuk melakukan penyusunan model, sementara data *testing* digunakan untuk validasi model.

Uji kestasioneran data dilakukan menggunakan uji formal statistik *Augmented Dickey-Fuller* maupun secara visual dengan melihat plot ACF dan PACF. Tabel 1 menunjukkan bahwa data produksi Sawit (CPO) stasioner pada *differencing* 1 (nilai uji mutlak > nilai kritis mutlak), namun berdasarkan plot ACF dan PACF pada Gambar 1, terlihat bahwa pada data *differencing* 1 belum stasioner sehingga perlu dilakukan pengujian data pada *differencing* 2. Berdasarkan plot ACF dan PACF pada *differencing* 2 (Gambar 1) terlihat data sudah stasioner atau dengan kata lain rata-rata dan variannya konstan (Kirchgassner, Gebhard dan Wolters, Jurgen. 2007).

**Tabel 1.** Hasil Uji ADF

Kondisi Data	Hasil Uji ADF Produksi Sawit (CPO)
Asli	Value of test-statistic is: 1.6423 9.5131 10.0679 Critical values for test statistics: 1pct 5pct 10pct tau3 -4.04 -3.45 -3.15 <b>Blm stasioner</b> phi2 6.50 4.88 4.16 phi3 8.73 6.49 5.47
Diff 1	Value of test-statistic is: -2.9912 Critical values for test statistics: 1pct 5pct 10pct <b>Stasioner</b> tau1 -2.6 -1.95 -1.61
Diff 2	Value of test-statistic is: -9.988 Critical values for test statistics: 1pct 5pct 10pct <b>Stasioner</b> tau1 -2.6 -1.95 -1.61



**Gambar 1.** Plot ACF dan Partial ACF data *differencing* produksi sawit (CPO)

Selanjutnya dilakukan pendugaan model ARIMA dengan hasil seperti tertera dalam tabel 2.

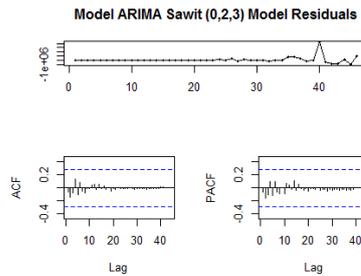
**Tabel 2.** Pendugaan Model ARIMA

---

```
Series: train[, "prod_cpo_ind"]
ARIMA(0,2,3)
sigma^2 estimated as 5.817e+11: log likelihood=-658.16
AIC=1324.32 AICc=1325.34 BIC=1331.46
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 143698.3 720033.1 276083.9 3.413579 4.385792 0.4760483 -0.0612822
```

---

Dari hasil *running* program model auto ARIMA menggunakan R Studio diperoleh model ARIMA (0,2,3) dengan hasil pemeriksaan *residual* yang sudah cukup stasioner pada rata-rata sebagaimana terlihat pada Gambar 2. Selain itu, hasil pemeriksaan kenormalan *residual* menggunakan Ljung Box juga menunjukkan bahwa *residual* sudah secara signifikan mendekati distribusi normal pada lag 5,10,15,20,25,30,35, dan 40. Hal tersebut terlihat dari nilai p yang signifikan atau di atas taraf 5 persen (Tabel 3).



**Gambar 2.** Grafik Pemeriksaan Model *Residual* ARIMA pada Data *Training*

**Tabel 3.** Hasil Uji L-Jung test *Residual* ARIMA pada Data *Training*

Lag	Box-Ljung test Produksi Sawit (CPO)	Hasil
5	X-squared=3.1548,df = 5,p-value = 0.6761	Sign normal
10	X-squared=3.9873,df = 10,p-value = 0.9479	Sign normal
15	X-squared=4.5624,df = 15,p-value = 0.9952	Sign normal
20	X-squared=4.9338,df = 20,p-value = 0.9998	Sign normal
25	X-squared = 5.059,df = 25, p-value = 1	Sign normal
30	X-squared = 5.1945, df = 30, p-value = 1	Sign normal
35	X-squared = 5.4247, df = 35, p-value = 1	Sign normal
40	X-squared = 5.6322, df = 40, p-value = 1	Sign normal
45	X-squared = 5.7853, df = 45, p-value = 1	Sign normal

Selanjutnya dilakukan peramalan data testing diikuti penghitungan nilai MAPE. Tabel 4 menunjukkan perbandingan nilai MAPE data *training* dan data *testing*, terlihat bahwa nilai MAPE produksi yaitu 4,39% untuk data *training* dan 4,66% untuk data *testing*.

**Tabel 4.** Perbandingan MAPE Data *Training* dan Data *Testing*

Produksi	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	143698.3	720033.1	276083.9	3.413579	<b>4.385792</b>	0.4760483	-0.0612822
Test set	1006691.8	2660352.4	1790191.7	2.135212	<b>4.658623</b>	3.0868068	NA

Hasil pengepasan model ARIMA (0,2,3) untuk produksi pada keseluruhan data *training* dan *testing* diperoleh model sebagai tertera pada Tabel 5 dengan nilai MAPE 6,19% untuk produksi.

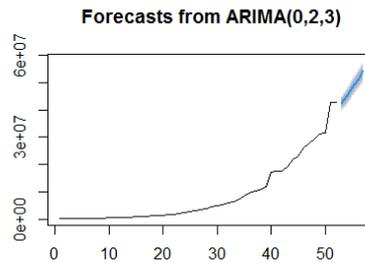
**Tabel 5.** Hasil Pengepasan Model ARIMA untuk Keseluruhan Data

Call:							
arima(x = total[, "prod_cpo_ind"], order = c(0, 2, 3))							
Coefficients:							
	ma1	ma2	ma3				
	-1.4356	0.1329	0.4579				
	s.e.	0.1725	0.2653	0.1451			
sigma^2 estimated as 1.832e+12: log likelihood = -780.65, aic = 1569.3							
Training set error measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	366045.6	1327388	537399.3	5.532319	<b>6.189547</b>	0.6376766	-0.1010958

Hasil peramalan 5 tahun ke depan dengan menggunakan model ARIMA (0,2,3) untuk produksi tertera pada Tabel 6. dengan plot yang terlihat pada gambar 3. Dari plot terlihat bahwa nilai ramalan cenderung mengalami kenaikan pada 5 tahun mendatang.

**Tabel 6.** Peramalan Produksi Sawit (CPO) ARIMA (0,2,3) 2019-2023

Point Forecast	Produksi
53	42313032
54	45185567
55	48186211
56	51186856
57	54187500



**Gambar 3.** Plot Peramalan Produksi Sawit (CPO) ARIMA (0,2,3)

### c. Regresi

Peubah bebas yang digunakan dalam melakukan pemodelan data produksi sawit (CPO) adalah peubah waktu yang dalam hal ini adalah tahun. Panjang *series* data pada data *training* adalah tahun 1967-2012, sementara data *testing* adalah periode tahun 2013-2018. Pendugaan model regresi awal dilakukan dengan menggunakan model regresi linier sederhana, namun karena nilai MAPE yang sangat tinggi (MAPE *training* 354,1, MAPE *testing*= 46,34), akhirnya dilakukan pendugaan model regresi kuadrat.

**Tabel 7.** Pendugaan Model Regresi Kuadrat dengan Peubah Waktu Data *Training*

```
Call:
lm(formula = prod_cpo_ind ~ t + t2, data = train)
Residuals:
    Min       1Q   Median       3Q      Max
-2307256 -1085103  217884  1043589  2988384
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2281239    553313   4.123 0.000168 ***
t            -491614    56873  -8.644 5.92e-11 ***
t2             21170     1222  17.323 < 2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1306000 on 43 degrees of freedom
Multiple R-squared:  0.9682,    Adjusted R-squared:  0.9668
F-statistic: 655.4 on 2 and 43 DF,  p-value: < 2.2e-16
```

Dapat dilihat bahwa peubah waktu dan peubah waktu dalam kuadrat memiliki koefisien yang signifikan. *R-Squared* sebesar 96,7% menunjukkan bahwa peubah waktu tersebut dapat menjelaskan 96,7% keragaman dari produksi sawit (CPO) di Indonesia. Selain itu, hasil uji normalitas menunjukkan bahwa residual berdistribusi normal, meskipun hasil uji autokorekasi dan heteroskedastisitas menunjukkan masih signifikan terjadinya kedua masalah tersebut. Namun hal tersebut sementara diabaikan karena tujuan utama adalah mencari model untuk peramalan.

**Tabel 8.** Hasil Pengujian Asumsi Klasik Model Regresi Kuadrat Data *Training*

```
Produksi Sawit (CPO)
Jarque-Bera test for normality
data:  reg2$residuals
JB = 1.1318, p-value = 0.4635
Durbin-watson test
```

```

data: reg2
DW = 0.39131, p-value = 6.583e-14
alternative hypothesis: true autocorrelation is greater than 0
studentized Breusch-Pagan test
data: reg2
BP = 15.293, df = 2, p-value = 0.0004778

```

Sedangkan hasil penghitungan MAPE menunjukkan bahwa nilai MAPE data *testing* lumayan baik sekitar 14,6% untuk model produksi sawit (CPO).

**Tabel 9.** MAPE Data *Training* dan *Testing* Model Regresi Kuadratik

Produksi Sawit (CPO)						
<b>Training</b>	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	0.7005	12.0536	25.9881	<b>124.5975</b>	151.6791	1260.5612
<b>Testing</b>	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	8.542	11.459	11.787	<b>14.600</b>	17.489	24.747

Hasil pengepasan model menggunakan keseluruhan data *training* dan *testing* menunjukkan bahwa keseluruhan koefisien signifikan dengan *R-Squared* sebesar 97,4% atau variabel peubah waktu dapat menjelaskan 97,4% keragaman produksi sawit (CPO) Indonesia.

**Tabel 10.** Pengepasan dan MAPE Pendugaaan Model Regresi Kuadratik dengan Peubah Waktu

```

Call:
lm(formula = prod_cpo_ind ~ t + t2, data = total)
Residuals:
    Min       1Q   Median       3Q      Max
-3339807 -1385377  153850  1173596  6279517
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3412741    733852   4.65 2.54e-05 ***
t            -685321    66545  -10.30 7.51e-14 ***
t2             26199     1262   20.76 < 2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 1832000 on 49 degrees of freedom
Multiple R-squared:  0.975,    Adjusted R-squared:  0.974
F-statistic: 954.4 on 2 and 49 DF,  p-value: < 2.2e-16
    Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0.004  8.528  28.350 169.075 189.252 1935.404

```

Hasil peramalan 5 tahun ke depan dengan menggunakan model regresi kuadratik tertera pada Tabel 11 dengan nilai ramalan cenderung mengalami kenaikan pada 5 tahun mendatang.

**Tabel 11.** Peramalan Produksi Sawit (CPO) dengan Regresi Kuadratik 2019-2023

Produksi Sawit (CPO)	
1	38617257
2	40682796
3	42800733
4	44971068
5	47193800

#### e. Fungsi Transfer

Peubah yang diasumsikan dapat mempengaruhi gejala produksi sawit (CPO) adalah harga CPO dunia sehingga akan digunakan untuk melakukan analisis dengan model fungsi transfer. Panjang *series* data pada data training adalah tahun 1967-2012, sementara data *testing* adalah periode tahun 2013-2018. Dari pola data dan uji *Augmented Dickey-Fuller* (Tabel 12 dan Gambar 4) harga CPO dunia terlihat bahwa data belum stasioner sehingga akan dilakukan *differencing* terlebih dahulu sebelum digunakan dalam permodelan. Hasil pengujian deret *differencing* (Tabel 13) menunjukkan bahwa data *differencing* secara nyata sudah stasioner (nilai uji *mutlak* > nilai kritis *mutlak*), oleh karena itu bisa dilanjutkan dengan identifikasi model ARIMA untuk harga CPO dunia sebagai peubah input.

**Tabel 12.** Uji *Augmented Dickey-Fuller* Harga CPO Dunia

---

```

Residual standard error: 9.532 on 40 degrees of freedom
Multiple R-squared: 0.08813, Adjusted R-squared: 0.01974
F-statistic: 1.289 on 3 and 40 DF, p-value: 0.2915
Value of test-statistic is: -0.3169 2.0409 1.351
Critical values for test statistics:
      1pct 5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
    
```

---

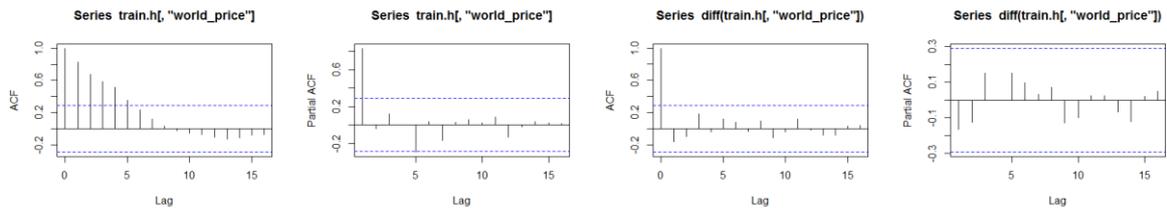
**Tabel 13.** Uji *Augmented Dickey-Fuller* untuk *differencing* harga CPO dunia

---

```

Residual standard error: 9.5 on 39 degrees of freedom
Multiple R-squared: 0.6205, Adjusted R-squared: 0.5914
F-statistic: 21.26 on 3 and 39 DF, p-value: 2.524e-08
Value of test-statistic is: -5.7704 11.1347 16.6727
Critical values for test statistics:
      1pct 5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
    
```

---



**Gambar 4.** ACF dan PACF Data Asli dan *Differencing* Harga CPO Dunia

Pendugaan model ARIMA dilakukan dengan memperhatikan hasil uji ADF dan pola dari ACF dan PACF yang signifikan. Model ARIMA deret input Harga CPO dunia data *training* ini akan digunakan untuk tahap *prewhitening*. Dengan menggunakan auto ARIMA untuk data *training* diperoleh model ARIMA (0,1,0) dengan MAPE 36,96 %. ARIMA (0,1,0) akan menghasilkan ramalan yang sama untuk periode ke depan sehingga dilakukan *trial error* dengan menggunakan ordo AR dan MA yang lain hingga diperoleh model ARIMA (1,1,0) dengan MAPE 19,04 %.

**Tabel 14.** Pendugaan Model ARIMA Deret Input Harga CPO Dunia Data Training

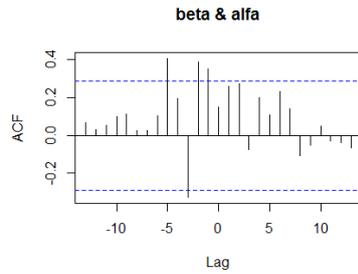
---

```

arima(x = train.h[, "world_price"], order = c(1, 1, 0))
Coefficients:
      ar1
      -0.0943
s.e.      0.1468
sigma^2 estimated as 93.14: log likelihood = -165.87, aic = 335.75
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 2.464373 9.5452 5.770456 5.503966 19.0398 0.9842841 -0.07858562
    
```

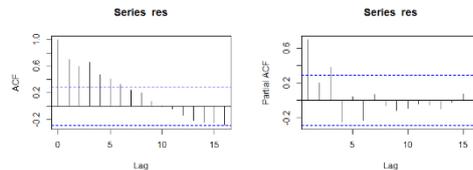
---

Selanjutnya dilakukan *prewhitening* dengan terlebih dahulu menentukan nilai alfa dan beta. Alfa merupakan nilai *residual* dari model arima variabel input yang dalam hal ini adalah harga CPO dunia, sedangkan beta merupakan nilai *residual* dari model arima output yang dalam hal ini adalah produksi sawit (CPO). Korelasi nilai beta dan alfa akan menentukan nilai r, b, dan s yang akan digunakan dalam permodelan fungsi transfer.



**Gambar 5.** *Cross Corelation* dari Beta dan Alfa Data *Training* produksi sawit (CPO)

Berdasarkan grafik *cross corelation* beta dan alfa pada Gambar 5. dapat dilihat bahwa pengaruh dari input harga dunia CPO kepada produksi sawit (CPO) Indonesia terjadi pada saat yang sama tanpa jeda, sehingga nilai koefisien  $b$  dan  $s$  adalah 0, begitu juga halnya dengan nilai  $r$  atau jeda pengaruh variabel output sendiri adalah 0. Selanjutnya dilakukan pengepasan model awal data *training* dengan membuat permodelan fungsi transfer awal menggunakan koefisien  $b$ ,  $s$ , dan  $r$  untuk mendapatkan nilai *residual*. Plot ACF dan PACF residual data *training* dapat dilihat pada Gambar 5 sedangkan hasil model hasil pendugaan residual dapat dilihat pada Tabel 15.



**Gambar 5.** Plot ACF dan PACF *Residual* Pendugaan Fungsi Transfer Data *Training*

**Tabel 15.** Hasil Pendugaan Model ARIMA Residual Data *Training*

---

```

Series: res
ARIMA(2,1,0) Produksi
Coefficients:
      ar1      ar2
    -0.4471  -0.4866
s.e.   0.1301   0.1325
sigma^2 estimated as 4.122e+12:  log likelihood=-716.72
AIC=1439.43  AICc=1440.02  BIC=1444.85
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 3112.08 1963062 1311614 -51.9274 121.4352 0.909169 0.07844336
z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1 -0.44711    0.13007  -3.4374 0.0005874 ***
ar2 -0.48656    0.13255  -3.6708 0.0002417 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

---

Diperoleh model ARIMA *residual* (2,1,0) untuk produksi. Nilai koefisien AR1 dan AR2 pada data produksi signifikan sehingga model ini dapat digunakan untuk permodelan fungsi transfer lebih lanjut meskipun MAPE cukup tinggi. Selanjutnya dilakukan pendugaan model fungsi transfer dengan menggunakan model ARIMA *residual* (2,1,0) untuk output produksi sawit (CPO) dan input harga CPO dunia pada data *training* (Tabel 16). Selain itu, dilakukan juga pendugaan model fungsi transfer dengan model ARIMA *residual* (1,1,0) untuk output produksi sawit (CPO) dan input harga CPO dunia pada data *training* (Tabel 17) sebagai pembandingan.

**Tabel 16.** Pendugaan Fungsi Transfer Arima *Residual* (2,1,0) Output Produksi Sawit (CPO) dengan Input Harga CPO Dunia Data *Training*

---

```

Series: train.h[, "prod_cpo_ind"]
Regression with ARIMA(2,1,0) errors
Coefficients:
      ar1      ar2      xreg
 0.4236  0.2328 -15182.39
s.e.  0.1496  0.1599  14316.85
sigma^2 estimated as 1.037e+12:  log likelihood=-685.06
AIC=1378.11  AICc=1379.11  BIC=1385.34
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 254253 973141.6 439907 5.354342 8.574113 0.7585265 -0.1246663
z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1  4.2365e-01 1.4961e-01 2.8317 0.004629 **
ar2  2.3282e-01 1.5990e-01 1.4560 0.145379
xreg -1.5182e+04 1.4317e+04 -1.0605 0.288937
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

---

**Tabel 17.** Pendugaan Fungsi Transfer Arima *Residual* (1,1,0) Output Produksi Sawit (CPO) dengan Input Harga CPO Dunia Data *Training*

---

```

Series: train.h[, "prod_cpo_ind"]
Regression with ARIMA(1,1,0) errors
Coefficients:
ar1      xreg
0.5153 -8186.316
s.e.  0.1461 14137.182
sigma^2 estimated as 1.064e+12:  log likelihood=-686.07
AIC=1378.14  AICc=1378.73  BIC=1383.56
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 314038.5 997137.4 443924.6 6.134815 7.541092 0.7654541 -0.1901787
z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1  0.51533  0.14614  3.5262 0.0004216 ***
xreg -8186.31647 14137.18186 -0.5791 0.5625468
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

---

Nilai MAPE pada tabel 16 dan 17 menunjukkan bahwa nilai MAPE pada ARIMA *residual* (1,1,0) dengan output produksi sawit (CPO) menghasilkan nilai MAPE yang lebih kecil sehingga model fungsi transfer ini yang akan digunakan selanjutnya. Berikutnya dilakukan peramalan dengan data *testing* menggunakan input nilai aktual dan nilai ramalan, hasil dapat dilihat pada Tabel 18.

**Tabel 18.** Peramalan Data *Testing* menggunakan Input Nilai Aktual dan Ramalan

---

```

Input Aktual Series: test.h[, "prod_cpo_ind"]
Regression with ARIMA(1,1,0) errors
Coefficients:
      ar1      xreg
 0.5153 -8186.316
s.e.  0.0000  0.000
sigma^2 estimated as 1.064e+12:  log likelihood=-81.61
AIC=165.21  AICc=166.54  BIC=164.82
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 1600261 2623747 1645449 4.344065 4.486476 0.5447921 -0.1653447
Input Ramalan Series: test.h[, "prod_cpo_ind"]
Regression with ARIMA(1,1,0) errors
Coefficients:
      ar1      xreg
 0.5153 -8186.316
s.e.  0.0000  0.000
sigma^2 estimated as 1.064e+12:  log likelihood=-81.55
AIC=165.1  AICc=166.43  BIC=164.71
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1

```

---

---

Training set 1611435 2594883 1698911 4.407061 **4.682741** 0.5624927 -0.233611

---

Selanjutnya dilakukan pendugaan ulang model input untuk keseluruhan data *training* dan *testing*, hasil dapat dilihat pada Tabel 19. Terlihat nilai MAPE masih cukup bagus yaitu 6,03% untuk model produksi sawit (CPO). Sedangkan ramalan untuk lima tahun ke depan dapat dilihat pada Tabel 20, terlihat bahwa nilai ramalan cenderung mengalami kenaikan.

**Tabel 19.** Pendugaan Ulang Model Fungsi Transfer Keseluruhan Data

---

```
Series: dataestimasi[, "prod_cpo_ind"]
Regression with ARIMA(1,1,0) errors
Coefficients:
      ar1      xreg
    0.6577 -3225.781
s.e. 0.1214 13606.905
sigma^2 estimated as 1.688e+12: log likelihood=-789.56
AIC=1585.13  AICc=1585.64  BIC=1590.92
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 345420.2 1261052 546321.9 4.29557 6.032967 0.6484784 -0.1917019
```

---

**Tabel 20.** Peramalan Fungsi Transfer (0,0,0)(1,1,0) Produksi Sawit (CPO) 2019-2023

---

Point Forecas	Produksi Sawit (CPO)
53	46148278
54	48297070
55	49710345
56	50639803
57	51251070

---

#### f. Model VAR

Pemodelan dengan menggunakan model VAR dilakukan dengan menggunakan peubah produksi sawit (CPO), luas areal, harga CPO dunia, volume ekspor CPO, dan volume impor CPO. Hal ini diasumsikan bahwa apabila terjadi gejolak luas areal, harga CPO dunia, volume ekspor karet, dan volume impor karet akan mempengaruhi pergerakan produksi sawit (CPO). Panjang series data pada data *training* adalah tahun 1981-2012, sementara data *testing* adalah periode tahun 2013-2018. Panjang *lag* (p) dalam model VAR ( $lag=p$ ) ditentukan dengan melakukan *trial error* menggunakan nilai p dari 1 s.d 6 lalu memperhatikan hasil signifikansi lag baik dengan maupun tanpa menggunakan konstanta dan *trend*. Dari hasil *trial error* diperoleh nilai p=2 dengan *type* menggunakan konstanta dan *trend*. Penggunaan konstanta dan *trend* dilakukan karena dari sekian banyak *trial error* diperoleh informasi bahwa kedua hal tersebut seringkali signifikan. Hasil pendugaan model VAR (2) dapat dilihat pada tabel 21. *R-squared* sebesar 99,1% menunjukkan bahwa peubah luas areal, harga CPO dunia, volume ekspor CPO, dan volume impor CPO dengan konstanta dan trend mampu menjelaskan 99,1% keragaman produksi sawit (CPO).

**Tabel 21.** Pendugaan Model VAR (2) Data *Training*

---

```
VAR Estimation Results:
=====
Endogenous variables: prod_cpo_ind, luas_areal, world_price, ekspor_cpo, impor_cpo
Deterministic variables: both
Sample size: 30
Log Likelihood: -1724.098
Roots of the characteristic polynomial:
1.001 0.8011 0.8011 0.6375 0.6375 0.5892 0.5892 0.545 0.4969 0.4969
Call:
VAR(y = sawit[15:46, c(3, 2, 4, 5, 6)], p = 2, type = "both")

Estimation results for equation prod_cpo_ind:
=====
```

---

prod\_cpo\_ind = prod\_cpo\_ind.l1 + luas\_areal.l1 + world\_price.l1 + ekspor\_cpo.l1 + impor\_cpo.l1 + prod\_cpo\_ind.l2 + luas\_areal.l2 + world\_price.l2 + ekspor\_cpo.l2 + impor\_cpo.l2 + const + trend

	Estimate	Std. Error	t value	Pr(> t )
prod_cpo_ind.l1	4.928e-02	2.751e-01	0.179	0.85981
luas_areal.l1	-2.919e-01	1.073e+00	-0.272	0.78860
<b>world_price.l1</b>	<b>5.499e+04</b>	<b>2.181e+04</b>	<b>2.521</b>	<b>0.02134 *</b>
ekspor_cpo.l1	1.242e-01	2.634e-01	0.472	0.64294
impor_cpo.l1	-2.004e-01	1.714e+00	-0.117	0.90822
prod_cpo_ind.l2	-3.943e-01	2.775e-01	-1.421	0.17235
luas_areal.l2	5.849e-01	1.202e+00	0.487	0.63243
world_price.l2	2.228e+04	2.680e+04	0.831	0.41669
<b>ekspor_cpo.l2</b>	<b>6.860e-01</b>	<b>2.528e-01</b>	<b>2.714</b>	<b>0.01422 *</b>
impor_cpo.l2	-3.691e-01	1.631e+00	-0.226	0.82353
<b>const</b>	<b>-2.800e+06</b>	<b>1.469e+06</b>	<b>-1.907</b>	<b>0.07265 .</b>
<b>trend</b>	<b>3.744e+05</b>	<b>1.235e+05</b>	<b>3.033</b>	<b>0.00716 **</b>

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 713200 on 18 degrees of freedom  
Multiple R-Squared: 0.9944, Adjusted R-squared: **0.991**  
F-statistic: 290 on 11 and 18 DF, p-value: < 2.2e-16

**Tabel 22.** Uji Asumsi Klasik Residual Model VAR (2) Data *Training* dan Keseluruhan Data

<b>Portmanteau Test (asymptotic)</b>	Chi-squared=321.65,df=350,p-value= <b>0.8592</b>	Chi-squared=313.08,df=350,p-value = <b>0.9225</b>
JB-Test (multivariate)	Chi-squared=25.954,df=10,p-value=0.003802	Chi-squared=42.618,df=10,p-value=5.819e-06
<b>Skewness only (multivariate)</b>	Chi-squared=11.941,df=5,p-value=0.03561	Chi-squared=10.881,df=5,p-value = <b>0.0538</b>
Kurtosis only (multivariate)	Chi-squared = 14.014,df=5,p-value=0.01552	Chi-squared=31.737,df=5,p-value=6.696e-06
<b>ARCH (multivariate)</b>	Chi-squared = 375, df = 1125, p-value = <b>1</b>	Chi-squared = 465, df = 1125, p-value = <b>1</b>

Dari hasil pengujian keberadaan autokorelasi dan heteroskedastisitas, diperoleh informasi bahwa *residual* yang diperoleh dari model VAR (2) data *training* terbebas dari kedua masalah tersebut, dibuktikan dengan nilai *p-value* yang tidak signifikan (di atas 0,05). Sedangkan hasil pengujian normalitas *residual* dari model VAR (2) data *training*, diperoleh informasi bahwa asumsi tersebut belum dapat terpenuhi (signifikan dengan nilai *p-value* di bawah 0,05). Akan tetapi, asumsi normalitas sementara dapat diabaikan karena data yang digunakan sudah cukup panjang. Selanjutnya dilakukan ramalan pada data *testing* dan pengepasan model pada seluruh data. Hasil pengujian normalitas, non autokorelasi, serta homoskedastisitas dengan model VAR (2) pada data keseluruhan menunjukkan bahwa *residual* yang diperoleh memenuhi ketiga asumsi tersebut yang ditunjukkan dengan nilai *p-value* yang tidak signifikan atau lebih dari 0,05 (Tabel 22).

**Tabel 23.** MAPE Data *Training*, *Testing*, dan Pengepasan Keseluruhan Data Model VAR (2)

Produksi Sawit (CPO)						
Training	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	0.01214	1.80731	5.49573	<b>5.42178</b>	8.33221	15.26919
Testing	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	2.157	2.939	5.981	<b>5.996</b>	7.258	12.247
Keseluruhan Data	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	0.1474	1.6033	3.8985	<b>5.7336</b>	6.8437	30.4707

Dari hasil penghitungan MAPE data *training*, *testing*, dan pengepasan keseluruhan data diperoleh informasi MAPE data *testing* sudah cukup baik, yaitu sebesar 5,96% dan 5,42 untuk data *training*. Sedangkan ramalan untuk lima tahun ke depan dapat dilihat pada Tabel 24, terlihat bahwa nilai ramalan cenderung mengalami kenaikan.

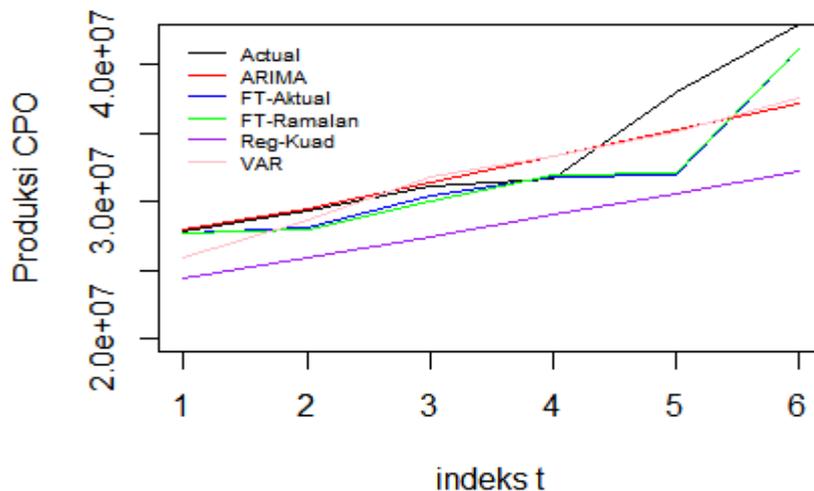
**Tabel 24.** Peramalan VAR (2) Produksi Sawit (CPO) 2019-2023

Fcst	produksi sawit (CPO)	CI
[1,]	44665772	
[2,]	48832689	
[3,]	52889350	
[4,]	58740730	
[5,]	63835557	

**Tabel 25.** Rekapitulasi MAPE *Training*, *Testing*, dan Pengepasan Keseluruhan Data Model ARIMA, Regresi, Fungsi Transfer, dan VAR

PRODUKSI SAWIT (CPO)							
ARIMA	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	143698.3	720033.1	276083.9	3.413579	<b>4.385792</b>	0.4760483	-0.0612822
Test set	1006691.8	2660352.4	1790191.7	2.135212	<b>4.658623</b>	3.0868068	NA
Total set	366045.6	1327388	537399.3	5.532319	<b>6.189547</b>	0.6376766	-0.1010958
Regresi Kuadrat		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Traning set		0.7005	12.0536	25.9881	<b>124.5975</b>	151.6791	1260.5612
Test set		8.542	11.459	11.787	<b>14.600</b>	17.489	24.747
Total set		0.004	8.528	28.350	<b>169.075</b>	189.252	1935.404
Fungsi Transfer	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	314038.5	997137.4	443924.6	6.134815	<b>7.541092</b>	0.7654541	-
0.1901787							
Test (input aktual)	1600261	2623747	1645449	4.344065	<b>4.486476</b>	0.5447921	-
0.1653447							
Test (input ramalan)	1611435	2594883	1698911	4.407061	<b>4.682741</b>	0.5624927	-0.233611
Total set	345420.2	1261052	546321.9	4.29557	<b>6.032967</b>	0.6484784	-
0.1917019							
VAR		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Training set		0.01214	1.80731	5.49573	<b>5.42178</b>	8.33221	15.26919
Testing set		2.157	2.939	5.981	<b>5.996</b>	7.258	12.247
Total set		0.1474	1.6033	3.8985	<b>5.7336</b>	6.8437	30.4707

Dari hasil rekapitulasi nilai MAPE *training*, *testing*, dan pengepasan keseluruhan data model ARIMA, Regresi, Fungsi Transfer, dan VAR, terlihat bahwa MAPE data *testing* model fungsi transfer memberikan nilai yang paling kecil yaitu 4,49% untuk peramalan produksi sawit (CPO).



**Gambar 6.** Perbandingan Plot Aktual dan Ramalan Data *Testing* menggunakan Metode ARIMA, Regresi Kuadrat, Fungsi Transfer, dan VAR Produksi Sawit (CPO) Indonesia

## KESIMPULAN DAN SARAN

### Kesimpulan

- Hasil analisis yang didapat dari empat model adalah sebagai berikut:  
Peramalan Produksi Sawit (CPO)
  - Model ARIMA terbaik adalah ARIMA (0,2,3) dengan MAPE data testing 4,66%.
  - Model regresi terbaik adalah regresi kuadratik dengan MAPE 14,6% .
  - Model Fungsi Transfer terbaik dengan peubah input harga CPO dunia adalah ARIMA (0,0,0)(1,1,0) dengan model input ARIMA (1,1,0).
  - Model VAR terbaik dengan memasukkan peubah luas areal, harga CPO dunia, volume ekspor CPO dan volume impor CPO adalah VAR(2) type “both”.
- Dari keempat model yang diujicobakan, model terbaik adalah fungsi transfer dengan nilai MAPE data testing terkecil yaitu 4,49% untuk produksi sawit (CPO).
- Hasil ramalan fungsi transfer produksi sawit (CPO) untuk lima tahun kedepan adalah :  
46 148 278 Ton, 48 297 070 Ton, 49 710 345 Ton, 50 639 803 Ton, dan 51 251 070 Ton.

### Saran

- Perlu dilakukan kajian mendalam dengan metode peramalan lainnya.
- Untuk model regresi, perlu diujicobakan menggunakan variabel bebas/peubah input selain waktu.
- Untuk model fungsi transfer dan VAR perlu diujicobakan dengan menggunakan variabel lain yang sekiranya lebih berpengaruh, baik secara teoritis maupun praktis.
- Perlu dikaji metode peramalan untuk data non parametrik, sehingga tidak terhalang oleh tidak terpenuhinya berbagai asumsi.

## DAFTAR PUSTAKA

- Direktorat Jenderal Perkebunan. 2019. Statistik Perkebunan Indonesia 2018-2020 Jakarta: Direktorat Jenderal Perkebunan, Kementerian Pertanian. Statistik Perkebunan Indonesia 2018-2020.
- Gujarati, Damodar. N dan Porter, Dawn. C. 2009. Basic Econometrics. Boston: Douglas Reiner.
- Kirchgassner, Gebhard dan Wolters, Jurgen. 2007. Introduction to Modern Time Series Analysis. Berlin: Springer Berlin Heidelberg Newyork.
- Wei, William WS. 2006. Time Series Analyis. Phladelphia: Department of Statistics The Fox School of Business and Management Temple University.
- Wooldridge, Jeffrey M. 2015. Introductory Econometrics:A Modern Approach. Boston:Cegage Learning

## KAJIAN METODE ESTIMASI PRODUKSI GULA INDONESIA

Efi Respati - *Statisticians*  
Center for Agricultural Data and Information System - Ministry of Agriculture  
Jln. Harsono RM 3 Gd D Lt 4, Ragunan-Jakarta Selatan, Indonesia  
E-mail: [efi@pertanian.go.id](mailto:efi@pertanian.go.id)

### ABSTRAK

Publikasi resmi data produksi gula di Indonesia dirilis oleh pemerintah setiap tahun dengan lag 2 tahun. Perumusan kebijakan produksi, impor, distribusi pasokan serta pengendalian harga, sangat memerlukan data terkini bahkan data hasil peramalan beberapa periode ke depan. Angka estimasi sudah disusun oleh Direktorat Jenderal Perkebunan, namun masih dengan model univariate. Dilakukan kajian metode estimasi produksi gula Indonesia dengan pendekatan metode ARIMA, Regresi, Fungsi Transfer dan model Vector Autoregression (VAR) menggunakan software R-Studio untuk pengolahan datanya. Hasil uji statistik menunjukkan bahwa Fungsi Transfer merupakan metode terbaik untuk estimasi produksi gula dengan nilai MAPE sebesar 5,30% dan data hasil estimasi produksi tahun 2020-2024 yang cukup realistis.

*Kata kunci: gula, ARIMA, regresi, fungsi transfer, var, MAPE*

### ABSTRACT

*The official publication of sugar production data in Indonesia is released by the government annually which has 2 years lag. Formulation of production, import, supply distribution and price control policies, requires the current data and even the forecast data for the next few periods. Estimated figures have been compiled by the Directorate General of Plantation, resulted from the univariate model. This paper compares the forecasting of sugar production in Indonesia by the ARIMA, Regression, Transfer Function and Vector Autoregression (VAR) model approaches with the R-Studio software. The best method is Fungsi Transfer with MAPE = 5,30% and the sugar production estimated is more realistic.*

*Keywords: projection model, arima, regression, transfer function, var.*

### PENDAHULUAN

Kegiatan pengumpulan tebu/gula untuk perkebunan rakyat di Indonesia saat ini mengacu pada Buku Pedoman Pelaksanaan Pengelolaan Data Komoditas Perkebunan yang diterbitkan oleh Direktorat Jenderal Perkebunan Tahun 2013 (Edisi Kedua). Berdasarkan buku pedoman tersebut, komoditas tebu/gula merupakan tanaman semusim, sehingga pengumpulan datanya dilakukan semesteran, namun demikian rilis data resmi dilakukan tahunan. Pengumpulan data tebu/gula untuk perkebunan besar, baik Perkebunan Besar Negara maupun Perkebunan Besar Swasta dilakukan oleh Badan Pusat Statistik melalui pelaporan oleh Pabrik Gula secara online. Rilis angka tahunan dilakukan secara bersama-sama antara Kementerian Pertanian dan BPS serta sinkronisasi akhir dengan Perusahaan Gula.

Ketersediaan data tebu/gula yang dirilis secara resmi oleh Direktorat Jenderal Perkebunan saat ini adalah data dengan lag t-2, yakni pada tahun 2020 baru tersedia angka tetap tahun 2018. Disamping angka tetap, Direktorat Jenderal Perkebunan merilis juga Angka Sementara dengan lag t-1, karena belum semua daerah/perusahaan gula melaporkan realisasi produksi pada periode tersebut. Pada

periode yang sama, Direktorat Jenderal Perkebunan juga merilis Angka Estimasi tahun ke-t, yang diperoleh dengan menggunakan metode peramalan univariate.

Makalah ini membahas kajian metode ARIMA, Regresi, Fungsi Transfer dan VAR yang bisa diterapkan dalam memodelkan produksi gula Indonesia, serta menentukan metode yang paling akurat dalam melakukan peramalan produksi gula 5 (lima) tahun kedepan.

## METODOLOGI

### 1. Sumber Data

**Tabel 1. Cakupan Data**

No	Variabel	Periode data	Sumber data	Level data
1	Produksi Gula	1972-2019	Ditjen Perkebunan	Nasional
2	Harga Internasional Gula - Pelelangan EU - Pelelangan US	1972-2019	World Bank	Internasional
3	Ekspor Molases	1972-2019	BPS	Nasional
4	Impor GKP	1972-2019	BPS	Nasional

### 2. Software

Software yang digunakan dalam menyusun makalah ini adalah software *R-Studio*. Keunggulan software R-Studio adalah software yang *open source*, sehingga tidak memerlukan biaya untuk pembelian maupun perpanjangan lisensi.

### 3. Tinjauan Literatur

Metode statistik yang dikembangkan oleh para ahli untuk melakukan peramalan data sangat beragam, baik peubah tunggal maupun peubah ganda, diantaranya metode ARIMA (*Auto Regressive Integrated Moving Average*), Regresi, Fungsi Transfer dan VAR (*Vector Autoregression*).

#### a. *Autoregressive Integrated Moving average (ARIMA)*

*Autoregressive Integrated Moving average (ARIMA)* atau biasa disebut juga dengan metode time series Box Jenkins, sangat sesuai digunakan untuk melakukan peramalan jangka pendek, sementara untuk peramalan jangka panjang kurang baik ketepatannya. Metode ARIMA merupakan metode yang hanya menggunakan peubah dependen dan mengabaikan peubah independen sewaktu melakukan peramalan.

Metode ARIMA dibagi kedalam tiga kelompok model, yaitu *autoregressive model (AR)*, *moving average model (MA)* dan model campuran yang memiliki karakteristik kedua model di atas yaitu *autoregressive integrated moving average (ARIMA)*.

#### 1) *Autoregressive Model (AR)*

AR adalah suatu model yang menjelaskan pergerakan suatu peubah melalui peubah itu sendiri di masa lalu.

Model *autoregressive* orde ke- $p$  dapat ditulis sebagai berikut:

ARIMA ( $p,0,0$ )

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t \dots \dots \dots (1)$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $Y_{t-p}$  = data *time series* pada kurun waktu ke (t-P)
- $\mu$  = suatu konstanta
- $\theta_1 \dots \theta_p$  = parameter autoregressive ke-p
- $\varepsilon_t$  = nilai kesalahan pada waktu ke t

**2) Moving Average Model (MA)**

MA adalah suatu model yang melihat pergerakan peubahnya melalui sisaannya di masa lalu.

Bentuk model MA dengan ordo q atau MA (q) atau model ARIMA (0,d,g) dapat ditulis sebagai berikut:

$$Y_t = \mu - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t \dots \dots \dots (2)$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $\phi_1 \dots \phi_q$  = parameter-parameter moving average
- $\varepsilon_{t-q}$  = nilai kesalahan pada waktu ke (t-q)

**3) Autoregressive Intergrated Moving Everage (ARIMA)**

ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampau. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t \dots \dots \dots (3)$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $Y_{t-p}$  = data *time series* pada kurun waktu ke (t-P)
- $\mu$  = suatu konstanta
- $\theta_1 \theta_q \phi_1 \phi_n$  = parameter-parameter model
- $\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

**b. Model Regresi**

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih peubah. Pada analisis regresi, dibedakan menjadi dua, yaitu peubah respon atau biasa juga disebut peubah bergantung (*dependent variable*) dan peubah *explanatory* atau biasa disebut penduga (*predictor variable*) atau disebut juga peubah bebas (*independent peubah*). Model regresi digunakan untuk meramalkan atau memprediksi nilai masa depan dari peubah respon berdasarkan nilai-nilai tertentu yang diberikan oleh peubah prediktor.

Asumsi yang mendasari penggunaan analisis regresi linier adalah:

5. Sisaan mengikuti fungsi distribusi normal;

6. Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas);
7. Tidak terdapat autokorelasi antara sisaan untuk setiap data pengamatan; dan
8. Tidak terdapat multikolinearitas antara peubah respon.

- **Analisis Regresi Linier**

a. **Analisis Regresi Linier Sederhana**

Model regresi linier sederhana melibatkan satu peubah prediktor dan diformulasikan dalam model matematis sebagai berikut:

$$y = \beta_0 + \beta_1 x + \varepsilon \dots\dots\dots(4)$$

dimana:

- $y$  = peubah respon/dependen/terikat
- $x$  = peubah prediktor/independen/bebas
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi peubah bebas
- $\varepsilon$  = error/residu/sisaan

Parameter model atau koefisien regresi yaitu  $\beta_0$  dan  $\beta_1$  masing-masing memiliki interpretasi sebagai intercept dan slope dari suatu garis lurus.  $\beta_1$  mengukur perubahan rata-rata dari peubah respon  $y$  untuk setiap perubahan dari peubah prediktor  $x$ . Parameter ini biasanya tidak diketahui dan harus diprediksi dari data sampel.  $\varepsilon$  merupakan kesalahan statistik, didefinisikan sebagai sisa atau perbedaan hasil antara nilai data pengamatan peubah respon terhadap nilai peubah respon hasil prediksi.

**Menaksir Parameter  $\beta_0$  dan  $\beta_1$**

Misalkan kita mempunyai  $n$  pasangan observasi, katakan  $(y_1, x_1), (y_2, x_2) \dots (y_n, x_n)$ , data ini dapat digunakan untuk memprediksi parameter  $\beta_0$  dan  $\beta_1$  sehingga jumlah kuadrat dari deviasi antara observasi-observasi dan garis regresi menjadi minimum. Kita gunakan persamaan diatas untuk menaksir nilai parameter tersebut,

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \dots\dots\dots(5)$$

dengan meminimumkan fungsi kuadrat tersebut maka akan diperoleh taksiran untuk  $\beta_0$  dan  $\beta_1$  sebagai berikut.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \left(\frac{\sum_{i=1}^n x_i}{n}\right) \left(\frac{\sum_{i=1}^n y_i}{n}\right)}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Sedangkan varian untuk intercept dan slope adalah:

$$v(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \quad v(\hat{\beta}_0) = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \right]$$

$$\sigma^2 = \frac{\left( \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \right) - \hat{\beta}_1 \left( \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n} \right)}{n - p}$$

dimana n = jumlah pengamatan

p = jumlah parameter (untuk regresi linier sederhana, p = 2)

- **Analisis Regresi Linier Berganda**

Jika terdapat k peubah prediktor, maka disebut dengan regresi linier berganda dengan model matematis sebagai berikut:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \dots\dots\dots(6)$$

dengan

- y** = peubah respon/dependen/terikat
- $x_1, x_2, \dots, x_k$  = peubah prediktor/independen/bebas
- $\beta_0$**  = konstanta/intercept
- $\beta_1, \beta_2, \dots, \beta_k$  = koefisien regresi peubah bebas
- $\varepsilon$**  = error/residu/sisaan

**c. Metode Fungsi Transfer**

Dalam bidang pertanian, produksi komoditas pertanian sangat dipengaruhi oleh peubah lainnya seperti serangan OPT, penggunaan saprodi, gejolak harga komoditas tersebut atau komoditas lainnya, dan lainnya. Apabila peramalan produksi hanya didasarkan pada besarnya produksi komoditas yang bersangkutan saja tanpa memperhatikan faktor-faktor yang mempengaruhinya, maka informasi untuk pembuatan perencanaan menjadi tidak lengkap, sehingga tujuan peramalan menjadi tidak tercapai secara utuh. Salah satu upaya menganalisis data deret waktu multivariat agar diperoleh hasil yang dapat memberikan informasi yang lengkap dan simultan, adalah dengan menggunakan model Fungsi Transfer.

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau Yt) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri (Yt) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau Xt) dengan deret output tersebut. Model fungsi transfer merupakan fungsi dinamis yang pengaruhnya tidak hanya pada hubungan linier antara deret input dengan deret output pada waktu ke-t, tetapi juga pada waktu t+1, t+2, ..., t+k. Hubungan seperti ini pada fungsi transfer dapat menimbulkan delay (waktu senjang) antara peubah input dan peubah output.

Tujuan pemodelan fungsi transfer adalah untuk menetapkan model yang sederhana, yang menghubungkan deret output (Yi) dengan deret input (Xi) dan gangguan/noise(ni). Wei (1994) juga menjelaskan bahwa di dalam fungsi transfer terdapat rangkaian output yang mungkin

dipengaruhi oleh rangkaian *multiple input*. Pada kasus *single input* peubah, dapat menggunakan metode korelasi silang yang dianjurkan oleh Box and Jenkins (1976). Teknik ini juga dapat digunakan ketika terdapat *single input* peubah yang lebih dari satu selama antar variable *input* tidak berkorelasi silang. Jika beberapa atau semua peubah input berkorelasi silang maka teknik prewhitening atau metode korelasi silang tidak dapat digunakan secara langsung. Alasan utama bagi perlunya suatu perencanaan atau peramalan adalah adanya tenggang waktu pengambilan keputusan yang dapat berkisar dari beberapa hari atau sampai beberapa tahun. Pada analisis fungsi transfer untuk peramalan deret berkala univariate, terdapat deret berkala output yang diperkirakan dipengaruhi oleh deret berkala input dan input-input lain yang digabungkan dalam satu kelompok yang disebut gangguan (noise). Deret input mempengaruhi deret output melalui sebuah fungsi transfer yang mendistribusikan pengaruhnya secara dinamis melalui beberapa periode waktu yang akan datang dengan persentase tertentu yang disebut sebagai bobot respons impuls atau bobot fungsi transfer.

Model umum Fungsi Transfer:

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \quad \dots\dots\dots(7)$$

Dimana:

- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$
- $q \rightarrow$  ordo MA bagi noise  $N_t$

**d. Vector Autoregression (VAR)**

*Vector Autoregression* (VAR) dikemukakan pertama kali oleh Christopher Sims (1980). Sims mengembangkan model ekonometri dengan mengabaikan pengujian asumsi secara apriori. VAR dikembangkan oleh Sims sebagai kritik atas metode simultan. Jumlah peubah yang besar dan klasifikasi endogen dan eksogen pada metode simultan merupakan dasar dari kritik tersebut. Penggunaan pendekatan struktural atas pemodelan persamaan simultan biasanya menerapkan teori ekonomi di dalam usahanya untuk mendeskripsikan hubungan antar peubah yang ingin diuji. Akan tetapi sering ditemukan bahwa teori ekonomi saja ternyata tidak cukup kaya di dalam menyediakan spesifikasi yang ketat dan tepat atas hubungan dinamis antar peubah (Gujarati, 2010). Model VAR merupakan jalan keluar atas permasalahan ini melalui pendekatan non-struktural atau tidak mendasarkan pada teori ekonomi tertentu (*atheoretical*). Metode VAR memperlakukan seluruh peubah secara simetris tanpa memperlakukan peubah dependen dan independen atau dengan kata lain model ini memperlakukan seluruh peubah sebagai peubah endogen., karena pada kenyataannya suatu peubah dapat bertindak sebagai endogen atau dependen dari suatu peubah yang lainnya (Gujarati, 2010).

Kelebihan dalam penggunaan metode VAR (Gujarati, 2010):

- e. Kemudahan dalam penggunaan, tidak perlu mengkhawatirkan tentang penentuan peubah endogen dan peubah eksogen.
- f. Kemudahan dalam estimasi, metode *Ordinary Least Square* (OLS) dapat diaplikasikan pada tiap persamaan secara terpisah.
- g. *Forecast* atau peramalan yang dihasilkan pada beberapa kasus ditemukan lebih baik daripada yang dihasilkan oleh model persamaan simultan yang kompleks.
- h. Perangkat estimasi yang digunakan adalah *Impulse Respon Function* (IRF) untuk melacak respon dari peubah dependen dalam sistem VAR terhadap *shock* dari *error term* dan *Variance*

*Decomposition* yang memberikan informasi mengenai pentingnya masing-masing *error term* dalam mempengaruhi peubah-peubah dalam VAR.

Di sisi lain, terdapat beberapa kritik terhadap model VAR menyangkut permasalahan berikut (Gujarati, 2010) :

- 6) Model VAR merupakan model yang *atheoretic* atau tidak berdasarkan teori, hal ini tidak seperti pada persamaan simultan. Pada persamaan simultan, pemilihan peubah yang akan dimasukkan dalam persamaan memegang peranan penting dalam mengidentifikasi model.
- 7) Pada model VAR penekanannya terletak pada *forecasting* atau peramalan sehingga model ini kurang cocok digunakan dalam menganalisis kebijakan.
- 8) Permasalahan yang besar dalam model VAR adalah pada pemilihan *lag length* atau panjang lag yang tepat. Karena semakin panjang lag, maka akan menambah jumlah parameter yang akan bermasalah pada *degrees of freedom*.
- 9) Peubah yang tergabung pada model VAR harus stasioner. Apabila tidak stasioner, perlu dilakukan transformasi bentuk data, misalnya melalui *first difference*.
- 10) Sering ditemui kesulitan dalam menginterpretasi tiap koefisien pada estimasi model VAR, sehingga sebagian besar peneliti melakukan interpretasi pada estimasi fungsi *impulse respon*.

### Estimasi Model VAR

Enders (2004) menjelaskan, ketika peneliti tidak memiliki kepastian untuk menentukan bahwa suatu peubah adalah eksogen, maka suatu perluasan analisis fungsi perpindahan alami akan memperlakukan masing-masing peubah secara simetris. Sebagai contoh, pada kasus-kasus peubah yang membiarkan alur waktu atau *time path*  $\{s_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{y_t\}$  dan membiarkan *time path*  $\{y_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{s_t\}$ .

Di dalam sistem *bivariate*, hubungan tersebut dapat digambarkan seperti pada persamaan (1.4) di bawah ini:

$$\begin{aligned} s_t &= b_{10} - b_{12}y_t + \gamma_{11}s_{t-1} + \gamma_{12}y_{t-1} + \varepsilon_{s_t} \\ y_t &= b_{20} - b_{21}s_t + \gamma_{21}s_{t-1} + \gamma_{22}y_{t-1} + \varepsilon_{y_t} \end{aligned} \quad \dots \dots \dots (8)$$

Dengan mengasumsikan bahwa kedua peubah  $s_t$  dan  $y_t$  adalah stasioner:  $\varepsilon_{s_t}$  dan  $\varepsilon_{y_t}$  adalah *disturbances* yang memiliki rata-rata nol dan matriks kovarians terbatas atau bersifat *white noise* dengan standar deviasi yang berurutan  $\sigma_s$  dan  $\sigma_y$ ; serta  $\{\varepsilon_{s_t}\}$  dan  $\{\varepsilon_{y_t}\}$  adalah *disturbances* yang independen dengan rata-rata nol dan kovarian terbatas (*uncorrelated white-noise disturbances*). Kedua persamaan di atas merupakan orde pertama VAR, karena panjang *lag* nya hanya satu. Agar persamaan (4.1) lebih mudah dipahami dan digunakan sebagai alat analisis maka ditransformasikan dengan menggunakan matriks aljabar, dan hasilnya dapat dituliskan secara bersama seperti pada persamaan di bawah ini:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} s_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix}$$

Atau dengan bentuk lain:

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t \dots\dots\dots(9)$$

Dimana:

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \quad x_t = \begin{bmatrix} s_t \\ y_t \end{bmatrix} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix}$$

Dengan melakukan pengalihan antara persamaan (4.2) dengan  $B^{-1}$  atau invers matriks B, maka akan dapat ditentukan model VAR dalam bentuk standar, seperti dituliskan pada persamaan di bawah ini:

$$x_t = A_0 + A_1 x_{t-1} + \ell_t \dots\dots\dots(10)$$

$$A_0 = B^{-1} \Gamma_0$$

dimana  $A_1 = B^{-1} \Gamma_1$

$$\ell_t = B^{-1} \varepsilon_t$$

Untuk tujuan notasi, maka  $\{a_{i0}\}$  dapat didefinisikan sebagai elemen ke-i dari vektor  $A_0$ ;  $\{a_{ij}\}$  sebagai elemen dalam baris ke-i dan baris ke-j dari matriks  $A_1$ ; dan  $\{e_{it}\}$  sebagai elemen ke-i dari vektor  $e_t$ . Dengan menggunakan notasi baru yang telah dijelaskan sebelumnya, maka persamaan (4.3) dapat ditulis menjadi:

$$\begin{aligned} s_t &= a_{10} + a_{11}s_{t-1} + a_{12}y_{t-1} + e_{1t} \\ y_t &= a_{20} + a_{21}s_{t-1} + a_{22}y_{t-1} + e_{2t} \dots\dots\dots(11) \end{aligned}$$

**Fungsi Impulse Response**

Fungsi *impulse response* menggambarkan tingkat laju dari *shock* peubah yang satu terhadap peubah yang lainnya pada suatu rentang periode tertentu. Sehingga dapat dilihat lamanya pengaruh dari *shock* suatu peubah terhadap peubah lain sampai pengaruhnya hilang atau kembali ke titik keseimbangan. Analisis fungsi *impulse respon* dapat dituliskan dalam bentuk *Vector Moving Avarage (VMA)* dari bentuk standar VAR pada persamaan (4.4).

$$\begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} - \\ s \\ - \\ y \end{bmatrix} + \sum_{i=1}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \dots\dots\dots(12)$$

dimana  $s_t$  dan  $y_t$  memiliki hubungan dengan  $e_{1t}$  dan  $e_{2t}$  secara berurutan. Selanjutnya dengan menggunakan operasi aljabar matriks maka *vector error* dapat ditentukan sebagai berikut :

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \frac{1}{1 - b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{s_{t-i}} \\ \varepsilon_{y_{t-i}} \end{bmatrix} \dots\dots\dots(13)$$

dengan menggabungkan persamaan (4.5) dan (4.6) akan didapat:

$$\begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} - \\ s_t \\ - \\ y_t \end{bmatrix} + \frac{1}{1 - b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{s_{t-i}} \\ \varepsilon_{y_{t-i}} \end{bmatrix} \dots\dots\dots(14)$$

Persamaan (4.7) dapat disederhanakan dengan mendefinisikan matriks 2x2  $\Phi_i$  dengan elemen  $\Phi_{jk}$  (i) seperti persamaan berikut :

$$\Phi_i = A_1^i / (1 - b_{12}b_{21}) \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \dots\dots\dots(15)$$

sehingga diperoleh bentuk matriks persamaan fungsi *impulse respon*:

$$\begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} \bar{s} \\ \bar{y} \end{bmatrix} + \sum_{i=0}^n \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{st-1} \\ \varepsilon_{yt-1} \end{bmatrix} \dots\dots\dots(16)$$

dimana :

$\Phi_{ij}(i)$  = efek dari *structural shock* pada s dan y

$\Phi_{ij}(0)$  = *impact multipliers*

$\sum \Phi_{ij}(i)$  = *cumulative multipliers*

$\sum \Phi_{ij}(i)$  pada saat  $n \rightarrow \infty$  = *long run multipliers*

### Variance Decomposition

*Variance decomposition* atau disebut juga *forecast error variance decomposition* merupakan perangkat pada model VAR yang akan memisahkan variasi dari sejumlah peubah yang diestimasi menjadi komponen-komponen *shock* atau menjadi peubah *innovation*, dengan asumsi bahwa peubah-peubah *innovation* tidak saling berkorelasi. Kemudian, *variance decomposition* akan memberikan informasi mengenai proporsi dari pergerakan pengaruh *shock* pada sebuah peubah terhadap *shock* peubah yang lain pada periode saat ini dan periode yang akan datang.

Bentuk VMA dari peubah x pada satu periode ke depan dapat dituliskan sbb.:

$$x_{t+1} = \bar{x} + \sum_{i=0}^{\infty} \phi_i \varepsilon_{t+1-i} \dots\dots\dots(17)$$

*Forecast error* pada satu periode kedepan adalah:

$$E_t x_{t+1} = \bar{x} + \sum_{i=1}^{\infty} \phi_i \varepsilon_{t+1-i} \dots\dots\dots(18)$$

*Forecast* satu periode ke depan dilambangkan dengan  $\Phi_0 \varepsilon_{t+1}$ . *Forecast error* pada periode n ke depan adalah:

$$x_{t+n} - E_t x_{t+n} = \bar{x} + \sum_{i=0}^{n-1} \phi_i \varepsilon_{t+n-i} \dots\dots\dots(19)$$

*Forecast error* pada n periode ke depan untuk peubah s adalah:

$$s_{t+n} - E_t s_{t+n} = \phi_{11}(0) \varepsilon_{st+n} + \phi_{11}(1) \varepsilon_{st+n-1} + \dots + \phi_{11}(n-1) \varepsilon_{yt+1} \\ + \phi_{12}(0) \varepsilon_{yt+n} + \phi_{12}(1) \varepsilon_{yt+n-1} + \dots + \phi_{12}(n-1) \varepsilon_{yt+1} \dots\dots\dots(20)$$

*Variance dari forecast error*  $s_{t+n}$  periode n ke depan adalah  $\sigma_s(n)^2$  dimana:

$$\sigma_s(n)^2 = \sigma_s^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2] + \sigma_y^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2] \dots\dots\dots(21)$$

Forecast error variance decomposition adalah proporsi dari  $\sigma_s(n)^2$  terhadap shock  $s$  dan shock  $y$ . Sehingga forecast error variance decomposition pada shock  $s$  adalah:

$$\sigma_s^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2] / \sigma_s(n)^2 \dots\dots\dots(22)$$

sedangkan forecast error variance decomposition pada shock  $y$  adalah:

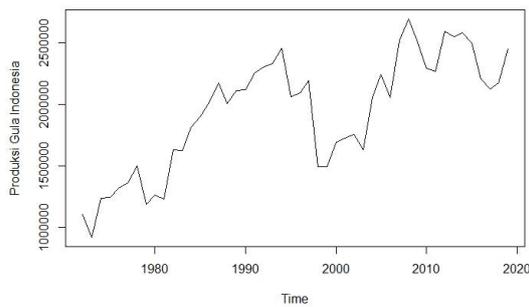
$$\sigma_y^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2] / \sigma_y(n)^2 \dots\dots\dots(23)$$

## HASIL PEMBAHASAN

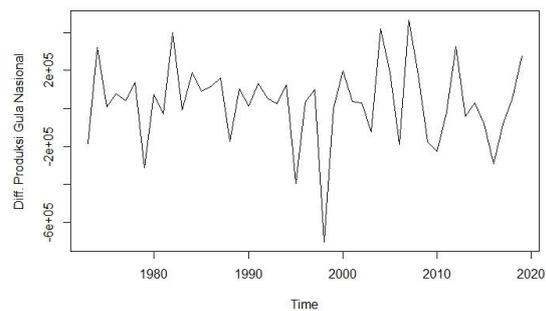
### 1. Estimasi Metode ARIMA

Eksplorasi data series produksi gula Indonesia tahun 1972-2019 untuk pemodelan ARIMA dilakukan dengan membuat plot data yang menunjukkan kecenderungan tren meningkat (Gambar 1). Hasil uji Augmented Dickey-Fuller untuk data produksi gula dengan pembedaan (*differencing*) hingga tingkat 2 baru menunjukkan data sudah stasioner seperti nilai statistic uji lebih kecil dari nilai kritis sbb.:

```
Augmented Dickey-Fuller Test Unit Root Test:
Value of test-statistic is: -5.5465
Critical values for test statistics:
    1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```



Gambar 1. Plot Produksi Gula, 1972-2019



Gambar 2. Plot Produksi Gebu dengan Differencing Tingkat 1, 1972-2019

Data series produksi untuk pemodelan dibagi menjadi series *data training* (1972 – 2013) dan series *data testing* (2014-2019). Series data *training* digunakan untuk mencari model terbaik dan akan divalidasi dengan menggunakan series data *testing*.

Penentuan ordo lag  $p$  (AR) dan  $q$  (MA) dilakukan melalui investigasi plot ACF dan PACF namun belum terlihat kriteria lag AR dan MA-nya, sehingga dilakukan dengan memanfaatkan script pada software R-Studio untuk menunjukkan 10 model tentative, dengan hasil seperti tersaji pada Tabel 2.

**Tabel 2. Sepuluh model tentative produksi gula dengan differencing tingkat 2**

No	p	q	sbc
[1,]	5	5	1,000.05
[2,]	5	4	1,000.12
[3,]	3	0	1,001.95
[4,]	2	0	1,002.01
[5,]	4	0	1,004.56
[6,]	5	2	1,004.75
[7,]	5	1	1,005.17
[8,]	5	3	1,005.89
[9,]	4	1	1,006.70
[10,]	5	0	1,008.59

Model akhir yang dipilih didasarkan pada signifikansi pada semua komponen AR dan MA, yakni model ARIMA (2,2,0).

Uji signifikansi terhadap komponen MA1 dan MA2 adalah sbb.

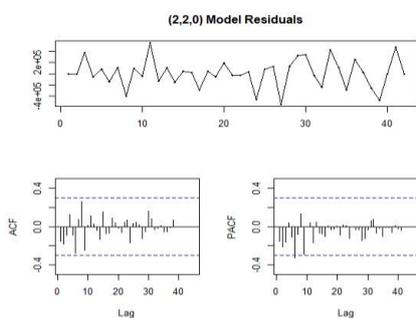
```
ARIMA(2,2,0)
Coefficients:
      ar1      ar2
s.e.  -0.7594  -0.5010
      0.1384   0.1379

sigma^2 estimated as 6.514e+10:  log likelihood=-554.17
AIC=1114.33  AICc=1115  BIC=1119.4

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  -0.75939    0.13840  -5.4870 4.087e-08 ***
ar2  -0.50105    0.13794  -3.6325 0.0002807 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pemeriksaan sisaan terhadap model ARIMA (2,2,0) menunjukkan pola terdistribusi normal serta pola ACF dan PACF sisaan yang tidak nyata, seperti tersaji pada Gambar 3. Hasil Uji Ljung-Box yang mengindikasikan autokorelasi sisaan tidak signifikan hingga 30 lag (Tabel 3).



**Gambar 3. Plot sisaan ARIMA (2,2,0) Peubah Produksi Gula**

**Tabel 3. Hasil Uji Ljung-Box Arima (0,1,2)**

lag	statistic	df	p-value
5	4.118287	5	0.53251
10	15.686355	10	0.10897
15	19.353877	15	0.19814
20	20.909063	20	0.40250
25	24.926748	25	0.46649
30	31.738055	30	0.37981

Setelah asumsi sisaan telah memenuhi syarat, maka dilakukan peramalan untuk set data testing yang menghasilkan nilai MAPE sebesar 10,07% untuk data training dan 25,65% untuk data testing.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	8663.095	242775.6	178651.2	0.3102701	10.06978	1.106790	-0.151378
Test set	-566284.459	644198.2	566284.5	-25.6505516	25.65055	3.508276	NA

Pengepasan model ARIMA (2,2,0) untuk seluruh data diperoleh nilai MAPE yang cukup bagus sebesar 9,73% untuk series data *training*.

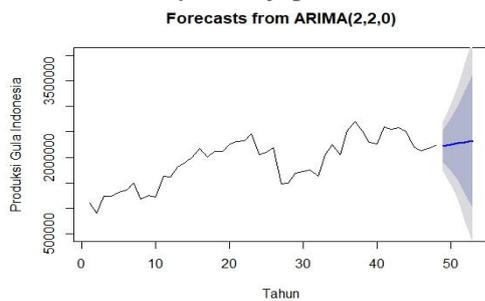
```
Series: oiahtebu[, "Prod_Total"]
ARIMA(2,2,0)

Coefficients:
      ar1      ar2
-0.7246 -0.4720
s.e.   0.1306   0.1293

sigma^2 estimated as 6.095e+10: log likelihood=-635.81
AIC=1277.61  AICc=1278.18  BIC=1283.1

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 4807.026 236371.7 176544.5 0.184322 9.727602 1.151006 -0.1223915
```

Hasil peramalan produksi gula tahun 2020 – 2024 menggunakan metode ARIMA (2,2,0) serta plot hasil ramalannya tersaji pada Tabel 4 dan Gambar 4.



**Tabel 4. Hasil ramalan produksi gula Indonesia dengan metode ARIMA (2,2,0)**

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
49	2,215,708	1,899,315	2,532,100	1,731,828	2,699,588
50	2,250,772	1,737,998	2,763,545	1,466,552	3,034,991
51	2,282,416	1,561,012	3,003,819	1,179,125	3,385,707
52	2,294,634	1,284,267	3,305,001	749,412	3,839,857
53	2,322,543	1,015,069	3,630,016	322,935	4,322,151

Gambar 4. Plot hasil ramalan produksi gula dengan metode ARIMA (2,2,0)

## 2. Estimasi Metode Regresi

Estimasi data produksi gula menggunakan metode regresi linear dan kuadratik dilakukan dengan peubah Tahun (t) sebagai peubah bebas.

### a. Regresi linear sederhana

Hasil regresi linear sederhana produksi gula dengan peubah tahun menunjukkan bahwa baik intercept maupun peubah tahun nyata pada taraf kepercayaan 100%, dengan  $R^2=60,25\%$ .

Persamaan regresi linearnya adalah : **Produksi = -57115822 + 29610Tahun**

```
Call:
lm(formula = Prod_Total ~ Tahun, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-580223 -201948  44617  222765  527774

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -57115822   7485750  -7.630 2.16e-09 ***
Tahun         29610       3756    7.883 9.61e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 305600 on 41 degrees of freedom
Multiple R-squared:  0.6025, Adjusted R-squared:  0.5928
F-statistic: 62.15 on 1 and 41 DF, p-value: 9.606e-10
```

Hasil permalan produksi gula tahun 2020 – 2024 dengan menggunakan metode regresi linear tersebut adalah sbb.:

**Tabel 5. Hasi ramalan produksi Gula Metode Regresi Linear**

No	Tahun	Produksi (Ton)
1	2020	2,696,378
2	2021	2,725,988
3	2022	2,755,598
4	2023	2,785,208
5	2024	2,814,818

Nilai MAPE data training sebesar 14,28%  
 Min. 1st Qu. Median Mean 3rd Qu. Max.  
 0.4493 5.4345 12.5968 14.2773 20.4877 42.5671

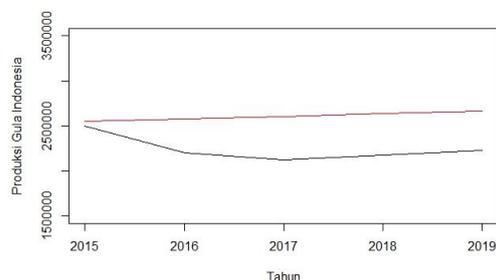
Nilai MAPE data testing sebesar 16,56%

Min. 1st Qu. Median Mean 3rd Qu. Max.  
 1.998 16.915 19.726 16.557 21.263 22.881

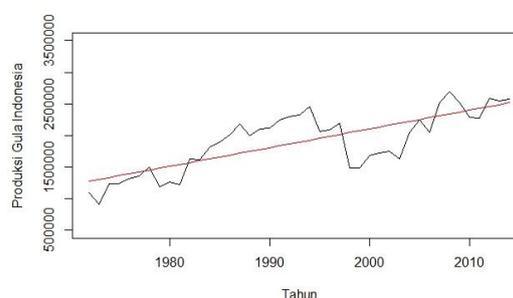
Uji normalitas dilakukan terhadap model regesi tersebut dan hasilnya adalah:

- Uji normalitas menggunakan Uji Jarque-Bera, diperoleh informasi bahwa data tidak menyebar normal, yakni nilai JB = 1.6417 dan p-value = 0.274.
- Terindikasi ada autokorelasi dengan ditunjukkan dari hasil uji Durbin-Watson sebesar 0.7425684, dan Statistic p-value = 0.5059807.
- Uji Chi-Square menunjukkan Non-constant Variance Score Test  
 Variance formula: ~ fitted.values  
 chisquare = 0.1753094, Df = 1, p = 0.67544

Grafik pola garis regresi pada data training dan testing adalah sbb.:



Gambar 5. Plot hasil regresi dengan data testing



Gambar 5. Plot hasil regresi dengan data training

## b. Regresi kuadrat

Hasil regresi kuadrat produksi gula dengan peubah bebas tahun menunjukkan bahwa baik intercept maupun peubah tahun nyata pada taraf kepercayaan 100%, sementara peubah tahun kuadrat tidak nyata, dengan R2 = 62,79%. Persamaan regresi kuadrat adalah :

$$\text{Produksi} = 1117457,2 + 52619,7\text{Tahun} - 547,9\text{Tahun}^2$$

Call:  
 lm(formula = Prod\_Total ~ t + t2, data = train)

Residuals:  
 Min 1Q Median 3Q Max  
 -644870 -205102 21680 223727 443952

Coefficients:  
 Estimate Std. Error t value Pr(>|t|)  
 (Intercept) 1117457.2 130842.0 8.541 1.49e-10 \*\*\*  
 t 52619.7 14411.0 3.651 0.000748 \*\*\*  
 t2 -547.9 331.7 -1.651 0.106482  
 ---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 299400 on 40 degrees of freedom

Multiple R-squared: 0.6279, Adjusted R-squared: 0.6093  
 F-statistic: 33.75 on 2 and 40 DF, p-value: 2.591e-09

Hasil permalan produksi gula tahun 2020 – 2024 dengan menggunakan metode regresi kuadratik adalah sbb.:

Tabel 6. Hasil Ramalan Produksi Gula Metode Regresi Kuadratik

No	Tahun	Produksi (Ton)
1	2020	2,380,315
2	2021	2,378,692
3	2022	2,375,974
4	2023	2,372,160
5	2024	2,367,250

Nilai MAPE data training= 14,28%

Min. 1st Qu. Median Mean 3rd Qu. Max.  
 0.4493 5.4345 12.5968 14.2773 20.4877 42.5671

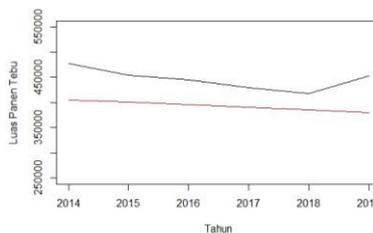
Nilai MAPE data testing= 8,22%

Min. 1st Qu. Median Mean 3rd Qu. Max.  
 5.239 6.885 7.596 8.220 9.396 11.984

Uji yang dilakukan terhadap model regresi tersebut dan hasilnya adalah:

- Uji normalitas menggunakan Uji Jarque-Bera, diperoleh informasi bahwa data menyebar normal, yakni nilai JB = 3.2481 dan p-value = 0.087.
- Terindikasi ada autokorelasi dengan ditunjukkan dari hasil uji Durbin-Watson sebesar 0.7227986, dan statistic p-value = 0.5410567.
- Uji Chi-Square menunjukkan Non-constant variance Score Test  
 Variance formula: ~ fitted.values  
 chisquare = 3.429739, Df = 1, p = 0.064032

Grafik pola garis regresi pada data training dan testing adalah sbb.:



Gambar 5. Plot hasil regresi kuadratik dengan data testing

### 3. Estimasi Metode Fungsi Transfer

Dengan asumsi bahwa produksi gula Indonesia sangat dipengaruhi oleh harga gula di pasar global (Pelelangan Amerika Serikat), maka dilakukan pemodelan Fungsi Transfer dengan menyertakan peubah harga gula global tersebut.

Data deret input harga gula global dan data deret output produksi gula mempunyai tren meningkat dari waktu ke waktu atau tidak stasioner sehingga harus dilakukan pembedaan/*differencing* sebagai syarat untuk melakukan pemodelan Fungsi Transfer. Pembedaan tingkat 1 pada kedua data series tersebut telah stasioner berdasarkan hasil uji Augmented Dickey-Fuller, sbb.:

Hasil uji Augmented Dickey_Fuller data series produksi gula <i>differencing</i> tingkat 1 untuk data training:  value of test-statistic is: -5.3081  Critical values for test statistics: 1pct 5pct 10pct tau1 -2.62 -1.95 -1.61	Hasil uji Augmented Dickey_Fuller data series harga gula global <i>differencing</i> tingkat 1:  value of test-statistic is: -5.0342  Critical values for test statistics: 1pct 5pct 10pct tau1 -2.62 -1.95 -1.61
--	--

Penelusuran model deret input Harga Gula Global oleh R-Studio ditunjukkan bahwa **ARIMA (2,1,1)** merupakan model terbaik dengan komponen AR dan MA yang nyata, sbb.:

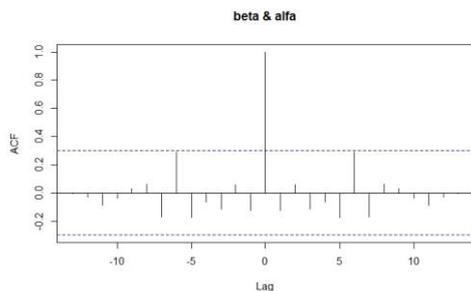
```
arima(x = train.h[, "Harga_US"], order = c(2, 1, 1))
Coefficients:
      ar1      ar2      ma1
 0.4329 -0.4655 -0.7285
s.e. 0.1707  0.1650  0.1448

sigma^2 estimated as 0.009365:  log likelihood = 37.89,  aic = -67.79

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.43286   0.17074  2.5351 0.011240 *
ar2 -0.46548   0.16503 -2.8206 0.004793 **
ma1 -0.72852   0.14478 -5.0321 4.852e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Kedua series data baik deret input (harga gula global) maupun deret output (produksi gula), mempunyai tren atau tidak stasioner, sehingga menghasilkan hubungan yang semu. Oleh karenanya, perlu dilakukan analisis korelasi silang terhadap deret input Harga Gula Global dengan Produksi Gula dengan hasil sbb.:



Gambar 6. Plot korelasi silang deret input Harga Gula Global dengan Produksi Gula

Plot korelasi disamping menunjukkan nyata pada lag=0 atau nilai b=0 yang berarti bahwa tidak ada jeda pengaruh harga gula global terhadap produksi gula Indonesia, dan nilai s=0 atau tidak ada panjang pengaruh harga gula global terhadap produksi gula Indonesia. Nilai r=0 atau tidak ada pengaruh produksi gula periode sebelumnya terhadap produksi gula periode ini karena adanya perubahan harga gula global. Hal ini mengingat data series produksi gula merupakan data tahunan.

Setelah dilakukan penelusuran model noise, diperoleh model **ARIMA (1,1,4)** sebagai model terbaik, dengan signifikansi masing-masing koefisien adalah sbb.:

```
Series: train.h[, "Prod_Total"]
Regression with ARIMA(1,1,4) errors

Coefficients:
      ar1      ma1      ma2      ma3      ma4      xreg
-0.8892  0.6427 -0.0760  0.4673  0.4078 -506999.1
s.e. 0.1900  0.2570  0.1876  0.2050  0.1939  257050.4

sigma^2 estimated as 4.746e+10:  log likelihood=-573.06
AIC=1160.11  AICc=1163.41  BIC=1172.28

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 32237.51 199330.9 158861.2 1.348258 8.987608 1.00392 -0.05838864

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1 -8.8924e-01 1.9005e-01 -4.6790 2.882e-06 ***
ma1  6.4267e-01 2.5696e-01  2.5011  0.01238 *
ma2 -7.5954e-02 1.8765e-01 -0.4048  0.68564
ma3  4.6733e-01 2.0503e-01  2.2794  0.02265 *
ma4  4.0776e-01 1.9385e-01  2.1034  0.03543 *
xreg -5.0700e+05 2.5705e+05 -1.9724  0.04857 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Peramalan berbasis Fungsi Transfer dengan derajat  $(r,b,s) = (0,0,0)$  dan model noise ARIMA (1,1,4) dilakukan terhadap data aktual (series data testing), diperoleh nilai **MAPE = 4,62%**

```
Series: test.h[, "Prod_Total"]
Regression with ARIMA(1,1,4) errors

Coefficients:
      ar1      ma1      ma2      ma3      ma4      xreg
s.e.  -0.8892  0.6427 -0.076  0.4673  0.4078 -506999.1
      0.0000  0.0000  0.000  0.0000  0.0000  0.0

sigma^2 estimated as 4.746e+10: log likelihood=-53.64
AIC=109.27 AICc=111.27 BIC=108.66
```

```
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -40302.8 132046.8 101120.8 -1.883667 4.615944 0.8396967 0.1555004
```

Untuk melakukan peramalan produksi gula Indonesia tahun 2020-2024, maka harus menggunakan metode Fungsi Transfer dimana sebelumnya dilakukan peramalan deret input harga gula global yang telah diperoleh pada penelusuran model ARIMA sebelumnya, yakni ARIMA (2,1,1), menghasilkan nilai **MAPE data testing** sebesar **5,30%** sbb.:

```
Series: test.h[, "Prod_Total"]
Regression with ARIMA(1,1,4) errors

Coefficients:
      ar1      ma1      ma2      ma3      ma4      xreg
s.e.  -0.8892  0.6427 -0.076  0.4673  0.4078 -506999.1
      0.0000  0.0000  0.000  0.0000  0.0000  0.0

sigma^2 estimated as 4.746e+10: log likelihood=-54.1
AIC=110.21 AICc=112.21 BIC=109.6
```

```
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -47017.32 148436.6 115967.2 -2.198028 5.300254 0.9629797 0.1763033
```

Hasil ramalan produksi gula tahun 2020 – 2024 menggunakan model Fungsi Transfer dengan derajat  $(r,b,s) = (0,0,0)$  dan model noise ARIMA (1,1,4) adalah sebagai berikut:

```
Time Series:
Start = 49
End = 53
Frequency = 1
[1] 2365373 2367447 2365938 2371460 2366035
```

#### 4. Estimasi Vector Auto Regression (VAR)

Model VAR merupakan alat analisis yang sangat berguna dalam memahami adanya hubungan timbal balik (interrelationship) antara peubah ekonomi maupun dalam pembentukan ekonomi yang berstruktur. Dalam kajian ini mengikutsertakan peubah produksi gula, harga gula global, ekspor molasses dan impor gula Indonesia ke dalam sistem persamaan VAR. Model VAR dapat digunakan untuk menjelaskan fenomena semua peubah yang masuk ke dalam sistem, namun dalam sub bab ini lebih dibahas fenomena terhadap peubah produksi gula Indonesia, dan menggunakan hasil model VAR untuk peramalan produksi tahun 2020-2024.

Hasil penelusuran diperoleh bahwa VAR dengan lag=2 dengan mengikutsertakan tren merupakan model terbaik, dengan signifikansi sebagai berikut:

VAR Estimation Results:

```

=====
Endogenous variables: Prod_Total, Harga_EU, Harga_US, Ekspor_Mol, Impor_Gula
Deterministic variables: both
Sample size: 46
Log Likelihood: -1723.853
Roots of the characteristic polynomial:
0.9829 0.8154 0.7234 0.7234 0.6709 0.6709 0.4754 0.4754 0.4721 0.4721
Call:
VAR(y = olahtebu[1:48, c(9, 10, 11, 12, 13)], p = 2, type = "both")

```

Estimation results for equation Prod\_Total:

```

=====
Prod_Total = Prod_Total.l1 + Harga_EU.l1 + Harga_US.l1 + Ekspor_Mol.l1 +
Impor_Gula.l1 + Prod_Total.l2 + Harga_EU.l2 + Harga_US.l2 + Ekspor_Mol.l2 +
Impor_Gula.l2 + const + trend

```

	Estimate	Std. Error	t value	Pr(> t )
Prod_Total.l1	4.206e-01	1.767e-01	2.381	0.0230 *
Harga_EU.l1	9.627e+05	6.363e+05	1.513	0.1395
Harga_US.l1	-4.254e+05	3.394e+05	-1.253	0.2186
Ekspor_Mol.l1	6.006e-01	2.453e-01	2.449	0.0196 *
Impor_Gula.l1	2.514e-02	6.489e-02	0.387	0.7009
Prod_Total.l2	-1.298e-01	1.786e-01	-0.727	0.4724
Harga_EU.l2	-9.441e+05	6.582e+05	-1.434	0.1606
Harga_US.l2	6.925e+05	3.077e+05	2.251	0.0310 *
Ekspor_Mol.l2	-1.697e-02	2.375e-01	-0.071	0.9434
Impor_Gula.l2	-1.420e-01	7.172e-02	-1.980	0.0559 .
const	6.034e+05	2.484e+05	2.429	0.0206 *
trend	2.112e+04	9.654e+03	2.187	0.0357 *

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 187900 on 34 degrees of freedom  
Multiple R-Squared: 0.8582, Adjusted R-squared: 0.8123  
F-statistic: 18.7 on 11 and 34 DF, p-value: 2.549e-11

Model VAR dengan p=2

Uji normalitas terhadap sisaan, homokedastisitas dan non autokorelasi sudah terpenuhi sebagai berikut:

```

Portmanteau Test (asymptotic)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 288.09, df = 350, p-value = 0.9932
JB-Test (multivariate)

data: Residuals of VAR object varhsheet.b2
Chi-squared = 29.747, df = 10, p-value = 0.0009423

$Skewness

Skewness only (multivariate)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 14.45, df = 5, p-value = 0.01299

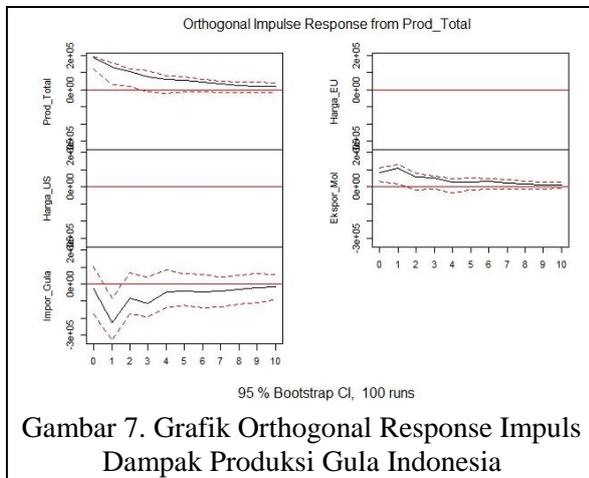
$Kurtosis

Kurtosis only (multivariate)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 15.296, df = 5, p-value = 0.009168

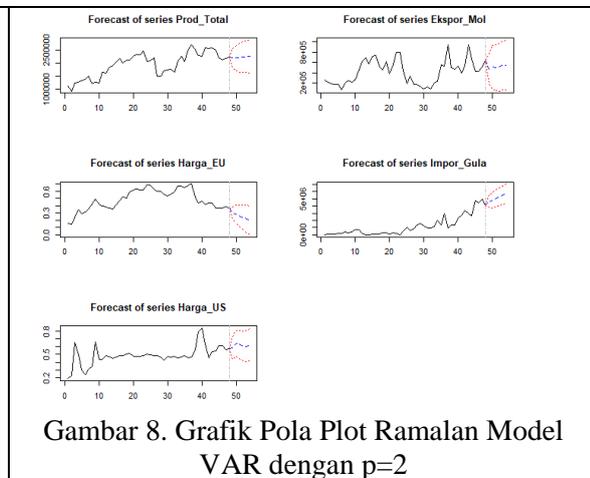
ARCH (multivariate)
data: Residuals of VAR object varhsheet.b2
Chi-squared = 615, df = 1125, p-value = 1

```

Hasil grafik orthogonal impulse response menunjukkan bahwa apabila ada perubahan produksi gula Indonesia pada tahun tertentu, maka produksi gula akan merespon hingga 3 tahun ke depan, ekspor molasses akan terpengaruh hingga 1 tahun ke depan. Perubahan produksi gula Indonesia tidak berdampak pada harga gula global maupun impor gula Indonesia.



Gambar 7. Grafik Orthogonal Response Impuls Dampak Produksi Gula Indonesia



Gambar 8. Grafik Pola Plot Ramalan Model VAR dengan p=2

Nilai MAPE data ser testing sebesar 5,97%, sementara nilai MAPE data set traning seebstar 7,32% sebagai berikut:

MAPE DATA TESTING					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.6659	1.6010	2.8525	5.9743	10.6953	14.9812
MAPE DATA TRAINING					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.3146	2.8863	5.3065	7.3280	8.6586	39.4954

Model VAR dengan p=2 dengan mempertimbangkan tren digunakan untuk melakukan peramalan data produksi gula Indonesia dengan hasil sebagai berikut:

Tabel 7. Hasil Ramalan Produksi Gula Metode VAR

No	Tahun	Produksi (Ton)	Pertumbuhan (%)
1	2019	2,227,046	
2	2020	2,192,783	-1.54
3	2021	2,173,840	-0.86
4	2022	2,171,179	-0.12
5	2023	2,203,382	1.48
6	2024	2,214,708	0.51

## SIMPULAN DAN SARAN

Ke-empat metode yang digunakan untuk memodelkan dan meramalkan data produksi gula Indonesia dapat ditabulasikan sebagai berikut:

Tabel 8. Perbandingan MAPE dan Hasil Ramalan dengan Empat Metode

No	Model	MAPE (%)		Hasil ramalan produksi gula (Ton)					Pertumbuhan 2020-2024 (%)
		Training	Testing	2020	2021	2022	2023	2024	
1	Arima (2,2,0)	10.07	25.65	2,215,708	2,250,772	2,282,416	2,294,634	2,322,543	1.19
2	Regresi								
	a. Linear Sederhana	14.28	16.56	2,696,378	2,725,988	2,755,598	2,785,208	2,814,818	1.08
	b. Kuadratik	14.28	8.22	2,380,315	2,378,692	2,375,974	2,372,160	2,367,250	-0.14
3	Fungsi Transfer	8.99	5.30	2,365,373	2,367,447	2,365,938	2,371,460	2,366,035	0.01
4	VAR	7.33	5.97	2,192,783	2,173,840	2,171,179	2,203,382	2,214,708	0.25

Berdasarkan atas keragaan diatas,dapat disimpulkan bahwa Model Fungsi Transfer merupakan model terbaik karena mempunyai nilai MAPE yang terkecil dengan hasil ramalan produksi gula yang cukup realistis.

Berdasarkan simpulan diatas, maka beberapa saran yang diajukan adalah sbb.;

- Model Fungsi Transfer dapat disarankan untuk memodelkan data produksi gula Indonesia dan meramalkan beberapa tahun ke depan.
- Model VAR dapat dipertimbangkan sebagai model alternatif terbaik karena mempunyai nilai MAPE yang tidak terlalu berbeda dengan model Fungsi Transfer.

## DAFTAR PUSTAKA

- Enders, W. 2010. Applied Econometric Time Series. USA: University of Alabama. Wiley, Third Edition.
- Guha, B and Bandyopadhyay, G. 2016. Gold Price Forecasting Using ARIMA Model. Journal of Advanced Management Science Vol. 4, No. 2, March 2016
- Gujarati, D.N. and D.C. Porter, 2010. Dasar-dasar Ekonometrika. Jakarta: Penerbit Salemba Empat, Buku 2



# KAJIAN PENGEMBANGAN METODE ESTIMASI PRODUKSI KAKAO

Yuliawati Rohmah

*Center for Agricultural Data and Information System-Ministry of Agriculture*

*Jalan Harsono RM No. 3 Gd D Lt 4, Ragunan-Jakarta Selatan, Indonesia*

*E-mail: yuliawati.rohmah@gmail.com*

## ABSTRAK

Kakao merupakan salah satu komoditas perkebunan strategis di Indonesia. Selama ini data produksi kakao yang dirilis oleh Direktorat Jenderal Perkebunan bersama BPS berupa data tahunan yang disajikan untuk kondisi 2 tahun yang lalu merupakan Angka Tetap (ATAP), satu tahun yang lalu merupakan Angka Sementara (ASEM) dan untuk tahun yang berjalan merupakan Angka Estimasi (AESTI). Selama ini metode yang digunakan untuk menyusun AESTI adalah Metode *Single Exponential Smoothing* (SES) dan *Double Exponential Smoothing* (DES). Metode untuk menghasilkan AESTI data perkebunan perlu dikaji kembali agar didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik. Sehingga kajian ini bertujuan untuk melakukan pemodelan, membandingkan efektivitas dari pemodelan tersebut dan menentukan metode terbaik dalam mengestimasi produksi kakao. Metode yang diterapkan adalah metode *Autoregressive Integrated Moving Average* (ARIMA), regresi, fungsi transfer dan *Vector Auto Regression* (VAR) dengan menggunakan *software program RStudio*. Pemilihan model terbaik dilakukan dengan membandingkan besaran nilai *Mean Absolute Percentage Error* (MAPE) dan kerealistisan hasil estimasi dibandingkan dengan data series sebelumnya. Berdasarkan dari hasil estimasi dan nilai MAPE disimpulkan bahwa ARIMA (1,1,0) adalah model terbaik untuk estimasi produksi kakao.

Kata kunci: ARIMA, regresi, fungsi transfer, VAR, produksi kakao

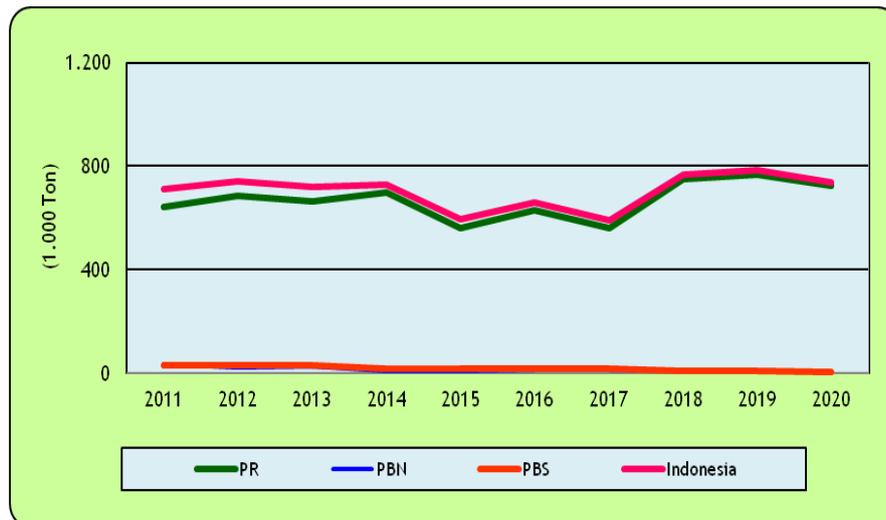
## ABSTRACT

*Cocoa is one of the strategic estate crop commodities in Indonesia. All this time the cocoa production data released by the Directorate General of Estate Crops together with BPS annually with a lag  $n-2$  year for Fixed Number (ATAP), a lag  $n-1$  year for Preliminary Number (ASEM) and the current year for Estimation Number (AESTI). All this time, the methods used to compile AESTI are the Single Exponential Smoothing (SES) and Double Exponential Smoothing (DES). The method for producing AESTI data needs to be reviewed in order to obtain a method that is more accurate, more objective and statistically better. The purpose of this paper are modeling, comparing the effectiveness of the models and determine the best method in estimating cocoa production. The method applied are Autoregressive Integrated Moving Average (ARIMA), regression, transfer function and Vector Auto Regression (VAR) using the RStudio software program. The best method will be selected by comparing the value of the Mean Absolute Percentage Error (MAPE) and the reliability of the estimation results to the previous data series. Based on the estimation results and MAPE value, it is concluded that ARIMA (1,1,0) is the best model for estimating cocoa production.*

*Keywords: ARIMA, regression, transfer function, VAR, cocoa production*

## PENDAHULUAN

Kakao merupakan salah satu komoditas ekspor perkebunan strategis yang ditetapkan pemerintah. Berdasarkan data FAO selama tahun 2014-2018, Indonesia menempati posisi ketiga sebagai negara produsen kakao dunia dengan kontribusi sebesar 13,03% atau rata-rata produksi 0,65 juta ton per tahun terhadap total produksi dunia setelah Pantai Gading dan Ghana dengan masing-masing kontribusi mencapai 36,46% dan 17,81%. Perkebunan kakao di Indonesia didominasi oleh Perkebunan Rakyat (PR) sebesar 95,06% dan sisanya dimiliki oleh Perkebunan Besar Nasional (PBN) sebanyak 2,21% dan Perkebunan Besar Swasta (PBS) mencapai 2,73%. Daerah sentra produksi kakao di Indonesia berdasarkan rata-rata produksi dalam periode 5 tahun terakhir (2014-2018) pada urutan 3 teratas yang memberikan kontribusi diatas 10% terhadap total produksi nasional adalah Provinsi Sulawesi Tengah (18,76%), Sulawesi Selatan (17,05%) dan Sulawesi Tenggara (16,33%). Perkembangan produksi kakao Indonesia dalam 10 tahun terakhir berfluktuatif dan cenderung menurun dengan rata-rata pertumbuhan sebesar negatif 0,41% per tahun (Gambar 1).



Gambar 1. Perkembangan Produksi Kakao Indonesia Tahun 2011-2020

Saat ini, rilis resmi data produksi kakao dikeluarkan oleh Direktorat Jenderal Perkebunan bersama BPS berupa data tahunan yang disajikan untuk kondisi 2 tahun yang lalu (*lag n-2*) merupakan Angka Tetap (ATAP), satu tahun yang lalu (*lag n-1*) merupakan Angka Sementara (ASEM) dan untuk tahun yang berjalan merupakan Angka Estimasi (AESTI). Data statistik perkebunan yang diperoleh merupakan hasil sinkronisasi dan validasi yang dilaksanakan secara berjenjang dari tingkat kabupaten sampai tingkat nasional dengan mengacu pada Pedoman Pengelolaan Data Komoditas Perkebunan (PDKP) yang diterbitkan oleh Direktorat Jenderal Perkebunan. Selama ini metode yang digunakan untuk menyusun AESTI adalah Metode Pemulusan Eksponensial Tunggal atau *Single Exponential Smoothing* (SES) dan Pemulusan Eksponensial Ganda atau *Double Exponential Smoothing* (DES). Pemilihan model terbaik dilakukan dengan membandingkan besaran nilai *Mean Absolute Percentage Error* (MAPE) dan kerealistisan hasil estimasi dibandingkan dengan data series sebelumnya. Metode untuk menghasilkan AESTI data perkebunan perlu dikaji kembali agar

didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik dibandingkan metode yang telah dilakukan selama ini.

Berdasarkan hal di atas, maka kajian ini bertujuan untuk:

- d. Melakukan analisa dan estimasi data produksi kakao menggunakan model *Auto Regressive Integrated Moving Average* (ARIMA), Regresi, Fungsi Transfer dan *Vector Auto Regression* (VAR).
- e. Membandingkan metode tersebut dalam memperoleh model estimasi data produksi kakao yang memiliki tingkat akurasi tertinggi.
- f. Menentukan metode terbaik dalam mengestimasi data produksi kakao.

## BAHAN DAN METODE

Data yang digunakan dalam kajian ini adalah data sekunder *time series* tahunan. Wujud produksi kakao yang dibahas berupa biji kakao kering. Variabel, satuan, level, periode dan sumber data yang digunakan dalam kajian ini adalah sebagai berikut:

Tabel 1. Jenis Variabel, Periode dan Sumber Data

No	Variabel Data	Satuan	Level	Periode	Sumber
1	Produksi	Ton	Nasional	1980-2019	Ditjen Perkebunan
2	Luas areal	Ha	Nasional	1980-2019	Ditjen Perkebunan
3	Hargadunia	US\$/Kg	Internasional	1980-2019	World Bank
4	Harga produsen	Rp/Kg	Nasional	1996-2019	Ditjen Perkebunan
5	Volume ekspor	Ton	Nasional	1980-2019	BPS
6	Volume impor	Ton	Nasional	1980-2019	BPS

Data yang digunakan dalam metode ARIMA dan regresi adalah produksi, sedangkan variabel harga produsen kakao nasional digunakan pada metode fungsi transfer sebagai variabel yang diduga dapat mempengaruhi produksi kakao nasional. Adapun pada metode VAR, data yang digunakan adalah produksi, luas areal, harga kakao dunia, volume ekspor dan volume impor karena kakao merupakan komoditas ekspor andalan. Pertimbangan lain dalam pemilihan dan penggunaan variabel data dalam model adalah ketersediaan series data dalam jangka waktu yang cukup panjang.

Pada proses pengolahan dan analisa, data *time series* dibagi menjadi dua bagian yakni data *training* untuk penyusunan model periode tahun 1980 - 2013 dan sisanya sebagai data *testing* untuk validasi model periode tahun 2014 - 2019. Kemudian dari hasil data *training* disusun model dan dilakukan estimasi sesuai periode data *testing*, setelah itu dilakukan evaluasi kesesuaian ramalannya. Model terbaik dipilih dari berbagai alternatif metode estimasi yang dicoba dengan melihat nilai *Mean Absolute Percentage Error* (MAPE) dan kesesuaian hasil estimasi dengan historis data aktualnya. Model estimasi terbaik yang terpilih kemudian dilakukan untuk estimasi 5 tahun ke depan yakni tahun 2020 – 2024 dengan menggabungkan seluruh data (*training* dan *testing*). Metode estimasi produksi kakao nasional

yang digunakan dalam kajian ini terdiri dari ARIMA, Regresi, Fungsi Transfer dan VAR menggunakan *software* program *RStudio*.

## 1. ARIMA

Metode ARIMA merupakan metode yang hanya menggunakan variabel dependen dan mengabaikan variabel independen sewaktu melakukan estimasi. ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampaunya. Bentuk modelnya adalah:

$$y_t = \mu + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

$y_t$	= data <i>time series</i> sebagai variable dependen pada waktu ke-t
$y_{t-p}$	= data <i>time series</i> pada kurun waktu ke (t-p)
$\mu$	= suatu konstanta
$\theta_1 \theta_q \phi_1 \phi_n$	= parameter-parameter model
$\varepsilon_{t-q}$	= nilai sisaan pada waktu ke-(t-q)

## 2. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih variabel.

### a. Regresi Linier Sederhana

Model regresi linier sederhana melibatkan satu variabel prediktor dan diformulasikan dalam model matematis sebagai berikut:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

dimana:

$y_t$	= variabel respon/dependen/terikat
$x_t$	= variabel prediktor/independen/bebas
$\beta_0$	= konstanta/intercept
$\beta_1$	= koefisien regresi variabel bebas
$\varepsilon_t$	= error/residu/sisaan

### b. Regresi Kuadratik

Model regresi kuadratik merupakan nilai variabel tak bebas dengan bentuk naik atau turun secara linier atau terjadi secara parabola dengan model matematis berikut ini:.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$$

dimana:

- $y_t$  = variabel respon/dependen/terikat
- $x_t$  = variabel prediktor/independen/bebas
- $\beta_0$  = konstanta/intercept
- $\beta_1, \beta_2$  = koefisien regresi variabel bebas
- $\varepsilon_t$  = error/residu/sisaan

### 3. Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret *output* atau  $y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $x_t$ ) dengan deret *output* tersebut.

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\varphi_p(B)} \varepsilon_t$$

dimana :

- $b$  = panjang jeda pengaruh  $x_t$  terhadap  $y_t$
- $r$  = panjang *lag*  $y$  periode sebelumnya yang masih mempengaruhi  $y_t$
- $s$  = panjang *lag*  $x$  periode sebelumnya yang masih mempengaruhi  $y_t$
- $p$  = ordo AR bagi *noise*  $N_t$
- $q$  = ordo MA bagi *noise*  $N_t$

### 4. VAR

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen., karena pada kenyataannya suatu variabel dapat bertindak sebagai endogen atau dependen dari suatu variabel yang lainnya (Gujarati, 2010).

$$x_t = \beta_{10} + \sum_{i=1}^p \alpha_{1i} x_{t-i} + \sum_{i=1}^p \beta_{1i} y_{t-i} + \sum_{i=1}^p \gamma_{1i} z_{t-i} + e_{1t}$$

$$y_t = \beta_{20} + \sum_{i=1}^p \alpha_{2i} x_{t-i} + \sum_{i=1}^p \beta_{2i} y_{t-i} + \sum_{i=1}^p \gamma_{2i} z_{t-i} + e_{2t}$$

$$z_t = \beta_{30} + \sum_{i=1}^p \alpha_{3i} x_{t-i} + \sum_{i=1}^p \beta_{3i} y_{t-i} + \sum_{i=1}^p \gamma_{3i} z_{t-i} + e_{3t}$$

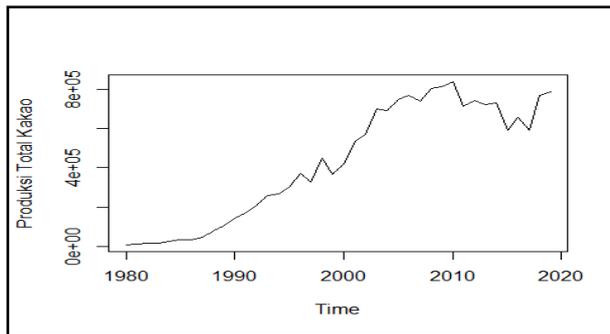
dimana :

- $x_t, y_t, z_t$  = variabel endogen
- $\beta_0$  = vektor konstanta  $n \times 1$ .
- $\alpha, \beta, \gamma$  = parameter dari  $x, y$ , dan  $z$
- $p$  = panjang *lag*
- $t$  = waktu
- $\varepsilon$  = vektor dari *shock* masing-masing variabel

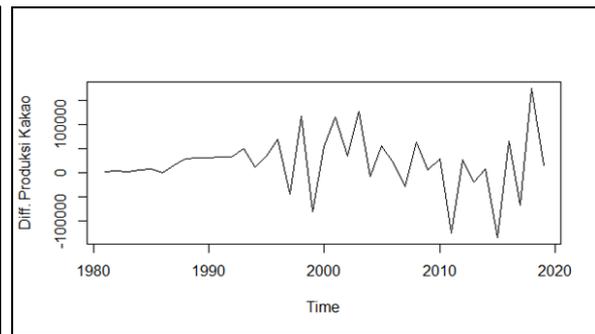
## HASIL DAN PEMBAHASAN

### 1. ARIMA

#### Eksplorasi Data Produksi Kakao Indonesia



Gambar 2. Produksi Kakao Tahun 1980-2019

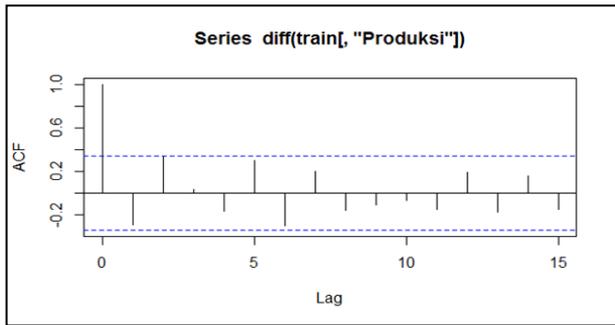


Gambar 3. Produksi Kakao *Differencing 1*

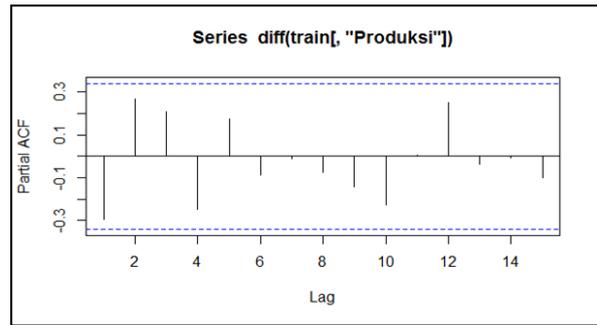
Produksi kakao dalam periode 40 tahun terakhir berfluktuatif dengan rata-rata pertumbuhan sebesar 13,15% per tahun. Berdasarkan uji kestasioneran data menggunakan *Augmented Dickey Fuller* (ADF) memperlihatkan data produksi kakao tidak stasioner, karena nilai *t-statistic* (-1,02) lebih besar dibandingkan *critical value* (-4,15; -3,50; -3,18) sehingga perlu dilakukan proses *differencing level 1*. Hal ini diperkuat dengan plot produksi kakao berdasarkan sebaran datanya (Gambar 1) yang tidak konstan di sekitar rata-rata. Hasil *differencing level 1* produksi kakao telah bersifat stasioner karena nilai *t-statistic* (-2,29) lebih kecil dibandingkan *critical value* pada alpha 5% (-1,95) dan sebaran datanya (Gambar 2) memiliki pola *single mean* atau konstan sekitar rata-rata bukan nol.

#### Model ARIMA Produksi Kakao

Tahap awal pada metode estimasi dengan ARIMA, setelah dipastikan data bersifat stasioner, maka dilakukan identifikasi model ARIMA berdasarkan hasil plot *Auto Correlation Function* (ACF) dan *Partial Auto Correlation Function* (PACF) seperti tampak pada Gambar 4 dan Gambar 5 dibawah ini. Dari hasil plot ACF bersifat *cut off* pada *lag 1*, sedangkan plot PACF, data sudah tidak memiliki pola khusus baik *tail off* ataupun *cut off* sehingga diperoleh dugaan awal untuk model ARIMA (1,1,0) atau ARIMA (0,1,1). Hal ini sesuai dengan hasil *autoarima* yakni ARIMA (1,1,0). Alternatif lain untuk mendapatkan model ARIMA dapat diperoleh dengan melakukan *overfitting* dari hasil *armaselect* dengan uji *Minimum Information Criterion* (Minic) seperti yang terdapat Tabel 2.



Gambar 4. Plot ACF Differencing Level 1



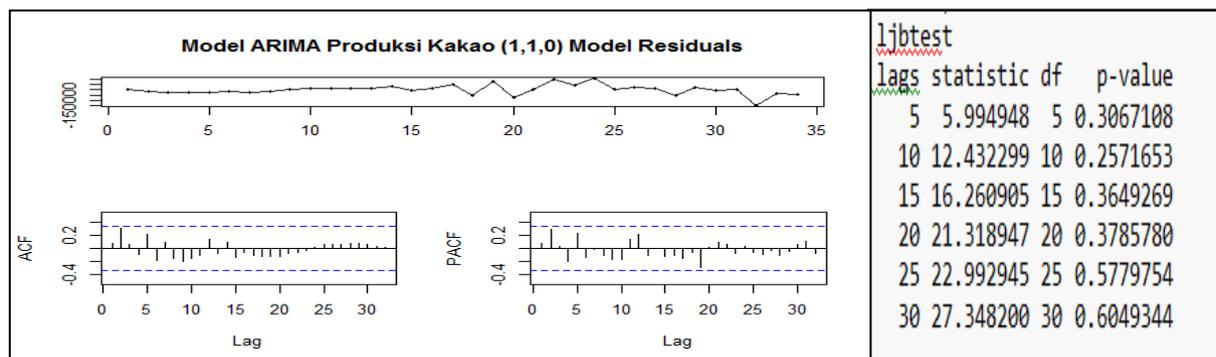
Gambar 5. Plot PACF Differencing Level 1

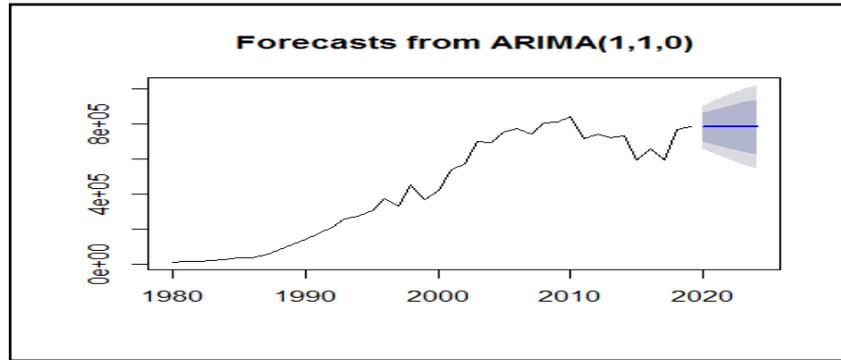
Tabel 2. Model ARIMA dan Estimasi Produksi Kakao Hasil Uji Minic

minic1	p	q	sbc	No	Model	MAPE (%)		Hasil Estimasi Produksi Kakao (Ton)				
						Training	Testing	2020	2021	2022	2023	2024
[1,]	0	0	714.7548	1	ARIMA (1,1,0)	24,93	18,81	780707	781348	781222	781247	781242
[2,]	1	0	716.0958	2	ARIMA (2,1,0)	11,68	11,5	866003	866499	905252	901989	920615
[3,]	2	0	717.9505	3	ARIMA (3,1,0)	10,93	11,42	860002	874712	910797	920291	937730
[4,]	3	0	720.6517	4	ARIMA (4,1,0)	10,75	11,62	860976	877557	909337	919906	933290
[5,]	4	0	722.8669	5	ARIMA (5,1,0)	10,55	12,12	832108	847262	852814	894379	894481
[6,]	5	0	725.9591	6	ARIMA (2,1,5)	10,75	11,71	846958	910637	905559	965873	971937
[7,]	2	5	733.9369	7	ARIMA (1,1,5)	10,71	11,03	850955	920072	916114	987489	994828
[8,]	1	5	734.1190	8	ARIMA (1,1,1)	14,66	11,36	813708	790162	808811	794041	805738
[9,]	1	1	736.6691	9	ARIMA (3,1,5)	9,79	11,42	843637	901496	901419	958909	983139
[10,]	3	5	737.0775									

Setelah model dan hasil estimasi diperoleh, maka tahap berikutnya adalah mengevaluasi hasil estimasi baik dengan menggunakan uji MAPE untuk data *training* dan data *testing* maupun dengan melihat kerealistisan hasil estimasi dengan data aktualnya. Model terbaik terpilih adalah ARIMA (1,1,0) dengan pertimbangan hasil estimasinya mendekati data aktual meskipun nilai MAPE bukan yang paling kecil, baik MAPE *training* 24,93% maupun MAPE *testing* 18,81%, sedangkan hasil estimasi model lain dianggap *overestimate* meski dengan nilai MAPE yang lebih kecil (Tabel 2). Langkah selanjutnya berupa pemeriksaan sisaan, plot sisaan terdistribusi normal dan dan plot ACF serta PACF sisaan tidak nyata. Sedangkan dari hasil uji *Ljung-Box*, autokorelasi sisaan tidak signifikan pada 30 *lag* (Tabel 3). Hasil estimasi produksi kakao dengan model ARIMA (1,1,0) untuk 5 tahun kedepan berfluktuatif dengan rata-rata pertumbuhan sebesar 0,02% per tahun (Gambar 6).

Tabel 3. Hasil Uji Pemeriksaan Sisaan dan Hasil Uji Ljung-Box ARIMA (1,1,0)





Gambar 6. Hasil Estimasi Produksi Kakao Model ARIMA (1,1,0) Tahun 2020-2024

## II. REGRESI

### a. Regresi Linier Sederhana

Dalam model regresi linier sederhana, variabel eksogen yang digunakan adalah tahun. Tahap awal untuk pembentukan model regresi adalah melakukan pendugaan model untuk mengetahui hubungan antara tahun dengan produksi kakao. Hasil uji *coefficients* memperlihatkan signifikansi baik pada intersep maupun variabel tahun pada alpha 0,1%. Model regresi yang dihasilkan juga memiliki selang kepercayaan yang sangat baik dimana nilai  $R^2$  dan *adjusted R<sup>2</sup>* masing-masing sebesar 94% (Tabel 4). Persamaan model regresi linier produksi kakao adalah sebagai berikut:

$$\text{Produksi} = -57783318 + 29135\text{Tahun}$$

Tabel 4. Hasil Test Coefficients Model Regresi Linier Sederhana Produksi Kakao

---

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-57783318	2527358	-22.86	<2e-16 ***
Tahun	29135	1266	23.02	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 72420 on 32 degrees of freedom  
 Multiple R-squared: 0.943, Adjusted R-squared: 0.9413  
 F-statistic: 529.7 on 1 and 32 DF, p-value: < 2.2e-16

---

Dari evaluasi terhadap hasil estimasi dengan menggunakan uji MAPE atau nilai kesalahan rata-rata dalam meramalkan sebesar 84,81% untuk *training* dan 42,30% untuk *testing* (Tabel 5). Tahap selanjutnya melakukan serangkaian pengujian terhadap beberapa asumsi model regresi yakni sebaran data, autokorelasi dan homoskedastisitas. Seperti yang terdapat pada Tabel 6, berdasarkan hasil Uji Jarque-Bera diperoleh data menyebar normal, terindikasi ada autokorelasi dari hasil Uji Durbin-Watson dan homoskedastisitas tidak

terpenuhi dari hasil Uji Chi-Square. Produksi kakao pada tahun 2020-2024 akan mengalami kenaikan yang sangat signifikan dengan rata-rata per tahun sebesar 2,62% berdasarkan hasil estimasi dengan menggunakan model regresi linier (Tabel 7).

Tabel 5. Nilai MAPE Data *Training* dan Data *Testing* Model Regresi Linier Produksi Kakao

Data	MAPE (%)
Training	84,81
Testing	42,30

Tabel 6. Hasil Uji Asumsi Model Regresi Linier Produksi Kakao

Uji Jarque-Bera			
Jarque-Bera test for normality			
data: reg1\$residuals			
JB = 1.5033, p-value = 0.2935			
Uji Durbin-watson			
lag	Autocorrelation	D-w Statistic	p-value
1	0.6552971	0.4975742	0
Alternative hypothesis: rho != 0			
Uji Chi-Square			
Non-constant Variance Score Test			
Variance formula: ~ fitted.values			
Chisquare = 3.536016, Df = 1, p = 0.06005			

Tabel 7. Hasil Estimasi Produksi Kakao Model Regresi Linier Tahun 2020-2024

Tahun	Produksi Kakao (Ton)
2020	1.069.243
2021	1.098.378
2022	1.127.513
2023	1.156.648
2024	1.185.783
Rata-rata Pertumbuhan (%)	2,62

## b. Regresi Kuadratik

Variabel eksogen yang digunakan dalam metode regresi kuadratik adalah tahun. Berdasarkan hasil pendugaan model regresi kuadratik untuk produksi kakao diperoleh signifikansi untuk intersep pada alpha 5% dan variabel tahun pada alpha 0,1%, sedangkan untuk variabel  $t^2$  tidak signifikan dari hasil uji *coefficient* dengan selang kepercayaan masing-masing sebesar 94% untuk *R-squared* dan *adjusted R-squared* (Tabel 8). Persamaan model regresi kuadratik produksi kakao adalah sebagai berikut:

$$\text{Produksi} = -104984,07 + 25750\text{Tahun} + 96,69\text{Tahun}^2$$

Tabel 8. Hasil Test Coefficients Model Regresi Kuadratik Produksi Kakao

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-104984.07	39912.51	-2.630	0.0132	*
t	25750.75	5257.94	4.897	2.88e-05	***
t2	96.69	145.73	0.664	0.5119	

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 73060 on 31 degrees of freedom  
 Multiple R-squared: 0.9438, Adjusted R-squared: 0.9402  
 F-statistic: 260.4 on 2 and 31 DF, p-value: < 2.2e-16

Tahap berikutnya yakni melakukan penghitungan MAPE data *training* dan data *testing*. Nilai MAPE untuk data *training* sangat besar yaitu 72,81%, sedangkan MAPE data *testing* meskipun lebih rendah dari MAPE *training* tetapi masih besar yakni 46,62% (Tabel 9).

Tabel 9. Nilai MAPE Data *Training* dan *Testing* Model Regresi Kuadratik Produksi Kakao

Data	MAPE (%)
Training	72,81
Testing	46,62

Tahap selanjutnya melakukan pengujian terhadap beberapa asumsi model regresi dengan Uji Jarque-Bera, Uji Durbin-Watson dan Uji Chi-Square. Berdasarkan serangkaian uji tersebut diperoleh hasil bahwa data tersebar normal, diduga terdapat autokorelasi dan heteroskedastisitas (Tabel 10). Tahapan akhir yakni melakukan estimasi sementara produksi kakao untuk tahun 2020-2024 dengan hasil estimasi seperti tertera pada Tabel 11 yang memperlihatkan kenaikan produksi kakao yang sangat signifikan setiap tahunnya dengan rata-rata 2,93%.

Tabel 10. Hasil Uji Asumsi Model Regresi Kuadratik Produksi Kakao

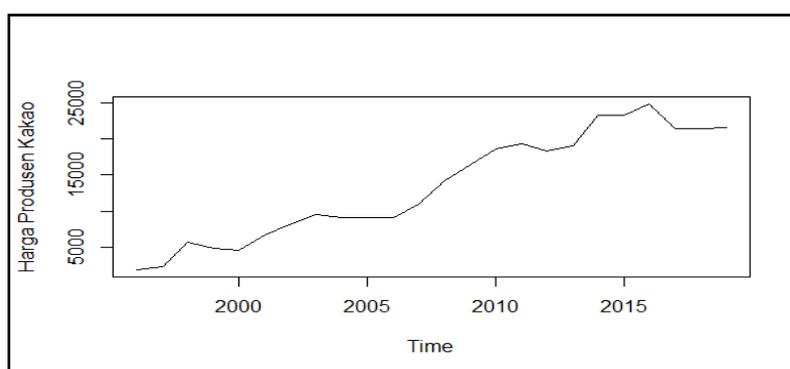
Uji Jarque-Bera			
Jarque-Bera test for normality			
data: reg2\$residuals			
JB = 0.56249, p-value = 0.7225			
Uji Durbin-Watson			
lag	Autocorrelation	D-w Statistic	p-value
1	0.6436503	0.5068265	0
Alternative hypothesis: rho != 0			
Uji Chi-Square			
Non-constant Variance Score Test			
Variance formula: ~ fitted.values			
Chisquare = 7.074551, Df = 1, p = 0.0078186			

Tabel 11. Hasil Estimasi Produksi Kakao Model Regresi Kuadratik Tahun 2020-2024

Tahun	Produksi Kakao (Ton)
2020	1.113.334
2021	1.147.110
2022	1.181.080
2023	1.215.243
2024	1.249.599
Rata-rata Pertumbuhan (%)	2,93

### III. FUNGSI TRANSFER

Pada metode fungsi transfer, peubah input yang digunakan adalah harga kakao di tingkat produsen (Gambar 7) dengan pertimbangan naik turunnya harga kakao di tingkat petani diduga sangat mempengaruhi produksi kakao di dalam negeri. Langkah pertama untuk proses analisa model adalah dengan mengidentifikasi model ARIMA peubah input berdasarkan hasil uji ADF dan plot ACF serta PACF. Berdasarkan hasil uji ADF diketahui data peubah input non stasioner dimana nilai *test-statistic* (-2,49) lebih besar dibanding *critical value* (-4,38; -3,60; -3,24), sehingga harus dilakukan proses *differencing*. Data harga kakao tingkat produsen stasioner setelah di *differencing level 2* dengan nilai *critical value* lebih tinggi pada (-2,66; -1,95; -1,6) dibandingkan nilai *test-statistic* (-4,12).

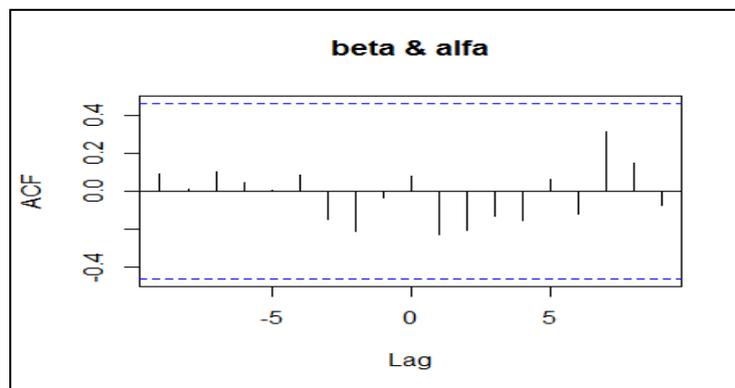


Gambar 7. Harga Kakao Tingkat Produsen Tahun 1996-2019

Langkah kedua menduga model ARIMA peubah input baik dengan *autoarima* ataupun *armaselect*. Setelah melakukan *overfitting* dari berbagai kemungkinan model ARIMA peubah input, maka dipilih ARIMA (4,2,3) dengan pertimbangan nilai MAPE paling kecil yakni 9,59%. Langkah ketiga yakni melakukan *prewhitening* dan korelasi silang antara deret input dengan produksi yang menghasilkan nilai *r,s,b* yakni (0,0,0) karena tidak ada yang nyata seperti tampak pada Gambar 8. Nilai *b* merupakan *lag* pertama kali dampak input berpengaruh terhadap output, *s lag* berikutnya setelah *b* dimana input berdampak terhadap output, dan *r* merupakan pengaruh output terhadap dirinya sendiri. Pada gambar 8 dapat dijelaskan bahwa nilai *b* dan *s* adalah 0 karena tidak ada yang nyata, sedangkan *r* dianggap 0 karena kakao merupakan tanaman tahunan.

Langkah keempat yaitu pengepasan model  $(r,s,b) = (0,0,0)$  yang menghasilkan nilai MAPE 14,21%. Identifikasi model *noise* atau residual dari peubah input merupakan langkah

kelima yang dilakukan dengan model ARIMA seperti langkah kedua yang menghasilkan model ARIMA (1,0,0) sebagai model terpilih untuk residual. Langkah selanjutnya melakukan pengepasan model  $(r,s,b) = (0,0,0)$  dan *noise* (1,0,0) dengan nilai MAPE 8,12% dan signifikan ar1, *intercept* dan *xreg* pada alpha 0,1% (Tabel 12).



Gambar 8. Plot Hasil *Prewhitening* dan Korelasi Silang Antara Deret Input dengan Output

Tabel 12. Hasil Test Coefficients Model Fungsi Transfer Produksi Kakao

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z )
ar1	7.9869e-01	1.2811e-01	6.2343	4.538e-10 ***
intercept	3.7843e+05	8.4711e+04	4.4673	7.922e-06 ***
xreg	2.1658e+01	6.1595e+00	3.5162	0.0004377 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1 '.'

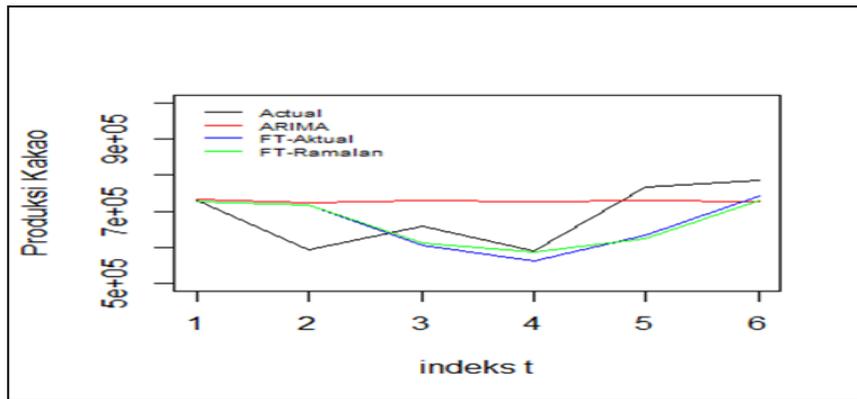
Langkah keenam melakukan serangkaian estimasi dan membandingkan nilai MAPE data *training* yang terdiri dari:

- Model fungsi transfer dengan data input nilai aktual.
- Model fungsi transfer dengan data input nilai estimasi.
- Model ARIMA output yang telah lebih dahulu dibahas di awal.

Nilai MAPE untuk model fungsi transfer yang menggunakan data input *testing* nilai estimasi lebih besar 13,69% dibanding yang menggunakan data input *testing* data aktual 11,58% dan model ARIMA biasa 11,39% (Tabel 13). Sedangkan dari plot hasil estimasi dengan fungsi transfer baik data input ramalan maupun data input aktual lebih mengikuti pola data aktual dibandingkan model ARIMA (Gambar 9).

Tabel 13. Nilai MAPE Data Training Model Fungsi Transfer Produksi Kakao

No	Model Estimasi	MAPE (%)	
		Training	Testing
1	FT ARIMA (1,0,0) xreg=harga kakao produsen aktual	8,12	11,58
2	FT ARIMA (1,0,0) xreg=harga kakao produsen ARIMA (4,2,3)	8,12	13,69
3	ARIMA (1,1,0)	9,21	11,39



Gambar 9. Nilai MAPE Hasil Estimasi

Langkah ketujuh yang merupakan langkah terakhir adalah menduga ulang model input harga kakao tingkat produsen dengan model ARIMA (4,2,3) dan menduga ulang fungsi transfer ARIMA (1,0,0) untuk melakukan estimasi produksi kakao 5 tahun kedepan yang menunjukkan pertumbuhan sekitar 0,98% per tahun secara rata-rata meskipun berfluktuatif setiap tahunnya (Tabel 14).

Tabel 14. Hasil Estimasi Produksi Kakao Model Fungsi Transfer Tahun 2020-2024

Tahun	Produksi Kakao (Ton)
2020	824.387
2021	855.868
2022	871.497
2023	862.331
2024	856.680
Rata-rata Pertumbuhan (%)	0,98

#### IV. VAR

Pada model VAR, variabel yang dimasukkan yakni produksi, luas areal, harga kakao dunia, volume ekspor dan volume impor. Tahap awal dalam penentuan model VAR adalah melakukan penelusuran model dari lag atau  $p=1$  sampai dengan  $p=4$  dengan dan tanpa tren (*type both and const*) dengan hasil model terpilih yakni VAR (1) tipe=*const* dengan pertimbangan karena merupakan model VAR dengan variabel yang signifikan paling banyak.

Tabel 15. Model VAR Produksi Kakao Terpilih

p=1, type=const								
Endogenous Variable	Luas Areal	Produksi	Harga Dunia	Voleks	Volimp	const	R <sup>2</sup>	R <sup>2</sup> Adjusted
Produksi		**	*		0	**	98%	98%
Luas Areal	***						99%	99%
Harga Dunia			***		*		81%	78%
Voleks		*	*	*		**	95%	94%
Volimp					***		95%	94%
12 variabel yg signifikan pada alpha = 10%								

Tahap selanjutnya melakukan serangkaian pengujian terhadap model yakni normalitas sebaran, autokorelasi dan keragaman. Dari hasil Uji Chi-squared, Uji Jarque-Bera dan ARCH dapat disimpulkan asumsi normalitas dan non autokorelasi terpenuhi serta ragam homogen (Tabel 16).

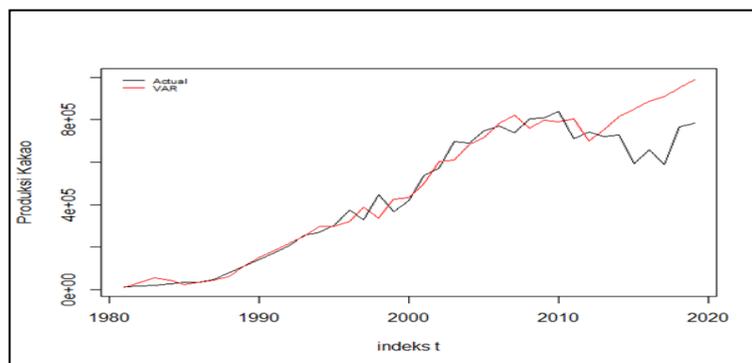
Tabel 16. Hasil Uji Asumsi Model VAR Produksi Kakao

Portmanteau Test (asymptotic)	
data: Residuals of VAR object varhsheet1	
Chi-squared = 341.62, df = 375, p-value = 0.8911	
JB-Test (multivariate)	
data: Residuals of VAR object varhsheet1	
Chi-squared = 33.713, df = 10, p-value = 0.0002066	
Skewness only (multivariate)	
data: Residuals of VAR object varhsheet1	
Chi-squared = 15.397, df = 5, p-value = 0.008793	
Kurtosis only (multivariate)	
data: Residuals of VAR object varhsheet1	
Chi-squared = 18.315, df = 5, p-value = 0.002576	
ARCH (multivariate)	
data: Residuals of VAR object varhsheet1	
Chi-squared = 420, df = 1125, p-value = 1	

Tahap berikutnya dalam model VAR adalah menghitung nilai MAPE dari data *training* dan data *testing* seperti yang tampak di Tabel 17. Untuk MAPE data *training* lebih besar 18,72% dibandingkan MAPE data *testing* yakni 13,94%. Pada Gambar 10 terlihat hasil plot estimasi data *training* dan data *testing* terhadap data aktual, dimana data *training* lebih mengikuti pola data aktual dibandingkan hasil estimasi data *testing*.

Tabel 17. Nilai MAPE Data Training dan Data Testing Model VAR Produksi Kakao

Data	MAPE (%)
Training	18,72
Testing	13,94



Gambar 10. Plot Data Ramalan Model VAR(1)  $type=const$  Terhadap Data Aktual Produksi Kakao Tahun 1980-2019

Tahap akhir dari serangkaian tahapan pada proses pemodelan dengan metode VAR berupa estimasi produksi kakao untuk periode tahun 2020-2024 yang menduga akan terjadi

kenaikan produksi kakao secara kontinyu pada 5 tahun mendatang dengan rata-rata bertumbuh sebesar 2,27% per tahun.

Tabel 18. Hasil Estimasi Produksi Kakao Model VAR (1) type=const Tahun 2020-2024

Tahun	Produksi Kakao (Ton)
2020	773.258
2021	802.688
2022	822.694
2023	836.528
2024	845.757
Rata-rata Pertumbuhan (%)	2,27

### Model Estimasi Terbaik

Hasil analisa dari 4 model estimasi yang digunakan pada kajian ini dapat diringkas sebagai berikut:

Tabel 19. Ringkasan Hasil Analisa Model Estimasi Produksi Kakao

No	Model	MAPE (%)		Produksi Kakao (Ton)					Rata-rata Pertumbuhan (%)
		Training	Testing	2020	2021	2022	2023	2024	
1	ARIMA(1,1,0)	24,93	18,81	780.707	781.348	781.222	781.247	781.242	0,02
2	a. Regresi Linier	84,81	42,30	1.069.243	1.098.378	1.127.513	1.156.648	1.185.783	2,62
	b. Regresi Kuadratik	72,81	46,62	1.113.334	1.147.110	1.181.080	1.215.243	1.249.599	2,93
3	FT ARIMA(1,0,0) xreg = harga produsen	13,69	8,12	824.387	855.868	871.497	862.331	856.680	0,98
4	VAR (1) type=const	18,72	13,94	773.258	802.688	822.694	836.528	845.757	2,27

Berdasarkan Uji MAPE dari Tabel 19 maka model terbaik dan terpilih untuk estimasi produksi kakao adalah model ARIMA (1,1,0) karena memiliki hasil estimasi dan rata-rata pertumbuhan yang dianggap lebih mendekati data aktualnya, meskipun nilai MAPE untuk *training* dan *testing* bukan yang paling kecil dibandingkan 4 model lainnya. Sedangkan hasil estimasi model lainnya dianggap *overestimate* terhadap historis data aktualnya. Hasil estimasi dari model terbaik untuk produksi kakao tahun 2020 sebesar 780.707 ton, turun 0,42% dibandingkan tahun 2019 yang mencapai 783.978 ton. Hasil estimasi produksi kakao yang dilakukan oleh Pusdatin lebih tinggi dibandingkan estimasi Ditjen Perkebunan sebesar 739.483 ton atau selisih 5,57%. Produksi kakao meningkat kembali sebesar 0,08% menjadi 781.348 ton di tahun 2021. Untuk tahun 2021-2024 berfluktuatif naik turun meskipun tidak signifikan. Rata-rata pertumbuhan produksi kakao 5 tahun kedepan sebesar 0,02% (Tabel 20).

Tabel 20. Hasil Estimasi Model ARIMA (1,1,0) Produksi Kakao Tahun 2020-2024

Tahun	Produksi (Ton)		Pertumbuhan (%)
	Ditjen Perkebunan	Pusdatin	
2019*	783.978		
2020**	739.483	780.707	
2021		781.348	0,08
2022		781.222	-0,02
2023		781.247	0,00
2024		781.242	0,00
Rata-rata Pertumbuhan (%)			0,02

### KESIMPULAN

Dari keempat metode estimasi yang digunakan dalam kajian ini yaitu ARIMA, Regresi, Fungsi Transfer dan VAR, metode estimasi terbaik untuk estimasi produksi kakao berdasarkan pertimbangan statistik dan kerealistisan hasil estimasi dengan historis data aktualnya adalah Model ARIMA (1,1,0).

Kekurangan dari kajian ini terutama pada model VAR, variabel produksi dan volume ekspor serta volume impor dalam wujud produk yang berbeda. Sehingga untuk pengembangan kajian berikutnya disarankan agar wujud produk untuk variabel produksi, volume ekspor dan volume impor adalah sama yakni biji kakao kering.

### DAFTAR PUSTAKA

- Enders, Walter. 2010. *Applied Econometric Time Series*. USA: University of Alabama. Wiley, Third Edition.
- Gujarati, D.N. and D.C. Porter, 2010. *Dasar-dasar Ekonometrika*. Jakarta: Penerbit Salemba Empat, Buku 2, Edisi 5.
- Kementerian Pertanian. 2019. *Statistik Perkebunan Indonesia 2018-2020*. Jakarta: Direktorat Jenderal Perkebunan Kementerian Pertanian.
- Rohmah, Yuliawati. 2019. *Outlook Komoditas Perkebunan Kakao*. Jakarta: Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian

# **KAJIAN MODEL PERAMALAN PRODUKSI KOPI DI INDONESIA: PENDEKATAN MODEL ARIMA, REGRESI, FUNGSI TRANSFER DAN VAR**

Oleh: Roch Widaningsih

## **ABSTRAK**

Kopi merupakan tanaman perkebunan unggulan di Indonesia. Tahun 2019, nilai ekspor kopi mencapai 883,12 juta USD, menempati urutan ke empat setelah kelapa sawit, karet, dan kelapa. Angka tetap (ATAP) data produksi kopi oleh Badan Pusat Statistik (BPS) adalah data tahunan dengan lag  $n-1$  tahun. Guna merumuskan kebijakan produksi, pasokan dan distribusi komoditas kopi dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Tahun 2020, Pusdatin mengkaji empat metode yaitu metode ARIMA, regresi, fungsi transfer dan VAR (*Vector Auto Regression*) dengan menggunakan *analisis R Studio* guna melakukan pemodelan data kopi di Indonesia. Metode yang dipilih guna melakukan peramalan data produksi kopi adalah metode yang terbaik dikaji dari nilai MAPE (*Mean Absolute Percentage Error*) terkecil dan kerealistikan hasil peramalan. Berdasarkan dari hasil nilai MAPE disimpulkan bahwa Fungsi Transfer adalah yang terbaik dengan MAPE 6,78%.

## **PENDAHULUAN**

Peluang industri kopi di Indonesia sangat bagus. Apalagi dengan semakin dikenalnya kopi Indonesia di Eropa dan Amerika, terutama kopi khusus (*specialty coffee*) seperti kopi Gayo, kopi Mandailing, kopi Lampung, kopi Bajawa dan lainnya. Menteri Pertanian. Andi Amran Sulaiman mempromosikan komoditas kopi Indonesia kepada dunia melalui forum World Coffee Producers Forum (WCPF) yang dihelat di Kolombia, Selasa (11/7/2017).

Fokus pemerintah saat ini adalah meningkatkan produksi komoditas kopi dalam negeri melalui penyediaan bibit berkualitas tinggi, pemupukan tepat waktu, manajemen air dan program peremajaan untuk mengganti tanaman kopi yang sudah tua. Pengembangan kopi ke depan juga akan lebih memperhatikan aspek kearifan lokal, sehingga dapat dihasilkan jenis-jenis kopi specialty yang bernilai tinggi dari berbagai daerah.

Ditjen Perkebunan dalam renstra menempatkan komoditas kopi menjadi salah satu komoditas yang menjadi sasaran pokok sub agenda prioritas peningkatan agroindustri yaitu peningkatan produksi komoditas andalan dan prospektif ekspor serta mendorong perkembangan agroindustri di pedesaan, selain komoditas kelapa sawit, kakao, teh dan kelapa (Direktorat Jenderal Perkebunan, 2015).

ATAP data produksi kopi oleh Badan Pusat Statistik (BPS) adalah data tahunan dengan lag  $n-1$  tahun. Guna merumuskan kebijakan produksi, pasokan dan distribusi komoditas kopi sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Metode ramalan produksi kopi dapat digunakan dengan menggunakan beberapa metode statistik yang telah dikembangkan oleh para ahli.

Makalah ini akan dikaji metode ARIMA, Regresi, Fungsi Transfer dan VAR dalam melakukan pemodelan dan peramalan produksi kopi di Indonesia.

Tujuan dari disusunnya kegiatan ini adalah:

- g. Melakukan analisis dan peramalan data produksi kopi menggunakan model ARIMA, Regresi, Fungsi Transfer dan VAR di Indonesia
- h. Membandingkan metode tersebut dalam memperoleh ramalan data produksi komoditas kopi.
- i. Menentukan metode terbaik dalam meramal data produksi komoditas kopi di Indonesia.

## BAHAN DAN METODE

Data yang digunakan dalam melakukan pemodelan produksi kopi adalah data series produksi kopi tahun 1980-2019. Peubah yang diasumsikan mempengaruhi besaran produksi kopi untuk Fungsi Transfer adalah volume impor dan untuk pemodelan VAR adalah luas areal, volume ekspor, volume impor, harga dunia, dan harga urea.

Peramalan data produksi kopi, dilakukan uji coba beberapa metode yakni mengkaji metode ARIMA, Regresi, Fungsi Transfer dan VAR (Vector Autoregression), menggunakan software *Analisis R Studio*.

Metode yang digunakan dalam melakukan peramalan data produksi kopi di Indonesia adalah :

### h. ARIMA (*Autoregressive Integrated Moving Average*)

ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampaunya. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

$Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t

$Y_{t-p}$  = data *time series* pada kurun waktu ke  $(t-P)$

$\mu$  = suatu konstanta

$\theta_1 \theta_q \phi_1 \phi_n$  = parameter-parameter model

$\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

### i. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih variabel.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

Dimana :

$y$  = variabel respon/dependen/terikat

$x_1, x_2, \dots, x_k$  = variabel prediktor/independen/bebas

$\beta_0$  = konstanta/intercept

$\beta_1, \beta_2, \dots, \beta_k$  = koefisien regresi variabel bebas

$\varepsilon$  = error/residu/sisaan

j. Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau  $Y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $Y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $X_t$ ) dengan deret output tersebut.

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t$$

Dimana:

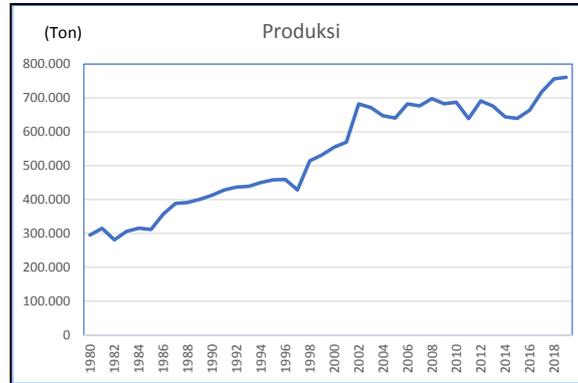
- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$

k. VAR (Vector )

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen., karena pada kenyataannya suatu variabel dapat bertindak sebagai endogen atau dependen dari suatu variabel yang lainnya (Gujarati, 2010).

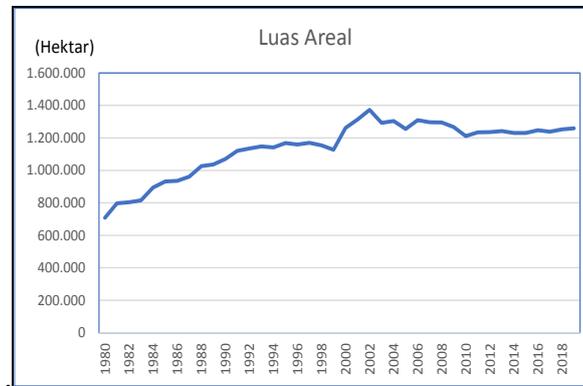
## HASIL DAN PEMBAHASAN

### d. Eksplorasi Data



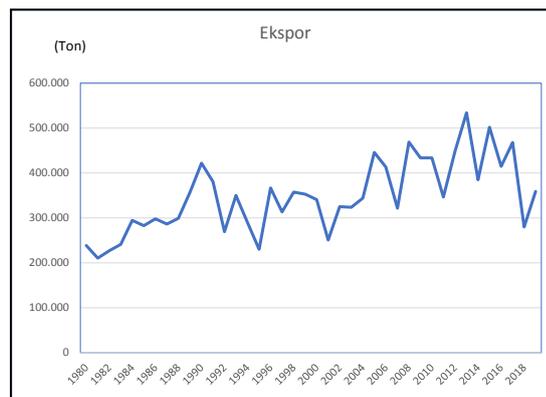
**Gambar 1.** Perkembangan Produksi Kopi di Indonesia, 1980-2019

Produksi kopi periode 2080-2019 terus meningkat dengan rata-rata pertumbuhan sebesar 2,64% (Gambar 1.)



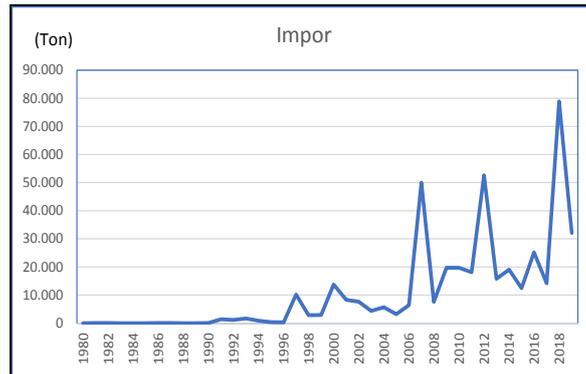
**Gambar 2.** Perkembangan Luas Areal Kopi di Indonesia, Tahun 1980-2019

Luas areal kopi periode 2080-2019 sedikit meningkat dengan rata-rata pertumbuhan sebesar 1,56% (Gambar 2.)



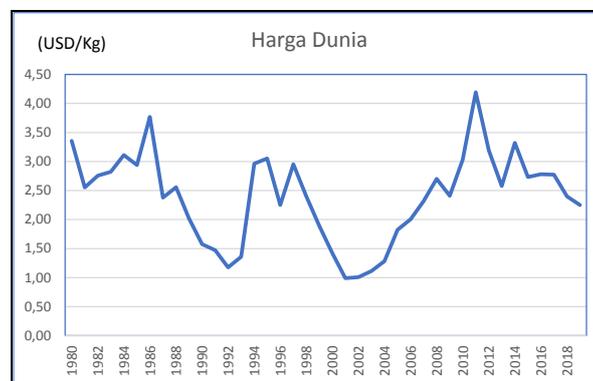
**Gambar 3.** Perkembangan Volume Ekspor Kopi di Indonesia, Tahun 1980-2019

Volume ekspor kopi di Indonesia periode 2080-2019 berfluktuasi, cenderung meningkat dengan rata-rata pertumbuhan sebesar 3,41% (Gambar 3.)



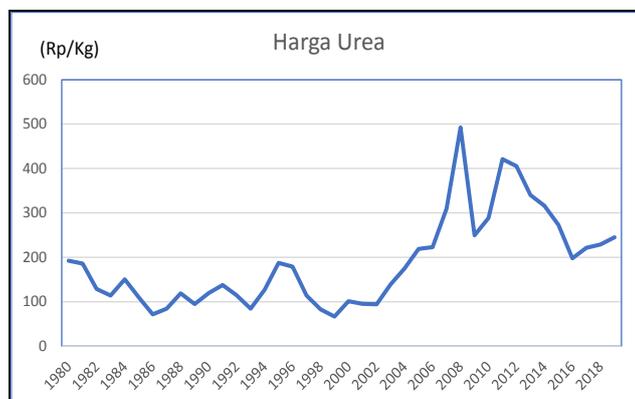
**Gambar 4.** Perkembangan Volume Impor Kopi di Indonesia, Tahun 1980-2019

Volume impor kopi di Indonesia periode 2080-2019 berfluktuasi, cenderung meningkat sangat tinggi dengan rata-rata pertumbuhan sebesar 161,74% (Gambar 4.)



**Gambar 5.** Perkembangan Harga Dunia Kopi di Indonesia, 1980-2019

Harga dunia kopi periode 2080-2019 berfluktuasi cenderung meningkat dengan rata-rata pertumbuhan sebesar 2,09% (Gambar 5.)



**Gambar 6.** Perkembangan Harga Urea di Indonesia, Tahun 1980-2019

Perkembangan harga urea periode 2080-2019 berfluktuasi cenderung meningkat dengan rata-rata pertumbuhan sebesar 4,66% (Gambar 6.)

**e. ARIMA**

Dalam melakukan pemodelan produksi kopi menggunakan model Autoregressive Integrated Average (ARIMA), data yang digunakan adalah periode 1980-2019. Periode data tersebut kemudian dipisahkan menjadi data set training dan testing. Panjang series data pada data set training adalah 1980-2013, dataset testing adalah periode 2014-2019. Dataset training digunakan untuk melakukan penyusunan model, sementara dataset testing digunakan untuk validasi model.

Uji kestasioneran data seperti yang disyaratkan apabila melakukan pemodelan ARIMA dilakukan secara visual menggunakan hasil plot data maupun uji formal statistik. Gambar 1 menunjukkan produksi terus meningkat dengan tahun tertentu sedikit berfluktuasi dengan peningkatan yang berbeda setiap tahunnya, sehingga dapat disimpulkan bahwa data kurang stasioner. Hasil uji statistikpun menunjukkan bahwa data kurang stasioner karena uji Augmented Dickey-Fuller (ADF) yang mengindikasikan bahwa data produksi kopi tidak stasioner, terlihat dari nilai *p* yang lebih besar dari taraf nyatanya ( 0.05) (Tabel 1).

**Tabel 1.** Hasil Uji Augmented Dickey-Fuller Produksi Kopi

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-52597 -10984   -3210   10532 106412

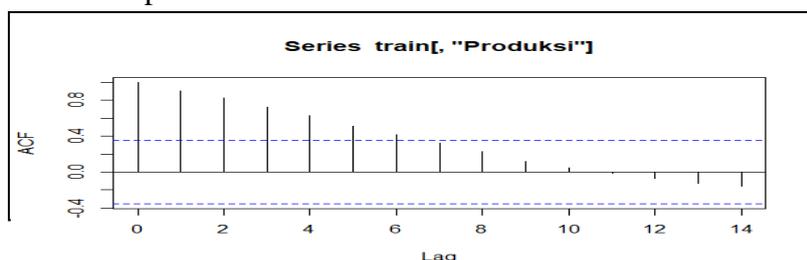
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.529e+04  3.457e+04   2.178  0.0365 *
z.lag.1      -2.190e-01  1.211e-01  -1.808  0.0794 .
tt           2.572e+03  1.574e+03   1.634  0.1114
z.diff.lag   -5.571e-02  1.733e-01  -0.322  0.7498
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30790 on 34 degrees of freedom
Multiple R-squared:  0.1159, Adjusted R-squared:  0.0379
F-statistic: 1.486 on 3 and 34 DF, p-value: 0.2358

value of test-statistic is: -1.8083 3.0557 1.6904

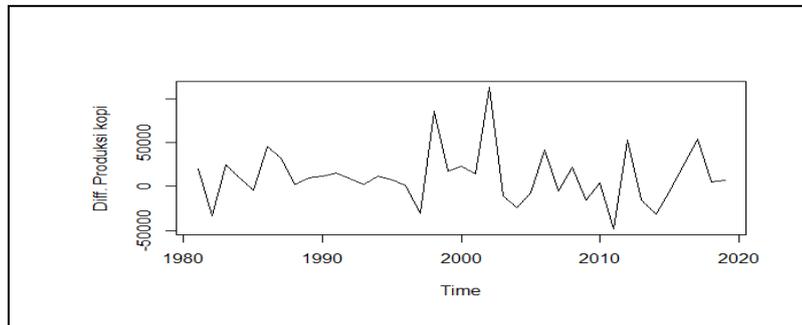
Critical values for test statistics:
      1pct   5pct 10pct
tau3  -4.15  -3.50 -3.18
phi2   7.02   5.13  4.31
phi3   9.31   6.73  5.61
```

➤ Plot ACF untuk pemeriksaan autokorelasi



**Gambar 7.** Plot ACF Data Produksi Kopi

- Plot ACF mengindikasikan data produksi kopi tidak stasioner (memperkuat hasil Uji ADF sebelumnya)
- Untuk menjadi stasioner maka data dideferencing, dan setelah differencing satu kali data sudah stasioner seperti ditunjukkan pada Tabel 2.
- Data Produksi Kopi Setelah *Differencing*



**Gambar 8.** Plot ACF *differencing*

**Tabel 2.** Hasil Uji Augmented Dickey-Fuller Differencing Produksi Kopi

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-47369 -4591    8250   20370  111787

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
z.lag.1      -0.96251    0.23923  -4.023 0.000292 ***
z.diff.lag  -0.05444    0.16583  -0.328 0.744631
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33680 on 35 degrees of freedom
Multiple R-squared:  0.5177, Adjusted R-squared:  0.4901
F-statistic: 18.78 on 2 and 35 DF, p-value: 2.873e-06

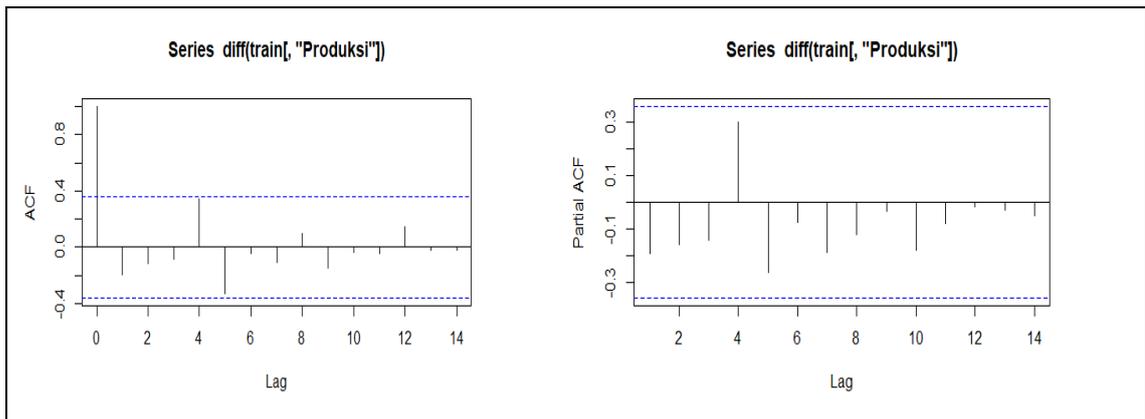
Value of test-statistic is: -4.0234

Critical values for test statistics:
      1pct   5pct  10pct
tau1 -2.62 -1.95 -1.61
```

- P-value hasil uji < taraf uji
- Statistik uji < critical value
- Disimpulkan: data STASIONER

Setelah mendapatkan data produksi kopi yang stasioner, maka langkah selanjutnya adalah menentukan ordo AR dan MA menggunakan fungsi auto arima yang terdapat pada RStudio. Dari hasil pendugaan auto arima pada Tabel 3, maka model ARIMA yang terpilih adalah ARIMA (0,1,0) yang berarti model ARIMA yang optimal untuk mengestimasi data yang diolah adalah ARIMA dengan ordo *Auto Regressive* (AR) = 0, ordo *differencing* = 1, dan ordo *Moving Average* (MA) = 0.

➤ **Pendugaan Model ARIMA**



**Gambar 9.** Plot ACF PACF hasil *differencing*

Plot ACF PACF hasil *differencing* tidak menunjukkan perilaku yang khas sesuai model ARIMA tertentu.

**Tabel 3.** Hasil Fungsi AUTOARIMA

```
Series: train[, "Produksi"]
ARIMA(0,1,0) with drift

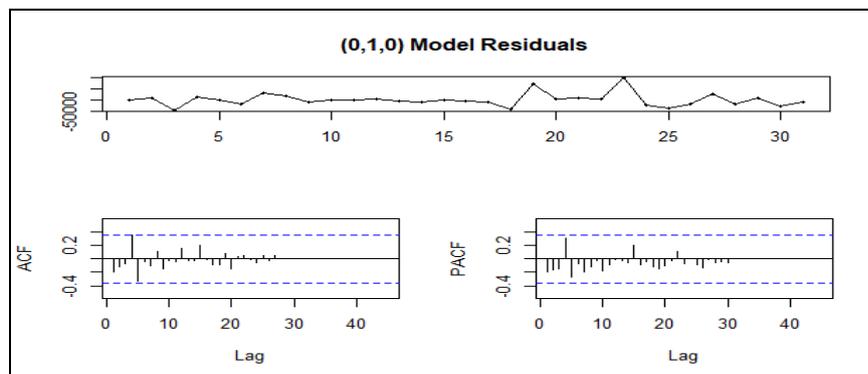
Coefficients:
      drift
11542.667
s.e.      5448.433

sigma^2 estimated as 1.01e+09: log likelihood=-388.42
AIC=780.84  AICC=781.24  BIC=783.83

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 8.336182 30835.15 20990.96 -0.1570285 4.124206 0.8991456 -0.2487313
```

Dengan fungsi `auto.arima` diperoleh model ARIMA dengan AIC terkecil adalah ARIMA(0,1,0)

➤ **Pemeriksaan Plot Sisaan**



**Gambar 10.** Plot ACF PACF Sisaan ARIMA (0,1,0)

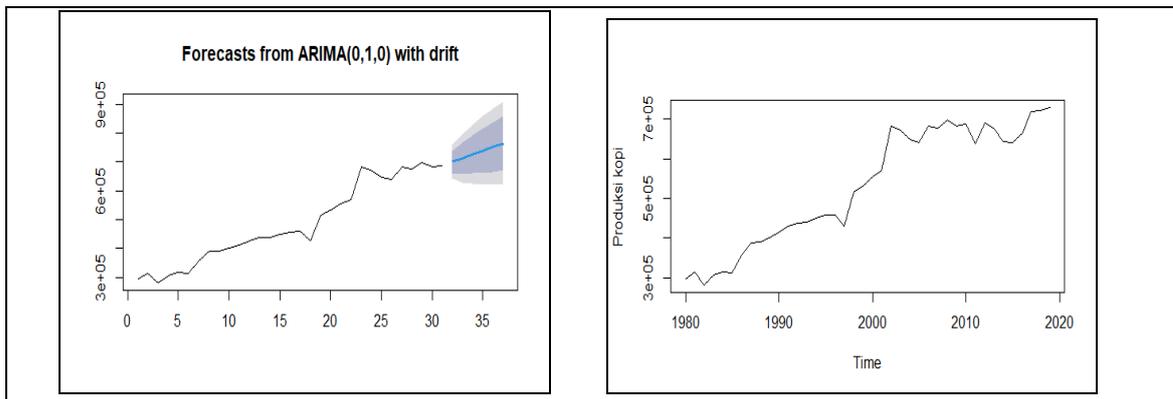
Plot ACF PACF sisaan tidak menunjukkan perilaku yang khas sesuai model ARIMA tertentu.

➤ **Peramalan data testing**

**Tabel 4.** Hasil Peramalan Data Testing

Time Series:								
Start = 32								
End = 37								
Frequency = 1								
[1] 699985.9 713050.9 726115.8 739180.7 752245.7 765310.6								
		ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set		9.093804	29116.72	19188.89	-0.256405	3.967808	0.8797856	-0.1901961
Test set		-55417.100000	59971.23	55417.10	-8.322250	8.322250	2.5408023	NA

➤ **Plot Ramalan ARIMA**



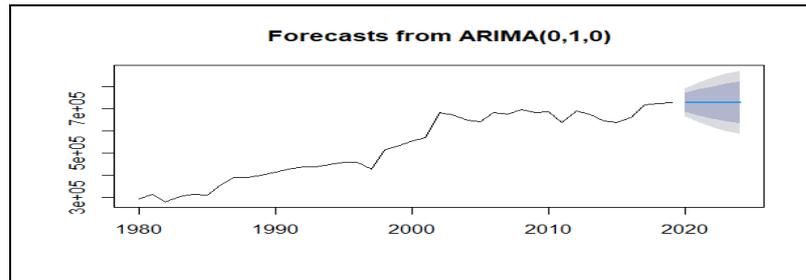
**Gambar 11.** Perbandingan plot ramalan model ARIMA (0,1,0) dan plot data asli produksi kopi

➤ **Pengepasan Model untuk Seluruh Data**

**Tabel 5.** Hasil Pengepasan Model Seluruh Data

Series: <u>serieskopi[, "Produksi"]</u>								
<u>ARIMA(0,1,0)</u>								
<u>sigma^2</u> estimated as 1.061e+09: log likelihood=-460.59								
AIC=923.18 <u>AICC</u> =923.28 BIC=924.84								
Training set error measures:								
		ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set		10859.9	32156.17	22423.5	2.071878	4.34356	0.9753207	-0.1691664

➤ **Peramalan ARIMA (0,1,0) untuk seluruh Data**



**Gambar 12.** Ramalan model ARIMA (0,1,0)

**Tabel 6.** Hasil Ramalan Produksi Kopi ARIMA (0,1,0)

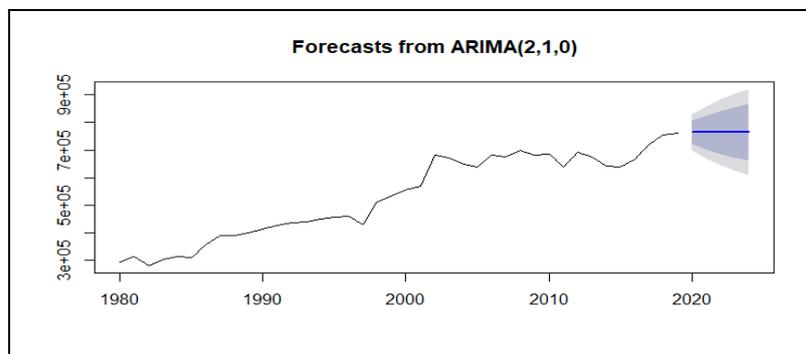
	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	729074	687339.2	770808.8	665246.2	792901.8
2021	729074	670052.1	788095.9	638807.8	819340.2
2022	729074	656787.2	801360.8	618521.0	839627.0
2023	729074	645604.4	812543.6	601418.3	856729.7
2024	729074	635752.2	822395.8	586350.6	871797.4

Karena hasil ramalan dari tahun 2020-2024 produksi kopi sama maka dicari ARIMA ordo lain, melalui fungsi armaselect (diff) dan terpilih ARIMA (2,1,0).

➤ **Peramalan ARIMA (2,1,0) untuk seluruh Data**

**Tabel 7.** Hasil Pengepasan Model Seluruh Data ARIMA (2,1,0)

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	10782.07	32611.51	22886.26	2.009599	<b>4.389805</b>	0.9612616	-0.1249047



**Gambar 13.** Ramalan model ARIMA (2,1,0)

**Tabel 8.** Hasil Ramalan Produksi Kopi ARIMA 2,1,0)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	763732.3	720277.7	807187.0	697274.2	830190.5
2021	764111.9	702334.5	825889.3	669631.5	858592.3
2022	764313.5	686680.0	841947.0	645583.3	883043.7
2023	764342.7	673548.4	855137.0	625484.8	903200.5
2024	764357.4	661979.5	866735.2	607784.0	920930.7

**f. Regresi Linier**

Peubah bebas yang digunakan dalam melakukan pemodelan data produksi kopi di Indonesia adalah tahun. Data yang digunakan adalah periode tahun 1980 hingga 2019. Periode data tersebut kemudian dipisahkan menjadi data set *training* dan *testing*. Panjang series data pada data set *training* adalah tahun 1980 hingga 2013, sementara dataset *testing* adalah periode tahun 2014-2019. Dataset *training* digunakan untuk melakukan penyusunan model, sementara dataset *testing* digunakan untuk validasi model.

Dari hasil pendugaan parameter (Tabel 9) menunjukkan model cukup bagus, terlihat dari nilai *p-value*  $2,2 \times 10^{-16}$  atau  $H_0$  ditolak, artinya koefisien signifikan (berbeda dari nol) model yang jauh lebih kecil dari 0,05 dan variabel tahun signifikan secara uji t dengan *p-value*  $< 2,2 \times 10^{-16}$ . Nilai  $R^2$  model bernilai 93,66% yang artinya peubah tahun dapat menjelaskan 93,66% keragaman dari produksi kopi di Indonesia.

**Tabel 9.** Penyusunan Model Regresi Linier Produksi Kopi Indonesia

```
Call:
lm(formula = Produksi ~ Tahun, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-82034 -17465    694   18356 101372

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.753e+07  1.289e+06  -21.36  <2e-16 ***
Tahun        1.404e+04  6.454e+02   21.75  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36920 on 32 degrees of freedom
Multiple R-squared:  0.9366, Adjusted R-squared:  0.9347
F-statistic: 473.1 on 1 and 32 DF, p-value: < 2.2e-16
```

Model luas areal kopi Indonesia diformulasikan sebagai berikut:

$$y_t = -2,75 \times 10^7 + 1,40 \times 10^4 T$$

Evaluasi nilai kesalahan model regresi yang terbentuk dilakukan dengan menghitung nilai MAPE atau nilai rata-rata persentase kesalahan peramalan. Nilai MAPE untuk dataset *training* dan *testing* masing-masing sebesar 6,08% dan 15,46% seperti tersaji pada Tabel 10.

**Tabel 10.** Nilai MAPE Data Training Model Regresi Linier Produksi Kopi Indonesia

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1425	0.9810	3.9891	4.9622	7.5571	19.1481

**Tabel 11.** Nilai MAPE Data Testing Model Regresi Linier Produksi Kopi Indonesia

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.21	11.69	14.36	14.47	16.89	19.36

Setelah diperoleh model regresi yang akan digunakan, maka perlu dilakukan pengujian asumsi regresi linier, yakni normalitas, non-autokorelasi, dan homoskedastisitas. Hasil uji asumsi normalitas menggunakan *Jarque Bera test* menunjukkan bahwa asumsi normalitas terpenuhi karena  $p\text{-value} > 0,05$  (Tabel 12). Uji asumsi selanjutnya yakni non-autokorelasi menggunakan *Durbin Watson test* juga menunjukkan bahwa asumsi terpenuhi, karena  $p\text{-value} < 0,05$  (Tabel 13) dan asumsi terakhir yakni homoskedas tidak terpenuhi karena uji NCV test menunjukkan bahwa  $p\text{-value} > 0,05$  (Tabel 13).

**Tabel 12.** Hasil Uji Asumsi Normalitas Regresi Linier Produksi Kopi menggunakan *Jarque Bera Test*

```
Jarque-Bera test for normality
data:  reg1$residuals
JB = 2.5487, p-value = 0.1235
```

- P-value = 0,1235 > 0,05, artinya NORMAL

**Tabel 13.** Hasil Uji Asumsi Non-Autokorelasi Regresi Linier Produksi Kopi menggunakan *Durbin Watson Test*

```
lag Autocorrelation D-w statistic p-value
1 0.5808205 0.7457178 0
Alternative hypothesis: rho != 0
```

- P-value = 0 < 0,05, artinya terdapat autokorelasi

**Tabel 14.** Hasil Uji Asumsi Normalitas Homoskedas Regresi Linier Produksi Kopi menggunakan *NCV test*

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 5.505276, Df = 1, p = 0.018959
```

- P-value = 0,019 < 0,05, artinya heteroskedastis

Setelah diperoleh model regresi linier yang optimal, maka model regresi tersebut digunakan untuk melakukan peramalan produksi kopi 5 tahun ke depan (2020-2024). Hasil peramalan produksi kopi Indonesia menggunakan regresi linier dengan peubah bebas yaitu tahun ditampilkan pada Tabel 15 berikut.

**Tabel 15.** Hasil Peramalan Produksi Kopi menggunakan Regresi Linier

2020	2021	2022	2023	2024
833.349	1.505.180	1.520.782	1.536.383	1.551.984

### g. Regresi Kuadratik

Peubah bebas yang digunakan dalam melakukan pemodelan data produksi kopi di Indonesia adalah peubah waktu (tahun) dan tahun kuadrat. Hasil pendugaan parameter (Tabel 16) menunjukkan model cukup bagus, terlihat dari nilai probabilitas  $< 2,2 \times 10^{-16}$  atau  $H_0$  ditolak, artinya koefisien signifikan (berbeda dari nol) model yang jauh lebih kecil dari 0,05.

Dengan  $R^2$  sebesar 93,76%, yang artinya peubah tahun dan tahun kuadrat dapat menjelaskan 93,76% keragaman dari produksi kopi di Indonesia.

**Tabel 16.** Pendugaan Parameter Model Regresi Produksi Kopi di Indonesia

```
Call:
lm(formula = Produksi ~ t + t2, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-87060 -18609  -1887   17806  97917

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 262574.23   18074.23   14.528 2.22e-15 ***
t           15766.69    2534.72    6.220 6.57e-07 ***
t2            -52.35      74.24   -0.705  0.486
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37220 on 31 degrees of freedom
Multiple R-squared:  0.9376, Adjusted R-squared:  0.9336
F-statistic: 233.1 on 2 and 31 DF, p-value: < 2.2e-16
```

➤ **Penghitungan Dugaan dengan Analisis Regresi Kuadrat**

**Tabel 17.** Hasil Penghitungan Dugaan dengan Analisis Regresi Kuadrat

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
7.843	8.966	12.037	12.151	14.783	17.349

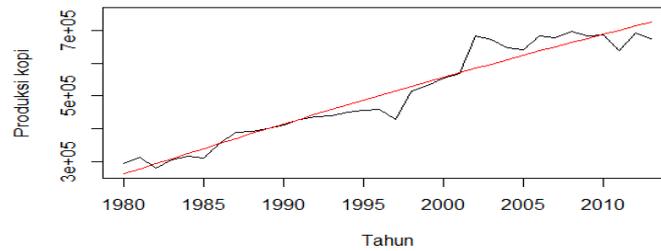
Model produksi kopi Indonesia diformulasikan sebagai berikut:

$$Y_t = 262574,23 + 15788,69 t - 52,35 T^2$$

Dengan menggunakan model regresi tersebut, dilakukan peramalan produksi kopi di Indonesia dengan hasil tersaji pada Tabel 18 dan plot peramalan tersaji pada Gambar 14.

**Tabel 18.** Hasil Ramalan Produksi Kopi di Indonesia, Tahun 2020-2024

No.	Tahun	Produksi Kopi (Ton)
1.	2020	809476,2
2.	2021	821002,3
3.	2022	832423,6
4.	2023	843740,3
5.	2024	854952,2



**Gambar 14.** Plot Data Aktual dan Ramalan Produksi Kopi di Indonesia

**g. Fungsi Transfer**

Peubah yang diasumsikan dapat mempengaruhi gejala produksi kopi adalah volume impor kopi, dan peubah ini akan digunakan untuk melakukan analisis menggunakan model Fungsi Transfer. Dari pola data volume impor dan uji Augmented Dickey-Fuller (Tabel 19) terlihat belum stasioner oleh karenanya sebelum melakukan pemodelan, series data volume impor ini dilakukan differencing terlebih dahulu (Tabel 20).

**Tabel 19.** Data Volume Impor Kopi yang belum Stasioner

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-12253  -4909  -1854   2365   32794

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.624e+03  4.407e+03  -1.957  0.06039 .
z.lag.1     -1.052e+00  3.101e-01  -3.393  0.00208 **
tt           9.793e+02  3.089e+02   3.171  0.00367 **
z.diff.lag  -8.864e-02  2.344e-01  -0.378  0.70813
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10100 on 28 degrees of freedom
Multiple R-squared:  0.5756, Adjusted R-squared:  0.5302
F-statistic: 12.66 on 3 and 28 DF, p-value: 2.069e-05

Value of test-statistic is: -3.3925 4.3658 6.1288

Critical values for test statistics:
      1pct  5pct 10pct
tau3  -4.15  -3.50 -3.18
phi2   7.02   5.13  4.31
phi3   9.31   6.73  5.61
```

**Tabel 20.** Data Volume Impor Kopi yang sudah stasioner

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-8408  -1448    -21    1371   45322

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
z.lag.1      -2.3476     0.3260  -7.201  6.3e-08 ***
z.diff.lag    0.4545     0.1991   2.283   0.03 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11070 on 29 degrees of freedom
Multiple R-squared:  0.8179, Adjusted R-squared:  0.8054
F-statistic: 65.14 on 2 and 29 DF, p-value: 1.877e-11

Value of test-statistic is: -7.2009

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Pencarian ordo model ARIMA input untuk volume impor dilakukan dengan fungsi auto arima. Dari hasil fungsi auto arima, diperoleh ordo ARIMA terbaik untuk volume impor adalah ARIMA(0,1,2).

**Tabel 21.** Hasil Fungsi Auto ARIMA Volume Impor Kopi

```
Series: train.h[, "Impor"]
ARIMA(0,1,2) with drift

Coefficients:
      ma1      ma2      drift
    -1.1438  0.3820  925.1664
s.e.    0.2080  0.2148  415.0943

sigma^2 estimated as 98803837:  log likelihood=-349.73
AIC=707.46  AICC=708.89  BIC=713.45

Training set error measures:
              ME          RMSE          MAE          MPE          MAPE          MASE          A
CF1
Training set -97.60567  9337.015  5296.416 -2021.593  2040.57  0.7969183  0.02067
)2
```

Model ARIMA (0,1,2) menghasilkan mape sangat tinggi (MAPE=2040,57) oleh sebab itu dicari model tentarif model input dengan fungsi armaselect.(Tabel.22).

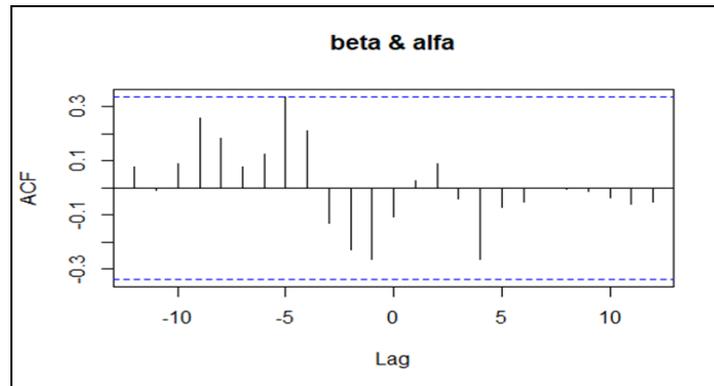
**Tabel 22.** Hasil Fungsi ARIMA Select Volume Impor Kopi

	p	q	sbc
[1,]	5	5	611.7790
[2,]	4	0	613.7263
[3,]	4	5	617.7032
[4,]	5	0	618.0143
[5,]	2	0	618.8429
[6,]	1	0	620.6314
[7,]	3	0	623.1914
[8,]	4	1	631.7059
[9,]	0	0	632.3897
[10,]	1	2	633.1285

Dari hasil pencarian ordo arima optimum volume impor kopi menggunakan fungsi auto arima, diputuskan model arima input volume impor kopi adalah ARIMA (0,1,2). Tahap selanjutnya adalah melakukan *prewhitening* dan analisis korelasi silang antara residual model arima volume impor kopi dan produksin kopi menggunakan ARIMA(0,1,2). Dari hasil plot *cfc* (*cross correlation function*) antara alfa (residual impor kopi) dan beta (residual produksi kopi), diperoleh hasil bahwa tidak terdapat *lag* dari pengaruh yang diberikan oleh impor kopi terhadap produksi kopi atau  $lag = 0$ . Hal tersebut dapat disimpulkan dari tidak adanya *bar* atau garis yang melewati ambang batas pada nilai *lag* positif (Gambar 15).

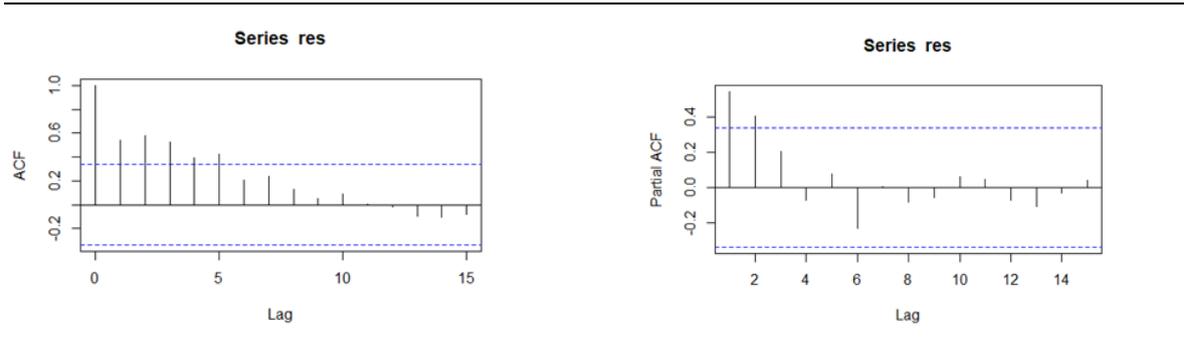
*Prewhitening* → alfa dan beta. Alfa adalah nilai residual dari model arima variabel input (vol impor kopi), sedangkan Beta adalah nilai residual dari model arima output (produksi kopi)

- pengaruh dari input volume impor kepada produksi kopi Indonesia terjadi saat yang sama, →  $b$  dan  $s = 0$ ,  $r$  atau jeda pengaruh variabel output = 0
- fungsi transfer alfa beta → residual → pendugaan ARIMA residual



**Gambar 15.** Plot ACF Alfa (Residual Impor Kopi) dan Beta (Residual Produksi Kopi)

Karena tidak adanya *bar* atau garis yang melewati ambang batas pada nilai *lag* positif, maka model ARIMA untuk pengepasan model fungsi transfer adalah ARIMA(0,0,0). Kemudian residual model ARIMA(0,0,0) tersebut diteliti stasioneritasnya. Dari plot ACF dan PCF residual, diperoleh hasil bahwa residual model cukup stasioner (Gambar 13) .



**Gambar 16.** Plot ACF dan PACF Residual Model ARIMA(0,0,0)

Langkah selanjutnya adalah pencarian model ARIMA terbaik untuk residual tersebut atau dapat disebut dengan *noise*. Hasil dari fungsi auto arima menunjukkan bahwa model ARIMA(0,1,1) merupakan model terbaik (Tabel 23). Namun karena ordo ARIMA(0,1,1) dinilai kurang baik untuk peramalan karena akan menghasilkan peramalan yang konstan, maka diperlukan referensi ordo ARIMA lain menggunakan fungsi *arma select* sebagaimana tampak pada Tabel 24.

**Tabel 23.** Hasil Fungsi Auto ARIMA Volume Impor Kopi

```

ARIMA(0,1,1)
Coefficients:
      ma1
      -0.6287
s.e.    0.1160

sigma^2 estimated as 7.224e+09:  log likelihood=-421.13
AIC=846.26  AICc=846.66  BIC=849.25

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      A
CF1
Training set 15870.96 82457.88 55566.1 29.83098 104.1706 0.9228574 -0.1701
856
  
```

**Tabel 24.** Hasil Pencarian Ordo ARIMA Optimum *Noise* menggunakan Fungsi Armaselect (diff)

	p	q	sbc
[1,]	1	0	755.2991
[2,]	2	0	755.7872
[3,]	3	0	760.3637
[4,]	4	0	763.2162
[5,]	0	0	763.2341
[6,]	5	0	767.1062
[7,]	0	1	772.0308
[8,]	1	1	775.3072
[9,]	0	2	776.4544
[10,]	2	1	778.2539

Setelah dilakukan uji signifikansi model ARIMA dan ARIMA-xreg pada pilihan ordo arima di atas, diperoleh ordo ARIMA *noise* yang signifikan adalah ARIMA(1,1,1) sebagaimana tampak pada Tabel 25. berikut.

**Tabel 25.** Hasil Uji Signifikansi ARIMA-Xreg (1,1,1)

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z )
ar1	-0.960026	0.065487	-14.6599	< 2.2e-16 ***
ma1	0.862428	0.109381	7.8846	3.156e-15 ***
xreg	0.101541	0.377266	0.2691	0.7878

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Oleh karena telah didapatkan model input serta *noise* yang terbaik untuk peramalan, maka langkah selanjutnya adalah melakukan peramalan. Peramalan dilakukan menggunakan dua jenis data input, yakni nilai aktual dan data ramalan. Dari hasil peramalan menggunakan nilai aktual sebagai input, diperoleh nilai MAPE sebesar 2,99% (Tabel 26 sementara peramalan menggunakan data ramalan sebagai input memiliki MAPE sebesar 3,08% (Tabel 27).

**Tabel 26.** Pengukuran Tingkat Kesalahan Model Fungsi Transfer dengan Nilai Aktual sebagai Input

Series: test.h[, "Produksi"]							
Regression with ARIMA(1,1,1) errors							
Coefficients:							
	ar1	ma1	xreg				
	-0.96	0.8624	0.1015				
s.e.	0.00	0.0000	0.0000				
sigma^2 estimated as 1.098e+09: log likelihood=-59.15							
AIC=120.29 AICC=121.63 BIC=119.9							
Training set error measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	AC
F1							
Training set	20305.25	29381.45	21509.71	2.80622	2.994606	0.8528154	0.33110
02							

**Tabel 27.** Pengukuran Tingkat Kesalahan Model Fungsi Transfer dengan Data Ramalan sebagai Input

```

Series: test.h[, "Produksi"]
Regression with ARIMA(1,1,1) errors

Coefficients:
      ar1      ma1      xreg
    -0.96  0.8624  0.1015
s.e.   0.00  0.0000  0.0000

sigma^2 estimated as 1.098e+09:  log likelihood=-59.32
AIC=120.64  AICC=121.97  BIC=120.25

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      AC
F1
Training set 20692.26 30405.06 22080.45 2.858231 3.075354 0.875444 0.32922
91

```

Setelah diperoleh model fungsi transfer yang optimal, maka model tersebut digunakan untuk melakukan peramalan produksi kopi 5 tahun ke depan (2020-2024). Hasil peramalan produksi kopi Indonesia menggunakan fungsi transfer dengan volume impor kopi sebagai peubah ditampilkan pada Tabel 28 berikut.

**Tabel 28.** Hasil Peramalan Produksi Kopi menggunakan Fungsi Transfer

No.	Tahun	Produksi Kopi (Ton)
1.	2020	763514,7
2.	2021	761924,3
3.	2022	763559,1
4.	2023	762054,4
5.	2024	763439,4

#### **h. Model Vector Auto Regression (VAR) untuk Peramalan Produksi Kopi**

Pemodelan dengan menggunakan model VAR dilakukan dengan memasukkan peubah produksi kopi, luas areal kopi, volume ekspor kopi, volume impor kopi, harga kopi dunia, dan harga pupuk urea dunia. Hal ini diasumsikan bahwa apabila terjadi gejolak pada variabel-variabel tersebut, akan mempengaruhi gejolak produksi kopi ataupun sebaliknya.

Pada tahap pertama, perlu dilakukan pemeriksaan apakah terdapat pengaruh trend dalam model. Hasil pengujian signifikansi trend dilakukan pada  $lag=1$  dan  $lag=2$ . Hasil pengujian trend pada  $lag=1$  menunjukkan bahwa *trend* berpengaruh signifikan pada persamaan produksi kopi (signifikan 5%) dan persamaan harga pupuk urea dunia (signifikan 10%). Sementara hasil pengujian *trend* pada  $lag=2$  menunjukkan bahwa *trend* berpengaruh signifikan hanya pada persamaan harga pupuk urea dunia (signifikan 5%).

Dari hasil *running* model menggunakan *lag* 1 dan 2, diperoleh informasi komponen *trend* tidak terlalu berpengaruh signifikan sehingga *trend* dikeluarkan dalam model untuk *running* model VAR (p) selanjutnya.

Langkah selanjutnya adalah menentukan *lag* (p) yang akan digunakan dalam model VAR. Penentuan *lag* (p) dilakukan dengan melakukan pengujian VAR(1) hingga VAR(4) dengan pertimbangan kecukupan series data. Dari hasil *running* menggunakan p=1 s.d p=4, diperoleh informasi bahwa variabel yang signifikan cukup banyak pada p=4, sehingga model VAR yang digunakan selanjutnya adalah VAR (4). Hal tersebut diperkuat juga dengan nilai *Adjusted R-squared*: 0,9683 atau dapat dikatakan bahwa 96,83% keragaman produksi kopi dapat dijelaskan oleh kelima variabel/peubah penjelas termasuk konstanta. Hasil pengujian model VAR(4) dapat dilihat dalam Tabel 29.

Model VAR(5) dan VAR(6) tidak dirun karena keterbatasan series. Jika dirun maka muncul error.

Dugaan model VAR(4) untuk produksi kopi adalah sebagai berikut :

**Tabel 29.** Hasil Estimasi VAR Produksi Kopi dengan Peubah Luas Areal, Volume Ekspor, Volume Impor, Harga Kopi Dunia, dan Harga Pupuk Urea Dunia pada Lag=4 tanpa *Trend*

```

VAR Estimation Results:
=====
Endogenous variables: Produksi, Luas, Ekspor, Impor, World.Price, Urea.Price
Deterministic variables: const
Sample size: 36
Log Likelihood: -1731.958
Roots of the characteristic polynomial:
1.014 1.008 1.008 1.007 1.007 0.9567 0.9567 0.9066 0.8967 0.8967 0.8838 0.8838 0.842 0.842 0.8
076 0.8076 0.8058 0.7772 0.7772 0.7485 0.7485 0.7111 0.7111 0.4808
Call:
VAR(y = kopi[1:40, c(2, 3, 4, 5, 6, 7)], p = 4, type = "const")

Estimation results for equation Produksi:
=====
Produksi = Produksi.l1 + Luas.l1 + Ekspor.l1 + Impor.l1 + world.Price.l1 + Urea.Price.l1 + Pro
duksi.l2 + Luas.l2 + Ekspor.l2 + Impor.l2 + world.Price.l2 + Urea.Price.l2 + Produksi.l3 + Lua
s.l3 + Ekspor.l3 + Impor.l3 + world.Price.l3 + Urea.Price.l3 + Produksi.l4 + Luas.l4 + Ekspor.
l4 + Impor.l4 + world.Price.l4 + Urea.Price.l4 + const

      Estimate Std. Error t value Pr(>|t|)
Produksi.l1    4.636e-01  2.804e-01  1.653  0.1265
Luas.l1        1.047e-01  1.827e-01  0.573  0.5780
Ekspor.l1     -2.923e-02  1.507e-01  -0.194  0.8497
Impor.l1       1.608e-01  5.854e-01  0.275  0.7886
world.Price.l1 1.906e+04  1.290e+04  1.478  0.1675
Urea.Price.l1  -4.296e+02  1.491e+02  -2.881  0.0149 *
Produksi.l2    1.712e-01  3.222e-01  0.531  0.6058
Luas.l2        2.401e-01  2.183e-01  1.100  0.2949
Ekspor.l2     3.136e-02  1.255e-01  0.250  0.8073
Impor.l2       7.107e-01  9.513e-01  0.747  0.4707
world.Price.l2 -2.942e+03  1.654e+04  -0.178  0.8621
Urea.Price.l2  8.536e+01  2.005e+02  0.426  0.6785
Produksi.l3    3.720e-01  2.840e-01  1.310  0.2169
Luas.l3       -2.931e-01  2.070e-01  -1.416  0.1844
Ekspor.l3     7.212e-02  1.222e-01  0.590  0.5671
Impor.l3     -1.292e+00  1.038e+00  -1.245  0.2390
world.Price.l3 7.363e+03  1.476e+04  0.499  0.6276
Urea.Price.l3 -2.706e+02  1.803e+02  -1.500  0.1617
Produksi.l4    3.659e-01  2.853e-01  1.283  0.2260
Luas.l4       -7.985e-02  1.728e-01  -0.462  0.6530
Ekspor.l4    -2.919e-02  1.126e-01  -0.259  0.8003
Impor.l4     -1.163e+00  8.975e-01  -1.295  0.2217
world.Price.l4 7.241e+03  1.453e+04  0.498  0.6281
Urea.Price.l4  1.429e+02  1.856e+02  0.770  0.4577
const        -1.292e+05  1.180e+05  -1.095  0.2970
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 24300 on 11 degrees of freedom
Multiple R-Squared:  0.99, Adjusted R-squared: 0.9683
F-statistic: 45.6 on 24 and 11 DF, p-value: 6.612e-08

```

**Tabel 30.** Covariance Matrix Residuals

Covariance matrix of residuals:

	Produksi	Luas	Ekspor	Impor	World.Price	Urea.Price
Produksi	590321215	379894008	-33999092	-8045157	-1514.0674	-296067.84
Luas	379894008	1241910229	-505979033	105290510	-5673.9450	225371.16
Ekspor	-33999092	-505979033	3821548538	-694345666	-9771.1332	559509.19
Impor	-8045157	105290510	-694345666	210536858	2279.1237	-144186.41
World.Price	-1514	-5674	-9771	2279	0.1691	1.13
Urea.Price	-296068	225371	559509	-144186	1.1302	2576.04

Correlation matrix of residuals:

	Produksi	Luas	Ekspor	Impor	World.Price	Urea.Price
Produksi	1.00000	0.4437	-0.02264	-0.02282	-0.15156	-0.24009
Luas	0.44368	1.00000	-0.23226	0.20591	-0.39158	0.12600
Ekspor	-0.02264	-0.2323	1.00000	-0.77409	-0.38442	0.17832
Impor	-0.02282	0.2059	-0.77409	1.00000	0.38202	-0.19579
World.Price	-0.15156	-0.3916	-0.38442	0.38202	1.00000	0.05416
Urea.Price	-0.24009	0.1260	0.17832	-0.19579	0.05416	1.00000

Setelah terpilih *lag* model VAR yang optimal, maka perlu dilakukan pengujian asumsi non-autokorelasi, normalitas, dan homoskedas pada sisaan model VAR (4). Dari hasil pengujian non autokorelasi, diperoleh informasi bahwa model tersebut terbebas dari autokorelasi dengan tingkat kesalahan 5% (Tabel 30). Residual memenuhi asumsi kenormalan dengan tingkat kesalahan 5% melalui *normality test* (*Jarque Bera Test*, pengujian skewness, dan kurtosis) sebagaimana tampak pada Tabel 32, sedangkan asumsi homoskedastisitas atau kehomogenan ragam signifikan terpenuhi dengan *p-value* =1 (Tabel 33).

**Tabel 31.** Hasil Uji Asumsi Non-Autokorelasi model VAR(2) menggunakan *Serial Test*

Portmanteau Test (asymptotic)

data: Residuals of VAR object varhsheet4  
Chi-squared = 442.73, df = 432, **p-value = 0.3502**

**Tabel 32.** Hasil Uji Asumsi Normalitas Regresi Linier Produksi Kopi menggunakan *Normality Test*

<pre> \$JB     JB-Test (multivariate) data: Residuals of VAR object varhsheet4 Chi-squared = 6.4972, df = 12, <b>p-value = 0.889</b>  \$Skewness     Skewness only (multivariate) data: Residuals of VAR object varhsheet4 Chi-squared = 2.6627, df = 6, <b>p-value = 0.8498</b>  \$Kurtosis     Kurtosis only (multivariate) data: Residuals of VAR object varhsheet4 Chi-squared = 3.8345, df = 6, <b>p-value = 0.6991</b> </pre>
---

**Tabel 32.** Hasil Uji Asumsi Homoskedas Regresi Linier Produksi Kopi menggunakan ARCH Test

<pre> ARCH (multivariate) data: Residuals of VAR object varhsheet4 Chi-squared = 651, df = 2205, <b>p-value = 1</b> </pre>
--

Sebelum melakukan peramalan, perlu dilakukan evaluasi nilai kesalahan model VAR yang terbentuk dengan menghitung nilai MAPE atau nilai rata-rata persentase kesalahan peramalan. Nilai MAPE VAR(4) untuk dataset *training* dan *testing* masing-masing sebesar 4,29% dan 1,41% seperti tersaji pada Tabel 34.

**Tabel 34.** Nilai MAPE Model VAR(4) Produksi Kopi Indonesia

Produksi	
Min. :28.20	Min. :0.1072
1st Qu.:35.49	1st Qu.:0.5621
Median :37.63	Median :1.4033
<b>Mean :37.56</b>	<b>Mean :2.0478</b>
3rd Qu.:41.73	3rd Qu.:2.7757
Max. :44.01	Max. :6.7181
MAPE data testing=37,56	MAPE data training=2,05

Setelah diperoleh model VAR yang optimal yakni VAR(4), maka model tersebut digunakan untuk melakukan peramalan produksi kopi 5 tahun ke depan (2020-2024). Hasil peramalan produksi kopi Indonesia menggunakan metode VAR dengan peubah produksi kopi, volume ekspor kopi, volume impor kopi, harga kopi dunia, dan harga pupuk urea dunia ditampilkan pada Tabel 35. berikut.

**Tabel 35.** Hasil Peramalan Produksi Kopi menggunakan VAR(4)

No.	Tahun	Produksi Kopi (Ton)
1.	2020	825.447
2.	2021	864.817
3.	2022	903.968
4.	2023	1.023.836
5.	2024	1.088.755

## KESIMPULAN DAN SARAN

### Kesimpulan

- Hasil analisis dari empat model dalam meramalkan produksi kopi adalah sebagai berikut:
  - Model ARIMA terbaik adalah ARIMA (2,1,0) dengan MAPE 8,32%
  - Model regresi linier terbaik menghasilkan MAPE 14,47% dengan nilai probabilitas < 0,05 dan  $R^2$  sebesar 93,66%.
  - Model regresi kuadratik terbaik memiliki MAPE sebesar 12,15% dengan nilai probabilitas < 0,05 dan  $R^2$  sebesar 93,76%.

- Model Fungsi Transfer terbaik dengan peubah input volume impor kopi adalah ARIMA (0,0,0)(1,1,1) dengan model input ARIMA (0,1,1). Model Fungsi Transfer tersebut memiliki MAPE sebesar 6,78%.
- Model VAR terbaik dengan memasukkan peubah luas areal kopi, volume ekspor kopi, volume impor kopi, harga kopi dunia, dan harga urea dunia adalah VAR(4) tanpa *trend* (hanya konstanta). MAPE yang dihasilkan adalah 37,56%.
- Dari keempat model yang dicobakan, model terbaik adalah model Fungsi Transfer dengan melihat MAPE terkecil yaitu 6,78%
- Hasil ramalan produksi kopi model fungsi transfer dengan peubah input volume impor kopi dengan ARIMA (0,0,0)(1,1,1) dan model input ARIMA (0,1,1) untuk 2020 sampai 2024 adalah 763.515 Ton, 761.924 Ton, 763.559 Ton, 762.054 Ton, 763.439 Ton.

### **Saran**

- Perlu dilakukan kajian mendalam dengan metode peramalan lainnya.
- Untuk model regresi, perlu diujicobakan menggunakan variabel bebas/peubah input selain waktu.
- Untuk model fungsi transfer dan VAR perlu diujicobakan dengan menggunakan variabel lain yang sekiranya lebih berpengaruh, baik secara teoritis maupun praktis.
- Perlu dikaji metode peramalan untuk data non parametrik, sehingga tidak terhalang oleh tidak terpenuhinya berbagai asumsi.

## **DAFTAR PUSTAKA**

- Direktorat Jenderal Perkebunan. 2019. Statistik Perkebunan Indonesia 2018-2020 Jakarta: Direktorat Jenderal Perkebunan, Kementerian Pertanian. Statistik Perkebunan Indonesia 2018-2020.
- Draper, N. R, dan Smith, H. 1992. Analisis Regresi Terapan, Jakarta : PT Gramedia Pustaka Utama.
- Gujarati, Damodar. N dan Porter, Dawn. C. 2009. Basic Econometrics. Boston: Douglas Reiner.
- Malian, Husni & Sudi Mardianto dan Mewa Ariani. 2004. Faktor-faktor yang Mempengaruhi Produksi, Konsumsi dan Harga Beras serta Inflasi Bahan Makanan. Jurnal Agro Ekonomi, Volume 22 No. 2: 119-146
- Siagian, Viktor dan Muchamad Yusron. 2015. Analisis Faktor-Faktor yang Mempengaruhi Luas Lahan Garapan Usahatani Padi Sawah di Provinsi Banten. Lampung: Politeknik Negeri Lampung
- Thamrin, Syahrini. 2014. Faktor-faktor yang Mempengaruhi Produksi Usahatani Kopi Arabika di Kabupaten Enrekang Sulawesi Selatan. Pangkajene Kepulauan: Politeknik Pertanian Negeri Pangkajene Kepulauan



# KAJIAN METODE ESTIMASI PRODUKSI TEH INDONESIA

Diah Indarti  
Fungsional Statistisi, Pusdatin-Kementan

## ABSTRAK

Komoditas teh memiliki peranan penting dalam perekonomian nasional yaitu sebagai sumber pendapatan petani, penyerapan tenaga kerja, sumber devisa Negara, mendorong agroindustri pengembangan wilayah dan pelestarian lingkungan. Pada umumnya tanaman teh nasional dikembangkan di Indonesia sejak jaman Belanda. Khusus untuk teh rakyat mulai dikembangkan sekitar Tahun 1980-an sehingga kondisi tanaman pada umumnya merupakan tanaman tua/rusak dengan produktivitas yang sudah menurun dan sudah saatnya dilakukan perbaikan budidaya melalui rehabilitasi dan intensifikasi tanaman (Kementerian Pertanian, 2013)

Pada Tahun 2020 ini Pusat Data dan Sistem Informasi Pertanian bersama dengan Direktorat Jenderal Perkebunan dan Badan Pusat Statistik mengkaji beberapa metode peramalan untuk menghasilkan angka estimasi yang nilainya cenderung mendekati angka aktualnya. Tujuan dari makalah ini melakukan pemodelan, membandingkan efektivitas dari pemodelan tersebut dan menentukan metode terbaik dalam meramalkan dengan model ARIMA, Regresi, Fungsi Transfer dan VAR pada produksi teh di Indonesia.

**Kata Kunci :** ARIMA, regresi, fungsi transfer, VAR, estimasi produksi

## PENDAHULUAN

### Latar Belakang

Estimasi/peramalan adalah ilmu yang mempelajari tentang prediksi peristiwa masa depan dengan pengambilan data historis dan memproyeksikannya ke masa depan dengan menggunakan model matematis (Heizer et al.,2011). Estimasi tidak akan menggambarkan kejadian atau peristiwa yang sebenarnya, tetapi dalam model matematis peramalan dapat dibuat sedemikian sehingga galat (error) peramalan (kesalahan peramalan) yang diperoleh adalah sekecil-kecilnya. Galat biasanya diukur dalam *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Root Mean Square Absolute Percentage Error* (RMSAPE), dan lain-lain.

Metode analisis yang digunakan untuk mencari pengaruh data historis terhadap data yang akan datang agar memperoleh hasil estimasi adalah analisis deret waktu. Data yang digunakan dalam analisis ini harus memiliki interval yang sama, misalkan dalam harian, bulanan, atau tahunan. Analisis deret waktu dapat digunakan untuk menganalisis data yang terdiri satu variabel (*univariate*) dan data yang terdiri lebih dari satu variabel (*multivariate*). Kelemahan dari analisis deret waktu univariat adalah tidak diperhi-tungkannya pengaruh variabel-variabel lain di luar model yang mungkin saja memiliki pengaruh signifikan terhadap model karena pada analisis deret waktu univariat hanya dianalisis satu variabel terhadap data

historisnya. Sedangkan pada analisis deret waktu multivariat dapat dianalisis hubungan dinamis antar variabel dan meramalkan data secara simultan (Tsay, 2014).

Pada tahun 2020 Pusdatin membahas dan mengkaji beberapa metode peramalan guna menghasilkan angka yang nilainya semakin mendekati nilai sebenarnya (aktual). Pemodelan yang diterapkan adalah pemodelan ARIMA, regresi, fungsi transfer dan *Vector Auto Regression* (VAR) dengan menggunakan *software Program R Studio*.

Makalah ini mengulas hasil penerapan metode tersebut untuk meramal angka produksi teh khususnya di Indonesia serta membahas kelebihan dan kekurangan masing-masing metode yang diterapkan.

## **Tujuan**

Tujuan dari makalah ini adalah:

- a. Melakukan analisis dan pemodelan ARIMA, Regresi, Fungsi Transfer dan VAR untuk meramalkan produksi teh di Indonesia.
- b. Membandingkan efektivitas metode tersebut dalam hasil peramalan produksi teh.
- c. Menentukan metode terbaik dalam mengestimasi data produksi komoditas teh di Indonesia.

## **METODOLOGI**

### **Sumber Data**

Data yang digunakan dalam kajian ini adalah data sekunder yang diperoleh dari BPS dan Direktorat Jenderal Perkebunan berupa data tahunan dari 1980 sampai dengan 2019. Datanya terdiri dari data produksi teh, luas areal teh, harga konsumen teh, volume ekspor dan volume impor (Badan Pusat Statistik) di Indonesia. Data tahun 1980 sampai dengan 2013 digunakan untuk membangun model sedangkan data tahun 2014 sampai dengan 2019 digunakan untuk validasi model.

### **Metode**

Tahapan yang dilakukan dalam kajian ini adalah:

1. Melakukan eksplorasi data untuk mengetahui pola data.
2. Bagi data menjadi dua bagian, untuk penyusunan model (*training*) dan untuk validasi model (*testing*).
3. Susun model dengan berbekal data training.
4. Lakukan peramalan dengan masing-masing metode sesuai dengan periode data *training* dan *testing*.

#### **a. Metode ARIMA**

Langkah-langkahnya adalah sebagai berikut:

- 1) Memeriksa kestasioneran data produksi teh. Data dikatakan stasioner jika berfluktuasi di sekitar rata-rata untuk seluruh periode waktu. Hal ini dapat dilakukan dengan melihat plot data, plot ACF dan PACF serta uji *Augmented Dickey Fuller*.

- 2) Identifikasi ordo AR dan MA pada deret yang stasioner dari hasil amatan plot ACF dan PACF. Model AR ( $p$ ) dicirikan dengan pola *tail off* pada plot ACF dan pola *cut off* setelah *lag* ke- $p$  pada plot PACF. Model MA ( $q$ ) dicirikan dengan pola *cut off* setelah *lag* ke- $q$  pada plot ACF dan pola *tail off* pada plot PACF. Apabila ACF dan PACF keduanya berpola *tail off*, maka dapat dikatakan model yang digunakan adalah ARMA ( $p,q$ ). Bila terdapat proses differencing karena proses penstasioneran, maka model yang terbentuk adalah ARIMA ( $p,d,q$ ) dengan  $d$  merupakan banyaknya proses *differencing* dari data yang tidak stasioner.
- 3) Pendugaan model tentatif. Dari data yang sudah stasioner dilakukan penentuan besarnya ordo  $p$  dan  $q$ . Nilai ordo  $p$  ditentukan dari banyaknya koefisien pada plot PACF pertama yang berbeda nyata. Sedangkan nilai ordo  $q$  ditentukan dari banyaknya koefisien pada plot ACF pertama yang berbeda nyata. Perlu diperhatikan juga perilaku musiman ( $S$ ). Sehingga model yang diperoleh dapat ditulis dengan ARIMA ( $p,d,q$ )( $P,D,Q$ ) <sup>$S$</sup> . Penentuan komponen  $P,D,Q$  musiman sama halnya dengan penentuan  $p,d,q$  non musiman hanya diterapkan pada lag berjeda  $S$ .
- 4) *Overfitting*, merupakan penambahan komponen model untuk evaluasi bila ada komponen sistematis yang terlewat dan tercakup dalam sisaan dengan tujuan untuk memperoleh model terbaik. Bila tambahan komponen di model-model *overfitting* tidak nyata, maka kembali ke model awal.
- 5) Peramalan dengan menggunakan model ARIMA terbaik. Perhitungan dilakukan secara rekursif, yaitu menghitung peramalan satu periode kemudian dua periode, dan seterusnya sampai  $n$  periode ke depan.
- 6) Setelah nilai ramalan diperoleh, langkah berikutnya adalah evaluasi hasil ramalan tersebut pada dua hal: a) *magnitude* (nilai data) dan b) *direction* (pergerakan data). Bila pergerakan hasil model baik untuk *training* maupun *testing* dapat mengikuti pergerakan data aktual dengan baik, nilai data MAPE (*Mean Absolute Percentage Error*) yang diperoleh pun cukup kecil, maka model tersebut layak digunakan untuk peramalan.

## **b. Metode Regresi**

Langkah-langkahnya adalah sebagai berikut:

- 1) Penentuan parameter untuk pembentukan model.
- 2) Pendugaan model.
- 3) Pengujian hipotesis dengan Analisis Varians.
- 4) Peramalan dengan model yang terbentuk.
- 7) Evaluasi secara *magnitude* (MAPE) dan *direction* (Korelasi). Bila hasil keduanya bagus, maka model tersebut layak digunakan untuk peramalan.

## **c. Metode Fungsi Transfer**

Langkah-langkahnya adalah sebagai berikut:

- 1) Memeriksa kestasioneran deret input luas tanaman menghasilkan dan deret output produksi teh.

- 2) Membuat model ARIMA untuk deret input.
- 3) Melakukan *prewhitening* deret input.
- 4) Melakukan *prewhitening* deret output berdasarkan hasil *prewhitening* deret input.
- 5) Menghitung korelasi silang deret input dengan deret output untuk melihat hubungannya.
- 6) Menetapkan nilai  $b$ ,  $r$  dan  $s$  dengan melihat plot korelasi silang deret input dan output dengan ketentuan sebagai berikut:
  - a) Nilai  $b$  ditentukan berdasarkan *lag* positif yang pertama kali nyata pada plot korelasi silang.
  - b) Nilai  $s$  dilihat dari lamanya deret input mempengaruhi deret output setelah nyata yang pertama (nilai  $b$ ).
  - c) Nilai  $r$  menunjukkan lamanya deret output berhubungan dengan nilai terdahulu dari deret output itu sendiri. Nilai  $r$  dilihat dari plot *ACF* deret output atau ditentukan berdasarkan pola *lag* ( $b+s$ ), jika memiliki pola eksponensial maka  $r = 1$  dan memiliki pola gelombang sinus maka  $r = 2$ .
- 7) Membuat model ARIMA deret sisaan dengan mengamati plot *ACF* dan *PACF* deret sisaan.
- 8) Pendugaan model fungsi transfer.
- 9) Diagnosa model. Jika terdapat parameter yang tidak nyata, maka kembali ke langkah 6.
- 10) Peramalan dengan model yang diperoleh.

#### d. Metode VAR

Langkah-langkahnya adalah sebagai berikut:

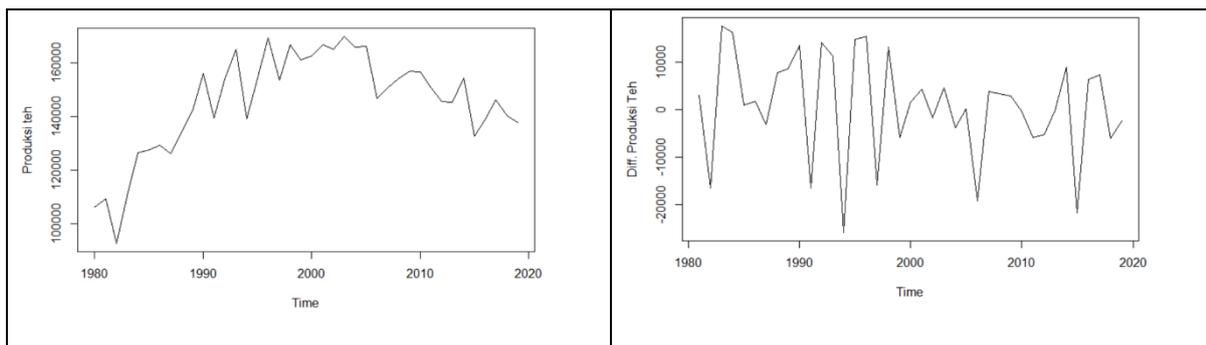
- 1) Menguji kestasioneran data.
  - 2) Menentukan panjang lag optimal. Jumlah lag ditentukan dengan menggunakan *Adjusted-R<sup>2</sup>* atau *AIC (Akaike Information Criterion)* terkecil. Bisa juga dengan mengambil Minimum Information Criterion terkecil.
  - 3) Dugaan parameter.
  - 4) Memeriksa korelasi silang antar sisaan.
  - 5) Diagnosis model satu per satu untuk model keseluruhan dan sisaan.
  - 6) Mengkaji respon impuls, yang merupakan informasi untuk pengaruh perubahan *shock* atau guncangan suatu peubah terhadap peramalan peubah lain. Dapat dilihat melalui plot respon impuls.
  - 7) Mengkaji dekomposisi ragam, yang berarti proporsi keragaman galat suatu peubah yang dijelaskan oleh galat masing-masing peubah dan galat peubah lain.
  - 8) Melakukan peramalan.
5. Memilih model terbaik dari keempat metode yang diterapkan. Semakin kecil nilai MAPE dan pola model mendekati garis aktual, menunjukkan data hasil peramalan semakin mendekati nilai aktual. Semakin kecil nilai AIC maka model semakin tepat.

6. Melakukan pendugaan ulang dengan model terbaik pada seluruh data (gabungan data *training* dan *testing*).
7. Menggunakan model terbaik untuk peramalan sesuai periode yang diinginkan. Dalam hal ini hingga tahun 2024.

## HASIL DAN PEMBAHASAN

### MODEL ARIMA

#### Eksplorasi Data



Gambar 1. Plot Produksi Teh, 1980-2019

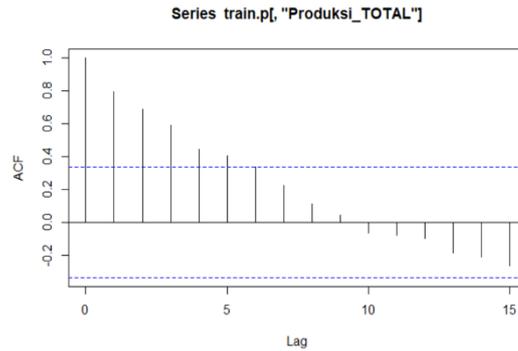
Gambar 2. Plot Produksi Teh *Differencing*

Gambar 1 adalah plot produksi teh yang belum stasioner, kemudian dilakukan *differencing* untuk mendapatkan plot yang stasioner (Gambar 2). Stasioneritas berarti bahwa tidak terdapat perubahan yang drastis pada data. Fluktuasi data berada disekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variansi dari fluktuasi tersebut (Makridakis, 19951). Data time series dikatakan stasioner jika rata-rata dan variansinya konstan, tidak ada unsur trend dalam data, dan tidak ada unsur musiman. Apabila data tidak stasioner, maka perlu dilakukan modifikasi untuk menghasilkan data yang stasioner. Salah satu cara yang umum dipakai adalah metode pembedaan (*differencing*). Untuk menentukan apakah series stasioner, nonstasioner dapat dibantu dengan melihat plot dari series atau bentuk *difference*-nya. Proses *differencing* dapat dilakukan untuk beberapa periode sampai data stasioner, yaitu dengan cara mengurangkan suatu data dengan data sebelumnya.

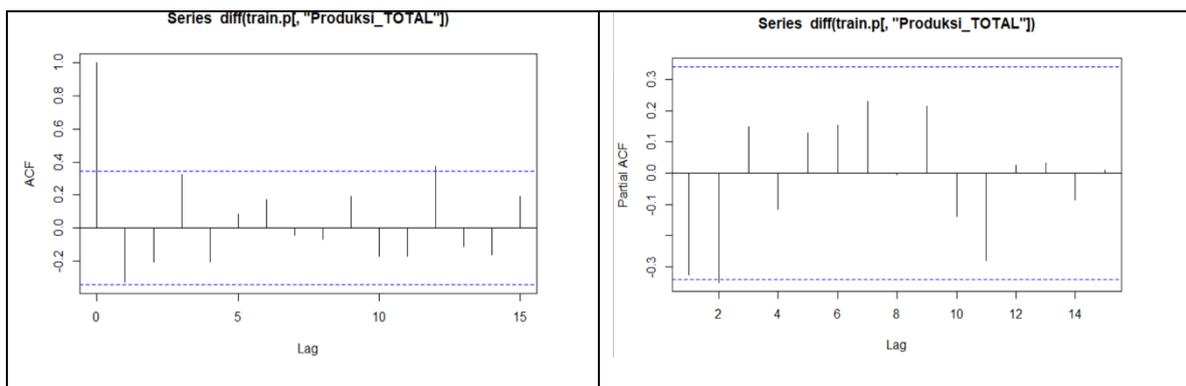
#### Uji Kestasioneran Data

Data deret waktu memerlukan transformasi dan pembedaan untuk mencapai kestasioneran data. Transformasi diperlukan agar deret waktu stasioner dalam ragam. Sedangkan pembedaan diperlukan agar deret waktu stasioner dalam rata-rata. Kestasioneran data dapat diketahui pengujian secara visual dan uji statistik. Uji secara visual dapat dilihat dari plot data ACF (Gambar 3-5). Agar tidak terjadi unsur subjektivitas, maka diperlukan uji statistik, yaitu menggunakan Uji *Augmented Dickey Fuller*. Hasil uji *Augmented Dickey Fuller* (Tabel 1) mengindikasikan bahwa data produksi teh konstan di sekitar rata-rata bukan nol atau berpola

Single Mean. Hal ini terlihat dari nilai  $p$  yang lebih besar dari taraf nyatanya ( $\alpha=0,05$ ). Oleh karena itu, data bisa dapat langsung digunakan untuk pembentukan model.



Gambar 3. Plot *Auto Correlation Function* (ACF) Data Training Produksi Teh



Gambar 4. Plot ACF Produksi Teh *Differencing*

Gambar 5. Plot ACF Produksi Teh *Differencing*

Tabel 1. Hasil Uji *Augmented Dickey-Fuller* Produksi Teh

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
    Min     1Q   Median     3Q    Max
-25174.1 -4028.3  786.4  6191.0 19648.1
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 28395.8615 14605.3997  1.944  0.0602 .
z.lag.1     -0.1691    0.1102  -1.534  0.1343
tt         -125.3212   178.9312  -0.700  0.4885
z.diff.lag  -0.3103    0.1618  -1.918  0.0636 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 10070 on 34 degrees of freedom  
 Multiple R-squared: 0.2403, Adjusted R-squared: 0.1732  
 F-statistic: 3.584 on 3 and 34 DF, p-value: 0.0236  
 Value of test-statistic is: -1.5339 1.9863 2.7699  
 Critical values for test statistics:

	1pct	5pct	10pct
tau3	-4.15	-3.50	-3.18
phi2	7.02	5.13	4.31
phi3	9.31	6.73	5.61

Berdasarkan Gambar 1 dan Tabel 1 belum terlihat stationer untuk itu dilakukan *differencing*. Hal ini dapat. Nilai test-statistic lebih besar dari critical values (nilai tau3) baik di taraf 1%, 5% maupun 10% maka  $H_0$  gagal ditolak. Terlihat bahwa data produksi teh belum stationer, maka dilakukan stasioneritas dengan melakukan differencing dan dilakukan pengujian ADF kembali.

Dari hasil *Augmented Dickey-Fuller test* pada Tabel 2, menunjukkan bahwa produksi teh telah stationer. Hal tersebut diketahui dari nilai statistik uji (-7,1641) yang lebih kecil dibandingkan nilai kritis taraf 5% yakni -1,95.

Tabel 2. Hasil Uji Augmented Dickey-Fuller *Differencing* Produksi Teh

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
Residuals:
  Min    1Q  Median    3Q   Max
-20543 -3697  2189  7887 20077
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1   -1.7978    0.2509  -7.164 2.35e-08 ***
z.diff.lag  0.3546    0.1536   2.308  0.027 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 9720 on 35 degrees of freedom
Multiple R-squared:  0.7193, Adjusted R-squared:  0.7033
F-statistic: 44.84 on 2 and 35 DF, p-value: 2.21e-10
Value of test-statistic is: -7.1641
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Data produksi teh telah stasioner maka langkah selanjutnya adalah menentukan ordo AR dan MA menggunakan fungsi auto arima. Dari hasil pendugaan auto arima pada Tabel 3, maka model ARIMA yang terpilih adalah ARIMA (0,1,1) yang berarti model ARIMA yang optimal untuk mengestimasi data yang diolah adalah ARIMA dengan ordo *Auto Regressive* (AR) = 0, ordo *differencing* = 1, dan ordo *Moving Average* (MA) = 1.

**Tabel 3.** Hasil Fungsi Auto ARIMA Produksi Teh

```
Series: train.p[, "Produksi_TOTAL"]
ARIMA(0,1,1)
Coefficients:
    ma1
   -0.3751
s.e.  0.1419
sigma^2 estimated as 104899113: log likelihood=-351.12
AIC=706.25  AICc=706.65  BIC=709.24
Training set error measures:
           ME  RMSE    MAE    MPE  MAPE
Training set 1897.353 9936.225 7702.993 1.14934 5.490948
           MASE    ACF1
Training set 0.9136643 -0.05934335
```

Dari hasil Pencarian Ordo ARIMA Optimum menggunakan Fungsi Armaselect, *differencing* 2 maka diperoleh ARIMA terbaik adalah ARIMA (1,2,2).

**Tabel 4.** Hasil Pencarian Ordo ARIMA Optimum menggunakan Fungsi Armaselect, *differencing* 2

```
   p q  sbc
[1,] 5 5 563.1253
[2,] 4 4 566.1299
[3,] 4 5 568.2073
[4,] 5 4 569.6058
[5,] 3 5 574.6094
[6,] 1 2 576.5459
[7,] 2 1 577.2052
[8,] 1 5 577.3801
[9,] 5 1 577.5171
[10,] 2 2 578.3250
```

**Tabel 5.** Hasil Pengujian ARIMA (1,2,2)

```
Series: teh[, "Produksi_TOTAL"]
ARIMA(1,2,2)
Coefficients:
    ar1  ma1  ma2
   -0.827  0.0096 -0.9904
s.e.  0.104  0.3583  0.3580
```

sigma<sup>2</sup> estimated as 1.19e+08: log likelihood=-408  
 AIC=823.99 AICc=825.21 BIC=830.54

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-1438.678	10203.56	7567.449	-1.113438	5.413107	0.8917245	-0.3992642
Test set	-67193.05	83919.37	67193.05	-10.21102	10.21102	2.8222219	NA

Hasil uji statistik terhadap model ARIMA(1,2,2) menunjukkan bahwa nilai MAPE data training sebesar 5,41% dan untuk data testing sebesar 10,21%. Nilai tersebut lebih besar daripada model ARIMA(0,1,1) namun model ARIMA(1,2,2) merupakan model yang ideal untuk melakukan peramalan karena tidak menghasilkan nilai ramalan yang konstan. Hal tersebut didukung oleh hasil uji *Ljung-Box* test yang menunjukkan bahwa residual model tidak memiliki autokorelasi, yang ditandai dengan nilai *p-value* lebih besar dari taraf uji yakni 5% (Tabel 6).

**Tabel 6.** Hasil Uji *Ljung-Box* test ARIMA (1,2,2)

Call:

```
arima(x = train.p[, "Produksi_TOTAL"], order = c(1, 2, 2))
```

Coefficients:

```
      ar1    ma1    ma2
-0.8169 0.0001 -0.9998
s.e. 0.1139 0.2208 0.2207
sigma^2 estimated as 107790757: log likelihood = -343.72, aic = 695.44
```

```
> library(lmtest)
```

```
> coefptest(model1)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z )
ar1	-8.1694e-01	1.1391e-01	-7.1720	7.394e-13 ***
ma1	6.7627e-05	2.2078e-01	0.0003	0.9998
ma2	-9.9981e-01	2.2074e-01	-4.5293	5.919e-06 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Setelah diperoleh model ARIMA yang optimal, maka model ARIMA (1,2,2) tersebut digunakan untuk melakukan peramalan produksi teh 5 tahun ke depan (2020-2024). Hasil peramalan produksi teh menggunakan ARIMA(1,2,2) ditampilkan pada Tabel 7.

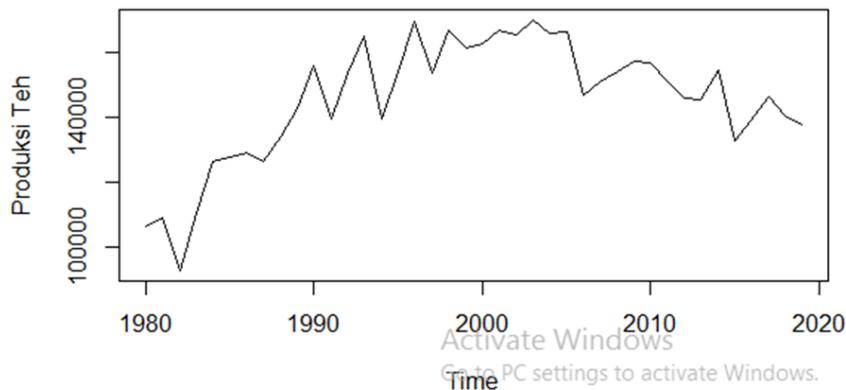
**Tabel 7.** Hasil Peramalan Produksi Teh dengan ARIMA (1,2,2)

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	136072.6	121837.2	150307.9	114301.50
2021	138832.8	116821.6	160844.1	105169.60
2022	137879.3	111020.5	164738.1	96802.32
2023	139997.1	108071.4	171922.8	91170.93
2024	139574.9	103682.5	175467.3	84682.17

## MODEL REGRESI

Analisis regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih variabel. Pada analisis regresi, variabel dibedakan menjadi dua bagian, yaitu variabel respon (*response variable*) atau biasa juga disebut variabel bergantung (*dependent variable*) dan variabel *explanatory* atau biasa disebut penduga (*predictor variable*) atau disebut juga variabel bebas (*independent variabel*). Hasil akhir dari studi analisis regresi adalah berupa suatu model yang dapat digunakan untuk meramalkan atau memprediksi nilai masa depan dari variabel respon berdasarkan nilai-nilai tertentu yang diberikan oleh variabel prediktor.

Model yang digunakan adalah regresi linear dan regresi kuadratis. Variabel data yang digunakan untuk permodelan adalah data waktu dan produksi. Pergerakan produksi teh terhadap waktu disajikan pada Gambar 6.



Gambar 6. Perkembangan Produksi Teh Terhadap Waktu

### Model Regresi Linear

Pendugaan model regresi linear berfungsi untuk mengetahui hubungan antara tahun dengan produksi. Tabel 8. menjelaskan bahwa hasil pendugaan model regresi linear mempunyai signifikan sebesar 1%, dengan MAPE untuk data training 8,60% dan MAPE untuk data testing 23,16%. Sementara uji *Jarque-Bera test for normality* dari model tersebut menunjukkan P-value >5%, H0 diterima, data terdistribusi normal.

#### Tabel 8. Pendugaan Model Regresi Linear

```
Call:
lm(formula = Produksi_TOTAL ~ Tahun, data = train)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-32981  -9687  -1198   11458  24081
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```

(Intercept) -2652388      513158  -5.169  1.22e-05 ***
Tahun        1402         257     5.453  5.31e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14700 on 32 degrees of freedom
Multiple R-squared:  0.4817, Adjusted R-squared:  0.4655
F-statistic: 29.74 on 1 and 32 DF,  p-value: 5.31e-06

```

**MAPE Training**

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.2413	4.8334	7.0245	8.6039	11.9361	35.5658

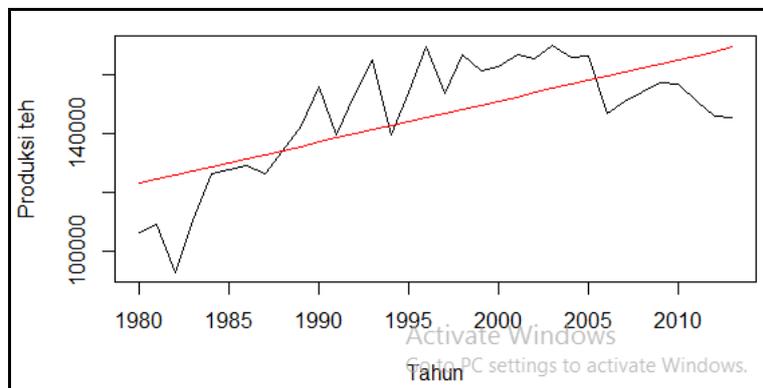
**MAPE Testing**

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
10.49	20.82	25.21	23.16	28.05	29.67

**Uji Kenormalan**

Jarque-Bera test for normality

data: reg1\$residuals  
 JB = 1.1139, p-value = 0.435



Gambar 7. Plot Regresi Linear Produksi Teh

**Model Regresi Kuadratis**

Pendugaan model regresi linear kuadratis dilakukan dengan memperhatikan signifikansi dari model dan parameter-parameternya. Jika dibandingkan hasil penelusuran model antara model regresi linier dengan kuadratis maka dapat disimpulkan bahwa hasil regresi kuadratis lebih memenuhi kaidah-kaidah statistik yang baik. Terlihat bahwa nilai probabilitas (<0,0001) model yang jauh lebih kecil dari 0,05. Dengan  $R^2$  sebesar 87,26% .

**Tabel 9.** Pendugaan Model Regresi Kuadratis

Call:  
 lm(formula = Produksi\_TOTAL ~ t + t2, data = train)

Residuals:

Min	1Q	Median	3Q	Max
-16556	-3702	1370	4520	11760

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	97551.15	3596.99	27.120	< 2e-16 ***
t	6156.36	504.44	12.204	2.25e-13 ***
t2	-144.08	14.77	-9.752	5.81e-11 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

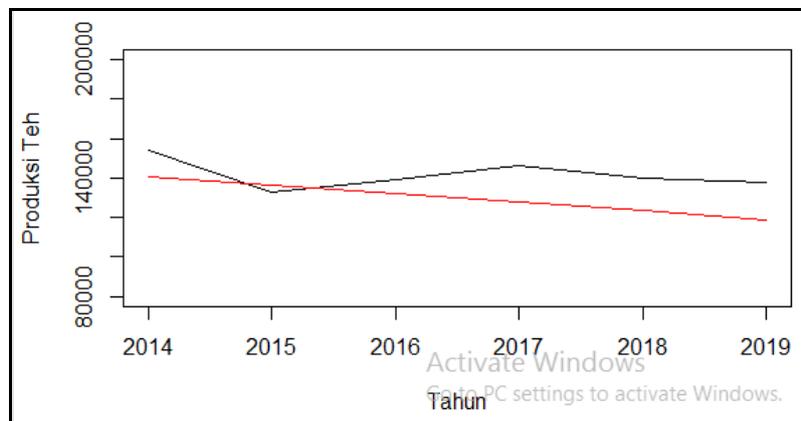
Residual standard error: 7407 on 31 degrees of freedom  
 Multiple R-squared: 0.8726, Adjusted R-squared: 0.8644  
 F-statistic: 106.2 on 2 and 31 DF, p-value: 1.351e-14

### MAPE Testing

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.947	5.778	10.543	9.188	12.308	14.007

### MAPE Training

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.09395	1.43713	2.82229	4.07128	5.80154	17.85311



Gambar 8. Plot Regresi Kuadratis Produksi Teh

Pada Tabel 10. dijelaskan bahwa hasil estimasi regresi linier pada tahun 2020 hingga tahun 2024 cenderung meningkat, sementara hasil estimasi regresi kuadratis cenderung menurun.

Tabel 10. Hasil Estimasi Regresi Linier dengan Regresi Kuadratis

Tahun	Regresi Linier	Regresi Kuadratis
2020	178.976	107.760
2021	180.377	101.958
2022	181.779	95.867
2023	183.181	89.488
2024	184.582	82.822

## MODEL FUNGSI TRANSFER

### Mempersiapkan Deret Input

Deret input yang digunakan adalah data luas tanaman menghasilkan teh. Terlebih dahulu data input tersebut harus dilakukan pemeriksaan kestasioneran. Hasil uji *Augmented Dickey Fuller* (Tabel 11) mengindikasikan bahwa data produksi teh belum stationer. Hal ini dilihat dari nilai *test-statistic* yang lebih besar dari *critical values* (nilai tau3), baik untuk taraf 1%, 5% maupun 10% menunjukkan bahwa  $H_0$  gagal ditolak. Maka akan dilakukan *differencing* satu kali dan kemudian dilakukan uji ADF kembali. Hasil uji ADF yang telah dilakukan *differencing* satu kali menunjukkan bahwa nilai *test-statistic* yaitu -6.0014 lebih kecil dari *critical values* menunjukkan bahwa  $H_0$  ditolak yang berarti data luas tanaman menghasilkan teh telah stasioner (Tabel 12).

**Tabel 11.** Hasil uji *Augmented Dickey Fuller* pada Luas Tanaman Teh

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
  Min    1Q  Median    3Q   Max
-21220.2 -2960.1 -755.9  2325.6 14244.4
Coefficients:
      Estimate Std. Error t value    Pr(>|t|)
(Intercept) 21501.0914 14192.5905  1.515  0.141
z.lag.1      -0.1964   0.1552 -1.265  0.216
tt           -33.6598  175.0514 -0.192  0.849
z.diff.lag   -0.2548   0.1926 -1.323  0.197
Residual standard error: 6579 on 28 degrees of freedom
Multiple R-squared:  0.2115, Adjusted R-squared:  0.127
F-statistic: 2.503 on 3 and 28 DF, p-value: 0.07967
Value of test-statistic is: -1.2651 1.3145 1. 8482
Critical values for test statistics:
      1pct 5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
```

**Tabel 12.** Hasil uji *Augmented Dickey Fuller* pada Luas Tanaman Teh  
(*Differencing* 1)

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none
```

Call:

```
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
```

Residuals: Ytopi

2020 178976.1

2021 180377.8

2022 181779.5

2023 183181.1

2024 184582.8

Min	1Q	Median	3Q	Max
-18198.6	-2196.7	931.9	3834.8	18325.1

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
z.lag.1	-1.7284	0.2880	-6.001	1.59e-06 ***
z.diff.lag	0.3070	0.1772	1.732	0.0938 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6573 on 29 degrees of freedom

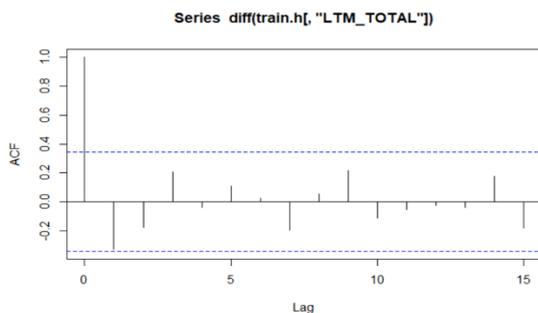
Multiple R-squared: 0.693, Adjusted R-squared: 0.6718

F-statistic: 32.73 on 2 and 29 DF, p-value: 3.658e-08

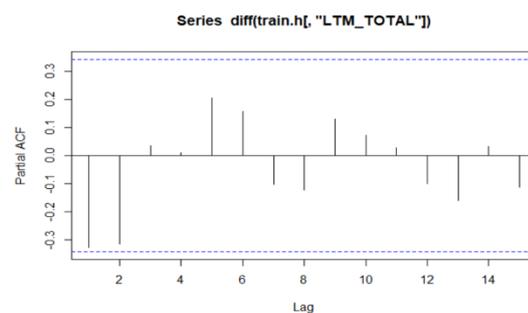
Value of test-statistic is: -6.0014

Critical values for test statistics:

	1pct	5pct	10pct
tau1	-2.62	-1.95	-1.61



Gambar 9. Plot ACF LTM Teh (*Differencing*)



Gambar 10. Plot PACF LTM Teh (*Differencing*)

## Membuat Model ARIMA untuk Deret Input

Pembuatan model ARIMA dilakukan dengan memperhatikan nilai awal dan pola dari ACF dan PACF yang signifikan. Model ARIMA deret input ini akan digunakan untuk tahap *prewhitening*. Pada pendugaan model input luas tanaman menghasilkan diperoleh ARIMA (0,1,1) dengan nilai MAPE 4.300691 (Tabel 13). Namun hasil dari estimasi luas tanaman menghasilkan untuk ARIMA (0,1,1) cenderung sama, maka dilakukan pemilihan model ARIMA melalui *armaselect*. Tabel 14 menunjukkan bahwa ARMA select telah dilakukan differencing 1 kali, sementara Tabel 15 menunjukkan bahwa ARMA select telah dilakukan differencing 2 kali.

Tabel 13. Auto Model ARIMA (0,1,1) Luas Tanaman Menghasilkan Teh

Series: train.h[, "LTM\_TOTAL"]  
ARIMA(0,1,1)

Coefficients:

ma1  
-0.4039  
s.e. 0.1417

sigma^2 estimated as 41069250: log likelihood=-335.66  
AIC=675.33 AICc=675.73 BIC=678.32

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	675.788	6217.187	4425.167	0.4604833	4.300691	0.9655485	-0.02837655

Tabel 14. ARMA Select Differencing 1

	p	q	sbc
[1,]	5	1	552.7390
[2,]	5	5	554.6447
[3,]	0	1	556.0142
[4,]	5	2	558.3862
[5,]	2	1	558.4723
[6,]	1	1	559.4557
[7,]	0	3	560.4417
[8,]	0	2	560.6025
[9,]	5	3	560.7322
[10,]	3	1	562.0332

Tabel 15. ARMA Select Differencing 2

	p	q	sbc
[1,]	3	5	556.9931
[2,]	4	5	559.3556
[3,]	5	5	563.7468
[4,]	1	5	578.7796
[5,]	2	5	581.8045
[6,]	0	5	582.1597
[7,]	0	3	583.9102
[8,]	1	3	583.9442
[9,]	1	1	586.5060
[10,]	1	2	587.8491

Tabel 16 menunjukkan bahwa model ARIMA (0,1,2) memiliki nilai AIC yang terkecil diantara model yang lainnya. Hasil uji *Ljung-Box* menunjukkan sisaan dari ketiga model tersebut saling bebas, didukung pula dengan plot ACF sisaan dan plot PACF sisaan yang tidak nyata pada semua *lag*. Sehingga, ketiga model ARIMA tersebut masih belum dapat digunakan untuk peramalan.

Tabel 16. Nilai AIC dan SBC Hasil Identifikasi Model ARIMA Data Luas Tanaman Menghasilkan

Model	AIC	SBC
ARIMA (2,1,1)*	677.82	558.4723
ARIMA (1,1,1)*	677.33	559.4557
ARIMA (0,1,2)*	677.32	560.6025

\* Terdapat parameter yang tidak nyata

Langkah selanjutnya melakukan *overfitting* untuk memperoleh model terbaik. Hasil *overfitting* pada Tabel 17 menunjukkan bahwa Model ARIMA (1,1,3) dan ARIMA (1,1,1)

memiliki parameter yang tidak nyata. Sementara model ARIMA (1,1,2) seluruh parameternya nyata namun mempunyai nilai AIC yang tidak lebih kecil dari ARIMA (1,1,3). Oleh karena itu, model ARIMA (1,1,2) ditetapkan sebagai model yang terbaik.

Tabel 17. Nilai AIC dan SBC Hasil *Overfitting* Model ARIMA Data Luas Tanaman Menghasilkan

Model	AIC	SBC
ARIMA (1,1,3)*	705.12	583.9442
ARIMA (1,1,1)*	707.13	586.5060
ARIMA (1,1,2)	705.56	587.8491

\* Terdapat parameter yang tidak nyata

### ***Prewhitening* Deret Input dan Output dan Korelasi Silang**

Tahap *prewhitening* dilakukan berdasarkan model ARIMA deret input. Dalam tahap ini digunakan unsur *white noise* model tersebut. Proses ini bertujuan untuk menghitung korelasi silang, sehingga dapat digunakan untuk menentukan hubungan antara deret input dan deret output. Jika hasil korelasi silang menunjukkan ada hubungan yang nyata, maka layak untuk dibuat model fungsi transfer.

### **Penentuan nilai $b$ , $r$ dan $s$**

Jika hasil korelasi silang nyata, maka dilanjutkan dengan mencari nilai  $b$ ,  $r$  dan  $s$  dengan melihat plot korelasi silang deret input dan output. Plot korelasi silang antara produksi teh dan luas tanaman menghasilkan pada (Lampiran 1 dan 2) menunjukkan nilai yang signifikan pada *lag* ke-  $b$ . Nilai  $s$  dilihat pada banyaknya *lag* korelasi silang yang signifikan setelah *lag* ke- $b$  yang diartikan dengan lamanya nilai luas tanaman menghasilkan mempengaruhi produksi teh setelah nyata yang pertama, dari diperoleh nilai  $s=0$ . Sedangkan nilai  $r$  dilihat dari plot ACF deret output yang signifikan setelah signifikan yang pertama dan diperoleh  $r=1$ .

### **Membuat Model ARIMA Deret Sisaan**

Pendugaan model ARIMA deret sisaan dilakukan dengan mengamati plot ACF dan PACF deret sisaan yang memperlihatkan bahwa masih terdapat *lag* yang nyata. Model yang baik adalah model yang tidak mempunyai *lag* yang nyata baik pada plot ACF dan PACF, di pengujian korelasi serta di korelasi silang deret sisaan. Bila pada plot ACF dan PACF masih terlihat pola sistematis, maka perlu dilakukan *overfitting*.

### **Peramalan Berbasis Fungsi Transfer**

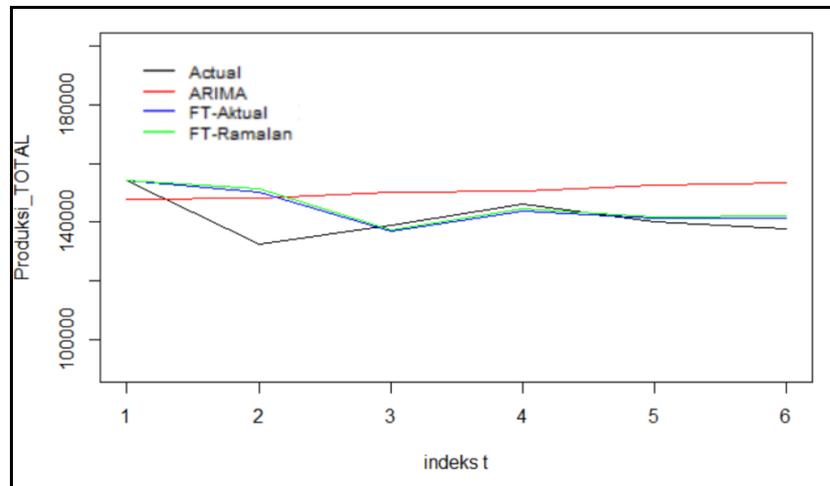
Model fungsi transfer didapat dari kombinasi model deret input awal dengan model ARIMA deret sisaan. Model deret input adalah data luas tanaman menghasilkan, sementara model ARIMA sisaan menggunakan data produksi.

Hasil pengujian dari model ARIMA (1,1,2) menghasilkan MAPE untuk data testing adalah 3,48 % dan MAPE untuk data training adalah 3,29% (Tabel 18). Plot Estimasi berbasis Fungsi Transfer, Data Input, Nilai Aktual dan Ramalan disajikan pada Gambar 11.

Tabel 18. Evaluasi *Magnitude* dan *Direction* Model ARIMA (1,1,2) Data Luas Tanaman Menghasilkan

Data	N	<i>Magnitude</i> (MAPE)
Testing	6	3.48
Training	34	3.29

<sup>a</sup> Probabilitas < 0,05



Gambar 11. Plot Estimasi berbasis Fungsi Transfer, Data Input, Nilai Aktual dan Ramalan

Untuk hasil estimasi pada fungsi transfer dengan ARIMA (1,1,2) disajikan pada Tabel 19 berikut.

Tabel 19. Hasil Estimasi Produksi Teh dengan Model Fungsi Transfer

Tahun	Produksi (Ton)	Pertumbuhan (%)
2020	135.853	
2021	138.815	2,18
2022	137.788	-0,74
2023	140.030	1,63
2024	139.606	-0,30

### MODEL VECTOR AUTO REGRESSION (VAR)

Model VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen. Model VAR berlaku pada saat nilai setiap variabel dalam sebuah system tidak hanya bergantung pada lag-nya sendiri, namun juga pada nilai lag variabel lain.

## Uji Lag Optimal

Penentuan panjang *lag* dalam model VAR menunjukkan derajat bebas. Pengujian ini dilakukan untuk menentukan jumlah *lag* optimal yang dapat digunakan dalam variabel yang akan dianalisis. Variabel yang digunakan pada mode VAR terdiri dari produksi, luas areal, harga konsumen, volume ekspor dan volume impor.

Untuk data teh ini, setelah dilakukan running model VAR dengan lag  $p=2$  type “both”, diperoleh informasi komponen konstanta dan trend signifikan, sehingga pada pengujian berikutnya dilakukan running model VAR dengan menggunakan lag  $(p) = 2$  s.d lag  $(p)=4$  dengan menyertakan komponen konstanta dan trend. Pemilihan lag  $p$  ditentukan dengan melihat banyaknya variabel yang signifikan dalam lag tersebut sekaligus memastikan harus ada variabel/peubah yang signifikan pada lag terpilih dimaksud.

Uji lag optimal dilakukan sampai lag  $p$  dimana pada lag tersebut jumlah variabel yang signifikan sudah berkurang dibandingkan dengan keseluruhan variabel yang digunakan dalam model atau bahkan sudah tidak ada lagi variabel yang signifikan.

## Uji Asumsi

Model VAR (2) type “both” dilakukan uji asumsi untuk autokorelasi, normalitas, dan homoskedastisitas. Hasil uji dengan menggunakan *Portmanteau test* menunjukkan bahwa nilai p-value lebih besar dari 0,05 maka **sisaan saling bebas atau asumsi non autokorelasi terpenuhi.**

Portmanteau Test (asymptotic)

data: Residuals of VAR object varhsheet.b2

Chi-squared = 278.8, df = 350, p-value = 0.998

Untuk pengujian normalitas VAR (2) type “both” menggunakan fungsi “normality.test” yang di dalamnya dilakukan pengujian Jarque-Bera tests Hasil pengujian yang terbentuk diberi nama “norm1”. Nilai p-value lebih kecil dari 0,05, namun karena jumlah data yang digunakan cukup banyak, maka series tersebut dapat dianggap normal..

## Tabel 20. Uji Asumsi *Vector Auto Regression*

JB-Test (multivariate)

data: Residuals of VAR object varhsheet.b2

Chi-squared = 44.63, df = 10, p-value = 2.536e-06

\$Skewness

Skewness only (multivariate)

data: Residuals of VAR object varhsheet.b2

Chi-squared = 19.02, df = 5, p-value = 0.001906

\$Kurtosis

Kurtosis only (multivariate)

data: Residuals of VAR object varhsheet.b2

Chi-squared = 25.61, df = 5, p-value = 0.0001062

Fungsi “arch.test” berfungsi untuk melakukan pemeriksaan heteroskedastisitas model VAR (2) type “both”. Nilai p-value lebih besar dari 0,05 **maka ragam sisaan homogen atau asumsi homoskedastisitas terpenuhi.**

ARCH (multivariate)

data: Residuals of VAR object varhsheet.b2  
Chi-squared = 405, df = 1125, p-value = 1

### **Estimasi VAR (2) type “both”**

Dari hasil estimasi data produksi teh dengan model VAR (2) type “both” menghasilkan MAPE testing 10,41% dan MAPE training 3,94%. Hasil estimasi model VAR (2) type “both” untuk 5 tahun kedepan adalah tahun 2020=145.908; tahun 2021=145.614; tahun 2022=144.463; tahun 2023=142.829 dan tahun 2024=140.946

MAPE Testing

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
3.440	7.243	11.062	10.406	13.009	17.276	

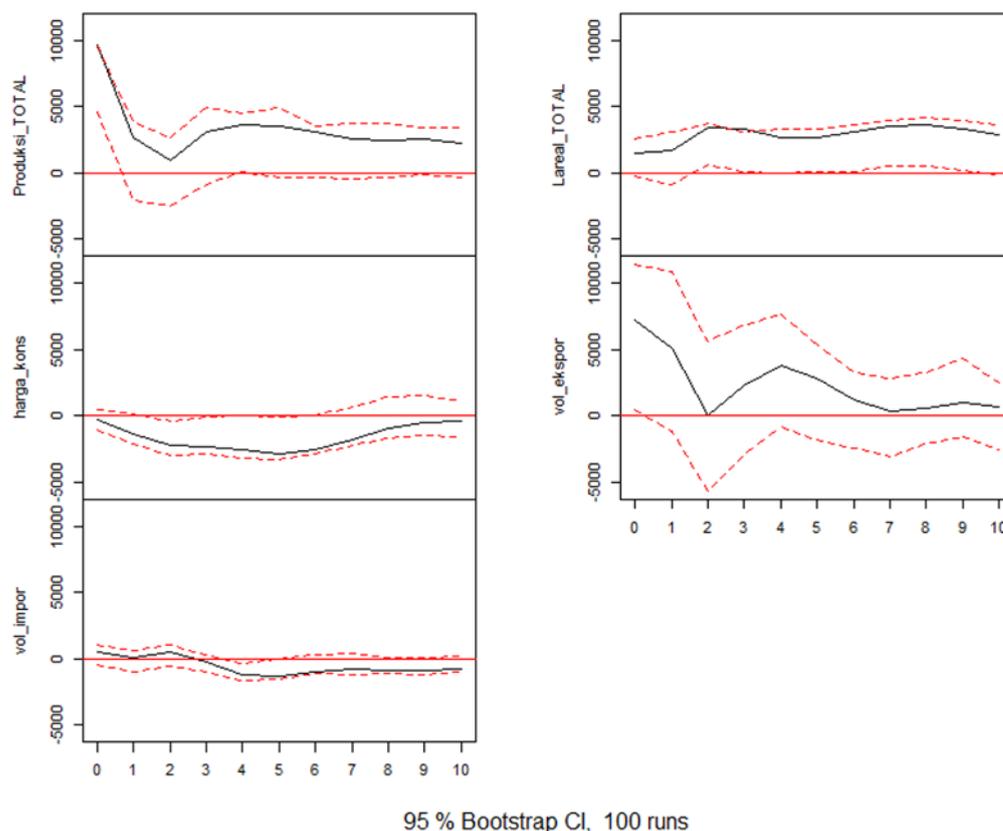
MAPE Training

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.04779	1.29469	2.26774	3.94332	5.35039	18.08124

### **Mengkaji Respon *Impulse* dan *Variance Decomposition***

*Impulse Response Function* (IRF) menjelaskan bagaimana perubahan atau gejala yang terjadi pada suatu variabel/peubah di tahun tertentu akan berdampak pada variabel/peubah lain di tahun tertentu dan tahun-tahun setelahnya.

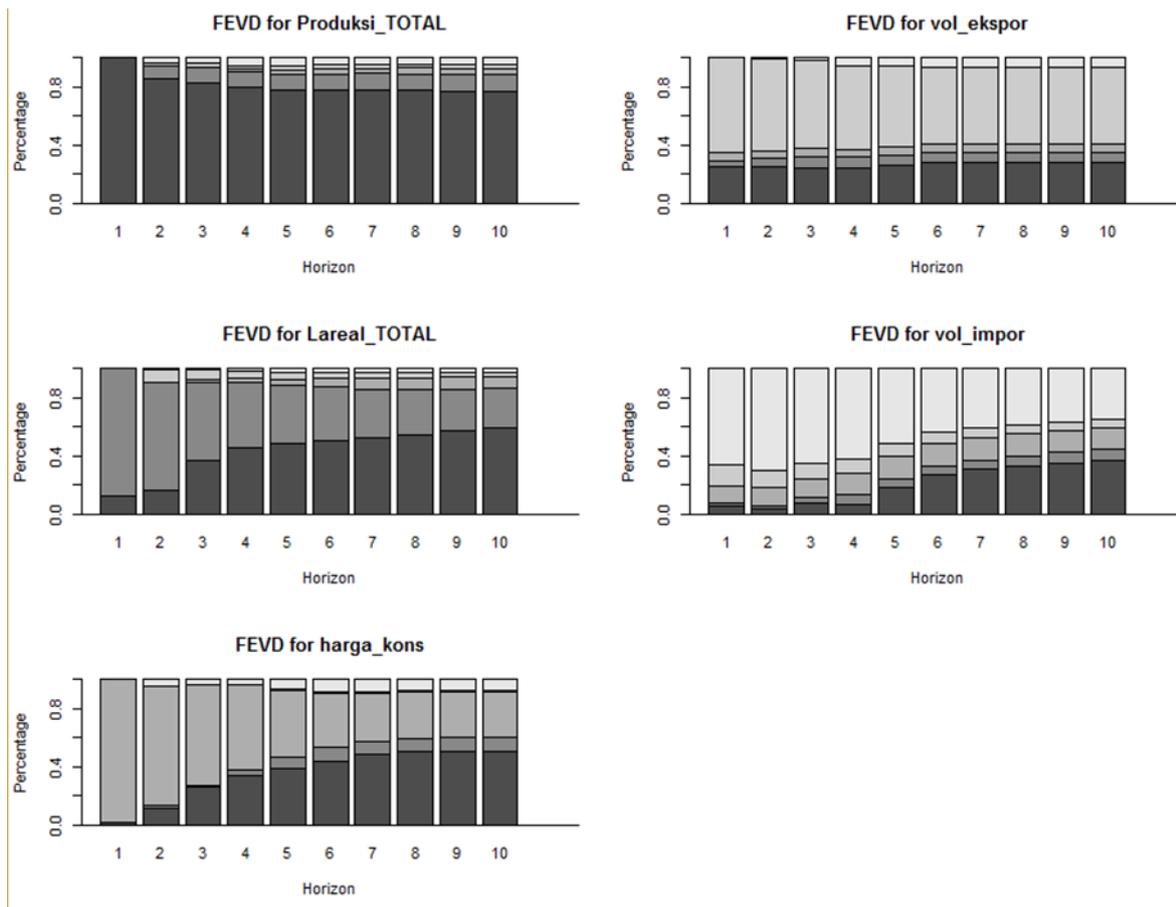
Orthogonal Impulse Response from Produksi\_TOTAL



Gambar 12. Orthogonal Impulse Response from Produksi Teh

Dari grafik *Impulse Response Function* produksi model terbaik VAR (2) type “both”, dapat dilihat bahwa jika terjadi perubahan pada produksi di tahun tertentu maka akan berdampak pada produksi itu sendiri sampai 3 tahun ke depan. Dampak tersebut akan hilang setelah 3 tahun. Sedangkan dampak perubahan produksi tidak berdampak pada luas areal, harga konsumen, volume ekspor dan volume impor. Perubahan produksi berdampak pada volume ekspor sampai dengan dua tahun dan akan hilang dampaknya setelah melewati tahun kedua tersebut.

*Variance Decomposition* atau dekomposisi keragaman digunakan untuk melihat variabel apa saja yang mempengaruhi komposisi keragaman suatu variabel/peubah. *Variance Decomposition* melakukan dekomposisi terhadap variabel *endogen* ke dalam *shocks* component bagi variabel endogen dalam VAR. VAR biasa digunakan untuk melakukan peramalan dari data yang saling berhubungan untuk menganalisa dampak dari gangguan random terhadap sistem dari variabel. *Variance Decomposition* melakukan pemecahan terhadap varians dari *forecast error* dari setiap variabel ke dalam komponen yang dapat mempengaruhi variabel *endogen*.



Gambar 13. Plot *Variance Decomposition* Model VAR

Dari grafik dekomposisi keragaman model terbaik VAR (2) type “both” di atas dapat dilihat bahwa komposisi produksi pada tahun pertama dipengaruhi sepenuhnya oleh produksi itu sendiri. Pada tahun kedua, komposisi produksi 90% dipengaruhi oleh produksi itu sendiri, 5% dipengaruhi oleh luas areal, dan 5% harga konsumen.

### Hasil Estimasi Dengan Model Terbaik

Penyusunan angka estimasi produksi teh dilakukan dengan penelusuran 4 model terbaik yaitu ARIMA, Regresi, Fungsi Transfer dan VAR. Model yang pertama adalah model time series atau ARIMA, model terbaik untuk ARIMA adalah pada orde ARIMA (1,2,2). Untuk model estimasi produksi teh nasional dengan ARIMA (1,2,2) menghasilkan MAPE untuk data training sebesar 5,29% dan MAPE untuk data testing sebesar 7,83%.

Model yang kedua adalah menggunakan regresi linier sebagai variabel bebas adalah waktu atau tahun. Untuk model regresi linier sederhana menghasilkan MAPE yang cukup besar yaitu 8,60% untuk data training dan 23,16% untuk data testing. Untuk model regresi kuadratik dengan variabel bebas waktu (tahun) menghasilkan MAPE yang lebih baik, yaitu sebesar 4,07% untuk data training dan 9,18% untuk data testing.

Model ketiga adalah menggunakan fungsi transfer, dengan data input nya adalah luas tanaman menghasilkan. Model ARIMA data input menggunakan ARIMA (1,1,2). Hasil dari

pengepasan fungsi transfer untuk MAPE data training 3,29% dan MAPE untuk data testing 3,48%.

Model keempat adalah *Vector Auto Regressive* (VAR) dengan menggunakan 5 variabel yaitu produksi, luas areal, harga konsumen, volume ekspor dan volume impor. Untuk Model VAR adalah nilai  $p=2$  dan  $type="both"$ ,  $p=2$  artinya menggunakan variabel bebas sampai lag-2 dan ada konstanta dan faktor trend. MAPE yang dihasilkan dari model VAR (2)  $type="both"$  adalah MAPE untuk data training 3,94% dan MAPE untuk data testing 10,41%.

**Tabel 21.** Hasil Estimasi Produksi Teh Dengan Model ARIMA, Regresi, Fungsi Transfer dan *Vector Auto Regression* (VAR)

No	Keterangan Model	MAPE		Hasil Estimasi Produksi Teh (Ton)					Rata-rata Pertumbuhan (%)
		Training	Testing	2020	2021	2022	2023	2024	Tahun 2020 -2025
1	ARIMA (1,2,2)	5,29	7,83	136.073	138.833	137.879	139.997	139.575	0,644
2	Regresi Linier	8,60	23,16	178.976	180.377	181.779	183.181	184.582	0,774
	Regresi Kuadratik	4,07	9,18	107.760	101.958	95.867	89.488	82.822	-6,365
3	Fungsi Transfer Arima (1,1,2) xreg=LTM	3,29	3,48	135.853	138.815	137.788	140.030	139.606	0,691
4	Var (2) $p=2$ $type=both$	3,94	10,41	145.908	145.614	144.463	142.829	140.946	-0,860

## KESIMPULAN

1. Model ARIMA (1,2,2) masih belum dapat digunakan karena MAPE ARIMA > MAPE Fungsi Transfer.
2. Model Regresi Linear dan Regresi Kuadratik tidak dapat digunakan karena MAPE nya masih terlalu besar dan hasil estimasinya tidak mendekati data aktualnya.
3. Model Fungsi Transfer ARIMA(1,1,2) dengan input variabel luas tanaman menghasilkan menjadi **model terbaik** karena memenuhi kaidah statistik dan hasil estimasinya mendekati angka aktual.
4. Model VAR(2)  $p=2$   $type=both$  bukan model terbaik untuk estimasi produksi teh karena MAPE nya masih cukup besar.

## DAFTAR PUSTAKA

- Budiar, Aindra. 2013. Peramalan Produksi TBS Kelapa Sawit Dengan Model Arima dan Model Fungsi Transfer Input Ganda. Institut Pertanian Bogor, Bogor
- Gujarati, D.N. and D.C. Porter, 2010. Dasar-dasar Ekonometrika. Jakarta: Penerbit Salemba Empat, Buku 2, Edisi 5.
- Heizer, J., Render, B. & Munson, C., 2011. Operations Management Sustainability and Supply Chain Management. Boston: Pearson.
- Kementerian Pertanian. 2013. Pedoman Teknis Pengembangan Tanaman Teh Tahun 2014. Jakarta: Direktorat Jenderal Perkebunan Kementerian Pertanian.
- Kementerian Pertanian. 2019. Statistik Perkebunan Indonesia 2018-2020 (Teh). Jakarta: Direktorat Jenderal Perkebunan.
- Mardiyanto, Atqo. 2000. Kajian Peramalan Dengan Model Struktural dan Non Struktural (VAR dan ARIMA). Institut Pertanian Bogor, Bogor
- Makridakis S, Wheelwright SC, McGee VE. 1999. Metode dan Aplikasi Peramalan. Suminto H, penerjemah. Binarupa Aksara, Jakarta.
- M. Firdaus 2020. Aplikasi Ekonometrika dengan E-Views, Stata dan R. Institut Pertanian Bogor (IPB) Press.
- Tsay, R.S., 2014. Multivariate Time Series Analysis With R and Financial Applications. New Jersey: John Wiley & Sons.



**ESTIMASI PRODUKSI LADA DI INDONESIA**  
*Estimation of Pepper Production in Indonesia*

**Roydatul Zikria<sup>1\*</sup>**

<sup>1</sup>*Pusat Data dan Sistem Informasi Pertanian, Kementerian Pertanian  
Jl. Harsono RM No. 3 Gedung D Lantai 4, Jakarta Selatan, Indonesia*

\*Korespondensi penulis. E-mail: roydatul\_z@pertanian.go.id

**ABSTRACT**

Indonesia pepper production has increased for the last five years. Estimation of pepper production in the next few years uses ARIMA model, regression, transfer function and VAR. Regarding estimation of those models, transfer function is selected as the best model to estimate pepper production. In detail, the model is transfer function ARIMA (3,1,3) using export quantity as input variable. Transfer model produces MAPE for training data and testing data which are respectively 7.16 and 2.54. Estimation of pepper production for the next five years (2020-2024) fluctuate. The estimation in 2020 is 84,611 ton while in 2024 is estimated 88,744 ton of production.

**Keywords:** *estimation, ARIMA, regression, transfer function, VAR*

**ABSTRAK**

Produksi lada di Indonesia cenderung meningkat selama lima tahun terakhir. Untuk mengestimasi produksi lada beberapa tahun ke depan, penelitian ini menggunakan model ARIMA, regresi, fungsi transfer dan VAR. Hasil estimasi dari keempat model tersebut dibandingkan dan terpilih fungsi transfer sebagai model estimasi terbaik. Model fungsi transfer yang terpilih adalah fungsi transfer ARIMA (3,1,3) dengan variabel input volume ekspor. Hasil estimasi dengan fungsi transfer menghasilkan MAPE data training sebesar 7,16 dan MAPE data testing 2,54. Produksi lada lima tahun ke depan (2020-2024) diramalkan fluktuatif. Tahun 2020 produksi lada diramalkan sebesar 84.611 ton kemudian meningkat di tahun 2024 menjadi 88.744 ton.

**Kata kunci:** *estimasi, ARIMA, regresi, fungsi transfer, VAR*

## PENDAHULUAN

Lada merupakan salah satu komoditas unggulan sub sektor perkebunan dengan nilai ekonomis tinggi. Menurut Kemala (2006) usahatani lada mampu menyumbang negara, menyediakan lapangan pekerjaan serta menjadi bahan baku industri maupun konsumsi langsung. Pengembangan usahatani lada di Indonesia sebagian besar berada di Luar Jawa seperti Kepulauan Bangka Belitung, Lampung, Sumatera Selatan, Sulawesi Selatan dan Kalimantan Timur. Berdasarkan data Kementerian Pertanian (2019) kelima provinsi tersebut berkontribusi sebesar 70% terhadap produksi lada di Indonesia. Jenis lada yang paling banyak diproduksi di Indonesia adalah lada hitam yang berasal dari Lampung atau dikenal dengan nama *Lampung black pepper* dan lada putih atau *Muntok white pepper* yang dihasilkan di Kepulauan Bangka Belitung (Nurdjannah, 2006).

Pengusahaan lada di Indonesia didominasi oleh Perkebunan Rakyat (PR) dan sebagian kecil diusahakan oleh Perkebunan Besar Swasta (PBS). Menurut Zikria (2019) perkebunan rakyat mendominasi 98,47% produksi lada di Indonesia selama periode 2010-2019, sedangkan kontribusi lada yang berasal dari PBS hanya 1,60%. Pengusahaan lada oleh perkebunan rakyat umumnya masih dilakukan secara tradisional (Damanik, 2001). Menurut Kardinan et al. (2018) budidaya lada secara tradisional dilakukan petani dimana penggunaan bahan kimia sebagai asupan tambahan menjadi andalan dan merupakan garansi keberhasilan petani. Oleh karena itu, ketika pupuk dan pestisida sulit diperoleh maka asupan tanaman lada berkurang yang mengakibatkan produksinya turun.

Produksi lada di Indonesia selama sepuluh tahun terakhir menunjukkan adanya trend peningkatan. Tahun 2010 produksi lada sebesar 83.664 ton dan meningkat menjadi 88.949 ton pada tahun 2019 atau terjadi peningkatan 0,86% per tahun (Kementerian Pertanian, 2019). Peningkatan tersebut mengindikasikan bahwa produksi beberapa tahun ke depan masih dapat diproyeksikan naik. Estimasi produksi lada beberapa tahun ke depan sangat diperlukan sebagai bahan penentu kebijakan maupun *Early Warning System (EWS)* mengingat data Angka Tetap (ATAP) Perkebunan memiliki lag yang cukup jauh dibandingkan data tahun berjalan.

Estimasi produksi pada tahun berjalan maupun beberapa periode ke depan sangat urgen untuk dilakukan. Hal ini karena informasi tersebut menjadi bahan untuk penentuan kebijakan di subsektor perkebunan. Direktorat Jenderal Perkebunan telah mengupayakan penyediaan data yang lebih *real time* dengan melakukan penyusunan Angka Estimasi (AESTI) pada tahun berjalan. Namun estimasi produksi komoditas perkebunan untuk lima tahun ke depan masih belum tersedia. Estimasi Ditjen Perkebunan hanya dilakukan untuk satu tahun ke depan menggunakan model *univariate* seperti *Double Exponential Smoothing (DES)*. Salah satu kelemahan dari model *univariate* yaitu variabel yang digunakan hanya satu misalnya produksi. Akibatnya hasil analisis hanya mampu memberikan gambaran terhadap satu variabel saja tanpa adanya intervensi dari variabel lain. Padahal produksi komoditas perkebunan tidak terlepas dari pengaruh variabel-variabel lain seperti harga, luas areal, ekspor-impor serta variabel lainnya. Selain itu, untuk menentukan kebijakan subsektor perkebunan seperti peningkatan produksi, diperlukan informasi variabel input lain yang diduga turut berpengaruh terhadap produksi komoditas perkebunan sehingga intervensi kebijakan dapat dilakukan pada variabel input

tersebut. Oleh karena itu diperlukan model yang mampu menyajikan analisis mendalam dalam mengestimasi produksi dengan melibatkan variabel input lain, misalnya model *multivariate*.

Penelitian ini akan menganalisis hasil estimasi produksi lada di Indonesia dengan model *univariate* maupun *multivariate*. Terdapat empat model yang digunakan dalam mengestimasi produksi lada antara lain *Autoregressive Integrated Moving Average (ARIMA)*, regresi, fungsi transfer dan *Vector Autoregression (VAR)*. Model ARIMA menghasilkan estimasi produksi lada tanpa ada pengaruh dari variabel lain. Model regresi mengestimasi produksi yang dipengaruhi oleh variabel waktu. Model fungsi transfer menghasilkan angka estimasi produksi dengan memasukkan intervensi dari satu variabel pendukung yang dianggap paling berpengaruh terhadap produksi. Model VAR mengestimasi produksi dengan mempertimbangkan pengaruh dari beberapa variabel lain atau terdapat lebih dari satu variabel pendukung yang diduga berpengaruh terhadap produksinya. Hasil estimasi dari keempat model tersebut akan dibandingkan untuk selanjutnya ditentukan model terbaik untuk meramalkan produksi lada di Indonesia beberapa tahun ke depan.

## METODE PENELITIAN

### Kerangka Pemikiran

Kerangka pemikiran dalam penelitian menggambarkan hasil estimasi produksi lada dengan empat model yaitu ARIMA, regresi, fungsi transfer dan VAR. Model ARIMA umumnya digunakan untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang kurang baik ketepatannya. Estimasi dengan model ARIMA hanya menggunakan variabel dependen dan mengabaikan variabel independennya. Dengan kata lain, untuk mengestimasi produksi lada beberapa tahun ke depan maka variabel yang digunakan hanya produksi itu sendiri.

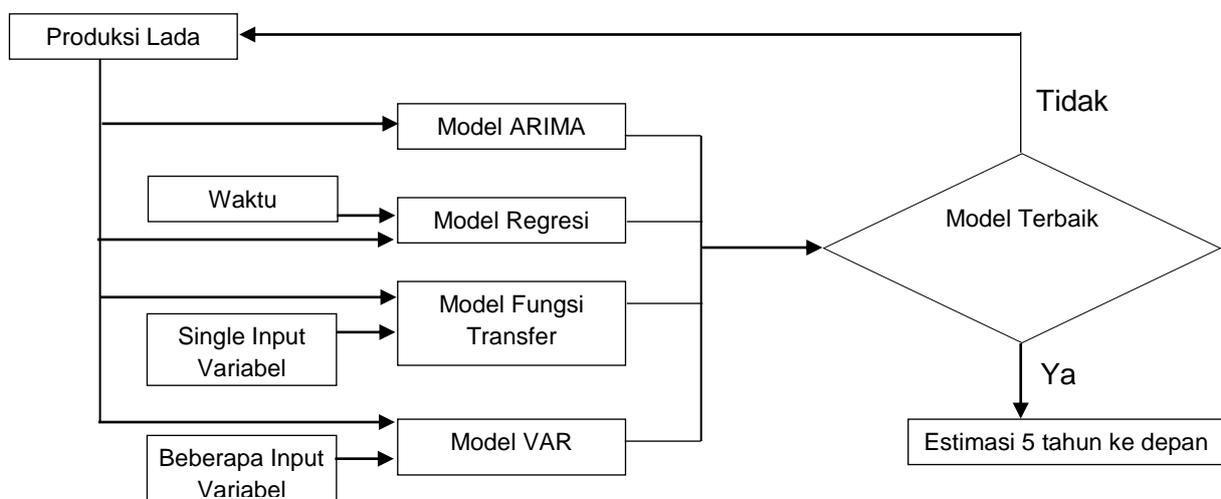
Estimasi dengan model regresi pada penelitian dibagi menjadi dua yaitu regresi linear sederhana dan regresi kuadratik. Kedua model regresi tersebut digunakan untuk menyelidiki pengaruh antara dua atau lebih variabel. Secara umum, estimasi dengan model regresi melibatkan variabel dependen dan variabel independen. Variabel dependen yang dimaksud pada analisis ini adalah produksi lada sedangkan variabel independen yang dipilih yaitu waktu ( $t$ ). Selanjutnya dilakukan peramalan produksi lada berdasarkan nilai-nilai tertentu yang diberikan oleh variabel independennya. Terdapat beberapa asumsi yang mendasari penggunaan model regresi linear sederhana maupun kuadratik antara lain sisaan mengikuti fungsi distribusi normal, varians sisaan konstan untuk setiap data amatan (homoskedastisitas) dan tidak terdapat autokorelasi antara sisaan untuk setiap data amatan.

Model fungsi transfer menggambarkan nilai ramalan masa depan dari suatu deret berkala (deret output) yang didasarkan pada nilai-nilai masa lalu dari deret itu sendiri serta didasarkan pula pada suatu deret berkala yang berhubungan (deret input). Model fungsi transfer merupakan fungsi dinamis yang pengaruhnya tidak hanya pada hubungan linear antara waktu ke- $t$  dengan deret/variabel input, tetapi juga terdapat hubungan antara variabel input dengan variabel output pada waktu ke- $t$ ,  $t+1$ , ...,  $t+k$ . Pada fungsi transfer terdapat rangkaian output

yang mungkin dipengaruhi oleh rangkaian *multiple* input. Untuk kasus *single input* variabel pada fungsi transfer, dapat menggunakan metode korelasi silang. Penelitian ini menggunakan *single input* variabel yaitu volume ekspor untuk meramalkan produksi lada sebagai variabel outputnya.

Model VAR menggunakan pendekatan non-struktural atau tidak mendasarkan pada teori ekonomi tertentu dalam melakukan peramalan. Model ini memperlakukan seluruh peubah secara simetris tanpa mempermasalahkan variabel dependen dan independen atau dengan kata lain model VAR memperlakukan seluruh variabel sebagai variabel dependen/endogen, karena pada kenyataannya suatu variabel dapat bertindak sebagai endogen atau dependen dari suatu peubah lain (Gujarati & Porter, 2010). Untuk meramalkan produksi lada beberapa tahun ke depan, penelitian ini menggunakan beberapa variabel antara lain produksi lada, luas areal, volume ekspor, harga ekspor, volume impor dan harga impor.

Pembentukan model estimasi produksi lada dilakukan dengan membagi series data aktual menjadi data training dan data testing. Data training digunakan untuk menentukan model estimasi dan meramalkan data testing yang sebenarnya sudah tersedia data aktualnya. Hasil ramalan data testing tersebut kemudian dibandingkan dengan data aktual untuk dihitung tingkat kesalahan (*error*) hasil ramalan. Model terbaik untuk estimasi adalah model dengan tingkat *error* yang paling kecil, dalam hal ini ditunjukkan oleh nilai *Mean Absolute Percentage Error* (MAPE) terkecil. Selain MAPE, pemilihan model terbaik juga mempertimbangkan kelogisan hasil ramalan dengan historis data sebelumnya. Berdasarkan hasil identifikasi model ARIMA, regresi, fungsi transfer dan VAR, dipilih model terbaik untuk meramalkan produksi lada di Indonesia selama lima tahun ke depan. Secara umum tahapan penelitian ini disajikan melalui kerangka konseptual penelitian pada Gambar 1.



Gambar 1. Kerangka Konseptual Penelitian

## **Pengumpulan Data**

Data yang digunakan pada penelitian ini bersumber dari Direktorat Jenderal Perkebunan, Kementerian Pertanian. Variabel yang digunakan antara lain produksi lada, luas areal lada, volume ekspor lada, harga ekspor lada, volume impor lada dan harga impor lada. Produksi lada merupakan total produksi lada yang diusahakan oleh Perkebunan Rakyat (PR), Perkebunan Besar Negara (PBN) dan Perkebunan Besar Swasta (PBS). Luas areal lada merupakan penjumlahan dari luas Tanaman Belum Menghasilkan (TBM), Tanaman Menghasilkan (TM) dan Tanaman Rusak/Tanaman Tidak Menghasilkan (TR/TTM). Volume ekspor maupun volume impor dihitung berdasarkan enam kode HS yaitu 09041110, 09041120, 09041190, 09041210, 09041220, 09041290. Harga ekspor didekati dari nilai ekspor dibagi volume ekspor, demikian halnya dengan harga impor yang dihitung dari nilai impor dibagi volume impor. Pemilihan variabel harga ekspor dan harga impor dilakukan dengan pertimbangan keterbatasan ketersediaan series data harga lada di tingkat domestik (harga produsen maupun harga konsumen) dan internasional (harga dunia). Series data yang digunakan pada penelitian ini yaitu data tahun 1976-2019, dimana data tahun 2019 merupakan Angka Sementara (ASEM) sedangkan sisanya merupakan Angka Tetap (ATAP). Berdasarkan series data tersebut, selanjutnya dilakukan pengelompokan data training untuk periode 1976-2013 dan data testing untuk periode 2014-2019, sehingga diperoleh total observasi sebanyak 44. Berdasarkan model terbaik yang terpilih, estimasi produksi lada di Indonesia dilakukan untuk lima tahun ke depan yaitu 2020-2024.

## **Analisis Data**

Secara empiris, penelitian ini membandingkan hasil estimasi produksi lada dengan empat model yaitu ARIMA, regresi, fungsi transfer dan VAR. Tahapan penelitian dimulai dengan mencari model estimasi berdasarkan historis data training untuk meramalkan data testing. Selanjutnya hasil estimasi data testing dibandingkan dengan nilai aktual produksinya untuk mengetahui tingkat kesalahan berdasarkan nilai MAPE. Berdasarkan nilai MAPE yang dihasilkan oleh keempat model estimasi tersebut dipilih model ramalan dengan MAPE terkecil. Model dengan MAPE terkecil dipilih sebagai model terbaik untuk mengestimasi produksi lada lima tahun ke depan. Model terbaik yang terpilih juga harus memenuhi asumsi statistik yang ditetapkan di masing-masing model. Pengolahan data untuk estimasi produksi lada baik dengan model ARIMA, regresi, fungsi transfer maupun VAR dilakukan dengan *software* RStudio.

## **Estimasi dengan Model ARIMA**

Model ARIMA dibagi ke dalam tiga kelompok model yaitu *Autoregressive Model (AR)*, *Moving Average Model (MA)* dan *Autoregressive Integrated Moving Average Model (ARIMA)*. Model AR menjelaskan pergerakan suatu peubah itu sendiri di masa lalu. Model AR ordo ke- $p$  untuk mengestimasi produksi lada atau dapat ditulis ARIMA ( $p, 0, 0$ ) sebagai berikut:

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t \dots \dots \dots (1)$$

dimana:

$Y_t$  = produksi lada pada waktu ke- $t$

$Y_{t-p}$  = produksi lada pada kurun waktu ke ( $t-p$ )

$\mu$  = suatu konstanta

$\theta_1 \dots \theta_p$  = parameter autoregressive ke- $p$

$\varepsilon_t$  = nilai kesalahan pada waktu ke- $t$

Model MA menjelaskan pergerakan peubahnya melalui sisaannya di masa lalu. Model MA dengan ordo  $q$  untuk mengestimasi produksi lada atau ARIMA (0,0, $q$ ) ditulis sebagai berikut:

$$Y_t = \mu - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t \dots \dots \dots (2)$$

dimana:

$Y_t$  = produksi lada pada waktu ke- $t$

$\phi_1 \dots \phi_q$  = parameter-parameter moving average

$\varepsilon_{t-q}$  = nilai kesalahan pada waktu ke ( $t-q$ )

Model ARIMA merupakan model dari fungsi linear nilai lampau beserta nilai sekarang dan sisaan lampainya. Bentuk model ARIMA ( $p,d,q$ ) untuk mengestimasi produksi lada ditulis sebagai berikut:

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t \dots \dots \dots (3)$$

dimana:

$Y_t$  = produksi lada pada waktu ke- $t$

$Y_{t-p}$  = produksi lada pada kurun waktu ke ( $t-p$ )

$\mu$  = suatu konstanta

$\theta_1 \theta_q \phi_1 \phi_n$  = parameter-parameter model

$\varepsilon_{t-q}$  = nilai sisaan pada waktu ke- $(t-q)$

Penggunaan model ARIMA mensyaratkan series data yang stasioner. Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan *differencing*. *Differencing* yaitu menghitung perubahan atau selisih nilai observasi. Data yang telah dilakukan *differencing* perlu dicek kembali apakah telah stasioner atau belum. Pengecekan stasioneritas data dapat dilihat dari beberapa cara antara lain melihat sebaran data, menggunakan uji *Augmented Dickey-Fuller Test (ADF) Test Unit Root Test* dan melihat dari perilaku autokorelasi berdasarkan plot *Autocorrelation Function (ACF)* dan *Partial Autocorrelation Function (PACF)*.

Berdasarkan sebaran datanya, data yang telah stasioner menyebar secara acak dan tidak memiliki pola-pola tertentu baik pola musiman maupun *trend*. Pengecekan stasioneritas dengan uji ADF memiliki hipotesis sebagai berikut:

Hipotesis: .....(4)

$H_0$  : Data tidak stasioner

$H_1$  : Data stasioner

Jika nilai *test-statistic* pada uji ADF lebih kecil dari *critical value for test-statistic* baik pada taraf ( $\alpha$ ) 1%, 5% atau 10% maka  $H_0$  ditolak yang berarti data telah stasioner. Pengecekan stasioneritas dari perilaku *autokorelasi* dilihat dari plot ACF dan PACF. Jika pada kedua plot tersebut tidak banyak lag yang keluar dari *confidence interval* maka data telah stasioner.

Data yang telah stasioner selanjutnya dilakukan tahapan pendugaan model ARIMA menggunakan fungsi *auto.arima* atau *armaselect* yang tersedia pada software RStudio. Software tersebut akan memberikan rekomendasi model terbaik untuk mengestimasi produksi lada. Berdasarkan model terbaik yang terpilih, kemudian dilakukan pemeriksaan sisaan menggunakan pengujian LJungBox. Jika autokorelasi sisaan tidak signifikan yang ditandai dengan nilai p-value yang lebih besar dari 5% atau 10%, maka model ARIMA tersebut sudah cukup baik untuk mengepas data produksi lada.

Model ARIMA yang terpilih digunakan untuk mengestimasi data testing. Hasil ramalan data testing selanjutnya dibandingkan dengan data aktualnya untuk mengecek akurasi hasil ramalan. Akurasi hasil ramalan model ARIMA ditunjukkan oleh MAPE data training dan data testing. Jika model terpilih dirasa telah menghasilkan MAPE yang kecil, maka model tersebut dapat digunakan untuk mengestimasi produksi lada untuk beberapa periode ke depan.

### **Estimasi dengan Model Regresi Linear Sederhana dan Regresi Kuadrat**

Model regresi linear sederhana dan regresi kuadrat pada penelitian ini menggunakan peubah prediktor waktu (t) sebagai variabel independen. Model regresi linear sederhana untuk mengestimasi produksi lada dituliskan sebagai berikut:

$$y = \beta_0 + \beta_1 t + \varepsilon \dots\dots\dots(5)$$

dimana:

- $y$  = produksi lada
- $t$  = tahun
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi variabel tahun
- $\varepsilon$  = error

Parameter  $\beta_1$  mengukur efek tahun terhadap rata-rata perubahan produksi lada. Parameter ini diprediksi dari data sampel.  $\varepsilon$  menggambarkan kesalahan statistik, didefinisikan sebagai sisa atau perbedaan hasil antara nilai data pengamatan variabel dependen terhadap nilai hasil prediksi.

Model regresi kuadratik yang digunakan untuk meramalkan produksi lada dituliskan sebagai berikut:

$$y = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon \dots\dots\dots(6)$$

dimana:

- $y$  = produksi lada
- $t$  = tahun
- $t^2$  = kuadrat dari tahun
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi variabel tahun
- $\beta_2$  = koefisien regresi variabel kuadrat tahun
- $\varepsilon$  = error

Dampak perubahan tahun terhadap rata-rata perubahan produksi lada ditunjukkan oleh parameter  $\beta_1$ . Parameter  $\beta_2$  menunjukkan efek perubahan kuadrat tahun terhadap rata-rata perubahan produksi lada.

Estimasi dengan model regresi linear sederhana dan regresi kuadratik mendasarkan pada beberapa asumsi antara lain:

**- Sisaan mengikuti fungsi distribusi normal**

Uji normalitas dilakukan dengan *Jarque-Bera (JB) test for normality*. Hipotesis yang mendasari yaitu:

Hipotesis: .....(7)

$H_0$  : Sisaan terdistribusi normal

$H_1$  : *Otherwise*

Jika nilai p-value hasil JB test lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi normalitas terpenuhi.

**- Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas)**

Asumsi homoskedastisitas diuji menggunakan *Non-constant Variance (NCV) Score Test* dengan hipotesis sebagai berikut:

Hipotesis: .....(8)

$H_0$  : Homoskedastisitas

$H_1$  : Heterokedastisitas

Jika nilai p-value pada NCV test lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi homoskedastisitas terpenuhi.

**- Tidak terdapat autokorelasi antar sisaan untuk setiap data pengamatan**

Ada tidaknya autokorelasi antar sisaan pada data amatan diuji berdasarkan nilai autokorelasi *D-W Statistic* dengan hipotesis sebagai berikut:

Hipotesis: .....(9)

$H_0$  : Tidak ada autokorelasi antar sisaan

$H_1$  : *Otherwise*

Jika nilai p-value *autocorrelation D-W Statistic* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi tidak adanya autokorelasi antar sisaan telah terpenuhi.

**Estimasi dengan Model Fungsi Transfer**

Produksi lada di Indonesia tidak terlepas dari pengaruh variabel lain. Menurut Hamdani et al. (2015) salah satu variabel yang berpengaruh terhadap produksi lada di Indonesia adalah volume ekspor. Penelitian ini menggunakan volume ekspor sebagai variabel input dalam mengestimasi produksi (variabel output) menggunakan model fungsi transfer. Model fungsi transfer pada penelitian ini menggambarkan ramalan produksi lada yang didasarkan pada nilai-nilai masa lalu dari produksi itu sendiri, serta didasarkan pada volume ekspor (variabel input) dan gangguan/noise. Model fungsi transfer untuk mengestimasi produksi lada dituliskan sebagai berikut:

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\theta_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \quad \dots(10)$$

dimana:

$y_t$  = produksi lada tahun ke-t

$x_t$  = volume ekspor lada tahun ke-t

$b$  = panjang jeda pengaruh volume ekspor terhadap produksi lada

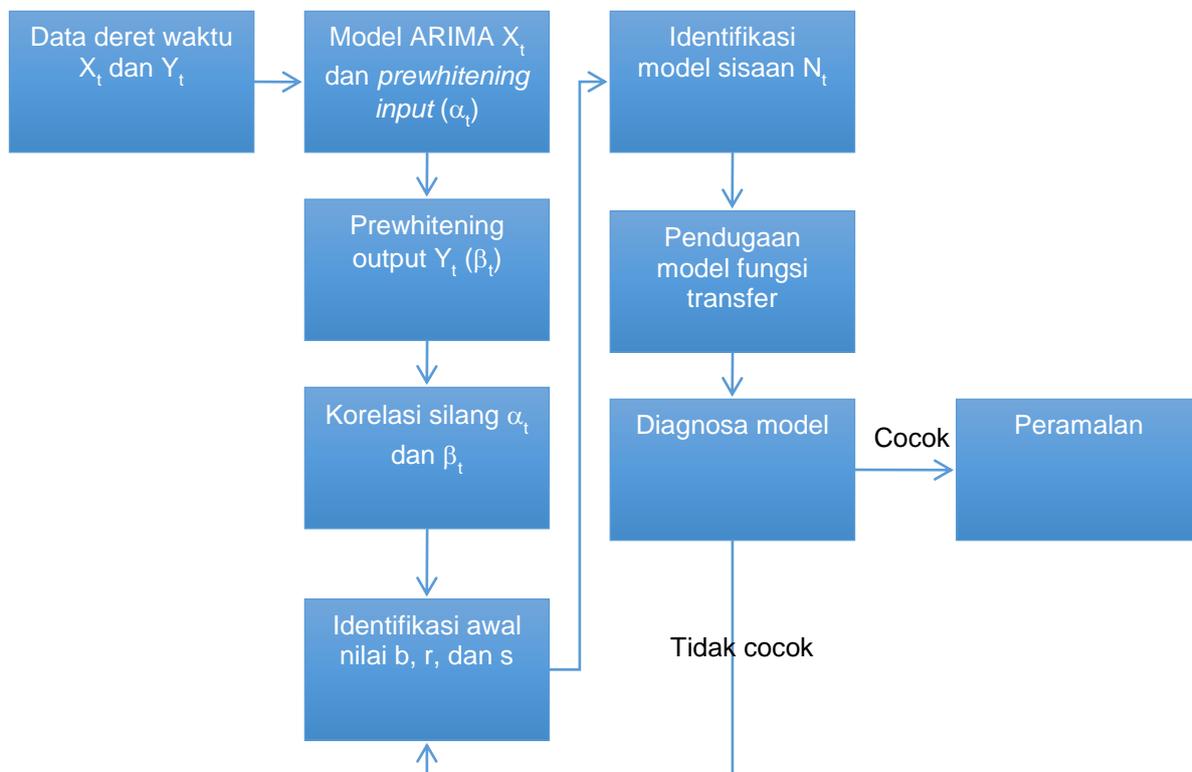
$r$  = panjang lag produksi lada periode sebelumnya yang masih mempengaruhi produksi lada tahun-t

$s$  = panjang jeda volume ekspor lada periode sebelumnya yang masih mempengaruhi produksi lada tahun-t

$p$  = ordo AR bagi noise  $N_t$

$q$  = ordo MA bagi noise  $N_t$

Pemodelan untuk estimasi produksi lada dengan fungsi transfer dilakukan melalui beberapa langkah sebagaimana disajikan pada Gambar 2 berikut:



Gambar 2. Langkah-langkah Estimasi Produksi Lada dengan Fungsi Transfer

## Estimasi dengan Model VAR

Pemodelan dengan *Vector Autoregression (VAR)* memperlakukan seluruh peubah secara simetris tanpa mempermasalahkan apakah variabel tersebut merupakan variabel dependen maupun independen. Penelitian ini menggunakan produksi lada sebagai variabel dependen. Produksi lada tersebut diduga dipengaruhi oleh beberapa variabel, tidak seperti model fungsi transfer yang hanya menggunakan volume ekspor saja sebagai variabel inputnya. Selain volume ekspor (Hamdani et al., 2015), variabel input lain yang diduga juga berpengaruh terhadap produksi lada di Indonesia antara lain luas areal (Fatma et al., 2020; Nursalam, 2020), dan harga (Fatma et al., 2020; Hamdani et al., 2015; Nursalam, 2020). Meskipun lada merupakan komoditas ekspor, namun tidak menutup kemungkinan Indonesia juga mengimpor lada dari negara lain. Selain itu, terdapat fenomena lada ekspor asal Indonesia yang ditolak (*reject*) oleh negara tujuan ekspor dikarenakan isu mutu dan keamanan pangan (Ditjenbun.pertanian.go.id, 2019). Oleh karena itu, penelitian ini menambahkan variabel volume impor dan harga impor untuk mengestimasi produksi lada di Indonesia.

Model VAR termasuk kategori model sistem, dimana ketika tidak ada kepastian untuk menentukan bahwa suatu peubah adalah eksogen (independen) maka suatu perluasan analisis fungsi perpindahan alami akan memperlakukan masing-masing peubah secara simetris (Enders, 2004). Sebagai contoh, pada kasus-kasus peubah yang membiarkan alur waktu atau *time path*  $\{s_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{y_t\}$  dan membiarkan *time path*  $\{y_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{s_t\}$ . Di dalam sistem *bivariate*, hubungan tersebut dapat digambarkan seperti pada persamaan berikut:

$$\begin{aligned} s_t &= b_{10} - b_{12}y_t + \gamma_{11}s_{t-1} + \gamma_{12}y_{t-1} + \varepsilon_{s_t} \\ y_t &= b_{20} - b_{21}s_t + \gamma_{21}s_{t-1} + \gamma_{22}y_{t-1} + \varepsilon_{y_t} \end{aligned} \dots(11)$$

Dengan mengasumsikan bahwa kedua peubah  $s_t$  dan  $y_t$  adalah stasioner:  $\varepsilon_{s_t}$  dan  $\varepsilon_{y_t}$  adalah *disturbances* yang memiliki rata-rata nol dan matriks kovarians terbatas atau bersifat *white noise* dengan standar deviasi yang berurutan  $\sigma_s$  dan  $\sigma_y$ : serta  $\{\varepsilon_{s_t}\}$  dan  $\{\varepsilon_{y_t}\}$  adalah *disturbances* yang independen dengan rata-rata nol dan kovarian terbatas (*uncorrelated white-noise disturbances*). Kedua persamaan di atas merupakan orde pertama VAR, karena panjang *lag* nya hanya satu. Agar Persamaan (11) lebih mudah dipahami dan digunakan sebagai alat analisis maka ditransformasikan dengan menggunakan matriks aljabar, dan hasilnya dapat dituliskan secara bersama seperti pada persamaan di bawah ini:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} s_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix} \text{ atau dengan bentuk lain:}$$

$$Bx_t = \Gamma_0 + \Gamma_1 X_{t-1} + \varepsilon_t \dots(12)$$

dimana:

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \quad x_t = \begin{bmatrix} s_t \\ y_t \end{bmatrix} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix}$$

Dengan melakukan pengalihan antara persamaan (12) dengan  $B^{-1}$  atau invers matriks B, maka akan dapat ditentukan model VAR dalam bentuk standar, seperti dituliskan pada persamaan di bawah ini:

$$x_t = A_0 + A_1 x_{t-1} + \ell_t \dots \dots \dots (13)$$

dimana:

$$A_0 = B^{-1} \Gamma_0$$

$$A_1 = B^{-1} \Gamma_1$$

$$\ell_t = B^{-1} \varepsilon_t$$

Pada penelitian ini  $X_t$  merupakan matriks yang dibentuk dari produksi lada, luas areal, volume ekspor, volume impor, harga ekspor dan harga impor.

Sebagaimana model regresi, model VAR juga didasarkan pada beberapa asumsi antara lain:

**- Sisaan mengikuti fungsi distribusi normal**

Uji normalitas pada model VAR didasarkan pada nilai *Jarque-Bera (JB) test (multivariate)*, *Skewness only (multivariate)*, dan *Kurtosis only (multivariate)*. Hipotesis yang mendasari yaitu:

Hipotesis: .....(14)

$H_0$  : Sisaan terdistribusi normal

$H_1$  : *Otherwise*

Jika nilai p-value dari *JB test, Skewness only dan Kurtosis only* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi normalitas terpenuhi.

**- Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas)**

Asumsi homoskedastisitas pada model VAR didasarkan pada nilai *ARCH (multivariate)* dengan hipotesis sebagai berikut:

Hipotesis: .....(15)

$H_0$  : Homoskedastisitas

$H_1$  : Heterokedastisitas

Jika nilai p-value pada *ARCH (multivariate)* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi homoskedastisitas terpenuhi.

**- Tidak terdapat autokorelasi antar sisaan untuk setiap data pengamatan**

*Pormanteau Test (asymptotic)* digunakan untuk menguji ada tidaknya autokorelasi antar sisaan pada data amatan. Hipotesis yang digunakan sebagai berikut:

Hipotesis: .....(16)

$H_0$  : Tidak ada autokorelasi antar sisaan

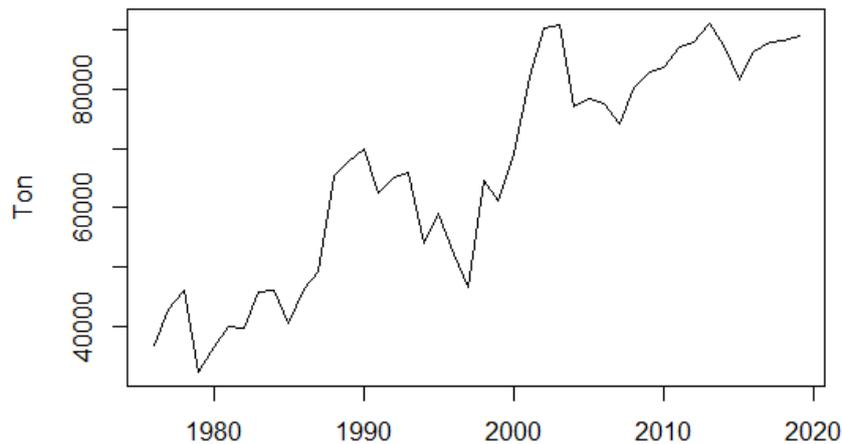
$H_1$  : *Otherwise*

Jika nilai p-value hasil uji *Pormanteau Test* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi tidak adanya autokorelasi antar sisaan telah terpenuhi.

## HASIL DAN PEMBAHASAN

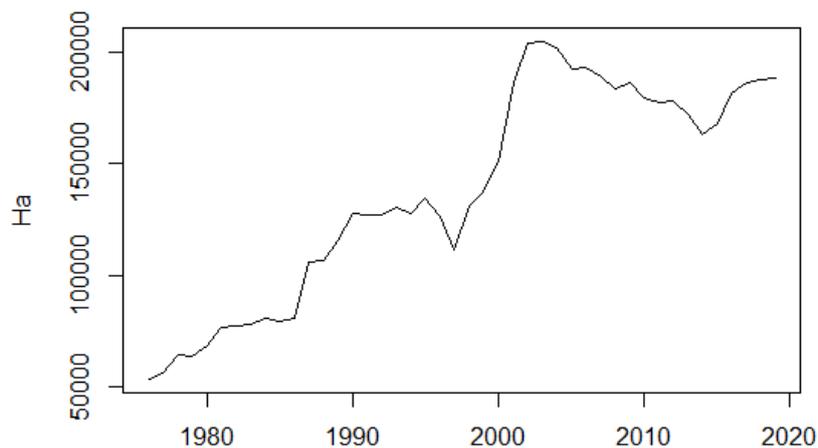
### Keragaan Lada di Indonesia

Lada merupakan salah satu komoditas unggulan sub sektor perkebunan dengan nilai ekonomis tinggi. Pengembangan usahatani lada di Indonesia sebagian besar berada di Luar Jawa seperti Kepulauan Bangka Belitung, Lampung, Sumatera Selatan, Sulawesi Selatan dan Kalimantan Timur. Berdasarkan data Kementerian Pertanian (2019) kelima provinsi tersebut berkontribusi sekitar 71% terhadap produksi lada di Indonesia. Jenis lada yang paling banyak diproduksi di Indonesia adalah lada hitam yang berasal dari Lampung atau dikenal dengan nama *Lampung black pepper* dan lada putih atau *Muntok white pepper* yang dihasilkan di Kepulauan Bangka Belitung (Nurdjannah, 2006). Perkembangan produksi lada dari tahun 1976 hingga tahun 2019 cenderung meningkat (Gambar 3). Pada tahun 1976 produksi lada sebesar 36.854 ton kemudian meningkat menjadi 88.949 ton pada tahun 2019. Selama sepuluh tahun terakhir (2010-2019) peningkatan produksi lada tercatat sebesar 1,01% per tahun, dengan produksi tertinggi dicapai pada tahun 2013 sebesar 91.039 ton.



Gambar 3. Perkembangan Produksi Lada di Indonesia Tahun 1976-2019

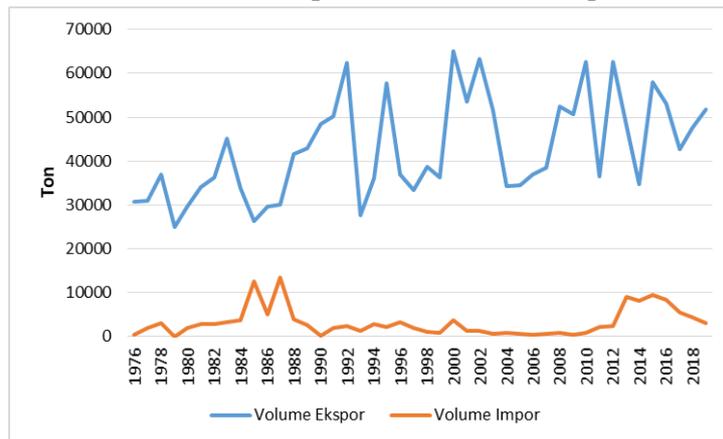
Peningkatan produksi lada seiring dengan perkembangan luas areal yang terus bertambah dari tahun ke tahun (Gambar 4). Pada tahun 1976 luas areal lada baik meliputi TBM, TM maupun TR/TTM sebesar 53.817 ha kemudian mengalami peningkatan menjadi 188.041 ha pada tahun 2019. Dengan rata-rata penambahan luas areal lada sebesar 0,18% per tahun selama sepuluh tahun terakhir, pengembangan areal lada melalui ekstensifikasi masih memungkinkan untuk dilakukan. Komposisi luasan lada di Indonesia sendiri didominasi oleh TM diikuti dengan TBM dan TR/TTM.



Gambar 4. Perkembangan Luas Areal Lada di Indonesia Tahun 1976-2019

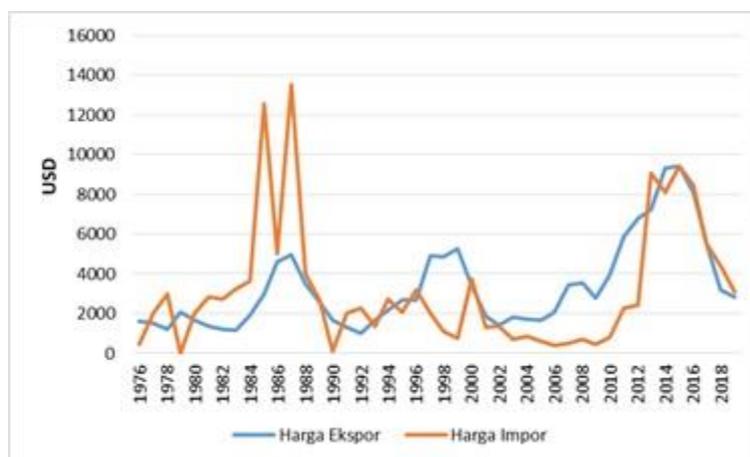
Sebagai salah satu produsen lada terbesar di dunia, Indonesia juga merupakan negara yang paling banyak mengekspor lada. Meskipun volume ekspor lada yang terdiri dari enam kode HS cenderung berfluktuatif dari tahun ke tahun, jika dibandingkan dengan volume impornya maka kuantitas lada yang diekspor masih jauh lebih banyak (Gambar 5). Pada tahun 1976 volume ekspor lada sebesar 30.831 ton dengan volume impornya hanya 128 ton. Di tahun 2019 volume ekspor lada meningkat menjadi 51.771 ton dengan volume impor sebesar 673 ton. Jika dibandingkan antara produksi dengan volume ekspor lada, terlihat bahwa lebih dari

50% produksi lada Indonesia diekspor untuk memenuhi permintaan lada dunia.



Gambar 5. Perkembangan Volume Ekspor dan Volume Impor Lada di Indonesia Tahun 1976-2019

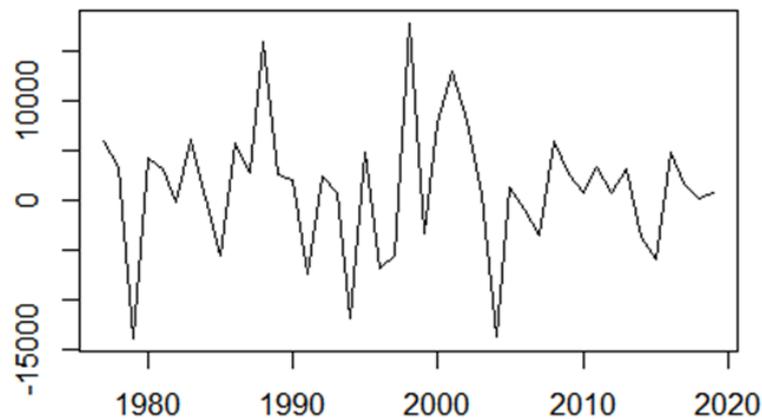
Harga ekspor dan harga impor merupakan pendekatan terhadap harga lada domestik maupun harga internasional yang ketersediaan series datanya terbatas. Harga ekspor dihitung dari nilai ekspor dibagi volume ekspor. Demikian halnya dengan harga impor yang dihitung dari nilai impor dibagi dengan volume impornya. Baik harga ekspor maupun harga impor lada selama periode 1976-2019 cenderung fluktuatif (Gambar 6). Harga ekspor lada tahun 1976 sebesar 1,6 ribu USD dan meningkat menjadi 2,8 ribu USD pada tahun 2019. Pada tahun 2015 harga lada mencapai nilai tertinggi yaitu 9,4 ribu USD. Selain pada harga ekspor, fluktuasi juga terjadi pada harga impor lada. Tahun 1976 harga impor lada tercatat 453 USD kemudian meningkat menjadi 3,1 ribu USD pada tahun 2019.



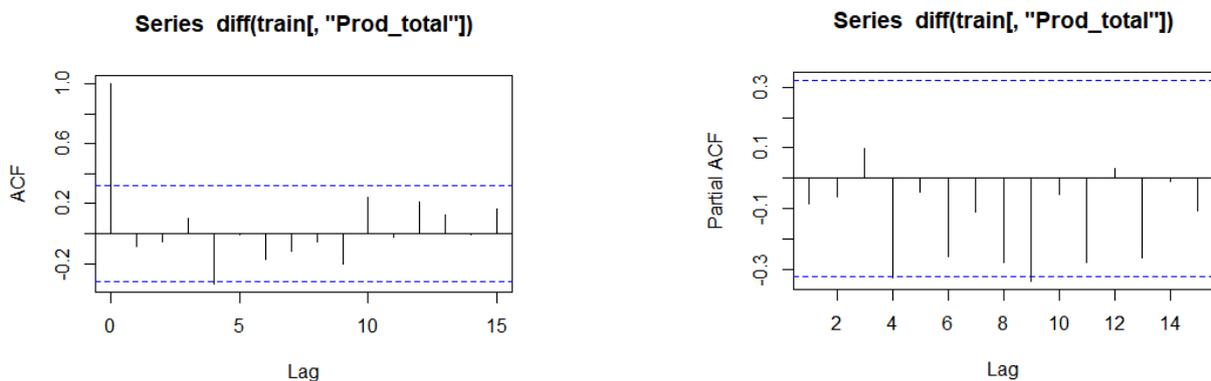
Gambar 6. Perkembangan Harga Ekspor dan Harga Impor Lada di Indonesia Tahun 1976-2019

## Hasil Estimasi dengan Model ARIMA

Pemodelan ARIMA dilakukan untuk data yang telah stasioner. Karena berdasarkan plot data serta hasil uji ADF terindikasi bahwa data produksi lada belum stasioner, dilakukan *differencing* satu kali. Setelah *differencing* data tersebut menjadi stasioner (Gambar 7) yang ditunjukkan dengan nilai *test-statistic* -4,79, dimana nilai tersebut lebih kecil dari *critical value* baik untuk  $\alpha$  sebesar 1%, 5% maupun 10%. Stasioneritas data tersebut juga didukung dengan plot ACF dan PACF, dimana pada plot tersebut tidak banyak lag yang keluar dari *confidence interval* (Gambar 8).



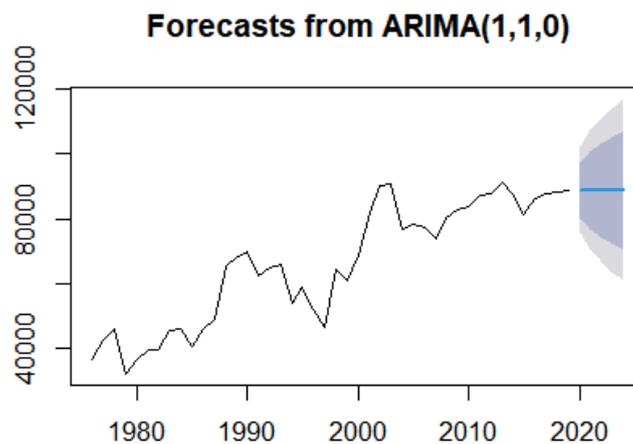
Gambar 7. Plot Produksi Lada Setelah *Differencing*



Gambar 8. Plot ACF dan PACF Setelah Dilakukan *Differencing* pada Data Produksi Lada

Dengan menggunakan software RStudio, diperoleh model terbaik untuk mengesimasi produksi lada yaitu ARIMA (1,1,0). Model ini menghasilkan MAPE data training sebesar 9,12 dan MAPE data testing sebesar 4,89. Berdasarkan hasil pemeriksaan sisaan dengan uji

LjungBox, ditemukan bahwa autokorelasi sisaan tidak signifikan pada seluruh lag yang ditandai dengan nilai p-value yang lebih besar dari 5%. Hal ini menunjukkan bahwa model ARIMA (1,1,0) sudah cukup baik untuk mengemas data produksi lada. Selanjutnya dengan model ARIMA (1,1,0) dilakukan estimasi produksi lada periode 2020-2024. Data Kementerian Pertanian (2019) menunjukkan bahwa pada tahun 2019 (ASEM) produksi lada sebesar 88.949 ton. Hasil estimasi lada dengan ARIMA (1,1,0) menghasilkan ramalan produksi pada tahun 2020 sebesar 88.916 ton kemudian meningkat menjadi 88.917 ton. Pada tahun 2021. Estimasi produksi lada sejak tahun 2021 hingga 2024 cenderung konstan yaitu 88.917 ton. Hasil estimasi produksi lada disajikan pada Gambar 9.



Gambar 9. Hasil Estimasi Produksi Lada Tahun 2020-2024

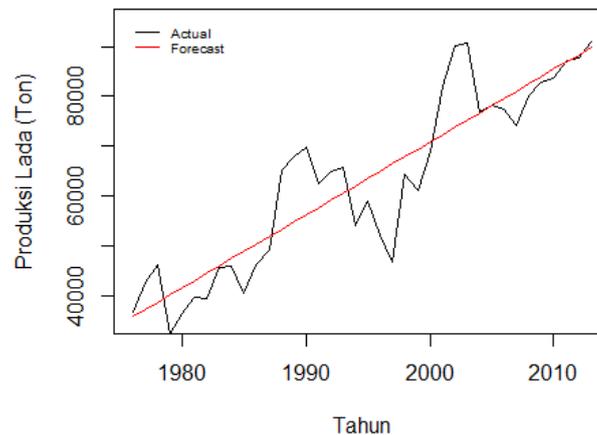
### Hasil Estimasi dengan Model Regresi Linear Sederhana dan Regresi Kuadratik

Model regresi linear sederhana untuk mengestimasi produksi lada menghasilkan persamaan regresi sebagai berikut:

$$\text{Produksi} = -2.844.606 + 1.458 t + \varepsilon \dots \dots \dots (17)$$

Secara parsial, intercept dan tahun masing-masing berkorelasi signifikan terhadap produksi lada yang ditunjukkan oleh nilai *p-value* lebih kecil dari 1%. Dengan *p-value* model yang juga lebih kecil dari 1% menunjukkan bahwa secara umum model regresi linear tersebut telah signifikan untuk mengestimasi produksi lada.  $R^2$  yang dihasilkan pada model ini sebesar 0,81 menunjukkan bahwa 81% produksi lada di Indonesia diduga dipengaruhi oleh tahun, sedangkan 19% lainnya dipengaruhi oleh variabel lain yang tidak masuk dalam model. Model (17) menunjukkan bahwa penambahan tahun sebanyak satu unit satuan akan meningkatkan produksi lada sebesar 1.458 ton. Dengan MAPE data training 10,27 dan MAPE data testing 9,46 model ini meramalkan produksi lada memiliki kecenderungan meningkat sebagaimana disajikan pada Gambar 10. Produksi lada tahun 2020 diramalkan sebesar 100.015 ton dan meningkat hingga mencapai 105.846 ton pada tahun 2024.

Berdasarkan hasil pengujian asumsi pada Hipotesis (7), (8) dan (9) ditemukan bahwa sisaan mengikuti fungsi distribusi normal yang ditunjukkan dengan nilai  $p$ -value hasil uji  $JB$  test sebesar 0,85. Namun asumsi homoskedastisitas pada model ini tidak terpenuhi dimana  $p$ -value dari uji NCV sebesar 0,99. Namun terdapat asumsi yang dilanggar yaitu terdapat autokorelasi antar sisaan yang ditunjukkan dengan nilai  $p$ -value dari *autocorrelation D-W Statistic* lebih kecil dari  $\alpha$ . Mengingat tidak semua asumsi terpenuhi, maka model regresi linear sederhana tidak direkomendasikan untuk dipilih sebagai model terbaik dalam mengestimasi produksi lada.



Gambar 10. Hasil Estimasi Produksi Lada dengan Model Regresi Linear Sederhana

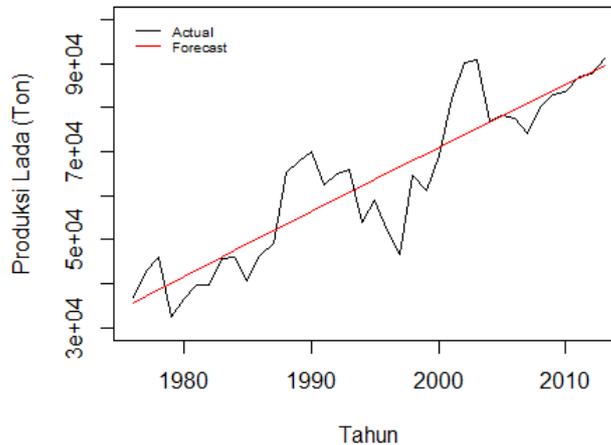
Model regresi kuadratik untuk mengestimasi produksi lada menghasilkan persamaan regresi sebagai berikut:

$$\text{Produksi} = 35.615 + 1.501 t - t^2 + \epsilon \dots (18)$$

Secara parsial, intercept dan tahun masing-masing berkorelasi signifikan terhadap produksi lada yang ditunjukkan oleh nilai  $p$ -value lebih kecil dari 1%, sedangkan kudrat tahun tidak signifikan. Dengan  $p$ -value model yang juga lebih kecil dari  $\alpha$  menunjukkan bahwa secara umum model regresi kuadratik telah signifikan untuk mengestimasi produksi lada. *Adjusted-R<sup>2</sup>* yang dihasilkan pada model ini sebesar 0,80 menunjukkan bahwa 80% produksi lada di Indonesia diduga dipengaruhi oleh tahun dan kuadrat tahun, sedangkan 20% lainnya dipengaruhi oleh variabel lain yang tidak masuk dalam model. Model (18) menunjukkan bahwa penambahan tahun sebanyak satu unit satuan akan meningkatkan produksi lada sebesar 1.501 ton, sedangkan penambahan kuadrat tahun sebanyak satu unit satuan akan menurunkan produksi lada sebesar 1 ton. Dengan MAPE data training 10,29 dan MAPE data testing 8,97 model ini meramalkan produksi memiliki kecenderungan meningkat sebagaimana disajikan pada Gambar 11. Produksi lada tahun 2020 diramalkan sebesar 104.969 ton dan meningkat hingga mencapai 110.505 ton pada tahun 2024.

Hasil pengujian asumsi pada Hipotesis (7), (8) dan (9) ditemukan bahwa sisaan mengikuti fungsi distribusi normal yang ditunjukkan dengan nilai  $p$ -value hasil uji  $JB$  test sebesar 0,89.

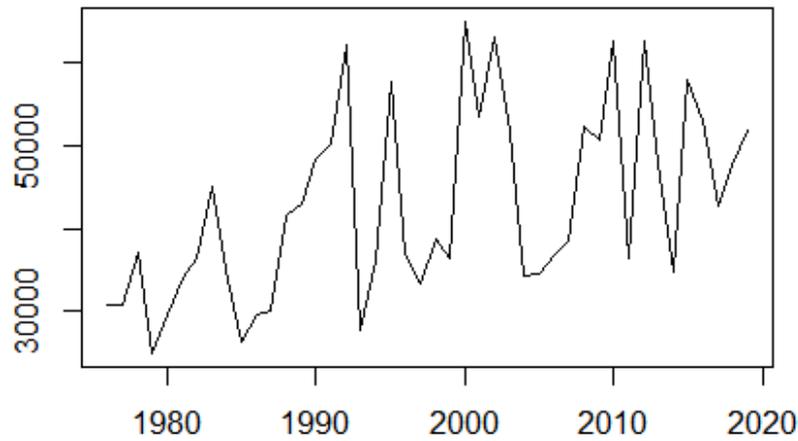
Namun asumsi homoskedastisitas pada model ini tidak terpenuhi dimana  $p$ -value dari uji NCV sebesar 0,98 atau lebih kecil dari  $\alpha$ . Selain itu, terdapat asumsi lain yang dilanggar yaitu ditemukan autokorelasi antar sisaan yang ditunjukkan dengan nilai  $p$ -value dari *autocorrelation D-W Statistic* lebih kecil dari  $\alpha$ . Mengingat banyak asumsi yang terlanggar, maka model regresi kuadratik juga tidak direkomendasikan untuk dipilih sebagai model terbaik dalam mengestimasi produksi lada.



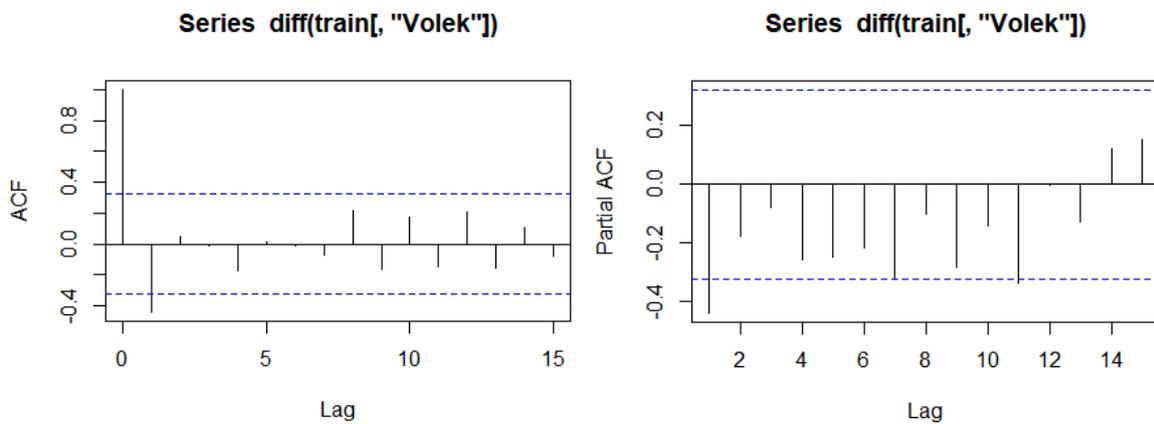
Gambar 11. Hasil Estimasi Produksi Lada dengan Model Regresi Kuadratik

### Hasil Estimasi dengan Model Fungsi Transfer

Estimasi dengan model fungsi transfer diawali dengan meramalkan nilai variabel input yaitu volume ekspor menggunakan model ARIMA. Dikarenakan data volume ekspor belum stasioner maka dilakukan *differencing* satu kali sehingga menghasilkan data yang telah stasioner (Gambar 12) yang ditunjukkan dengan nilai *test-statistic* -5,69, dimana nilai tersebut lebih kecil dari *critical value* baik untuk  $\alpha$  sebesar 1%, 5% maupun 10%. Stasioneritas volume ekspor lada juga didukung dengan plot ACF dan PACF, dimana pada plot tersebut tidak banyak lag yang keluar dari *confidence interval* (Gambar 13). Model ARIMA terbaik untuk mengestimasi volume ekspor lada adalah ARIMA (0,1,4) dengan AIC model sebesar 802,94.



Gambar 12. Plot Volume Ekspor Lada Setelah *Differencing*

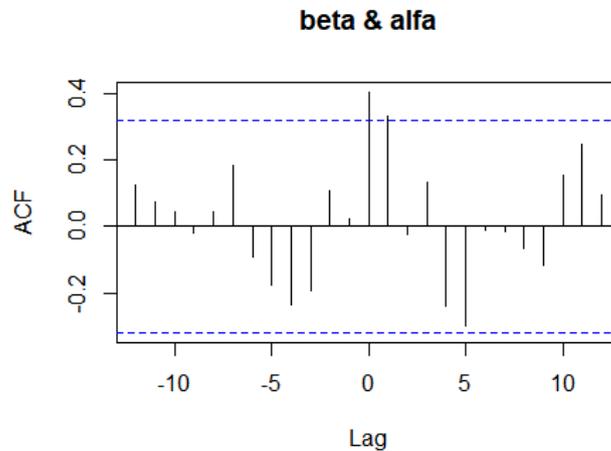


Gambar 13. Plot ACF dan PACF Setelah Dilakukan *Differencing* pada Data Volume Ekspor Lada

Tahapan selanjutnya dalam pemodelan fungsi transfer adalah prewhitening dan korelasi silang antar residual volume ekspor dengan residual produksi lada yang menghasilkan plot ACF (Gambar 14). Berdasarkan plot ACF tersebut diperoleh nilai  $b=0$  dimana lag pertama kali signifikan pada lag 0. Interpretasi dari nilai  $b=0$  yaitu tidak ada jeda pengaruh dampak volume ekspor terhadap produksi lada. Selain itu, diperoleh nilai  $s=0$  karena tidak ada tambahan lag yang signifikan setelah lag 0, yang berarti korelasi antara volume ekspor dengan produksi terjadi di tahun yang sama. Nilai  $r$  diasumsikan 0 karena data produksi maupun volume ekspor lada merupakan data tahunan yang tidak mengandung pola musiman.

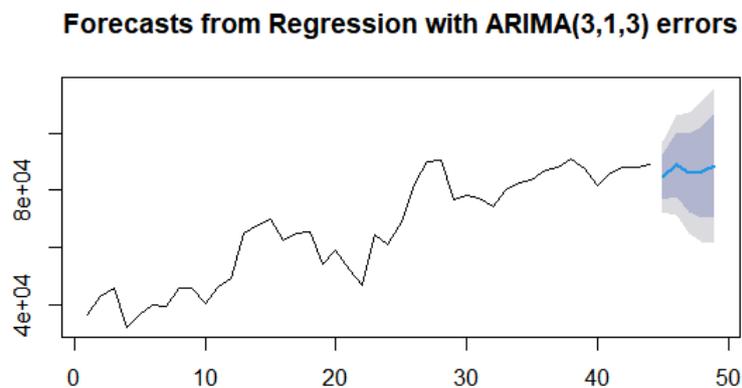
Setelah pengepasan model berdasarkan nilai  $r, s$ , dan  $b$ , dilakukan identifikasi model noise dengan memanfaatkan fungsi *auto.arima* maupun *armaselect* pada RStudio. Model noise

yang direkomendasikan yaitu ARIMA (3,1,4) namun karena model ini tidak signifikan di MA(0,1,4) maka dilakukan *overfitting*. Hasil *overfitting* memberikan model terbaik ARIMA (3,1,3) dengan MAPE data training 7,16 dan MAPE data testing 2,54.



Gambar 14. Plot ACF Hasil Korelasi Silang Residual

Secara umum, model fungsi transfer terbaik untuk mengestimasi produksi lada adalah fungsi transfer ARIMA (3,1,3,) dimana variabel input (volume ekspor) terlebih dahulu diestimasi dengan ARIMA (0,1,3). Estimasi produksi lada lima tahun ke depan cenderung fluktuatif (Gambar 15). Hasil estimasi lada pada tahun 2020 sebesar 84.611 ton kemudian meningkat menjadi 88.939 ton pada tahun 2021. Tahun 2022 produksi lada diramalkan turun menjadi 86.142 ton kemudian naik menjadi 86,475 ton pada tahun 2023 dan 88.744 pada tahun 2024. Dengan nilai MAPE data testing yang cukup kecil, model fungsi transfer ini direkomendasikan sebagai salah satu model untuk mengestimasi produksi lada beberapa periode ke depan.

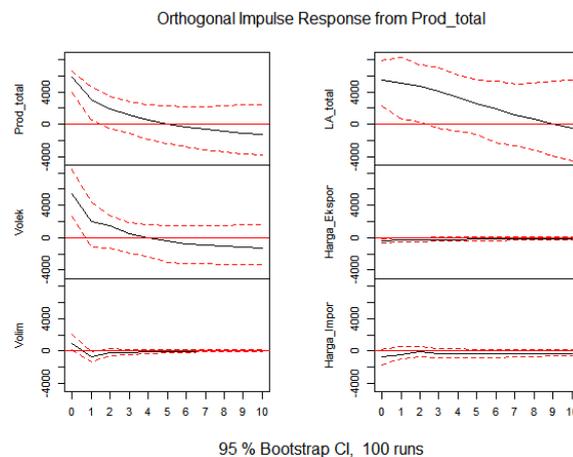


Gambar 15. Hasil Estimasi Produksi Lada dengan Model Fungsi Transfer

## Hasil Estimasi dengan Model VAR

Estimasi dengan model VAR diawali dengan pemilihan ordo- $p$  serta type model VAR yang paling baik. Model VAR terbaik ditandai dengan banyaknya variabel yang signifikan pada ordo yang telah ditentukan. Model VAR yang terpilih yaitu VAR dengan ordo-1 atau VAR(1) tanpa memasukkan konstanta maupun trend. Pada model VAR(1) beberapa variabel yang signifikan yaitu produksi  $t-1$ , luas areal  $t-1$ , volume ekspor  $t-1$  dan harga ekspor  $t-1$ , sedangkan volume impor  $t-1$  dan harga impor  $t-1$  tidak signifikan.

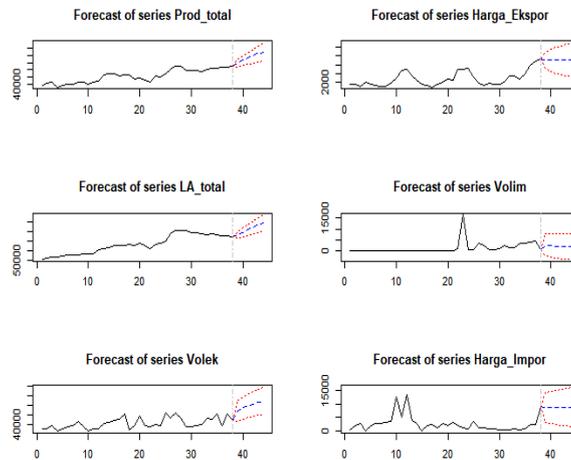
Salah satu kelebihan dari model VAR adalah dapat menampilkan *impulse respon* antar variabel. *Impulse respon* menggambarkan tingkat laju dari *shock* peubah yang satu terhadap peubah yang lainnya pada suatu rentang periode tertentu. Sehingga dapat dilihat lamanya pengaruh dari *shock* suatu peubah terhadap peubah lain sampai pengaruhnya hilang atau kembali ke titik keseimbangan. *Impulse respon* produksi pada model VAR(1) ditunjukkan pada Gambar 16. Berdasarkan gambar tersebut terlihat bahwa perubahan produksi berdampak pada produksi itu sendiri sampai 3 tahun ke depan, kemudian dampaknya hilang setelah 3 tahun. Perubahan produksi lada juga berdampak pada volume ekspor dan luas areal, namun tidak berdampak pada volume impor, harga ekspor dan harga impor.



Gambar 16. *Impulse Respon* Produksi pada Model VAR(1)

Model VAR(1) menghasilkan MAPE data training sebesar 50,31 dan MAPE data testing sebesar 34,22. Berdasarkan hasil pengujian asumsi pada Hipotesis (14), (15) dan (16) ditemukan bahwa sisaan tidak terdistribusi secara normal yang ditunjukkan dengan nilai  $p$ -value lebih kecil dari  $\alpha$  berdasarkan hasil uji *JB test*, *Skewness*, dan *Kurtosis*. Asumsi homoskedastisitas pada model ini terpenuhi dimana  $p$ -value pada ARCH sebesar 1. Selain itu, tidak terdapat autokorelasi antar sisaan yang ditunjukkan dengan nilai  $p$ -value dari *Portmanteu Test* sebesar 0,96. Mengingat fokus pemodelan pada penelitian ini adalah peramalan dengan series data yang cukup banyak, sehingga meskipun asumsi normalitas tidak terpenuhi, model VAR(1) dapat dipertimbangkan untuk digunakan dalam mengestimasi produksi lada.

Berdasarkan model VAR(1), estimasi produksi lada selama lima tahun ke depan diramalkan naik. Pada tahun 2020 produksi lada diprediksi sebesar 86.938 ton dan terus meningkat hingga mencapai 87.390 ton pada tahun 2024. Selain ramalan produksi lada, model VAR(1) ini juga meramalkan variabel input lain yaitu luas areal, volume ekspor, volume impor, harga ekspor dan harga impor sebagaimana disajikan pada Gambar 17.



Gambar 17. Estimasi Variabel Input pada Model VAR(1)

### Pemilihan Model Estimasi Terbaik

Kriteria pemilihan model estimasi terbaik pada penelitian ini didasarkan pada nilai terkecil dari MAPE data training dan data testing serta kelogisan hasil ramalan jika dibandingkan dengan perkembangan produksi lada beberapa tahun terakhir. Dari keempat model estimasi yang dibandingkan, diperoleh informasi bahwa model fungsi transfer memberikan MAPE data training dan data testing paling kecil dibandingkan model lain. Dengan produksi lada pada tahun 2019 sebesar 88.949 ton (Kementerian Pertanian, 2019) dan rata-rata pertumbuhan produksi selama lima tahun terakhir (2015-2019) sebesar 2,24%, maka produksi lada pada periode 2020-2024 diramalkan naik 1,25%. Perbandingan hasil estimasi produksi lada selama lima tahun ke depan dengan model ARIMA, regresi, fungsi transfer dan VAR disajikan pada Tabel 1

Tabel 1. Perbandingan Hasil Estimasi Produksi Lada dengan Model ARIMA, Regresi, Fungsi Transfer dan VAR

No	Model	MAPE		Estimasi Produksi Lada (Ton)					Pertumb. (%)
		Train	Test	2020	2021	2022	2023	2024	
1	ARIMA(1,1,0)	9,12	4,89	88.916	88.917	88.917	88.917	88.917	0,00
2	Regresi Linear	10,27	9,46	100.015	101.472	102.930	104.388	105.846	1,43
	Regresi Kuadratik	106,41	132,4	211.632	212.516	213.294	213.966	214.531	0,34
3	Fungsi Transfer ARIMA(3,1,3), xreg=volek	7,16	2,54	84.611	88.939	86.142	86.475	88.744	1,25
4	VAR(1) type=none	50,31	34,22	86.938	86.582	86.627	86.916	87.390	0,13

## KESIMPULAN DAN SARAN

### Kesimpulan

Model terbaik yang terpilih untuk mengestimasi produksi lada adalah fungsi transfer ARIMA (3,1,3) dengan variabel inputnya volume ekspor. Pemilihan fungsi transfer sebagai model estimasi dengan pertimbangan MAPE data training dan MAPE data testing yang dihasilkan merupakan yang terkecil dibandingkan model lain. Selain itu, hasil estimasi produksi lada selama lima tahun ke depan fluktuatif namun cenderung naik sebesar 1,25% per tahun. Hal ini seiring dengan rata-rata pertumbuhan produksi lada lima tahun terakhir yang mengalami peningkatan 2,24% per tahun. Dengan kata lain, estimasi produksi lada periode 2020-2024 masih logis jika dibandingkan dengan historis produksi lada beberapa tahun ke belakang. Hasil estimasi dengan fungsi transfer menunjukkan pada tahun 2020 produksi lada di Indonesia sebesar 84.611 ton kemudian meningkat di tahun 2024 menjadi 88.744 ton.

### Saran

Penelitian ini membatasi estimasi produksi lada menggunakan empat model yaitu ARIMA, regresi, fungsi transfer dan VAR. Meskipun model fungsi transfer dipilih menjadi model terbaik untuk mengestimasi produksi lada, akan lebih baik jika pada penelitian berikutnya mencoba model lain untuk mengestimasi produksi lada baik itu *univariate* atau *multivariate*. Selain itu, variabel lain selain yang telah digunakan pada penelitian ini dapat juga dicobakan untuk mengestimasi produksi lada.

## DAFTAR PUSTAKA

- Damanik, Sabarman. 2001. Analisis Penawaran dan Permintaan Lada Indonesia di Pasar Internasional. *Jurnal Litri*. 7(4):113-119.
- Ditjenbun.pertanian.go.id. (2019, 13 Agustus). IPC Pintu Masuk Negosiasi Perdagangan Lada Indonesia. Diakses pada 2 Desember 2020, dari <http://ditjenbun.pertanian.go.id/ipc-pintu-masuk-negoisasi-perdagangan-lada-indonesia/>
- Enders, W. 2004. *Applied Econometric Time Series*. USA: University of Alabama.
- Fatma, Hikmah, N., & Usman. 2020. Faktor-faktor yang Memengaruhi Produksi Lada di Desa Kongkomas Kecamatan Basidondo Kabupaten Tolitoli. *Jurnal Agrotech*. 10(1):35-40.
- Gujarati, D.N., & Porter, D.C. 2010. *Dasar-dasar Ekonometrika*. Jakarta: Salemba Empat, Buku 2, Edisi 5.
- Hamdani, Tety, E., & Eliza. 2015. Faktor-faktor yang Mempengaruhi Produksi Lada di Indonesia. *Jom Faperta*. 7(2):1-7.
- Kardinan, A., Laba, I.W., & Rismayani. 2018. Peningkatan Daya Saing Lada Melalui Budidaya Organik. *Perspektif*. 7(1):26-39.
- Kemala, Syarif. 2006. Strategi Pengembangan Sistem Agribisnis Lada untuk Meningkatkan Pendapatan Petani. *Perspektif*. 5(10): 48-54.
- Kementerian Pertanian. 2019. *Statistik Perkebunan Indonesia 2018-2020 (Lada)*. Jakarta: Direktorat Jenderal Perkebunan Kementerian Pertanian.
- Nurdjannah, Nanan. 2006. Perbaikan Mutu Lada Dalam Rangka Meningkatkan Daya Saing di Pasar Dunia. *Perspektif*. 5(1): 13-25.
- Nursalam. 2020. Analisis Produksi dan Efisiensi Alokatif Usahatani Lada di Desa Ameroro Kecamatan Tinondo Kabupaten Kolaka Timur. *Agrimor*. 5(3):57-59.
- Zikria, Roydatul. 2019. *Outlook Lada*. Jakarta: Pusat Data dan Sistem Informasi Pertanian.



## PENGEMBANGAN METODOLOGI ESTIMASI DATA PRODUKSI CENGKEH 2020

Vera Junita Siagian, Fungsional Statistisi, Pusdatin Kementan

### ABSTRAK

Pada tahun 2020, Pusdatin bekerjasama dengan Ditjen Perkebunan dan BPS mengkaji 4 metode estimasi yang nantinya didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik dari metode sebelumnya dan diharapkan hasil estimasinya cenderung mendekati angka aktualnya. Melalui kegiatan pengembangan metode estimasi data perkebunan tahun 2020, diharapkan akan dihasilkan metode yang paling sesuai untuk mengestimasi produksi cengkeh nasional dengan membandingkan besaran MAPE (Mean Absolut Percentage Error) dan kerealistisan hasil permalan dengan data series sebelumnya

Tujuan dari makalah ini adalah untuk membandingkan efektivitas metode tersebut dalam hasil estimasi produksi cengkeh dan menentukan metode terbaik dalam estimasi produksi cengkeh nasional. Metode yang diterapkan adalah ARIMA (*Autoregressive Integrated Moving Average*), Regresi, Fungsi Transfer dan VAR (*Vector Auto Reggression*) dengan menggunakan software R Studio

### ABSTRACT

*In 2020, Center for Agriculture Data and Informasi System (CADIS) together with the Directorate General of Plantation and the BPS-Statistic Indonesia examined 4 methods of estimation that will be obtained more accurately, more objectively and statistically better than the previous method and it is expected that the estimated results tend to be close to the actual number. Through the development of plantation data estimation method in 2020, it is expected that the most suitable method will be produced to estimate national clove production by comparing the amount of MAPE (Mean Absolut Percentage Error) and the reality of the results of the permalan with the previous series data*

*The purpose of this paper is to compare the effectiveness of such methods in the estimated production of cloves and determine the best method in estimating national clove production. The methods applied are ARIMA (Autoregressive Integrated Moving Average), Regression, Transfer Function and VAR (Vector Auto Reggression) using R Studio software*

## I. PENDAHULUAN

### A. Latar Belakang

Metode yang dilakukan selama ini untuk menyusun angka estimasi (AESTI) produksi cengkeh nasional menurut Buku Pedoman Pelaksanaan Pengeolaan Data Komoditas Perkebunan (PDKP) adalah metode pemulusan eksponensial tunggal atau Single Exponential Smoothing (SES) dan pemulusan eksponensial ganda atau DSS (Double Exponential Smoothing). Untuk pemilihan model terbaik yaitu dengan membandingkan besaran MAPE (Mean Absolut Percentage Error) dan kerealistisan hasil permalan dibandingkan dengan data series sebelumnya. Menurut T. Hani Handoko (2011), *Exponential Smoothing* adalah suatu tipe teknik peramalan rata-rata bergerak yang melakukan penimbangan terhadap data masa lalu dengan cara eksponensial sehingga data paling akhir mempunyai bobot atau timbangan lebih besar dalam rata-rata bergerak.

Pada tahun 2020 ini, Pusdatin mengkaji 4 (empat) metode estimasi guna menghasilkan angka yang nilainya semakin mendekati nilai sebenarnya. Ke empat metode yang di kaji yaitu ARIMA (*Autoregressive Integrated Moving Average*), Regresi, Fungsi Transfer dan VAR (*Vector Auto Regression*) dengan menggunakan software R Studio

### B. Tujuan

Tujuan dari makalah ini adalah:

- Melakukan analisis dan pemodelan dengan ARIMA, Regresi, Fungsi Transfer dan VAR untuk meramalkan produksi cengkeh nasional
- Membandingkan efektivitas metode tersebut dalam hasil estimasi produksi cengkeh
- Menentukan metode terbaik dalam meramal data produksi cengkeh nasional

## II. Sumber data dan Struktur Data

Data yang digunakan dalam kajian ini adalah data sekunder yang bersumber dari Direktorat Jenderal Perkebunan dan Badan Pusat Statistik (BPS). Untuk data luas areal dan produksi cengkeh bersumber dari Direktorat Jenderal Perkebunan sementara untuk data harga produsen dan volume ekspor bersumber dari BPS dengan series data tahunan yaitu dari tahun 1983 sampai tahun 2019.

## III. Metodologi

### a. ARIMA

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung flat (mendatar/konstan) untuk periode yang cukup panjang. Model Autoregresif Integrated Moving Average (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam

membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (time series) secara statistik berhubungan satu sama lain (dependent) [https://daps.bps.go.id/ile\\_artikel/77/arima.pdf](https://daps.bps.go.id/ile_artikel/77/arima.pdf)

b. Regresi

Analisis regresi mempelajari bentuk hubungan antara satu atau lebih peubah/variabel bebas (X) dengan satu peubah tak bebas (Y). Pada regresi harus ada variabel yang ditentukan dan variabel yang menentukan atau dengan kata lain adanya ketergantungan variabel yang satu dengan variabel yang lainnya. Untuk menentukan bentuk hubungan (model) diperlukan pemisahan yang tegas antara variabel bebas yang diberi simbol X dan variabel tak bebas yang diberi simbol Y. Dalam bentuk yang paling sederhana yaitu satu peubah bebas (X) dengan satu peubah tak bebas (Y) mempunyai persamaan:

$$Y = a + bX + \varepsilon$$

Dimana:

a : intersept

b: koefisien arah atau koefisien beta

c. Transfer

Model fungsi transfer merupakan gabungan dari karakteristik analisis regresi berganda dengan karakteristik deret berkala ARIMA (*Autoregressive Integrated Moving Average*). Dalam model ini, selain mengandung keterkaitan dengan kejadian pada waktu-waktu sebelumnya, juga keterkaitan yang sifatnya kausal dengan variabel lain sebagaimana pada model regresi. Model fungsi transfer dengan variabel input lebih dari dua deret berkala disebut dengan model fungsi transfer multivariat. Prosedur pembentukan model fungsi transfer multivariat melalui dua tahap, yaitu pembentukan model fungsi transfer tunggal dari masing-masing input, baru dilakukan pembentukan model fungsi transfer secara simultan dari semua variabel.

Beberapa hal yang berkaitan dengan model fungsi transfer antara lain deret berkala output, disebut  $Y_t$ , yang diperkirakan akan dipengaruhi oleh deret berkala input, disebut  $X_t$ , dan input-input lain yang digabungkan dalam satu kelompok yang disebut gangguan (noise)  $N_t$ . Model fungsi transfer bivariat ditulis dalam dua bentuk umum. Bentuk pertama adalah sebagai berikut: (Makridakis, dkk:1999:448)

$$Y_t = v(B)X_t + N_t$$

d. Model Var

VAR (*Vector Auto Regression*) digunakan untuk memproyeksikan sebuah sistem dengan variabel runtut waktu dan untuk menganalisis dampak dinamis dari faktor gangguan yang terdapat dalam sistem variabel tersebut. Pada dasarnya Analisis VAR sama dengan suatu model persamaan simultan, karena dalam Analisis VAR kita mempertimbangkan beberapa variabel endogen secara bersama-sama dalam suatu model. Sebenarnya analisis ini mirip dengan model persamaan simultan biasa. Hanya

saja di dalam Analisis VAR masing-masing variabel selain diterangkan oleh nilainya di masa lampau, juga dipengaruhi oleh nilai masa lalu dari semua variabel endogen lainnya dalam model yang diamati. Selain itu, dalam analisis VAR biasanya tidak ada variabel eksogen di dalam model.

## IV. Hasil Pembahasan

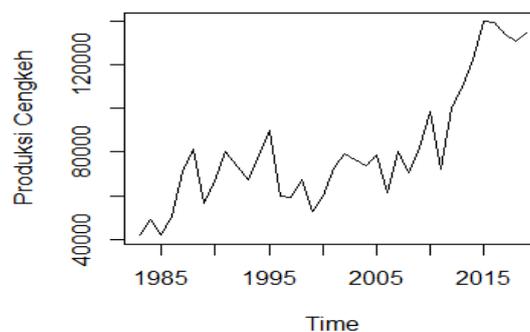
### IV.I. Model ARIMA

#### a. Penyusunan Model Menggunakan R Studio

R Studio adalah bahasa pemrograman dan sistem perangkat lunak yang dirancang khusus untuk mengerjakan segala hal terkait komputasi statistik. Bahasa pemrograman ini pertama kali dikembangkan pada tahun 1993 oleh dua orang pakar statistik yaitu Ross Ihaka dan Robert Gentleman di Auckland University, New Zealand. Sampai saat ini, bahasa pemrograman R terus berkembang secara pesat seiring dengan semakin populernya terminologi “*Big Data*” dan meningkatnya kebutuhan perusahaan akan *data scientist* untuk mengolah dan menganalisis data di perusahaan tersebut sebagai dasar pengambilan kebijakan dan mengotomatisasi proses bisnis menjadi *data driven*. Bahasa pemrograman seperti Python dan R telah menjadi pilihan utama bagi para peneliti maupun praktisi di bidang *data science* untuk mengolah dan menganalisis data baik itu untuk kepentingan penelitian maupun bisnis. Oleh karena itu, bagi seorang pemula di bidang *data science*, R merupakan bahasa pemrograman yang sangat *recommended* untuk dikuasai

#### b. Hasil Run Model Arima dan Eksplorasi data

Syarat utama dalam melakukan pemodelan ARIMA adalah kestasioneran data. Kestasioneran data dapat diketahui secara visual (plot datanya) dan juga dengan uji statistik. Dari plot data gambar 1 terlihat produksi cengkeh tahun 1983-2019 memiliki trend sehingga terindikasi tidak stasioner.



Gambar 1. Plot Produksi cengkeh tahun 1983-2019

Dengan uji statistik dapat dilakukan salah satunya yaitu dengan uji *Augmented Dickey-Fuller* yaitu jika test statistics lebih besar dari critical value maka data tersebut tidak stasioner. Untuk menstasionerkan data dapat dilakukan differencing. Differencing adalah menghitung perubahan atau selisih nilai observasi. Berdasarkan uji *Augmented Dickey-Fuller* dapat dilihat seperti hasil dibawah ini.

#### Hasil Uji Augmented Dickey-Fuller

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-27694  -6162   1818   8938  20181

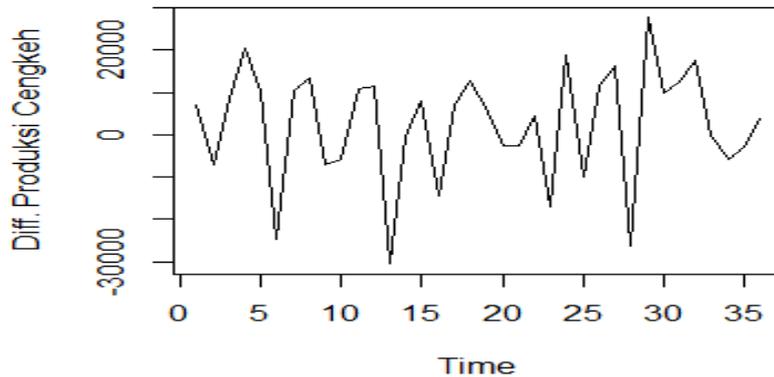
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 12667.5538  7931.8653   1.597  0.1204
z.lag.1      -0.2738    0.1496  -1.831  0.0768 .
tt           644.5555   362.0245   1.780  0.0848 .
z.diff.lag   -0.1377    0.1810  -0.761  0.4526
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13030 on 31 degrees of freedom
Multiple R-squared:  0.1736,    Adjusted R-squared:  0.09367
```

```
F-statistic: 2.171 on 3 and 31 DF, p-value: 0.1114

Value of test-statistic is: -1.8308 1.8733 1.8169
Critical values for test statistics:
      1pct  5pct 10pct
tau3  -4.15 -3.50 -3.18
phi2   7.02  5.13  4.31
phi3   9.31  6.73  5.61
```

Dari hasil uji Augmented Dickey-Fuller diatas terlihat bahwa nilai tes statistic adalah -1,8308 dan nilai critical tau 3 untuk 1% sebesar -4,15, taraf uji 5% sebesar -3,50 dan taraf uji 10% sebesar -3,18. Dari hasil uji Augmented Dickey-Fuller ini dapat disimpulkan data tidak stasioner hingga pada taraf 1%, 5% dan 10% yang berarti tolak  $H_0$ . Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan differencing, maka akan dilakukan differencing 1 kali. Setelah differencing satu kali, secara visual sudah terlihat data stasioner (Gambar 2) dan dari hasil uji Augmented Dickey-Fuller terlihat bahwa nilai tes statistic adalah -5.6269 dan nilai critical tau 1 untuk 1% sebesar -2,62, taraf uji 5% sebesar -1,95 dan taraf uji 10% sebesar -1,61% sehingga dapat disimpulkan data sudah stasioner pada taraf 1%, 5% dan 10%.



Gambar 2. Plot Produksi Cengkeh Tahun 1983-2019 dengan Differencing 1 kali

#### Hasil Uji Augmented Dickey-Fuller Differencing 1

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
Residuals:
    Min       1Q   Median       3Q      Max
-24685  -3796   5327  12691  24083

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
z.lag.1      -1.5147     0.2692  -5.627 3.21e-06 ***
z.diff.lag    0.2319     0.1711   1.355  0.185
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

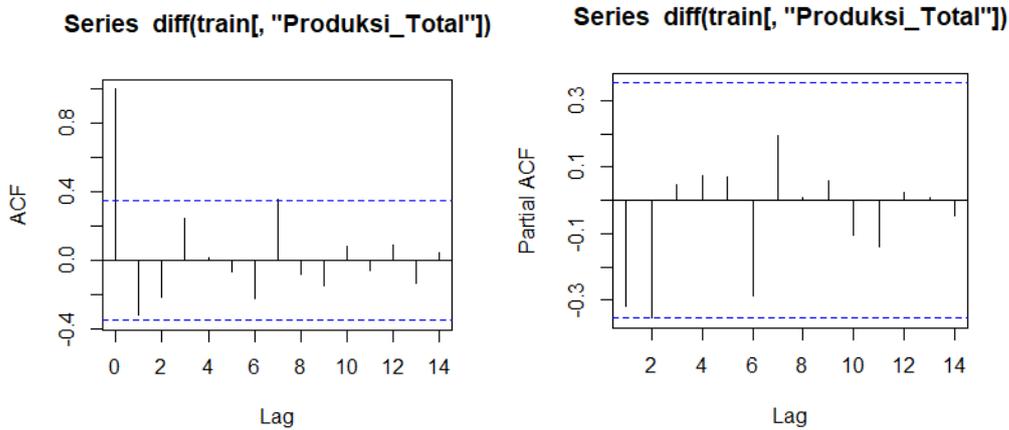
Residual standard error: 13530 on 32 degrees of freedom
Multiple R-squared:  0.6364, Adjusted R-squared:  0.6137
F-statistic: 28 on 2 and 32 DF,  p-value: 9.342e-08

Value of test-statistic is: -5.6269
Critical values for test statistics:
    1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

#### c. Identifikasi Model

Pada tahap ini akan ditentukan nilai p,d dan q dimana penentuan p dan q dapat dilakukan dengan beberapa cara yaitu dengan melihat korelogram autokorelasi (ACF) dan korelogram autokorelasi parsial (PACF), autoarima serta ar maselect. Sebelum menentukan nilai p,d,dan q terlebih dahulu data produksi cengkeh dibagi menjadi data training yaitu data dari tahun 1983-2013 dan data testing mulai tahun 2014-2019. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk validasi dan pengepasan model. Dari gambar plot ACF pada Gambar

1 menunjukkan nyata pada lag=0 , sementara untuk lag=1 dan seterusnya berada dibawah garis selang kepercayaan dan tidak menunjukkan pola *cut off* maupun *tail off* sehingga agak kesulitan untuk mengidentifikasi model ARIMA nya.



Gambar 3. Plot ACF dan PACF Produksi Cengkeh

Apabila terjadi kesulitan mengidentifikasi model tentatif ARIMA menggunakan plot ACF dan PACFnya maka dapat digunakan *autoarima*. Hasil *autoarima* adalah ARIMA (0,1,1) dengan Standard error model sebesar 0,1744 dan nilai MAPE data training sebesar 14,35475% seperti terlihat pada tabel dibawah ini.

```
Series: train[, "Produksi_Total"]
ARIMA (0,1,1) with drift
Coefficients:
      ma1      drift
    -0.5038  2234.673
s.e.  0.1744  1187.244
sigma^2 estimated as 171248635:  log likelihood=-336.96
AIC=679.92  AICC=680.81  BIC=684.22
Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
ACF1
Training set 92.08421 12457.69 10180.13 -2.311454 14.35475 0.8392568
0.01689235
```

Selain cara diatas dapat juga dijalankan dengan perintah *armaselect*, maka R-Studio akan menunjukkan sepuluh (10) model tentative yang diurutkan berdasarkan nilai SBC terkecil. Sepuluh model tentative yang ditampilkan berdasarkan *armaselect* yaitu:

	p	q	sbc
[1,]	5	5	501.9701
[2,]	4	5	541.0991
[3,]	2	1	582.5727
[4,]	3	1	583.3990
[5,]	1	2	584.6655
[6,]	4	1	585.0510
[7,]	4	3	585.3204
[8,]	1	1	585.6550
[9,]	1	3	586.0668
[10,]	4	2	586.3882

Dari hasil model autoarima dan 10 model tentative hasil armaselect dicobakan satu per satu dengan melihat mape terkecil dan kelayakan modelnya serta kelogisan hasil ramalan datanya. Setelah dicobakan satu persatu maka mape yang terkecil adalah ARIMA (1,1,2)

```
Call:
arima(x = train[, "Produksi_Total"], order = c(1, 1, 2))

Coefficients:
      ar1      ma1      ma2
    -0.6601  0.5310 -0.4690
s.e.   0.1705  0.1905  0.1746

sigma^2 estimated as 149054201:  log likelihood = -336.71,  aic = 681.42
> coeftest(model3)

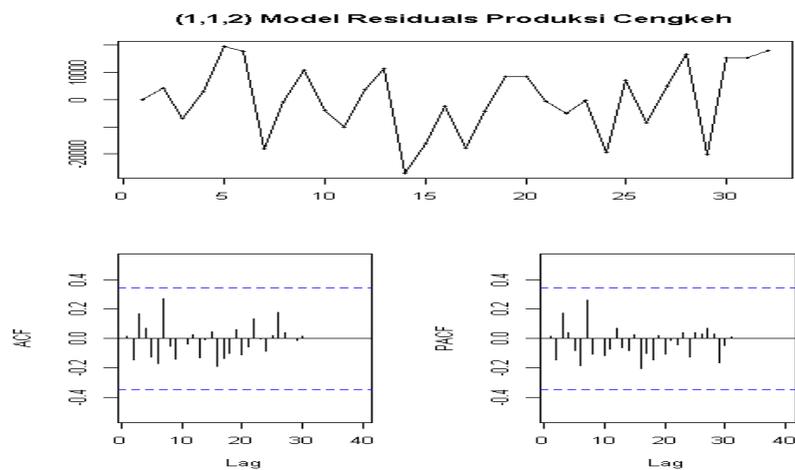
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1 -0.66014    0.17047  -3.8725 0.0001077 ***
ma1  0.53101    0.19049   2.7876 0.0053097 **
ma2 -0.46897    0.17458  -2.6863 0.0072255 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Dari Hasil penelusuran model ARIMA (1,1,2) diatas menunjukkan parameter ar 1, ma 1 dan ma 2 adalah nyata.

**d. Pemeriksaan sisaan**

Suatu model layak dipakai apabila sisaan terdistribusi normal. Pemeriksaan sisaan bisa dilakukan dengan plot sisaan dan Uji Ljung-Box. Hasil plot plot sisaan adalah sebagai berikut



Gambar 4. Plot Sisaan Model Arima (1,1,2)

Hasil plot sisaan model ARIMA (1,1,2) menunjukkan bahwa plot ACF dan PACF sudah tidak menunjukkan pola *cut off* atau *tail off* yang merujuk ke model ARIMA tertentu sehingga model ARIMA (1,1,2) sudah cukup layak. Berdasarkan hasil Uji

*Ljung-Box* kan mengindikasikan autokorelasi sisaan tidak signifikan hingga 30 lag. Oleh karenanya model ARIMA(1,1,2) sudah cukup baik mengepas data produksi cengkeh Indonesia.

lags	statistic	df	p-value
5	2.669662	5	0.7507545
10	8.143653	10	0.6148074
15	9.381370	15	0.8567478
20	15.448294	20	0.7502074
25	18.888783	25	0.8024217
30	24.958626	30	0.7270401

**e. Peramalan set data testing dan training dengan model ARIMA (1,1,2)**

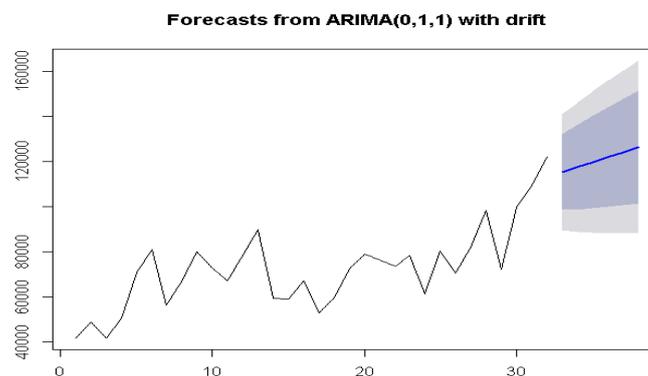
Hasil peramalan data testing selama lima tahun kedepan (tahun 2020 – 2024) tersaji di bawah ini. Nilai MAPE yang yang diperoleh yaitu 11,65% untuk data set testing dan 14,35% untuk data training seperti dibawah ini

Time Series:  
 Start = 33  
 End = 38  
 Frequency = 1

[1] 115343.7 117578.3 119813.0 122047.7 124282.4 126517.0

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	92.08421	12457.69	10180.13	-2.311454	14.35475	0.8392568	0.01689235
Test set	15957.59370	17100.54	15957.59	11.654473	11.65447	1.3155547	NA

Plot data dari tahun 1983-2013 (set data training) dan garis berwarna biru merupakan hasil peramalan data set testing (2014-2019), yang dilengkapi dengan arsiran selang kepercayaannya (daerah biru muda).



Gambar 5. Plot Ramalan Produksi Cengkeh Data Testing

**f. Pengepasan model untuk seluruh data**

Dengan model ARIMA (1,1,2) dilakukan pengepasan seluruh data dengan hasil sebagai berikut:

```
Call:
  arima(x = data_produksi[, "Produksi_Total"], order = c(1, 1, 2))

Coefficients:
```

```

      ar1      ma1      ma2
-0.6403  0.6233 -0.3767
s.e.    0.1566  0.1689  0.1537

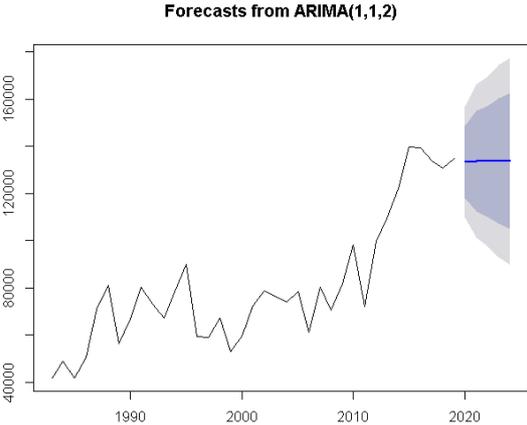
sigma^2 estimated as 140805058:  log likelihood = -389.89,  aic = 787.77
Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
ACF1
Training set 3260.297 11704.68 9631.251 2.519568 12.85771 0.8541287 -
0.1477134

```

Dengan menggunakan model ARIMA (1,1,2) diperoleh nilai mape yaitu sebesar 12,86% dan kemudian dilakukan peramalan untuk tahun 2020-2024. Hasil peramalan dengan ARIMA (1,1,2) adalah sebagai berikut:

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	133529.6	118138.7	148920.5	109991.24	157068.0
2021	133893.8	112569.4	155218.2	101280.96	166506.7
2022	133660.6	110313.4	157007.8	97954.20	169367.0
2023	133809.9	107150.1	160469.7	93037.29	174582.6
2024	133714.3	104981.2	162447.4	89770.84	177657.8

Hasil estimasi data produksi cengkeh nasional untuk 5 tahun kedepan dari pengepasan model arima (1,1,2) adalah, pada tahun 2020 sebesar 133.529 ton, tahun 2021 sebesar 133.893 ton, tahun 2022 sebesar 133.660 ton, tahun 2023 sebesar 133.809 ton dan tahun 2024 sebesar 133.714 ton. Untuk plot estimasi produksi cengkeh tahun 2020-2024 seperti gambar 6.



Gambar 6. Plot Hasil Estimasi Tahun 2020-2024

g. Hasil Estimasi

Setelah dilakukan beberapa tahapan penelusuran model maka model ARIMA yang dipilih adalah ARIMA (1.1.2) dengan nilai AIC 787,77 dengan mape 12,85. Hasil estimasi produksi cengkeh tahun 2020-2024 adalah sebagai berikut:

Tabel 3.1. Hasil Estimasi Produksi Cengkeh 2020-2024, Model ARIMA(1,1,2)

Tahun	Produksi (ton)	Pertumb. (%)
2019	134.792	
2020	133.529	-0,94
2021	133.893	0,27
2022	133.660	-0,17
2023	133.809	0,11
2024	133.714	-0,07

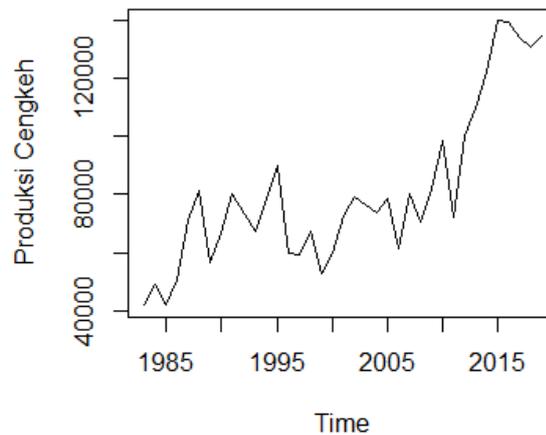
Keterangan : Tahun 2019 Angka Sementara Ditjen Perkebunan

## IV.2. Model Regresi

Dalam kajian ini, regresi yang akan dilakukan adalah regresi linier dan regresi kuadrat

### a. Regresi Sederhana

Series data yang digunakan untuk regresi sederhana dan kwadratik dimulai dari tahun 1983-2019. Plot data series produksi cengkeh tahun 1983-2019



Gambar 7. Plot Produksi Cengkeh Tahun 1983-2019

### b. Pembagian Data Testing dan Training

Data dibagi menjadi 2 bagian yaitu data training dimulai dari tahun 1983-2013 dan data testing dimulai dari tahun 2014-2019. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk validasi model.

### c. Pendugaan Model

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dengan produksi cengkeh. Data input yang digunakan untuk pendugaan model adalah data produksi cengkeh dari tahun 1983-2019. Hasil pendugaan model adalah sebagai berikut

```

Call:
lm(formula = Produksi_Total ~ Tahun, data = train)
Residuals:
    Min     1Q   Median     3Q      Max
-21262 -9837 -3379  9508 28738

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2607054.3  511390.1  -5.098 1.77e-05 ***
Tahun              1340.8    255.9   5.240 1.18e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13360 on 30 degrees of freedom
Multiple R-squared:  0.4779,    Adjusted R-squared:  0.4605
F-statistic: 27.46 on 1 and 30 DF,  p-value: 1.183e-05

```

Dari hasil pendugaan model diatas dijelaskan bahwa tahun signifikan terhadap produksi cengkeh. Nilai estimatenya adalah 1340,8 yang artinya ada kenaikan produksi cengkeh rata-rata 1340,8 ton per tahun. Model ini kurang bagus dimana  $R^2$  Adjustednya sebesar 46,08%. Dari model diatas dapat dibuat model produksinya yaitu  $Produksi = -2607054.3 * 1340.8Tahun$

Mape Data Training, Data Testing dan Plot

Mape data training adalah 15.27 sementara mape data testing adalah 28,18 seperti dibawah ini:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.155	6.862	13.836	15.277	23.591	38.525
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
24.62	25.74	27.19	28.18	31.18	32.16

### Evaluasi Asumsi

- Uji kenormalan dengan Jarque Bera (JB) yaitu salah satu uji normalitas jenis *gonness of fit test* yang mana mengukur apakah skewness dan kurtosis [sampel](#) sesuai dengan distribusi normal. Hasil uji JB= 1,5113, p-value=0,289 yaitu data terdistribusi normal dimana p-value > 5% (Ho ditolak)
- Uji Durbin Watson (DW) yaitu untuk mengetahui ada tidaknya autokoreasi. DW = 1,140429, p-value = 0,012 yaitu data ada autokorealsi antar residual (Ho ditolak)

Uji Non Constant Varians Score Test (NCVT) yaitu untuk melihat kehomogenan data.  $chisquare = 0,9524303$ ,  $Df = 1$ ,  $p = 0,3291$  yaitu data homogeny

Hasil Estimasi produksi cengkeh

Tabel 3.2. Hasil Estimasi Produksi Cengkeh Tahun 2020-2024

Tahun	Produksi (ton)	Pertumb. (%)
2019	134.792	
2020	101.441	-24,74
2021	102.782	1,32
2022	104.123	1,30
2023	105.463	1,29
2024	106.804	1,27

Keterangan : 2019 ASEM Ditjen Bun  
2020-2024 imasi Pusdatin

### Regresi kwadratik

#### f. Pendugaan Model

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dengan produksi cengkeh. Data input yang digunakan untuk pendugaan model adalah data produksi cengkeh dari tahun 1983-2019. Hasil pendugan model adalah sebagai berikut

Call:				
lm(formula = Produksi_Total ~ t + t2, data = train)				
Residuals:				
Min	1Q	Median	3Q	Max
-20142	-8432	-1402	9125	24904
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	57812.96	6603.60	8.755	1.23e-09 ***
t	144.32	985.96	0.146	0.885
t2	38.60	30.74	1.256	0.219
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 13240 on 29 degrees of freedom				
Multiple R-squared: 0.5048, Adjusted R-squared: 0.4707				
F-statistic: 14.78 on 2 and 29 DF, p-value: 3.752e-05				

Dari hasil pendugaan model diatas dijelaskan bahwa tahun tidak signifikan terhadap produksi cengkeh. Nilai estimatonya adalah 57812,96 yang artinya ada kenaikan produksi karet rata-rata 57812,96 ton per tahun. Model ini kurang bagus dimana R<sup>2</sup> Adjustednya sebesar 47%. Dari model diatas dapat dibuat model produksinya yaitu  $Produksi = 57812,96x144,32Tahun+38,60Tahun$

g. Mape Data data testing adalah sebagai berikut:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
15.93	16.	19.77	20.78	25.07	26.99

#### Evaluasi Asumsi

- Uji kenormalan dengan Jarque Bera (JB) = 1,299, p-value=0,3565 yaitu data terdistribusi normal dimana p-value > 5% (Ho ditolak)
- Uji Durbin Watson (DW) = 1,299, p-value = 0,008 yaitu data ada autokorealsi antar residual (Ho ditolak)
- Uji Non Constant Varians Score Test (NCVT) = 0,0311, Df =1, p=0,8554 yaitu data homogeny

#### Hasil Estimasi

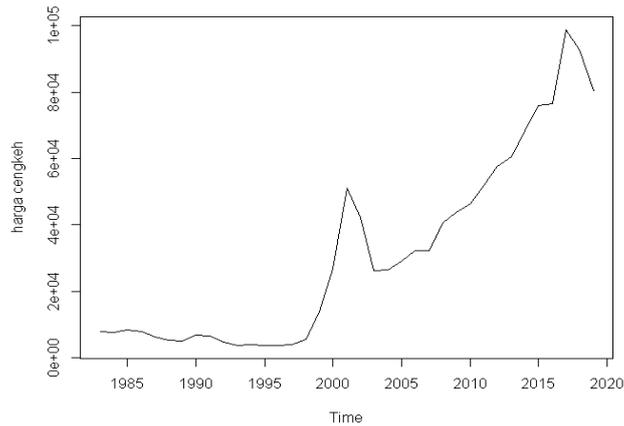
Tabel 3.2. Hasil Estimasi Produksi Cengkeh Tahun 2020-2024

Tahun	Produksi (ton)	Pertumb. (%)
2019	134.792	
2020	165.565	22,83
2021	169.684	2,49
2022	173.882	2,47
2023	178.156	2,46
2024	182.507	2,44

Keterangan: Tahun 2019 Angka Estimasi Ditjen Perkebunan  
Tahun 2020-2024: Angka Estimasi

### IV. 3. Model Fungsi Transfer

Seperti sudah dijelaskan diatas bahwa model fungsi transfer merupakan gabungan dari karakteristik analisis regresi berganda dengan karakteristik deret berkala ARIMA (*Autoregressive Integrated Moving Average*). Untuk model fungsi transfer ini menggunakan data harga untuk varibel inputnya. Plot data harga cengkeh.



Gambar. Plot data harga Cengkeh tahun 1983-2019

Dari plot data harga cengkeh terlihat bahwa adanya peningkatan dari tahun ke tahun sehingga terindikasi data tidak stasioner. Untuk pengujiannya dapat dilakukan dengan uji Augment Dickey-Fuller. Hasil Augmen Dickey Fuller adalah sebagai berikut:

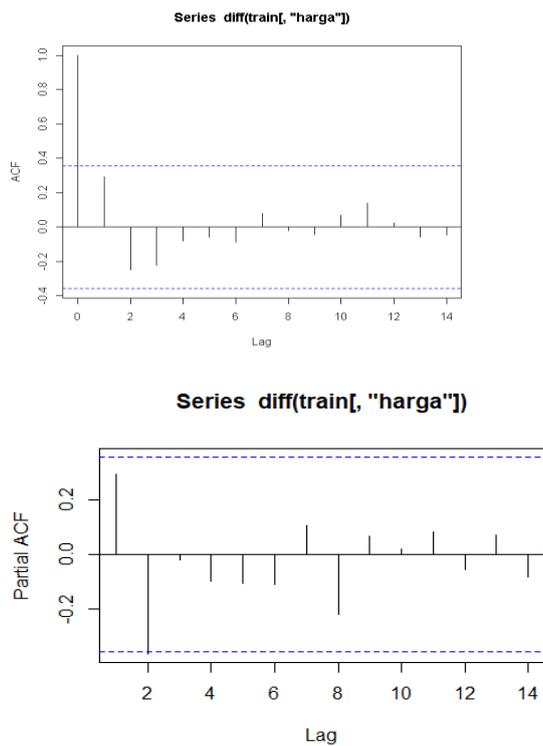
```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
    Min     1Q   Median     3Q    Max
-12092.3 -2864.2  296.1  3140.8 18435.3
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3354.6614  2541.3747  -1.320  0.19879
z.lag.1      -0.3464   0.1228  -2.819  0.00928 **
tt           728.1875  251.1117   2.900  0.00767 **
z.diff.lag   0.4279   0.1792   2.388  0.02484 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 5859 on 25 degrees of freedom
Multiple R-squared:  0.3257,    Adjusted R-squared:  0.2448
F-statistic: 4.026 on 3 and 25 DF, p-value: 0.01822
Value of test-statistic is: -2.8194 3.4211 4.4387
Critical values for test statistics:
    1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Dari hasil uji Augmented Dickey-Fuller terlihat bahwa data tidak stasioner karena nilai *tes statistic* lebih besar dari nilai *critical* dimana nilai *tes statistic nya* adalah -2,8194 dan nilai *critical tau 3* untuk 1% sebesar -4,15, taraf uji 5% sebesar -3,50 dan taraf uji 10% sebesar -3,18. Dari hasil uji Augmented Dickey-Fuller ini dapat disimpulkan data tidak stasioner hingga pada taraf 1%, 5% dan 10% yang berarti tolak  $H_0$  sehingga dilakukan differencing 1. Hasil uji Augmented Dickey-Fuller Diff 1 adalah sebagai berikut:

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
Residuals:
    Min     1Q   Median     3Q    Max
-15671.8 -1012.8  -31.1  4121.8 21398.0
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1   -0.8789   0.2153  -4.082 0.000377 ***
z.diff.lag  0.3297   0.1876   1.757 0.090651 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 6448 on 26 degrees of freedom
Multiple R-squared:  0.399, Adjusted R-squared:  0.3528
F-statistic: 8.63 on 2 and 26 DF, p-value: 0.001335
Value of test-statistic is: -4.0821
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Dari hasil uji Augmented Dickey-Fuller setelah di differencing 1 kali terlihat bahwa nilai *tes statistic* sudah lebih kecil dibandingkan nilai *critical*nya dimana nilai *tes statistic* adalah -4,0821 sementara nilai *critical tau 1* untuk 1% sebesar -2,62, taraf uji 5% sebesar -1,95 dan taraf uji 10% sebesar -1,61%. Hal ini dapat disimpulkan data tidak sudah stasioner hingga pada taraf uji 1%, 5% dan 10%.

Stasioneritas data dapat juga dilihat dari plot ACF dan PACF. Jika lag pada ACF maupun PACF banyak yang keluar dari garis *confidence interval* maka data tidak stasioner. Berikut adalah plot ACF dan PACF harga cengkeh setelah differencing 1.



Gambar. Plot ACF dan PACF Harga Cengkeh

Berdasarkan plot di atas, terlihat bahwa plot ACF cenderung cut off dimana hanya lag 0 yang keluar dari *confidence interval*. Plot PACF menunjukkan cut off pada lag 2, ditunjukkan berada di luar garis *confidence interval*. Hal tersebut memperkuat hasil uji ADF yaitu data harga cengkeh telah stasioner.

- **Pencarian Model Tentatif Variabel Input**

Untuk mencari model tentative variable harga, dapat dilakukan dengan *autoarima* maupun *arma selec*. Data yang digunakan untuk memilih model terbaik adalah data trainingnya. Hasil dari *autoarima* adalah sebagai berikut:

```
Series: train[, "harga"]
ARIMA(0,1,1)
Coefficients:
    ma1
    0.5113
s.e. 0.1733
sigma^2 estimated as 38183820: log likelihood=-304.08
AIC=612.16 AICc=612.6 BIC=614.96
Training set error measures:
      ME      RMSE     MAE      MPE      MAPE      MASE      ACF1
Training set 1135.225 5976.65 3397.916 3.092626 15.18112 0.867686 -0.06233682
```

Hasil *autoarima* untuk harga cengkeh adalah ARIMA (0,1,1) dengan Standard error sebesar 0,1733 dan nilai MAPE data training sebesar 15.18112% seperti terlihat pada tabel diatas. Selanjutnya dilakukan pengujian *coeftest* pada model ARIMA(0,1,1) untuk mengetahui signifikansi dari koefisien MA pada model tersebut, Output *coeftest*nya adalah

```
z test of coefficients:
  Estimate Std. Error z value Pr(>|z|)
ma1 0.51133 0.17328 2.9508 0.003169 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

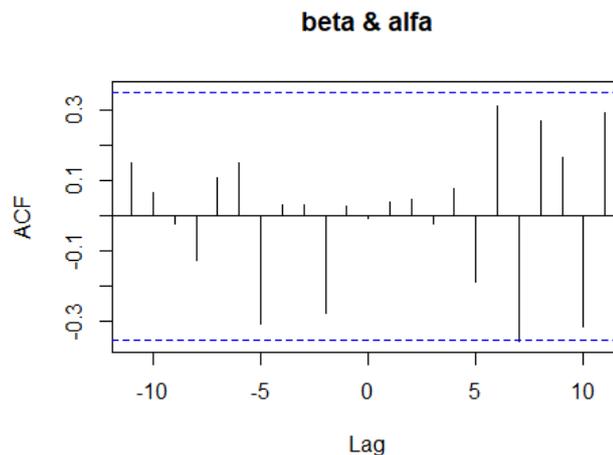
Berdasarkan output di atas terlihat bahwa Ma1 signifikan, pada taraf 1% Selain menggunakan *script auto.arima* dapat juga dijalankan dengan perintah *armaselect* dari ordo p=1 sd 5 dan q= 1sd 5 maka R-Studio akan menunjukkan sepuluh (10) model tentative yang diurutkan berdasarkan nilai AIC terkecil. Sepuluh model tentative yang ditampilkan adalah sebagai berikut:

```
      p q      sbc
[1,] 5 5 264.8169
[2,] 1 5 477.0809
[3,] 2 5 480.6898
[4,] 4 5 482.5022
[5,] 0 5 483.2223
[6,] 3 5 484.4578
[7,] 4 4 495.3675
[8,] 5 4 496.8568
[9,] 3 4 499.4510
[10,] 2 4 504.1149
```

Untuk mengetahui model terbaik dari 10 model tentative *auto.arima* dan model *armaselect* dapat diuji cobakan dengan membandingkan nilai AIC. Nilai AIC terkecil merupakan model terbaik. Setelah semua diuji cobakan maka nilai AIC terkecil adalah ARMA (0,1,1)

• **Prewhitening dan korelasi silang**

Korelasi silang menggambarkan struktur hubungan antara  $X_t$  dengan  $Y_t$ . Untuk mengidentifikasi pengaruh  $X_t$  terhadap  $Y_t$  maka deret  $X_t$  harus stasioner atau sudah distasionerkan. Dalam konteks pemodelan  $X_t$  terhadap  $Y_t$ , untuk membuat  $X_t$  stasioner tidak dengan pembedaan (*differencing*) namun dengan mengambil komponen *white noise* dari  $X_t$  (*prewhitening*). *Prewhitening* dilakukan terhadap deret input  $X_t$  yang didefinisikan sebagai alfa serta deret input  $Y_t$  yang didefinisikan sebagai beta. Hasil grafic ACF untuk beta dan alfa.



Gambar ACF untuk Beta dan alfa

Berdasarkan grafik ACF di atas, terlihat bahwa tidak ada yang keluar dari garis signifikansi. Karena lag signifikan pertama kali pada lag 0 maka nilai  $b=0$ . Selain itu, tidak ada tambahan lagi lag yang signifikan maka nilai  $s=0$ . Mengingat data produksi cengkeh dan harga cengkeh merupakan data tahunan yang tidak mengandung musiman maka diasumsikan nilai  $r=0$ . Nilai  $b=0$  menunjukkan tidak ada jeda pengaruh antara harga pada waktu  $t$  terhadap produksi pada waktu  $t$ . Nilai  $s=0$  berarti ada korelasi antara produksi dan harga cengkeh pada tahun yang sama. Dengan kata lain, dampak dari harga cengkeh terhadap produksi dirasakan pada waktu yang sama ( $t$ ).

- **Pengepasan Model**

Pengepasan model dilakukan untuk mencari residual model noisenya. Sehingga order ar dan ma nya masih (0,0,0). Hasil pengepasan model sebagai berikut:

```
Series: train[, "Produksi_Total"]
Regression with ARIMA(0,0,0) errors
Coefficients:
  intercept  xreg
  60638.170  0.4791
          s.e.    3542.986  0.1234
sigma^2 estimated as 179350971: log likelihood=-337.53
AIC=681.06  AICc=681.95  BIC=685.36
Training set error measures:
      ME  RMSE  MAE  MPE  MAPE  MASE  ACF1
Training set -4.459916e-12 12952.99 10792.34 -3.77367 16.51153 0.8904838 0.3164029
```

- Identifikasi Model Noise

Identifikasi model noise dilakukan untuk menentukan model ARIMA terbaik pada data residual hasil pengepasan model.. Arima yang direkomendasikan adalah (0,0,0):

```

Series: res
ARIMA(0,0,0) with zero mean

sigma^2 estimated as 102692901: log likelihood=-329.92
AIC=661.84 AICc=661.98 BIC=663.27

Training set error measures:
      ME  RMSE  MAE      MPE  MAPE  MASE  ACF1
Training set 321.8076    10133.75  8560.288  100  100  0.7672982  -
0.0009293487

```

Berdasarkan data residual, model ARIMA yang direkomendasikan oleh fungsi *auto.arima* yaitu ARIMA(0,0,0), namun model ini menghasilkan MAPE yang cukup besar yaitu 100%. Maka akan dicari model alternatif dengan dengan *armaselect*. Ada 10 model tentative terbaik yang ditampilkan hasil dari *armaselect* yaitu sebagai berikut:

```

      p q      sbc
[1,] 5 5 501.3197
[2,] 4 5 505.5649
[3,] 5 4 539.9377
[4,] 3 5 541.2054
[5,] 2 5 544.5020
[6,] 2 4 557.8965
[7,] 3 4 559.1056
[8,] 4 4 560.2821
[9,] 1 4 562.9512
[10,] 1 5 564.8626

```

Model alternative pertama yang disarankan untuk data residual adalah ARIMA(5,1,5). Kemudian di ujikan semua model yang direkomendasikan. Model yang dipakai adalah model ARIMA (1,1,4)

Output untuk model residual ARIMA (1,1,4)

```

modelres
Series: res
ARIMA(1,1,4)

Coefficients:
      ar1      ma1      ma2      ma3      ma4
      0.4836 -1.0105  0.0001  1.0102 -0.9997
s.e.  0.1793  0.2754  0.4331  0.3794  0.3454

sigma^2 estimated as 111911008: log likelihood=-322.14

AIC=656.28 AICc=659.93 BIC=664.68
> coeftest(modelres)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.4835611  0.1792812  2.6972 0.0069921 **
ma1 -1.0105164  0.2754057 -3.6692 0.0002433 ***

```

```

ma2  0.0001403  0.4331091  0.0003  0.9997415
ma3  1.0101849  0.3794309  2.6624  0.0077593  **
ma4  -0.9997162  0.3454042  -2.8943  0.0037996  **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Berdasarkan output di atas, maka model residual untuk ARIMA(1,1,4) signifikan pada ar1, ma1, ma 3 dan ma 4 tapi tidak signifikan si ma 2.

- **Pengepasan model dengan noise**

Pengepasan model dengan noise dilakukan dengan menggabungkan nilai r,s,b dan hasil identifikasi model noise yaitu ARIMA(5,1,5). Setelah dihasilkan model noisenya maka dilakukan pengepasan dengan arima output yaitu ARIMA (1,1,2).

- **Peramalan Fungsi Transfer dengan r,s,b (0,0,0) dan model noise (5,1,5)**

Berdasarkan model fungsi transfer dengan noise ARIMA (5,1,5), dilakukan peramalan berbasis nilai aktual dimana produksi cengkeh diestimasi menggunakan data aktual harga cengkeh periode 2014-2020 dengan mape 2,483371 seperti berikut:

```

Series: test[, "Produksi_Total"]
Regression with ARIMA(5,1,5) errors
Coefficients:
ar1  ar2  ar3  ar4  ar5  ma1  ma2  ma3  ma4  ma5
xreg
-0.7757 -0.2252 0.493  0.7272  0.7548  0.487 -0.1286  0.1667 -
0.4394 -0.9757 0.0979
s.e.  0.0000  0.0000  0.000  0.0000  0.0000  0.0000  0.000  0.0000
0.0000  0.0000  0.0000  0.0000

sigma^2 estimated as 113226926:  log likelihood=-52.15
AIC=106.3  AICC=107.63  BIC=105.91

Training set error measures:
MASE      ACF1      ME      RMSE      MAE      MPE      MAPE
Training set 1027.342 4881.245 3391.185 0.6850798 2.483371
0.5668602 0.2400872

```

Untuk membandingkan ketepatan model ramalan, dilakukan perbandingan hasil ramalan terhadap data aktual produksi cengkeh pada tahun 2014-2019. Hasil ramalan yang dibandingkan adalah ramalan dengan model ARIMA(1,1,2), fungsi transfer ARIMA(1,1,4) dimana input harga cengkeh yang digunakan adalah data aktual maupun ramalan. Hasil Estimasi produksi cengkeh 2020-2024

Berikut adalah hasil ramalan untuk lima tahun kedepan yaitu tahun 2020-2024

```

Time Series:
Start = 32
End = 36
Frequency = 1
[1] 136825.5 137825.0 137492.2 137603.0 137566.1
2019 = 134792

```

Tabel 3.2. Hasil Estimasi Produksi Cengkeh Tahun 2020-2024

Tahun	Produksi (Ton)	Pertumbuhan (%)
2019	134.792	
2020	136.825	1,51
2021	137.825	0,73
2022	137.492	-0,24
2023	137.603	0,08
2024	137.566	-0,03

#### IV.4. Model VAR

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen. Dalam pemodelan ini menggunakan empat variabel yaitu produksi cengkeh, luas areal cengkeh, harga cengkeh dan volume ekspor cengkeh. Data yang digunakan adalah dari series tahun 1983-2019. Untuk mencari model var terbaik maka dilakukan uji coba dari VAR lag p=1 sampai dengan lag p=6. dengan type both dan trend. Setelah dilakukan uji coba maka model terpilih adalah VAR lag 1 dengan type cons. Hasil dari VAR lag p=1 type cons adalah sebagai berikut:

##### VAR Estimation Results:

```
=====
Endogenous variables: Produksi_Total, Lareal_Total, harga, volek
Deterministic variables: const
Sample size: 30
Log Likelihood: -1251.825
Roots of the characteristic polynomial:
0.924 0.924 0.6103 0.07135
Call:
VAR(y = cengkeh_prod[2:32, c(2, 3, 4, 5)], p = 1, type = "const")
```

##### Estimation results for equation Produksi\_Total:

```
=====
Produksi_Total = Produksi_Total.l1 + Lareal_Total.l1 + harga.l1 +
volek.l1 + const
```

	Estimate	Std. Error	t value	Pr(> t )
Produksi_Total.l1	3.569e-01	1.709e-01	2.088	0.047111 *
Lareal_Total.l1	2.783e-02	2.768e-02	1.005	0.324400 *
harga.l1	5.965e-01	1.584e-01	3.766	0.000902 ***
volek.l1	-6.352e-01	5.180e-01	-1.226	0.231501
const	2.398e+04	1.932e+04	1.241	0.225977 *

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11590 on 25 degrees of freedom  
 Multiple R-Squared: 0.7117, Adjusted R-squared: 0.8496  
 F-statistic: 9.846 on 4 and 25 DF, p-value: 6.33e-05

Estimation results for equation Lareal\_Total:

=====  
 Lareal\_Total = Produksi\_Total.l1 + Lareal\_Total.l1 + harga.l1 +  
 volek.l1 + const

	Estimate	Std. Error	t value	Pr(> t )	
Produksi_Total.l1	-9.562e-01	3.443e-01	-2.777	0.01024	*
Lareal_Total.l1	1.044e+00	5.577e-02	18.724	3.18e-16	***
harga.l1	1.008e+00	3.191e-01	3.160	0.00409	**
volek.l1	6.821e-01	1.044e+00	0.654	0.51931	
const	1.652e+04	3.892e+04	0.424	0.67487	

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23360 on 25 degrees of freedom  
 Multiple R-Squared: 0.9562, Adjusted R-squared: 0.9491  
 F-statistic: 136.3 on 4 and 25 DF, p-value: < 2.2e-16

Estimation results for equation harga:

=====  
 harga = Produksi\_Total.l1 + Lareal\_Total.l1 + harga.l1 + volek.l1 +  
 const

	Estimate	Std. Error	t value	Pr(> t )	
Produksi_Total.l1	-2.511e-02	1.018e-01	-0.247	0.807	
Lareal_Total.l1	-2.279e-02	1.649e-02	-1.382	0.179	
harga.l1	9.718e-01	9.435e-02	10.301	1.76e-10	***
volek.l1	-4.785e-02	3.086e-01	-0.155	0.878	
const	1.673e+04	1.151e+04	1.454	0.159	

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6906 on 25 degrees of freedom  
 Multiple R-Squared: 0.9057, Adjusted R-squared: 0.8906  
 F-statistic: 60.01 on 4 and 25 DF, p-value: 1.877e-12

Estimation results for equation volek:

=====  
 volek = Produksi\_Total.l1 + Lareal\_Total.l1 + harga.l1 + volek.l1 +  
 const

	Estimate	Std. Error	t value	Pr(> t )	
Produksi_Total.l1	-6.063e-02	6.257e-02	-0.969	0.3418	
Lareal_Total.l1	-2.069e-02	1.014e-02	-2.042	0.0519	.
harga.l1	9.060e-02	5.799e-02	1.562	0.1308	
volek.l1	1.076e-02	1.897e-01	0.057	0.9552	
const	1.840e+04	7.073e+03	2.602	0.0154	*

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4245 on 25 degrees of freedom  
 Multiple R-Squared: 0.3958, Adjusted R-squared: 0.2991  
 F-statistic: 4.094 on 4 and 25 DF, p-value: 0.01095

Covariance matrix of residuals:

	Produksi_Total	Lareal_Total	harga	volek
Produksi_Total	134405054	-110625538	8872360	4301273
Lareal_Total	-110625538	545552607	17462609	-13681760
harga	8872360	17462609	47694490	-10665715
volek	4301273	-13681760	-10665715	18017856

Correlation matrix of residuals:

	Produksi_Total	Lareal_Total	harga	volek
Produksi_Total	1.00000	-0.4085	0.1108	0.08741
Lareal_Total	-0.40854	1.0000	0.1083	-0.13800
harga	0.11081	0.1083	1.0000	-0.36383
volek	0.08741	-0.1380	-0.3638	1.00000

Kemudian dilakukan uji asumsi untuk lag  $p=1$  type cons

- Pemeriksaan autokorelasi residual

Portmanteau Test (asymptotic)

data: Residuals of VAR object varhsheet1

Chi-squared = 172.26, df = 240, p-value = 0.9997

Nilai p-value lebih besar dari 0,05 maka sisaan saling bebas atau asumsi non autokorelasi terpenuhi.

pemeriksaan kenormalan dengan normality.test” yang dengan uji Jarque-Bera tests

JB-Test (multivariate)

data: Residuals of VAR object varhsheet1

Chi-squared = 56.877, df = 8, p-value = 1.903e-09

Skewness only (multivariate)

data: Residuals of VAR object varhsheet1

Chi-squared = 17.659, df = 4, p-value = 0.001439

\$Kurtosis

kurtosis only (multivariate)

data: Residuals of VAR object varhsheet1

Chi-squared = 39.218, df = 4, p-value = 6.279e-08

Nilai pvalue lebih kecil dari 0,05, namun karena jumlah data yang digunakan cukup banyak, maka series tersebut dapat dianggap normal.

- Pemeriksaan heteroskedastisitas dengan fungsi “arch.test”

ARCH (multivariate)

data: Residuals of VAR object varhsheet1

Chi-squared = 250, df = 500, p-value = 1

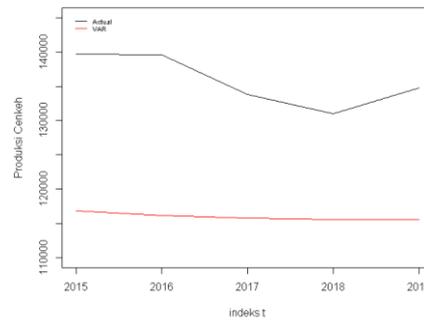
Nilai p-value lebih besar dari 0,05 maka ragam sisaan homogen atau asumsi homoskedastisitas terpenuhi.

- Mape data testing adalah sebagai berikut:

Mape data testing

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
1.74	13.47	14.23	14.51	16.33	16.77	1

Data testing hasil ramalan produksi cengkeh dengan nilai  $p=1$  MAPE =14,23%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data testing adalah 14,51%. Dari grafik di dibawah ini dapat dilihat bahwa pergerakan data var  $p=1$  tidak mengikuti pergerakan data testing.

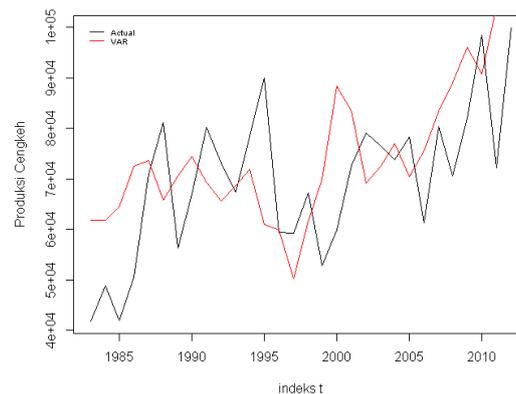


Plot Ramalan Dta Aktual dan Data Testing Var  $p=1$  type cons

- Mape data training adalah sebagai berikut

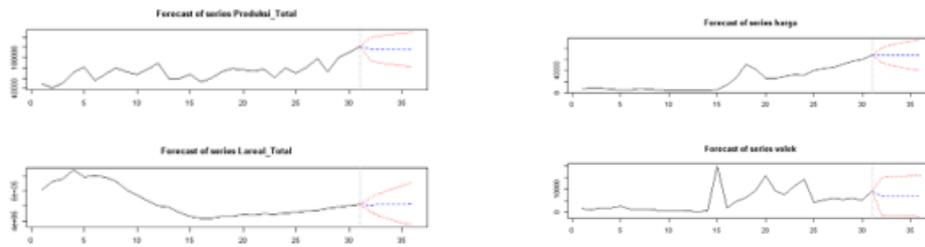
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.6985	8.2454	14.2054	19.2807	26.3376	53.6272

Data training hasil ramalan produksi cengkeh dengan nilai  $p=1$  MAPE =19,28%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data training adalah 19,28%. Dari grafik di dibawah ini dapat dilihat bahwa pergerakan data var  $p=1$  hampir mengikuti pergerakan data trainingnya



Plot ramalan dan Aktual data Testing Var  $p=1$  type cons

- Plot hasil ramalan produksi cengkeh, luas areal cengkeh, harga produsen cengkeh dan volume ekspor cengkeh dengan VAR  $p=1$  type cons



- Hasil ramalan produksi cengkeh tahun 2020-2024

Tahun	Produksi (Ton)	Pertumbuhan (%)
2019	134.792	
2020	126.835	-5,90
2021	126.199	-0,50
2022	125.779	-0,33
2023	125.633	-0,12
2024	125.606	-0,02

#### IV.5. Model Terbaik

##### a. Kriteria Pemilihan Model Terbaik

- Nilai MAPE Kecil
- Nilai AIC dan SIC kecil
- Uji asumsi terpenuhi

##### b. Model Terbaik

No	Model	Mape		Hasil estimasi Produksi Cengkeh (Ton)					Pertumb. 2020-2004 (%)
		training	testing	2020	2021	2022	2023	2024	
1	Arima	14,35	11,65	133.529	133.893	133.660	133.809	133.714	0,03
2	Regresi								
	- Linier sederhana	15,27	28,18	101.441	102.782	104.123	105.463	106.804	1,30
	- Kwadratik	15,27	20,78	165.565	169.684	173.882	178.156	182.507	2,47
3	Fungsi Transfer	2,48	2,83	136.825	137.825	137.492	137.603	137.566	0,14
4	VAR	9,28	4,51	126.835	126.199	125.779	125.633	125.606	-0,24

Dari table diatas di dapatkan model terbaik adalah Fungsi transfer

#### v. Kesimpulan dan Saran

Dari ke empat metode estimasi yang dikaji maka dapat disimpulkan bahwa model fungsi transfer adalah yang terbaik karena memiliki mape testing dan mape training terkecil dan angka estimasinya mendekati angka aktualnya. Mape training sebesar 2,48 persen dan mape testing sebesar 2,83 persen, sementara hasil estimasi tahun 2020 adalah 136.825 ton dan tahun 2019 (Angka Sementara) sebesar 134.792 ton.

# **KAJIAN MODEL ESTIMASI LUAS AREAL KARET NASIONAL PENDEKATAN MODEL REGRESI, ARIMA, FUNGSI TRANSFER DAN VAR (*Vector Auto Regressive*)**

**Mohammad Chafid**

*Center for Agricultural Data and Information System-Ministry of Agriculture  
Jalan Harsono RM No. 3 Gd D Lt 4, Ragunan-Jakarta Selatan, Indonesia  
E-mail: mohammad.hafidz1@gmail.com*

## **ABSTRAK**

Komoditas karet merupakan salah satu komoditas strategis penghasil devisa negara karena sebagian besar produksi karet nasional untuk diekspor. Status Angka statistik perkebunan terdiri dari Angka Tetap, Angka Sementara dan Angka Estimasi. Tujuan penulisan ini adalah mencari model alternatif lain untuk menyusun angka estimasi luas areal karet sehingga akurasi menjadi lebih baik yang ditandai dengan semakin kecilnya MAPE baik untuk data training maupun testing.

Model yang digunakan untuk menyusun angka estimasi luas areal karet meliputi, Metode Arima, Metode Regresi linier dan kuadratis dengan variabel bebas tahun, Metode Fungsi Transfer dengan peubah input harga karet sheet nasional, Metode VAR (*Vector Autoregressive*) dengan variabel luas areal, harga karet sheet dalam negeri, harga karet dunia, volume ekspor karet dan volume impor karet. Sumber data yang digunakan untuk variabel luas areal karet (1970 – 2019), harga karet nasional (1971-2019), volume ekspor dan impor karet (1970 – 2019) berasal dari Ditjen Perkebunan. Untuk variabel harga karet dunia (1971 - 2019) berasal dari World Bank. Run model menggunakan software RStudio.

Untuk analisis ini data dibagi menjadi 2 kelompok, yaitu data training tahun 1970 – 2013, dan data testing tahun 2014 – 2019. Data training untuk penyusunan model, sedangkan data testing untuk uji coba model dalam melakukan estimasi 6 tahun kedepan. Untuk estimasi luas areal karet alternatif model pertama adalah Model ARIMA. Model ARIMA terbaik adalah ARIMA (1,1,0), menghasilkan MAPE untuk data training 1,34%, dan MAPE data testing 0,69%. Model regresi linier menghasilkan MAPE data training sebesar 5,38% dan MAPE data testing 5,74%. Model regresi kuadratik menghasilkan MAPE data training sebesar 6,61%, MAPE data testing 4,25%. Untuk model yang ketiga dengan menggunakan Fungsi Transfer dengan variabel input harga karet sheet, menghasilkan MAPE untuk data training sebesar 0,23% dan MAPE data testing 0,19%. Untuk model yang keempat model VAR(2) type 'constant', menghasilkan MAPE data training 1,25% dan data MAPE data testing 3,63%.

Berdasarkan perbandingan besarnya MAPE baik data testing maupun data training dan hasil estimasi luas areal 5 tahun kedepan, maka model terbaik yang terpilih adalah model Fungsi Transfer ARIMA(1,1,1) dengan factor input harga karet sheet karena menghasilkan akurasi yang cukup tinggi sehingga MAPE rata-rata data testing sebesar 0,19%. Hasil estimasi luas areal karet nasional untuk model Fungsi Transfer ARIMA(1,1,1) dengan factor input harga karet sheet untuk tahun 2020 sebesar 3.694.629 hektar, tahun 2021 sebesar 3.703.988 hektar, tahun 2022 sebesar 3.712.086 hektar, tahun 2023 sebesar 3.719.032 hektar, dan tahun 2024 sebesar 3.725.004 hektar. Laju pertumbuhan estimasi luas areal karet nasional selama 5 tahun kedepan (2020 – 2024) rata-rata 0,21% per tahun.

***Kata Kunci : Luas areal Karet, Regresi, Arima, Fungsi Transfer, VAR (Vector Autoregressive)***

## PENDAHULUAN

Karet (*Hevea brasiliensis*) termasuk dalam genus *Hevea* dari familia *Euphorbiaceae*, yang merupakan pohon kayu tropis yang berasal dari hutan Amazon. Di dunia, setidaknya 2.500 spesies tanaman diakui dapat memproduksi lateks, tetapi *Hevea brasiliensis* saat ini merupakan satu-satunya sumber komersial produksi karet alam. Karet alam mewakili hampir separuh dari total produksi karet dunia karena sifat unik mekanik, seperti ketahanan sobek, dibandingkan dengan karet sintetis.

Data statistik perkebunan yang disajikan merupakan hasil sinkronisasi dan validasi yang dilaksanakan secara berjenjang dari tingkat kabupaten sampai tingkat nasional dengan mengacu pada Pedoman Pengelolaan Data Komoditas Perkebunan yang diterbitkan oleh Direktorat Jenderal Perkebunan. Secara periodik, data perkebunan disajikan dalam 3 status angka yaitu Angka Tetap (ATAP), Angka Sementara (ASEM) dan Angka Estimasi (AESTI). Angka Tetap merupakan angka hasil rekapitulasi dari pelaporan yang sudah tetap, sehingga tidak dilakukan estimasi, sedangkan untuk penentuan Angka Sementara dan Estimasi perlu dilakukan estimasi dengan metode estimasi yang paling relevan dan tepat (PDKP, 2013). Data Angka Tetap (ATAP) merupakan data 2 tahun yang lalu ( $n-2$ ), Angka Sementara (ASEM) merupakan data tahun lalu ( $n-1$ ), dan Angka Estimasi (AESTI) merupakan data tahun yang berjalan ( $n$ ).

Metode estimasi yang digunakan adalah Metode *Exponential Smoothing* (Peramalan Pemulusan Eksponensial) yang merupakan salah satu kategori metode time series yang menggunakan pembobotan data masa lalu secara eksponensial. Dalam kategori ini terdapat dua metode yang umum dipakai yaitu metode Pemulusan Eksponensial Tunggal (*Single Exponential Smoothing*) dan metode Pemulusan Eksponensial Ganda (*Double Exponential Smoothing*). Pemilihan model *Single Exponential Smoothing* atau *Double Exponential Smoothing* harus mempertimbangkan nilai *Mean Absolute Percentage Error (MAPE)* serta kerealistisan hasil estimasi bila dibandingkan dengan series data sebelumnya. MAPE adalah pengukur tingkat akurasi (ketepatan) nilai dugaan yang dihasilkan oleh model dalam bentuk presentase. Model yang mempunyai nilai MAPE lebih kecil dianggap sebagai model yang lebih baik. Keunggulan dari metode estimasi ini adalah dapat digunakan untuk meramalkan data yang berisi trend atau pola musiman. Namun metode estimasi ini juga memerlukan keahlian khusus dalam menginterpretasikan hasil estimasi yang diperoleh (PDKP, 2013).

Metode untuk menghasilkan angka estimasi (AESTI) yang diliris oleh Ditjen. Perkebunan pada waktu tahun berjalan ( $n$ ) perlu dikaji kembali, agar didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik dari metode sebelumnya, sehingga tingkat kesalahannya lebih kecil. Guna merumuskan kebijakan produksi, pasokan dan distribusi komoditas karet sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Metode ramalan luas areal karet dapat digunakan dengan menggunakan beberapa metode statistik yang biasa digunakan untuk melakukan peramalan.

Pada analisis ini akan dikaji metode ARIMA, Regresi, Fungsi Transfer dan VAR dalam melakukan pemodelan dan peramalan luas areal karet nasional.

Oleh karenanya, tujuan dari disusunnya analisis ini adalah:

- a. Melakukan analisis dan peramalan data luas areal karet nasional menggunakan model ARIMA, Regresi, Fungsi Transfer dan VAR.
- b. Menentukan metode terbaik dalam meramal data luas areal karet nasional.

## METODOLOGI

Data yang digunakan dalam melakukan pemodelan luas areal karet nasional adalah data series luas areal karet nasional tahun 1970 - 2019. Untuk analisis ini data dibagi menjadi 2 kelompok, yaitu data training tahun 1970 – 2013, dan data testing tahun 2014 – 2019. Data training untuk penyusunan model, sedangkan data testing untuk uji coba model dalam melakukan estimasi 6 tahun kedepan.

Peubah yang diasumsikan mempengaruhi besaran luas areal karet untuk Fungsi Transfer adalah harga karet sheet lokal. Harga karet lokal diasumsikan akan berpengaruh terhadap luas areal, karena jika terjadi kenaikan harga karet maka luas areal karet cenderung akan terus ditingkatkan, kondisi sebaliknya maka harga karet cenderung stagnan, pertumbuhan luas areal rendah, atau cenderung turun.

Peubah yang berpengaruh terhadap luas areal untuk pemodelan VAR adalah luas areal, harga karet domestik, harga karet dunia, volume ekspor dan impor karet. Harga karet dunia dan harga karet domestik diduga kuat berpengaruh pada tingkat luas areal. Untuk volume ekspor dan impor juga berpengaruh terhadap luas areal, dimana semakin tinggi permintaan ekspor karet maka luas areal cenderung akan semakin tinggi.

Peramalan data luas areal karet, dilakukan uji coba beberapa metode yakni mengkaji metode ARIMA, Regresi, Fungsi Transfer dan VAR (Vector Autoregression), menggunakan software *R Studio*.

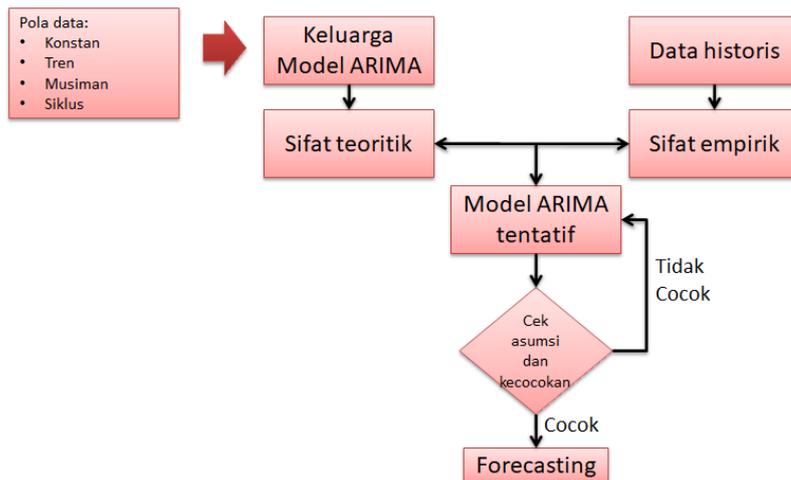
### a. ARIMA (*Autoregressive Integrated Moving Average*)

Metode Autoregressive Integrated Moving Average (ARIMA) atau biasa disebut juga sebagai metode Box-Jenkins merupakan metode yang secara intensif dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1970 (Iriawan, 2006).

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung flat (mendatar/konstan) untuk periode yang cukup panjang.

Model *Autoregresif Integrated Moving Average* (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (time series) secara statistik berhubungan satu sama lain (dependent).

Model ARIMA terdiri dari tiga langkah dasar, yaitu tahap identifikasi, tahap penaksiran dan pengujian, dan pemeriksaan diagnostik. Selanjutnya model ARIMA dapat digunakan untuk melakukan peramalan jika model yang diperoleh memadai.



Gambar 1. Prosedur Peramalan Model Arima (Box- Jenkins)

Hal yang perlu diperhatikan adalah bahwa kebanyakan deret berkala bersifat stasioner. Stasioner berarti tidak terdapat pertumbuhan atau penurunan pada data. Data secara kasarnya harus horizontal sepanjang sumbu waktu. Dengan kata lain, fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan varians dari fluktuasi tersebut pada pokoknya tetap konstan setiap waktu. Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan differencing. Yang dimaksud dengan differencing adalah menghitung perubahan atau selisih nilai observasi. Nilai selisih yang diperoleh dicek lagi apakah stasioner atau tidak. Jika belum stasioner maka dilakukan differencing lagi. Jika varians tidak stasioner, maka dilakukan transformasi logaritma.

Model Box-Jenkins (ARIMA) dibagi kedalam 3 kelompok, yaitu: model autoregressive (AR), moving average (MA), dan model campuran ARIMA (autoregressive moving average) yang mempunyai karakteristik dari dua model pertama. ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampainya. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $Y_{t-p}$  = data *time series* pada kurun waktu ke (t-p)
- $\mu$  = suatu konstanta
- $\theta_1, \theta_q, \phi_1, \phi_n$  = parameter-parameter model
- $\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

Model dikatakan baik jika nilai error bersifat random, artinya sudah tidak mempunyai pola tertentu lagi. Dengan kata lain model yang diperoleh dapat menangkap dengan baik pola data yang ada. Untuk melihat kerandoman nilai error dilakukan pengujian terhadap nilai koefisien autokorelasi dari error, dengan menggunakan salah satu dari dua statistik berikut, yaitu Uji Q-Box and Pierce dan uji Ljung-Box.

## b. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih peubah. Pada analisis regresi, dibedakan menjadi dua, yaitu peubah respon atau biasa juga disebut peubah bergantung (*dependent variable*) dan peubah *explanatory* atau biasa disebut penduga (*predictor variable*) atau disebut juga peubah bebas (*independent peubah*). Model regresi digunakan untuk meramalkan atau memprediksi nilai masa depan dari peubah respon berdasarkan nilai-nilai tertentu yang diberikan oleh peubah prediktor.

Asumsi yang mendasari penggunaan analisis regresi linier adalah:

1. Sisaan mengikuti fungsi distribusi normal;
2. Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas);
3. Tidak terdapat autokorelasi antara sisaan untuk setiap data pengamatan; dan
4. Tidak terdapat multikolinearitas antara peubah respon.

Model regresi linier sederhana melibatkan satu peubah prediktor dan diformulasikan dalam model matematis sebagai berikut:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

dimana:

- $y$  = peubah respon/dependen/terikat
- $x$  = peubah prediktor/independen/bebas
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi peubah bebas
- $\varepsilon$  = error/residu/sisaan

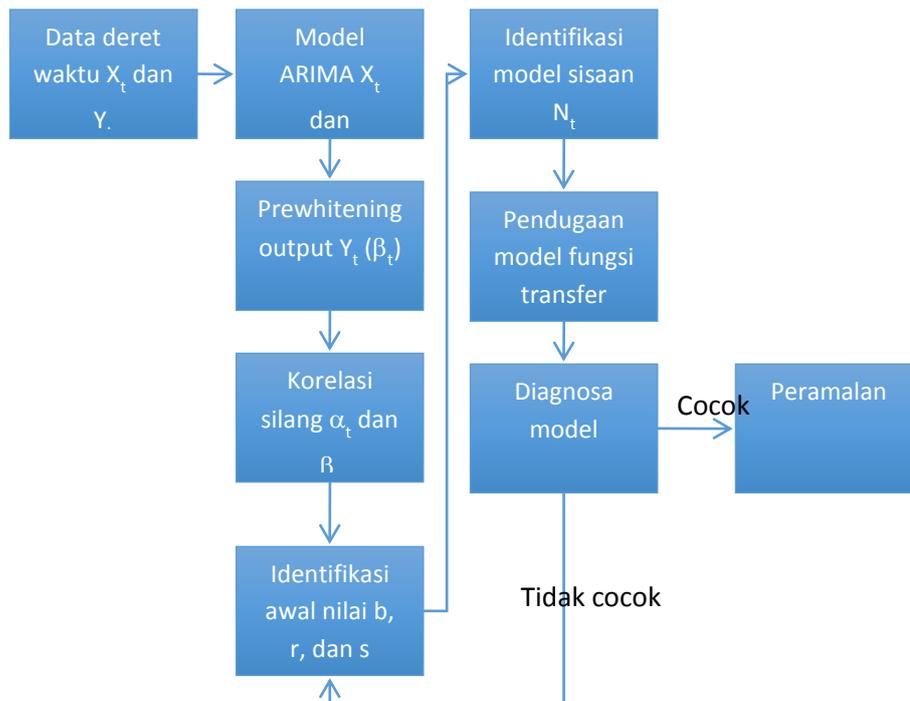
Parameter model atau koefisien regresi yaitu  $\beta_0$  dan  $\beta_1$  masing-masing memiliki interpretasi sebagai intercept dan slope dari suatu garis lurus.  $\beta_1$  mengukur perubahan rata-rata dari peubah respon  $y$  untuk setiap perubahan dari peubah prediktor  $x$ . Parameter ini biasanya tidak diketahui dan harus diprediksi dari data sampel.  $\varepsilon$  merupakan kesalahan statistik, didefinisikan sebagai sisa atau perbedaan hasil antara nilai data pengamatan peubah respon terhadap nilai peubah respon hasil prediksi.

## c. Model Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau  $Y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $Y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $X_t$ ) dengan deret output

tersebut. Model fungsi transfer merupakan fungsi dinamis yang pengaruhnya tidak hanya pada hubungan linier antara deret input dengan deret output pada waktu ke- $t$ , tetapi juga pada waktu  $t+1, t+2, \dots, t+k$ . Hubungan seperti ini pada fungsi transfer dapat menimbulkan delay (waktu senjang) antara peubah *input* dan peubah *output*.

Tujuan pemodelan fungsi transfer adalah untuk menetapkan model yang sederhana, yang menghubungkan deret output ( $Y_t$ ) dengan deret input ( $X_t$ ) dan gangguan/noise ( $n_t$ ). Wei (1994) juga menjelaskan bahwa di dalam fungsi transfer terdapat rangkaian output yang mungkin dipengaruhi oleh rangkaian multiple input. Pada kasus single input peubah, dapat menggunakan metode korelasi silang yang dianjurkan oleh Box and Jenkins (1976). Teknik ini juga dapat digunakan ketika terdapat single input peubah yang lebih dari satu selama antar variable input tidak berkorelasi silang. Jika beberapa atau semua peubah input berkorelasi silang maka teknik prewhitening atau metode korelasi silang tidak dapat digunakan secara langsung. Alasan utama bagi perlunya suatu perencanaan atau peramalan adalah adanya tenggang waktu pengambilan keputusan yang dapat berkisar dari beberapa hari atau sampai beberapa tahun. Pada analisis fungsi transfer untuk peramalan deret berkala univariate, terdapat deret berkala output yang diperkirakan dipengaruhi oleh deret berkala input dan input-input lain yang digabungkan dalam satu kelompok yang disebut gangguan (noise). Deret input mempengaruhi deret output melalui sebuah fungsi transfer yang mendistribusikan pengaruhnya secara dinamis melalui beberapa periode waktu yang akan datang dengan persentase tertentu yang disebut sebagai bobot respons impuls atau bobot fungsi transfer.



Gambar 2. Langkah-langkah melakukan pemodelan Fungsi Transfer

Model umum Fungsi Transfer:

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\varphi_p(B)} \varepsilon_t \quad \dots\dots\dots(7)$$

Dimana:

- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$
- $q \rightarrow$  ordo MA bagi noise  $N_t$

**d. Model Vector Autoregression (VAR)**

*Vector Autoregression* (VAR) dikemukakan pertama kali oleh Christopher Sims (1980). Sims mengembangkan model ekonometri dengan mengabaikan pengujian asumsi secara apriori. VAR dikembangkan oleh Sims sebagai kritik atas metode simultan. Jumlah peubah yang besar dan klasifikasi endogen dan eksogen pada metode simultan merupakan dasar dari kritik tersebut. Penggunaan pendekatan struktural atas pemodelan persamaan simultan biasanya menerapkan teori ekonomi di dalam usahanya untuk mendeskripsikan hubungan antar peubah yang ingin diuji. Akan tetapi sering ditemukan bahwa teori ekonomi saja ternyata tidak cukup kaya di dalam menyediakan spesifikasi yang ketat dan tepat atas hubungan dinamis antar peubah (Gujarati, 2010). Model VAR merupakan jalan keluar atas permasalahan ini melalui pendekatan non-struktural atau tidak mendasarkan pada teori ekonomi tertentu (*atheoretical*). Metode VAR memperlakukan seluruh peubah secara simetris tanpa memperlakukan peubah dependen dan independen atau dengan kata lain model ini memperlakukan seluruh peubah sebagai peubah endogen., karena pada kenyataannya suatu peubah dapat bertindak sebagai endogen atau dependen dari suatu peubah yang lainnya (Gujarati, 2010).

Kelebihan dalam penggunaan metode VAR (Gujarati, 2010):

- a. Kemudahan dalam penggunaan, tidak perlu mengkhawatirkan tentang penentuan peubah endogen dan peubah eksogen.
- b. Kemudahan dalam estimasi, metode *Ordinary Least Square* (OLS) dapat diaplikasikan pada tiap persamaan secara terpisah.
- c. *Forecast* atau peramalan yang dihasilkan pada beberapa kasus ditemukan lebih baik daripada yang dihasilkan oleh model persamaan simultan yang kompleks.
- d. Perangkat estimasi yang digunakan adalah *Impulse Respon Function* (IRF) untuk melacak respon dari peubah dependen dalam sistem VAR terhadap *shock* dari *error term* dan *Variance Decomposition* yang memberikan informasi mengenai pentingnya masing-masing *error term* dalam mempengaruhi peubah-peubah dalam VAR.

Di sisi lain, terdapat beberapa kritik terhadap model VAR menyangkut permasalahan berikut (Gujarati, 2010) :

- 1) Model VAR merupakan model yang *atheoritic* atau tidak berdasarkan teori, hal ini tidak seperti pada persamaan simultan. Pada persamaan simultan, pemilihan peubah yang akan dimasukkan dalam persamaan memegang peranan penting dalam mengidentifikasi model.
- 2) Pada model VAR penekanannya terletak pada *forecasting* atau peramalan sehingga model ini kurang cocok digunakan dalam menganalisis kebijakan.
- 3) Permasalahan yang besar dalam model VAR adalah pada pemilihan *lag length* atau panjang lag yang tepat. Karena semakin panjang lag, maka akan menambah jumlah parameter yang akan bermasalah pada *degrees of freedom*.
- 4) Peubah yang tergabung pada model VAR harus stasioner. Apabila tidak stasioner, perlu dilakukan transformasi bentuk data, misalnya melalui *first difference*.
- 5) Sering ditemui kesulitan dalam menginterpretasi tiap koefisien pada estimasi model VAR, sehingga sebagian besar peneliti melakukan interpretasi pada estimasi fungsi *impulse respon*.

#### e. Kriteria Pemilihan Model Terbaik

Untuk menguji kebaikan suatu model ada beberapa kriteria yang digunakan. Pada buku pedoman teknis ini kriteria yang digunakan adalah MAPE. Model time series/arima, model regresi, model fungsi transfer, dan model VAR dapat digunakan untuk melakukan estimasi variabel untuk beberapa tahun ke depan. Untuk model *time series* baik analisis ARIMA, model regresi, model fungsi transfer, dan model VAR, ukuran kelayakan model berdasarkan nilai kesalahan dengan menggunakan statistik MAPE (*mean absolute percentage error*) atau kesalahan persentase absolut rata-rata yang diformulasikan sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - F_t}{X_t} \right| \cdot 100$$

Dimana :  $X_t$  adalah data aktual

$F_t$  adalah nilai ramalan.

Semakin kecil nilai MAPE maka model yang diperoleh semakin baik, karena makin mendekati nilai aktual.

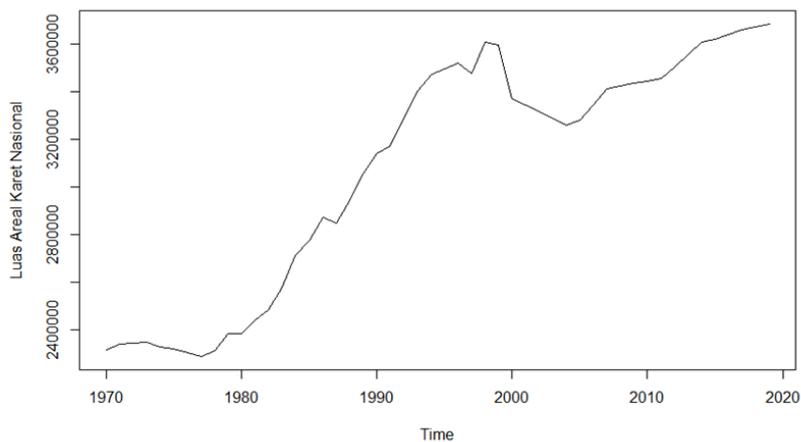
#### f. Sumber Data

Data yang digunakan untuk analisis ini bersumber dari Ditjen. Perkebunan, Kementerian Pertanian, meliputi data : produksi karet dalam bentuk karet kering satuan ton tahun 1967 – 2019, luas areal karet satuan hektar tahun 1970 – 2019, harga riil karet sheet satuan Rp/kg tahun 1971 – 2019, volume ekspor dan impor karet nasional satuan ton tahun 1969 – 2019. Disamping itu juga ada data harga karet dunia yang bersumber dari World Bank, tahun 1970 – 2019.

## HASIL DAN PEMBAHASAN

### A. Model ARIMA

Eksplorasi data luas areal karet nasional dalam bentuk karet kering berupa data tahunan dari tahun 1970 sampai 2019, seperti yang terlihat pada Gambar 1. Pada gambar 1 terlihat luas areal karet pada tahun 1970 sebesar 2,31 juta hektar, luas areal karet terus naik secara landai sehingga pada tahun 1980 luas areal karet mencapai 2,38 juta ton. Stelah tahun 1980 pertumbuhan luas areal cenderung lebih cepat, pertumbuhan luas areal karet tahun 1980 – 1999 rata-rata sebesar 2,20 %/tahun. Pada tahun 1999 sampai 2004 luas areal karet nasional cenderung terus turun, tetapi sejak tahun 2005 sampai 2019 luas areal kembali meningkat. Jika tahun 2005 luas areal karet sebesar 3,28 juta ha, maka tahun 2019 luas areal karet sudah mencapai 3,68 juta ha, rata-rata pertumbuhan mencapai 0,81%/tahun. Berdasarkan Gambar 1 juga bisa terlihat bahwa data belum stasioner karena masih mengalami perubahan seiring perubahan waktu.



Gambar 3. Perkembangan Luas areal Karet Nasional Tahun 1970 - 2019

Dalam melakukan pemodelan luas areal karet menggunakan model Autoregressive Integrated Average (ARIMA), data yang digunakan adalah periode tahun 1970 sampai 2019. Periode data tersebut kemudian dipisahkan menjadi data set training dan testing. Perlunya pemisahan data training dan testing adalah untuk menguji tingkat akurasi dalam melakukan peramalan. Panjang series data pada data set training adalah tahun 1970 sampai 2013, sementara dataset testing adalah periode 2014 sampai 2019. Dataset training digunakan untuk melakukan penyusunan model, sementara dataset testing digunakan untuk validasi model.

Uji kestasioneran data seperti yang disyaratkan apabila melakukan pemodelan ARIMA dilakukan secara visual menggunakan hasil plot data maupun uji formal statistik. Gambar 3 menunjukkan tidak ada fluktuasi yang muncul secara regular setiap bulannya, sehingga dapat disimpulkan bahwa data belum stasioner atau nilai rata-rata dan varian dari data time series

karet mengalami perubahan secara stokastik sepanjang waktu atau sebagian ahli menyatakan rata-rata dan variannya belum konstan (Narchrowi dan Haridus usman, 2006).

Tabel 1. Hasil Uji Augmented Dickey-Fuller Luas Areal Karet

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-212900  -28525   -7807   26983  150935

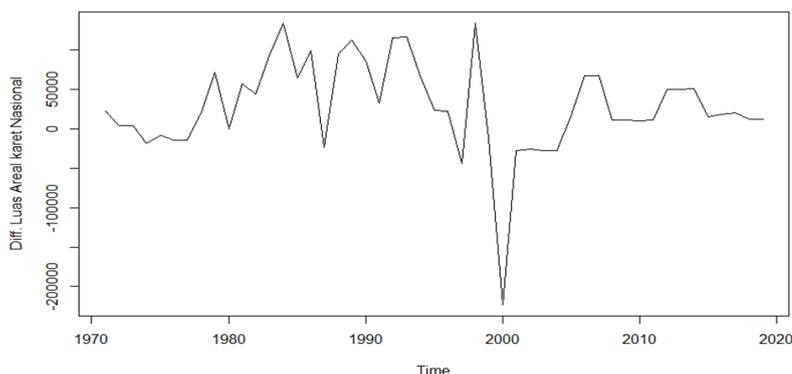
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.535e+05  9.644e+04  1.591  0.11871
z.lag.1      -5.687e-02  4.227e-02  -1.346  0.18535
tt           1.531e+03  1.486e+03  1.030  0.30851
z.diff.lag   3.806e-01  1.406e-01  2.706  0.00965 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56450 on 44 degrees of freedom
Multiple R-squared:  0.1643, Adjusted R-squared:  0.1073
F-statistic: 2.884 on 3 and 44 DF, p-value: 0.0463

value of test-statistic is: -1.3455 2.0193 1.0375

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61
```

Hal ini diperkuat oleh hasil uji formal statistik yaitu dengan uji Augmented Dickey-Fuller yang mengindikasikan bahwa data luas areal karet adalah belum stasioner, terlihat dari hasil uji tes statistik sebesar  $-1,3455$  sementara nilai kritis pada tingkat kepercayaan  $5\% = -3,50$  ( $\tau_3$ ) atau lebih kecil dari nilai uji statistik sehingga  $H_0$  tidak ditolak, atau data luas areal karet belum stasioner. Oleh karena itu, selanjutnya data luas areal karet dilakukan pembedaan (differencing) satu kali. Hasil plot setelah dilakukan differencing satu kali seperti pada Gambar 4.



Gambar 4. Perkembangan Luas Areal Karet Nasional Setelah Difference 1

Hasil plot luas areal karet setelah dilakukan differencing 1, menunjukkan bahwa data sudah terlihat stasioner untuk rata-rata. Hal ini juga didukung dengan uji Augmented Dickey-Fuller yang mengindikasikan bahwa data luas areal karet setelah differencing 1 sudah stasioner, terlihat dari hasil uji tes statistik sebesar  $-3,0955$  sementara nilai kritis pada tingkat kepercayaan 95%  $= -1,95$  ( $\tau_1$ ) dan tingkat kepercayaan 99%  $= -2,62$  ( $\tau_1$ ) atau lebih besar dari nilai uji statistik sehingga  $H_0$  ditolak, atau data luas areal karet setelah differencing 1 sudah stasioner.

Tabel 2. Hasil Uji Augmented Dickey-Fuller Luas areal Karet Difference 1

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-231542  -10602   6172   46647  149128

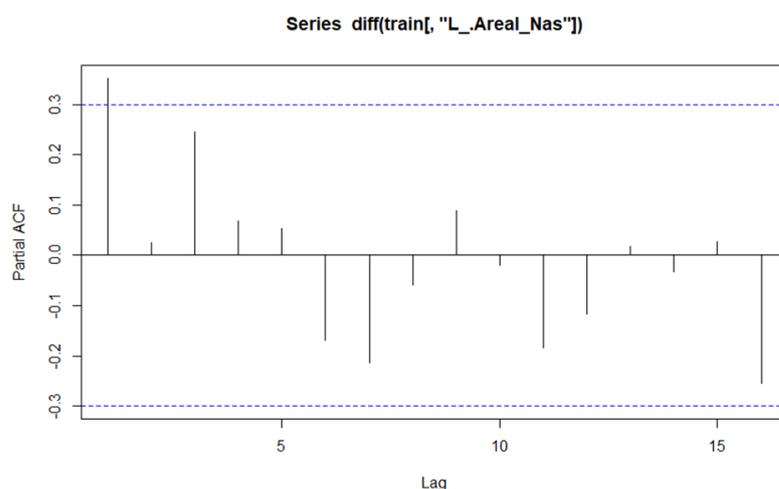
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1     -0.4712     0.1522  -3.095  0.00338 **
z.diff.lag  -0.1059     0.1481  -0.715  0.47816
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

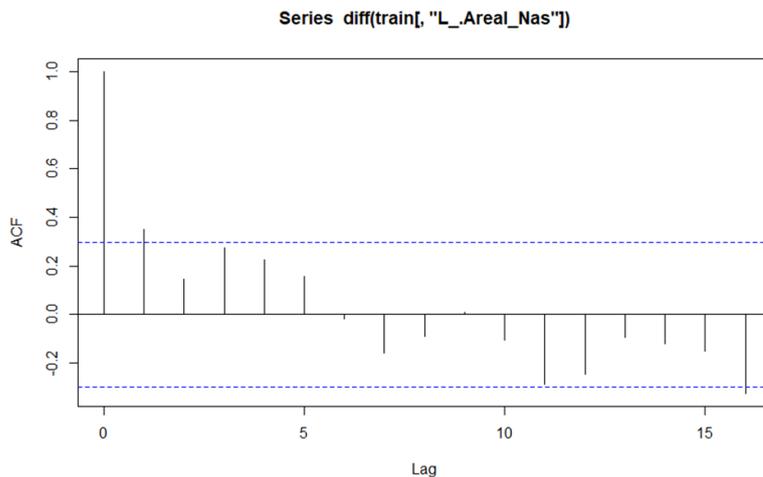
Residual standard error: 59200 on 45 degrees of freedom
Multiple R-squared:  0.2717, Adjusted R-squared:  0.2393
F-statistic: 8.392 on 2 and 45 DF, p-value: 0.0007991

Value of test-statistic is: -3.0955

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Identifikasi ordo AR dan MA untuk data luas areal karet sebelum ada difference menunjukkan pola ACF *tail off* sementara pola PACF tidak ada yang signifikan. Hal ini menunjukkan bahwa benar data karet belum signifikan, sehingga perlu dilakukan differencing 1 agar pola ACF dan PACF lebih jelas terlihat.





Gambar 5. Plot ACF dan PACF Luas Areal Karet Difference 1

Pada Gambar 5 menunjukkan bahwa baik plot ACF maupun plot ACF setelah *Difference 1* keduanya sudah *cut off*. Identifikasi ordo AR dan MA seperti tersaji pada Gambar 5 menunjukkan pola ACF *cut off* setelah lag 0, sementara pola PACF tidak ada yang menunjukkan signifikan, sehingga model tentatif belum dapat ditentukan.

Tabel 3. Model Arima Tentatif Berdasarkan Automodel

```
Series: train[, "L_.Areal_Nas"]
ARIMA(1,1,0) with drift

Coefficients:
      ar1      drift
      0.3443 28978.3
s.e.    0.1409 13384.6

sigma^2 estimated as 3.558e+09:  log likelihood=-532.89
AIC=1071.78  AICC=1072.4  BIC=1077.07

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 113.3363 57580.92 40544.13 0.002042352 1.338272 0.8003456 -0.001108081
```

Pengamatan secara visual pada plot ACF dan PACF sulit menentukan orde ARIMA, setelah dilakukan run model dengan menggunakan *auto arima* maka orde ARIMA yang disarankan adalah ARIMA (1,1,0), artinya model ARIMA tentative terbaik untuk melakukan estimasi luas areal karet nasional adalah untuk orde AR nilai  $p=1$ , untuk orde MA nilai  $q=0$ , dan difference  $d=1$ . Berdasarkan Tabel 3 dengan menggunakan ARIMA (1,1,0) maka untuk data training, akan menghasilkan MAPE = 1,34% artinya data berdasarkan model arima akan menyimpang rata-rata sekitar -1,34% sampai +1,34% dari data aktual.

Disamping metode pemilihan model Arima berdasarkan automodel, digunakan juga metode lain untuk mendapatkan orde ARIMA terbaik, yaitu dengan metode *Arima selection*. Pada metode ini akan dikeluarkan beberapa model terbaik. Setelah dilakukan pemilihan model metode terbaik tetap pada *differencing 1*. Model terbaik pertama menurut metode ini adalah ARIMA (1,1,0) menghasilkan nilai sbc paling kecil yaitu sebesar 949,42.

Tabel 4. Model Arima Tentatif Berdasarkan Arima Selection Differencing 1

	p	q	sbc
[1,]	1	0	949.4225
[2,]	0	0	950.3166
[3,]	2	0	954.0594
[4,]	3	0	956.1457
[5,]	4	0	960.2634
[6,]	1	1	963.6508
[7,]	1	2	963.9910
[8,]	0	1	964.2835
[9,]	0	4	964.3736
[10,]	1	3	964.5325

Selanjutnya dilakukan pengujian model ARIMA (1,1,0) apakah koefisien sudah signifikan. Untuk model ARIMA (1,1,0) koefisien ar1 sebesar 0,4557 dan koefisien ini signifikan pada taraf alpha 0,1%. Sehingga model ARIMA (1,1,0) layak digunakan.

Tabel 5. Uji Koefisien Model Arima (1,1,0)

```
Call:
arima(x = train[, "Areal_Nas"], order = c(1, 1, 0))

Coefficients:
      ar1
    0.4558
s.e.    0.1338

sigma^2 estimated as 3.694e+09:  log likelihood = -534.77,  aic = 1073.55

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.45578    0.13377   3.4071 0.0006566 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Salah satu syarat kebaikan model ARIMA adalah sebaran sisaan Ljung-Box. Hasil pengujian nilai p-value pada lag 5 sampai dengan lag 30 tidak ada yang signifikan, sehingga dapat disimpulkan bahwa sisaan bersifat random dan tidak ada autorelasi. Hal ini menunjukkan bahwa ARIMA (1,1,0) layak digunakan.

Tabel 6. Uji Sisaan Ljung Box Model Arima (1,1,0)

lags	statistic	df	p-value
5	3.571993	5	0.6125233
10	5.633401	10	0.8450659
15	11.073179	15	0.7473898
20	18.280982	20	0.5689045
25	20.572225	25	0.7162027
30	22.052382	30	0.8521318

Selanjutnya dilakukan uji kemampuan model ARIMA (1,1,0) apakah memiliki akurasi yang tinggi dalam melakukan peramalan. Untuk itu dilakukan uji coba peramalan dengan

menggunakan data testing, yaitu luas areal karet tahun 2014 – 2019. Sementara data training digunakan untuk menghasilkan model ARIMA (1,1,0). Dari hasil pengujian pada Tabel 6, menunjukkan MAPE untuk data training sebesar 1,34%, sementara MAPE data testing 0,69%. Hal ini menunjukkan Model Arima (1,1,0) sudah cukup akurat dalam melakukan peramalan karena hasil peramalan rata-rata menyimpang tidak lebih dari 2%.

Tabel 7. Hasil Pengujian Data Training dan Testing Arima (1,1,0)

Time Series:								
Start = 45								
End = 50								
Frequency = 1								
[1]	3592075	3623515	3653341	3682611	3711690	3740703		
		ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	113.3363	57580.92	40544.13	0.002042352	1.3382724	0.8003456	-0.001108081	
Test set	-20596.7946	31258.72	25320.04	-0.560081223	0.6910554	0.4998205	NA	

Selanjutnya dilakukan pengepasan model untuk seluruh data. Untuk Model ARIMA (1,1,0) koefisien ar1 sebesar 0,4640. Jika melakukan run model ARIMA (1,1,0) untuk seluruh data yaitu dari tahun 1967 – 2019 maka akan dihasilkan MAPE sebesar 1,23%. Hal ini menunjukkan bahwa antara data estimasi dengan data actual akan berbeda rata-rata berkisar antara -1,23% sampai +1,23%. Untuk metode estimasi dengan bias masih dibawah 2,5% dianggap masih cukup baik dan akurat.

Tabel 8. Model Arima (1,1,0) untuk Seluruh Data

Series: karet[, "L_.Area1_Nas"]							
ARIMA(1,1,0)							
Coefficients:							
	ar1						
	0.4640						
s.e.	0.1241						
sigma^2 estimated as 3.332e+09: log likelihood=-606.35							
AIC=1216.7 AICC=1216.96 BIC=1220.49							
Training set error measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	14753.45	56557.99	38056.52	0.4981572	1.238533	0.8087155	-0.1171866

Dengan menggunakan model ARIMA (1,1,0) menghasilkan angka estimasi luas areal karet untuk 5 tahun ke depan. Hasil Estimasi dengan model ARIMA ini pada tahun 2020 luas areal karet sebesar 3,689 juta ha. Pada tahun 2021 luas areal karet diestimasi akan naik sebesar 1,91% menjadi 3,691 juta. Pada tahun 2022 sampai 2024 juga menunjukkan kenaikan. Model Arima rata-rata menghasilkan angka estimasi yang cenderung terus naik atau terus turun. Rata-rata pertumbuhan luas areal karet tahun 2020 – 2024 rata-rata sebesar 0,058%. Jika dibandingkan pertumbuhan luas areal karet selama 5 tahun terakhir (tahun 2014 -2019) sebesar 0,42%, sementara hasil estimasi lima tahun kedepan rata-rata pertumbuhan hanya 0,031% per tahun atau jauh lebih rendah dari data historisnya.

Tabel 9. Ouput Peramalan Model Arima (1,1,0) untuk Luas Areal Karet

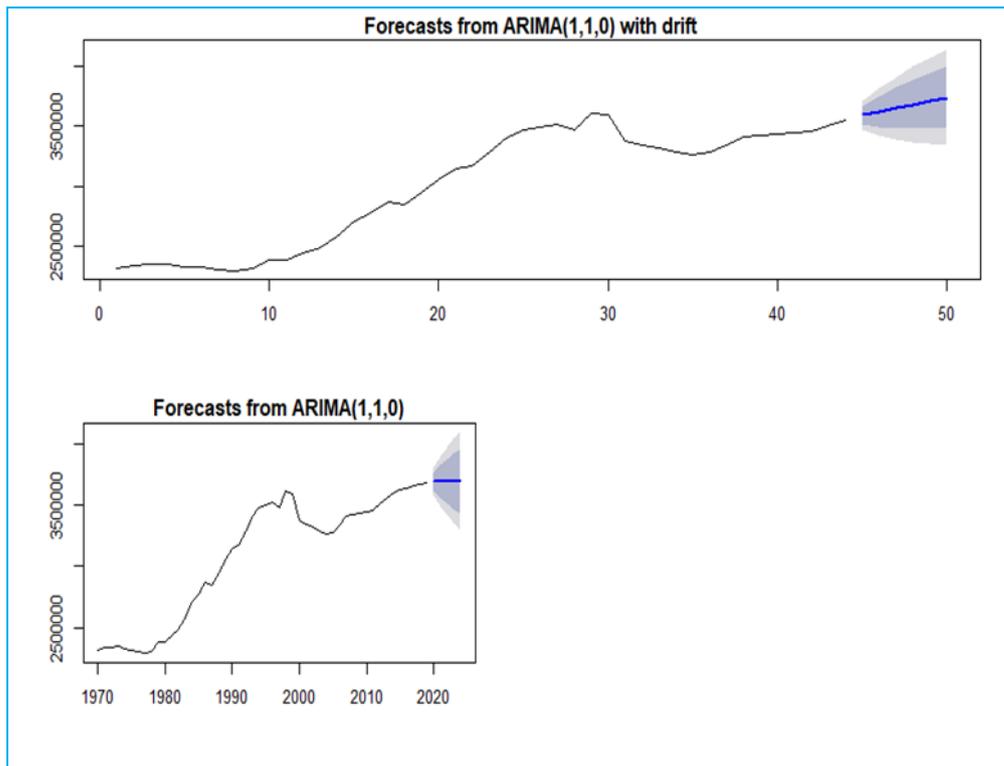
Point Forecast	Lo	Hi	Lo	Hi	Lo	Hi
2020	3689094	3615117	3763071	3575956	3802231	
2021	3691698	3560543	3822853	3491113	3892282	
2022	3692906	3512257	3873555	3416627	3969185	
2023	3693467	3469956	3916977	3351636	4035297	
2024	3693727	3432586	3954867	3294347	4093107	

Tabel 10. Hasil Estimasi Luas Areal Karet dengan Model ARIMA (1,1,0)

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	3.683.482	
2020	3.689.094	0,152
2021	3.691.698	0,071
2022	3.692.906	0,033
2023	3.693.467	0,015
2024	3.693.727	0,007
<b>Rata Rata Pertumbuhan</b>		<b>0,031</b>

Tahun 2019 : Angka Sementara (Ditjenbun)

Tahun 2020 - 2024 : Estimasi Berdasarkan Model Arima



Gambar 6. Hasil Estimasi Luas Areal Karet Tahun 2021 – 2024 Model Arima (1,1,0)

## B. Model Regresi

Model kedua untuk meramalkan luas areal karet adalah model regresi. Model regresi yang diuji adalah model regresi linier sederhana dan model regresi kuadratis. Untuk regresi linier sederhana maupun regresi kuadratis sebagai variabel tak bebas bebas adalah luas areal, sementara untuk variabel bebas adalah waktu.

Sebelum dilakukan analisis data dibagi dulu menjadi data training dan data testing. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk validasi model. Dari 50 observasi data luas areal karet dilakukan pembagian data training dan testing, dimana periode 1970-2013 menjadi data training sedangkan 2014-2019 menjadi data testing (enam tahun terakhir).

### Regresi Linier Sederhana

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dengan luas areal karet. Data input yang digunakan untuk pendugaan model adalah data training luas areal karet dari tahun 1970-2013.

Tabel 11. Ouput Model Regresi Linier

```
lm(formula = L_.Areal_Nas ~ Tahun, data = train)
Hasil Run Model Regresi
Residuals:
    Min       1Q   Median       3Q      Max
-225511 -169683 -36844   96135  386502
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -65197421   4768668  -13.67  <2e-16 ***
Tahun         34244       2394   14.30  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 201700 on 42 degrees of freedom
Multiple R-squared:  0.8296,    Adjusted R-squared:  0.8256
F-statistic: 204.5 on 1 and 42 DF,  p-value: < 2.2e-16
```

Dari hasil pendugaan model diatas dijelaskan bahwa intersep dan variabel tahun signifikan terhadap luas areal karet. Model ini layak dengan nilai F hitung = 204,5 dan nilai p-Value < 0.05 artinya model signifikan pada tingkat kepercayaan 95%. Koefisien regresi untuk variabel bebas tahun, juga menunjukkan nilai yang signifikan pada alpha = 0.1%. Model ini cukup layak dimana R<sup>2</sup> Adjusted sebesar 82,56%. Dari model diatas dapat dibuat model luas arealnya yaitu :

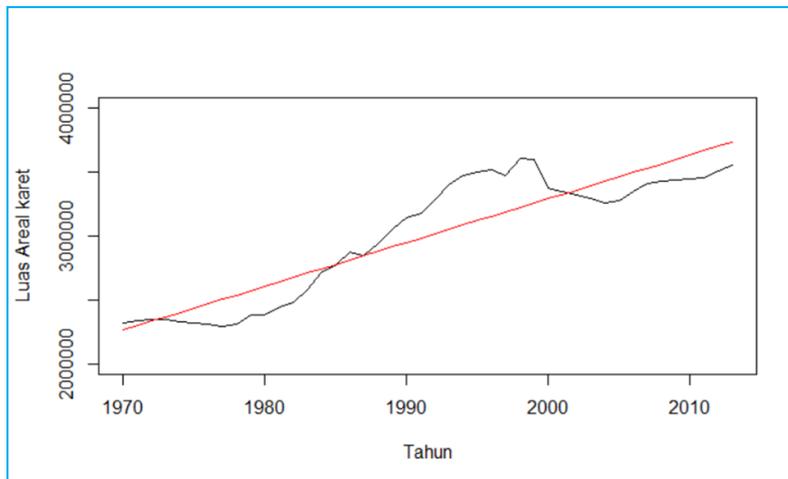
$$\text{Luas areal} = -65197421 + (34244 * \text{Tahun})$$

Untuk menguji kelayakan model dalam melakukan estimasi maka dilakukan pengujian perbandingan MAPE untuk data training dan data testing. Hasil pengujian model menunjukkan untuk regresi sederhana ini menghasilkan MAPE untuk data training rata-rata sebesar 5,37%, sedangkan untuk MAPE data testing lebih besar lagi yaitu 5,74%.

Tabel 10. MAPE Data Training dan Testing Model Regresi Linier

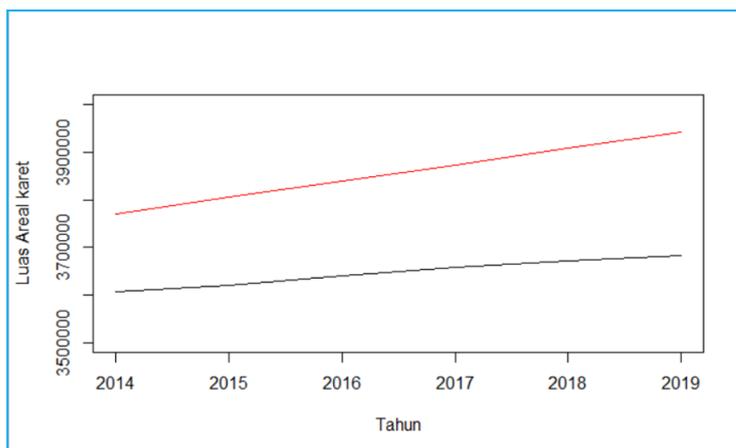
MAPE Data Training						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
0.08695	2.37788	5.08945	5.37622	8.22143	11.13077	
MAPE Data Testing						
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
4.562	5.184	5.681	5.743	6.293	7.018	

Untuk melihat tingkat akurasi dari hasil pemodelan regresi ini, selanjutnya dilakukan plot data training dengan data actual. Data training yang digunakan adalah luas areal karet dari tahun 1970-2013. Hasil plot pada Gambar 13 menunjukkan hasil plot regresi cenderung lurus, sementara data actual berflutuasi bisa di atas atau di bawah garis regresi linier. MAPE untuk data training ini cukup besar yaitu 5,37%. Antara tahun 1990 – 2000 hasil estimasi dengan model linier ini cenderung *underestimate*, tetapi setelah tahun 2000 hasil estimasi cenderung *overestimate* karena angka actual dibawah angka estimasi.



Gambar 7. Plot Data Training Regresi Linier dan Data Aktual Luas areal Karet 1970-2019

Hal yang sama juga dilakukan untuk plot data testing dan data aktual luas areal karet tahun 2014 – 2019. Jika model yang dihasilkan dari pola data data tahun 1970 – 2013, digunakan untuk mengestimasi luas areal tahun 2014 – 2019, menunjukkan akan dihasilkan MAPE data testing sebesar 5,74%, artinya rata-rata perbedaan data aktual dan data estimasi sebesar 5,74%. Hal ini ditunjukkan pada Gambar 8, data hasil estimasi model selalu di atas dari data aktualnya, karena MAPE sebesar 5,74%, maka hasil pengujian model menunjukkan hasil estimasi rata-rata lebih tinggi 5,74% dari data aktualnya.



Gambar 8. Plot Data Testing Regresi Linier dan Data Aktual Luas Areal Karet 2014-2019

Hasil estimasi menggunakan regresi sederhana menunjukkan hasil estimasi yang cenderung *over estimate*. Hasil estimasi selama 5 tahun kedepan luas areal karet nasional berkisar antara 3,97 juta ha sampai 4,11 juta ha. Hasil estimasi luas areal ini lebih tinggi dari luas areal Angka Sementara tahun 2019 yaitu sebesar 3,68 juta ha. Hal ini juga didukung nilai MAPE data testing yang terlampaui besar 5,74%, sehingga akurasi menjadi rendah. Hasil estimasi luas areal tahun 2020 sebesar 3,976 juta hektar, lebih tinggi 7,95% dari Angka Sementara tahun 2019, kenaikan cukup tinggi dibandingkan dengan data historisnya dimana sejak sepuluh tahun terakhir kenaikan luas areal karet tidak lebih dari 1,5%. Hal ini menunjukkan meskipun MAPE masih cenderung rendah tetapi hasil estimasi terlampaui tinggi dibandingkan data historisnya.

Tabel 12. Hasil Estimasi Model Regresi Linier Luas areal Karet Tahun 2020-2024

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	3.683.482	
2020	3.976.231	7,948
2021	4.010.475	0,861
2022	4.044.719	0,854
2023	4.078.964	0,847
2024	4.113.208	0,840
Rata Rata Pertumbuhan		0,850

Tahun 2019 : Angka Sementara (Ditjenbun)

Tahun 2020 - 2024 : Estimasi Berdasarkan Model Arima

## Regresi Kuadratik

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dan tahun kuadrat dengan luas areal karet. Data input yang digunakan untuk pendugaan model adalah

data training luas areal karet dari tahun 1970-2013. Hasil output run model untuk regresi kuadratik seperti pada Tabel 13.

Tabel 13. Model Regresi Kuadratik untuk Luas Areal Karet

```
Call:
lm(formula = L_.Areal_Nas ~ t + t2, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-246523 -112632 -40725  128510  291417

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1960076.7   79326.0  24.709 < 2e-16 ***
t             69421.4    8131.0   8.538 1.22e-10 ***
t2            -781.7     175.2   -4.462 6.21e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 167500 on 41 degrees of freedom
Multiple R-squared:  0.8853,    Adjusted R-squared:  0.8797
F-statistic: 158.3 on 2 and 41 DF,  p-value: < 2.2e-16
```

Persamaan regresi kuadratik yang dihasilkan adalah :

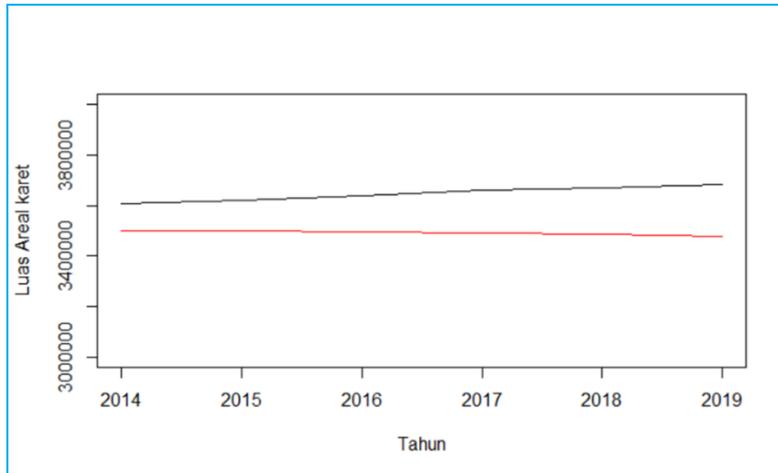
$$\text{Luas areal} = 1960076,7 + 69421,4*t) + (-781,7*t^2)$$

Model regresi kuadratik menghasilkan model yang layak dengan nilai F-hitung= 158,3 dan nilai p-value < 0.05 . Nilai R<sup>2</sup>-Adjusted lebih tinggi dari regresi linier yaitu sebesar 87,97%. Untuk melihat apakah koefisien regresi signifikan atau tidak digunakan uji t, hasil uji t menunjukkan intersep, koefisien t dan koefisien t<sup>2</sup> signifikan pada tingkat kepercayaan 99%.

Untuk menguji kelayakan model regresi kuadratik dalam melakukan estimasi maka dilakukan pengujian perbandingan MAPE untuk data training dan data testing. Hasil pengujian model menunjukkan untuk regresi kuadratik ini menghasilkan MAPE untuk data training rata-rata sebesar 6,61%, sedangkan untuk MAPE data testing lebih baik lagi yaitu 4,24%. Jika dibandingkan dengan regresi linier biasa, maka MAPE regresi kuadratik lebih baik, sehingga selisih antara data estimasi dan aktual lebih kecil. Untuk menggambarkan plot data hasil estimasi dan data actual pada Gambar 9, menunjukkan, garis warna merah merupakan hasil estimasi sedangkan warna hitam data actual, kedua garis ini relative lebih berdekatan dibandingkan dengan regresi linier. Hal ini menunjukkan bahwa regresi kuadratis menghasilkan model dan estimasi yang lebih baik dibandingkan regresi linier. Dari hasil plot itu menunjukkan bahwa hasil estimasi cenderung lebih rendah 4,24% dari data aktualnya, sehingga Ketika melakukan estimasi maka hasilnya akan cenderung *under estimate*.

Tabel 14. Perbandingan MAPE Data Training dan Data Testing Regresi Kuadratik

MAPE data Training					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4081	3.8529	5.0921	6.6129	8.8691	22.3774
MAPE data Testing					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.916	3.504	4.258	4.247	4.958	5.609



Gambar 9. Plot Data Testing Regresi Kuadratis dan Data Aktual Luas Areal Karet 2014 -2019

Hasil analisis estimasi menggunakan regresi kuadratik menunjukkan pada tahun 2020 luas areal karet mencapai 3,467 juta ton, atau turun sebesar 5,87%. Hasil estimasi ini diperkirakan terjadi *under estimate*, karena terjadi penurunan luas areal yang sangat tinggi. Angka pertumbuhan luas areal karet hasil estimasi dengan regresi kuadratik rata-rata turun sebesar 0,39% per tahun, atau rata-rata luas areal karet setiap tahun turun sekitar 10 - 15 ribu hektar. Meskipun model regresi kuadratik menghasilkan MAPE yang data testing yang lebih kecil 4,24% tetapi hasil estimasi diduga terlampau rendah (*Under Estimate*). Hasil estimasi luas areal tahun 2020 sebesar 3,467 juta hektar, lebih rendah 5,87% dari Angka Sementara tahun 2019, suatu penurunan cukup tinggi dibandingkan dengan data historisnya dimana sejak sepuluh tahun terakhir luas areal karet nasional tidak pernah mengalami penurunan. Hal ini menunjukkan meskipun MAPE masih cenderung kecil tetapi hasil estimasi terlampau rendah dibandingkan data historisnya.

Tabel 15. Hasil Estimasi Model Regresi Kuadratik Luas areal Karet Tahun 2020-2024

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	3.683.482	
2020	3.467.337	-5,868
2021	3.456.242	-0,320
2022	3.443.583	-0,366
2023	3.429.362	-0,413
2024	3.413.577	-0,460
Rata Rata Pertumbuhan		-0,390

Tahun 2019 : Angka Sementara (Ditjenbun)

Tahun 2020 - 2024 : Estimasi Berdasarkan Model Arima

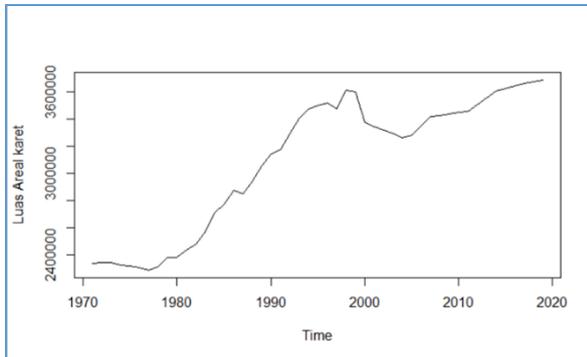
### C. Model Fungsi Transfer

Model fungsi transfer merupakan gabungan dari karakteristik analisis regresi berganda dengan karakteristik deret berkala ARIMA (*Autoregressive Integrated Moving Average*). Dalam model ini, selain mengandung keterkaitan dengan kejadian pada waktu-waktu sebelumnya, juga keterkaitan yang sifatnya kausal dengan variabel lain sebagaimana pada model regresi. Model fungsi transfer merupakan suatu model yang menggambarkan nilai prediksi dari suatu time series (deret output atau  $Y_t$ ) berdasarkan nilai-nilai dari deret itu sendiri ( $Y_t$ ) dan berdasarkan pula pada data time series yang mempunyai hubungan (deret input atau  $X_t$ ) dengan deret output. Model fungsi transfer yang bersifat dinamis berpengaruh tidak hanya pada hubungan linier antara waktu ke- $t$  input  $X_t$  dan waktu ke- $t$  output  $Y_t$ , tetapi juga saat input  $X_t$  dengan saat  $t, t+1, \dots, t+k$  pada output  $Y_t$  (Fitriani et al., 2013). Dalam melakukan estimasi dengan fungsi transfer diperlukan dua variabel, yaitu output  $Y_t$  dan input  $X_t$ . Input  $X_t$  merupakan variabel yang diduga paling berpengaruh terhadap output  $Y_t$ . Pada analisis ini dicontohkan estimasi untuk luas areal karet (output  $Y_t$ ) yang diduga dipengaruhi oleh harga karet dunia.

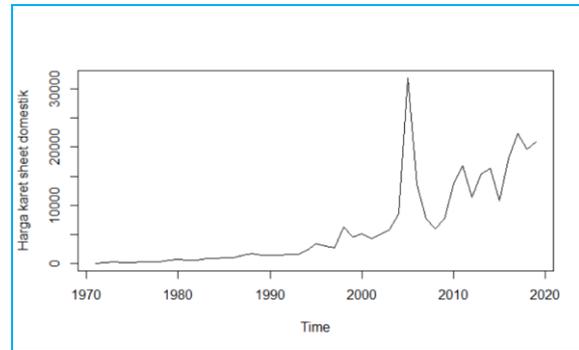
Dalam bahasan ini, model fungsi transfer variabel output luas areal karet akan melibatkan variabel harga karet sheet lokal sebagai variabel input. Langkah awal dalam mencari model fungsi transfer menggunakan software RStudio adalah penyiapan data, baik output  $Y_t$  (luas areal karet) maupun input  $X_t$  (harga karet sheet). Luas areal karet dalam satuan hektar bersumber dari Direktorat Jenderal Perkebunan, Kementerian Pertanian, sedangkan harga karet sheet dalam satuan Rp/Kg bersumber dari Ditjen Perkebunan. Banyaknya series data yang digunakan baik untuk luas areal maupun harga karet dunia harus seragam. Pada bahasan ini, series data yang digunakan yaitu 1971-2019 atau sebanyak 49 observasi.

Eksplorasi data dilakukan dengan menampilkan plot data luas areal maupun harga karet dunia. Berdasarkan plot data dapat diketahui pola data series 49 tahun yang akan digunakan untuk pemodelan. Berdasarkan Gambar 10 dan Gambar 11, terlihat bahwa terdapat data luas

areal karet nasional memiliki tren meningkat dari tahun ke tahun sedangkan harga karet sheet cenderung berfluktuasi. Harga karet sheet cenderung turun pada sekitar 10 tahun terakhir, akibat persaingan harga dengan produk karet sintetis. Luas areal karet nasional maupun harga karet sheet terindikasi tidak stasioner berdasarkan plotnya.



Gambar 10. Plot Data Luas areal Karet, 1971-2019



Gambar 11. Plot Data Harga Sheet, 1971-2019

Tahapan penyusunan model Fungsi Transfer luas areal karet dengan variabel input harga karet dunia adalah sebagai berikut :

- a. Pembagian series data awal menjadi series data *training* dan *testing*
- b. Pemeriksaan kestasioneran
- c. Pencarian model tentatif untuk variabel input
- d. *Prewhitening* dan korelasi silang
- e. Identifikasi model noise
- f. Pengepasan model
- g. Peramalan berbasis fungsi transfer

Data luas areal dan harga karet tahun 1971-2019 sebanyak 49 series akan dibagi menjadi series data *training* untuk periode 1971-2013 dan series data *testing* untuk periode 2014-2019. Selanjutnya dilakukan uji kestasioneran data untuk data input  $X_t$  yaitu harga karet dunia menggunakan uji Augmented Dickey-Fuller (ADF).

Hipotesis pada uji ADF ini adalah:

$H_0$  : data tidak stasioner

$H_1$  : data stasioner

Nilai test-statistic yang lebih besar dari critical values (nilai tau3), baik untuk taraf 1%, 5% maupun 10% menunjukkan bahwa  $H_0$  gagal ditolak, atau series data harga karet dunia belum stasioner. Oleh karena itu akan dilakukan pembedaan/differencing satu kali dan kemudian dilakukan uji ADF. Output hasil uji Augmented Dickey-Fuller sebelum dilakukan *differencing* menunjukkan belum stasioner. Setelah dilakukan Differencing 1 hasil uji menunjukkan sudah stasioner seperti pada Tabel 16.

Tabel 16. Hasil uji Kestationeran untuk Harga Karet Sheet Setelah Differencing 1

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-10156.3   -83.2   108.8   635.3  24321.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1     -1.5657     0.2511  -6.236  2.7e-07 ***
z.diff.lag   0.2453     0.1604   1.529   0.134
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

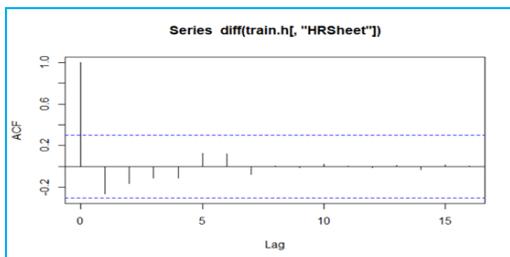
Residual standard error: 4886 on 38 degrees of freedom
Multiple R-squared:  0.6485,    Adjusted R-squared:  0.63
F-statistic: 35.05 on 2 and 38 DF,  p-value: 2.358e-09

Value of test-statistic is: -6.2363

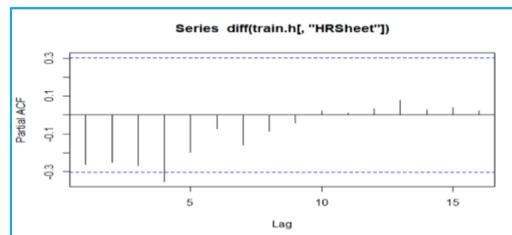
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Uji ADF pada data yang telah dilakukan *differencing* satu kali menunjukkan bahwa nilai *test-statistic* yaitu -6.2363 lebih kecil dari *critical values* (nilai tau1) menunjukkan bahwa  $H_0$  ditolak yang berarti data harga karet sheet sudah stasioner.

Stasioneritas data dapat juga dilihat dari plot ACF dan PACF. Jika lag pada ACF maupun PACF menunjukkan *tail off* maka data tidak stasioner. Berikut adalah script untuk menunjukkan plot ACF dan PACF harga karet dunia setelah differencing 1.



Gambar 12. Plot ACF Harga Karet Sheet dengan *Differencing* satu kali



Gambar 13. Plot PACF Harga Karet Sheet dengan *Differencing* satu kali

Plot ACF cenderung cut off dimana hanya lag 0 yang keluar dari *confidence interval*. Plot PACF menunjukkan cut off pada lag 4, ditunjukkan berada di luar garis *confidence interval*. Hal tersebut memperkuat hasil uji ADF yaitu data harga karet sheet telah stasioner.

Selain itu, plot ACF dan PACF juga dapat digunakan untuk mengidentifikasi ordo AR dan MA, dengan cara melihat lag yang nyata atau keluar dari garis *confidence interval*. Namun demikian, pada plot ACF diatas, terlihat nyata pada lag=0 yang sering terjadi pada

data riil, sehingga sulit untuk mengidentifikasi ordo AR dan MA melalui plot ACF dan PACF.

Pencarian model tentatif variabel input harga karet sheet dilakukan melalui penelusuran menggunakan model ARIMA. Model terbaik dapat dipilih menggunakan script *auto.arima* yang tersedia pada RStudio. Data yang digunakan untuk memilih model terbaik adalah series data *training*.

Tabel 17. Model Auto Arima Harga Karet Sheet Lokal

```
Series: train.h[, "HRSheet"]
ARIMA(0,1,2) with drift

Coefficients:
      ma1      ma2      drift
-0.5934 -0.2889 356.0193
s.e.    0.1688   0.1814 101.5680

sigma^2 estimated as 18249935:  log likelihood=-409.84
AIC=827.67  AICc=828.75  BIC=834.62

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -244.3634 4068.448 2050.902 -93.50906 102.7643 1.004758 -0.008115604
```

Berdasarkan pemilihan orde ARIMA menggunakan *automodel* menyarankan bahwa model terbaik untuk harga karet sheet adalah ARIMA (0,1,2) dengan MAPE data training yang masih sangat tinggi yaitu sebesar 102,76%. Model ARIMA (0,1,2) untuk data harga karet dunia tersebut menghasilkan nilai AIC = 827,67 sebagaimana output yang disajikan pada Tabel 17.

Tabel 18. Pengujian Koefisien Model Arima (0,1,2) Harga Karet Sheet

```
Call:
arima(x = train.h[, "HRSheet"], order = c(0, 1, 2))

Coefficients:
      ma1      ma2
-0.4715 -0.2142
s.e.    0.1505   0.1344

sigma^2 estimated as 18994630:  log likelihood = -411.82,  aic = 829.64

z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ma1 -0.47145   0.15052  -3.1321 0.001736 **
ma2 -0.21417   0.13437  -1.5939 0.110968
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Berdasarkan output di atas terlihat bahwa komponen *ma1* signifikan pada taraf 1%, *ma2* tidak signifikan. Selain menggunakan script *auto.arima* model tentatif dapat juga dipilih dengan *arima selection*. Pada Tabel 19 output yang dihasilkan untuk memilih model tentative terbaik untuk factor input *Xt* yaitu harga karet sheet.

Tabel 19. Model Arima Tentatif untuk Harga Karet Sheet (Differencing 1)

	p	q	sb
[1,]	0	0	715.5215
[2,]	1	0	717.2417
[3,]	2	0	719.2766
[4,]	4	0	719.7729
[5,]	3	0	720.9175
[6,]	5	0	722.5195
[7,]	1	1	728.3986
[8,]	0	1	728.8569
[9,]	0	2	730.3249
[10,]	2	1	731.8738

Hasil output R-Studio akan menunjukkan sepuluh model tentatif dimana idealnya model terbaik adalah model yang memiliki nilai SBC terkecil. Model ARIMA yang direkomendasikan ditunjukkan dari nilai p,d,q. Sebagai contoh model pertama dengan nilai p=0 dan q=0. Karena data harga karet sheet telah dilakukan *differencing* satu kali berarti d=1, artinya model yang direkomendasikan adalah ARIMA (0,1,0). Model terbaik kedua adalah ARIMA (1,1,0), untuk mengetahui apakah model ARIMA (1,1,0) lebih baik dari hasil *auto.arima* yaitu ARIMA (0,1,2) maka dapat dibandingkan dari nilai AIC dan koefisien ar1.

Tabel 20. Pengujian Koefisien Model Arima (1,1,0) Harga Karet Sheet

```
Call:
arima(x = train.h[, "HRSheet"], order = c(1, 1, 0))

Coefficients:
      ar1
      -0.2554
s.e.      0.1482

sigma^2 estimated as 22926969:  log likelihood = -415.53,  aic = 835.07

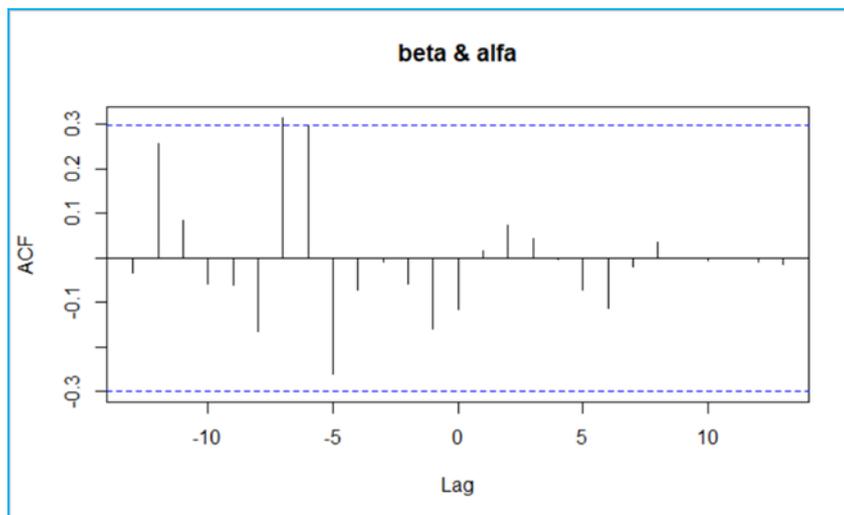
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1 -0.25536     0.14821  -1.723  0.08489 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model ARIMA (1,1,0) menghasilkan nilai AIC = 835,07 tidak berbeda jauh dari model ARIMA (0,1,2) dengan nilai AIC = 829,54. Pada model ARIMA (1,1,0) ditemukan bahwa ar1 signifikan pada taraf 10%. Oleh karena ARIMA (1,1,0) memiliki SBC yang paling kecil maka selanjutnya yang digunakan untuk peramalan harga karet sheet lokal menggunakan ARIMA (1,1,0).

Tahap selanjutnya untuk penyusunan model fungsi transfer ini adalah *prewhitening* dan korelasi silang. Korelasi silang menggambarkan struktur hubungan antara  $X_t$  dengan  $Y_t$ . Untuk mengidentifikasi pengaruh  $X_t$  terhadap  $Y_t$  maka deret  $X_t$  harus stasioner atau sudah distasionerkan. Dalam konteks pemodelan  $X_t$  terhadap  $Y_t$ , untuk membuat  $X_t$  stasioner tidak dengan pembedaan (*differencing*) namun dengan mengambil komponen *white noise* dari  $X_t$  (*prewhitening*). *Prewhitening* dilakukan terhadap deret input  $X_t$  yang didefinisikan sebagai

alfa serta deret input  $Y_t$  yang didefinisikan sebagai beta. Hasil output untuk prewhitening dan korelasi silang berupa grafik ACF untuk beta dan alfa.



Gambar 14. Plot korelasi silang Luas areal Karet dengan Harga Karet Sheet Lokal

Hasil plot korelasi silang digunakan untuk mengidentifikasi ordo  $r, s$ , dan  $b$ . Ordo  $r$  adalah panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$ , ordo  $s$  adalah panjang lag  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$ , dan ordo  $b$  adalah panjang jeda pengaruh  $X_t$  terhadap  $Y_t$ . Identifikasi ordo  $r, s$  dan  $b$  hanya dilihat pada lag yang positif.

Plot korelasi silang diatas menunjukkan bahwa tidak ada yang signifikan ditunjukkan dengan tidak ada yang keluar dari garis signifikansi, maka nilai  $b=0$ . Kemudian, tidak ada tambahan lagi lag yang signifikan maka nilai  $s=0$ . Mengingat data luas areal karet dan harga karet sheet merupakan data tahunan yang tidak mengandung musiman maka diasumsikan nilai  $r=0$ . Nilai  $b=0$  menunjukkan tidak ada jeda pengaruh antara harga karet sheet pada waktu  $t$  terhadap luas areal karet pada waktu  $t$ . Nilai  $s=0$  berarti ada korelasi antara luas areal dan harga karet sheet pada tahun yang sama. Dengan kata lain, dampak dari harga karet sheet terhadap luas areal dirasakan pada waktu yang sama ( $t$ ).

Untuk menghasilkan orde yang paling tepat untuk menentukan orde Arima fungsi transfer dengan melakukan identifikasi model noise. Untuk menghasilkan model terbaik dengan menggunakan *auto-arima* pada R Studio, model maka noise yang disarankan adalah ARIMA (0,1,0). Tahap selanjutnya dilakukan pengepasan model, untuk nilai  $r, s$  dan  $b$ . Hasil pengujian fungsi transfer dengan nilai  $r=0, s=0$ , dan  $b=0$  dengan model noise ARIMA (0,1,0) menghasilkan nilai MAPE yang cukup besar yaitu 220,20%, sehingga perlu dicari model alternatif lain.

Tabel 21. Pengujian Orde  $r=0$ ,  $s=0$ , dan  $b=0$  ARIMA (0,1,0) Luas areal Karet

```
Series: res
ARIMA(0,1,0)

sigma^2 estimated as 4.862e+10: log likelihood=-576.35
AIC=1154.69 AICC=1154.79 BIC=1156.43

Training set error measures:
Training set ME RMSE MAE MPE MAPE MASE ACF1
13586.14 217912.2 113713.4 77.4051 220.2055 0.9768427 -0.1708369
```

Tabel 22. Model Residual Tentatif ARIMA Untuk Luas areal Karet

```

p q sbc
[1,] 0 0 1034.348
[2,] 1 0 1037.863
[3,] 2 0 1041.348
[4,] 3 0 1045.216
[5,] 4 0 1048.784
[6,] 5 0 1052.569
[7,] 0 1 1056.889
[8,] 1 1 1059.959
[9,] 0 2 1060.750
[10,] 2 1 1062.982
```

Oleh karena model *autoarima* disarankan differencing tingkat 1, maka solusinya akan dicari model alternative. Model alternative yang diberikan untuk *model noise* seperti Tabel 20. Model tentatif pertama yang disarankan untuk data residual adalah ARIMA (0,1,0). Selanjutnya model tersebut didefinisikan sebagai *modelres* dan dilihat signifikansi AR. Oleh karena model noise ARIMA (0,1,0) sudah diujicobakan dan menghasilkan MAPE yang tinggi, maka selanjutnya dilakukan uji untuk model noise untuk residual dengan Arima (1,1,0). Untuk model noise ARIMA (1,1,0) menghasilkan komponen ar1 yang tidak signifikan (Tabel 23).

Tabel 23. Model Residual Tentatif ARIMA (1,1,0) Untuk Luas areal Karet

```
Series: res
ARIMA(1,1,0)

Coefficients:
      ar1
      -0.1626
s.e.    0.1508

sigma^2 estimated as 4.843e+10: log likelihood=-575.77
AIC=1155.55 AICC=1155.86 BIC=1159.02
z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1 -0.16263    0.15081 -1.0784  0.2809
```

Oleh karena model noise ARIMA (1,1,0) menghasilkan komponen ar1 yang tidak signifikan maka perlu dilakukan pengujian untuk model noise yang lain. Setelah dilakukan pengujian terhadap seluruh model tentative, ternyata sebagian besar tidak signifikan baik komponen ar maupun komponen ma, hanya satu model noise yang cukup baik yaitu ARIMA (1,1,1). Model noise ARIMA (1,1,1) menghasilkan komponen ma1 yang signifikan (Tabel

24). Model tentative yang perlu diuji besarnya MAPE adalah Model Noise ARIMA (1,1,0) dan ARIMA (1,1,1).

Tabel 24. Model Residual Tentatif ARIMA (1,1,1) Untuk Luas areal Karet

```
Series: res
ARIMA(1,1,1)
Coefficients:
      ar1      ma1
      0.3694 -0.6072
s.e.  0.3453  0.2832

sigma^2 estimated as 4.775e+10: log likelihood=-574.99
AIC=1155.99  AICC=1156.62  BIC=1161.2
> coeftest(modelres)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.36938    0.34531  1.0697  0.28475
ma1 -0.60716    0.28320 -2.1439  0.03204 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tahap berikutnya adalah dilakukan pengepasan model dengan noise. Pengepasan model dengan noise dilakukan dengan menggabungkan nilai  $r=0$ ,  $s=0$ ,  $b=0$  dan hasil identifikasi model noise yaitu ARIMA(1,1,0). Selanjutnya diperoleh output model fungsi transfer ARIMA (1,1,0) seperti pada Tabel 24.

Tabel 25. Model Fungsi Transfer ARIMA (1,1,0) Untuk Luas Areal Karet

```
Series: train.h[, "Areal"]
Regression with ARIMA(1,1,0) errors
Coefficients:
      ar1      xreg
      0.4550 -0.4689
s.e.  0.1353  1.5927

sigma^2 estimated as 3.952e+09: log likelihood=-522.74
AIC=1051.47  AICC=1052.1  BIC=1056.68

Training set error measures:
Training set  ME  RMSE  MAE  MPE  MAPE  MASE  ACF1
16097.44  60636  42247.68  0.5483252  1.377677  0.8230174  -0.1133076
> coeftest(tf.arima1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.45497    0.13528  3.3632  0.0007703 ***
xreg -0.46886    1.59275 -0.2944  0.7684719
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model fungsi transfer luas areal karet dengan ARIMA(1,1,0) menghasilkan MAPE yang sangat baik karena nilainya sangat kecil yaitu sebesar 1,38%. Untuk mengetahui *performance* dari model fungsi transfer tersebut dilakukan pengujian *coeftest* untuk mengetahui tingkat signifikansi dari masing- masing koefisien.

Berdasarkan hasil uji Z tes untuk fungsi transfer ARIMA (1,1,0), untuk koefisien ar1 signifikan dengan tingkat kepercayaan 99%, dan komponen fungsi transfer xreg yaitu variabel harga karet sheet tidak signifikan. Oleh karena itu komponen xreg tidak signifikan, maka akan dicobakan model fungsi transfer dengan model tentatif komponen noise ARIMA (1,1,1). Hasil output untuk fungsi transfer dengan model noise ARIMA (1,1,1) seperti pada Tabel 26.

Tabel 26. Model Tentatif Fungsi Transfer ARIMA (1,1,1) Untuk Luas Areal Karet

```
Series: train.h[, "Areal"]
Regression with ARIMA(1,1,1) errors

Coefficients:
      ar1      ma1      xreg
0.8581 -0.5697 -0.7694
s.e. 0.1199 0.1941 1.6327

sigma^2 estimated as 3.805e+09: log likelihood=-521.48
AIC=1050.97 AICC=1052.05 BIC=1057.92

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 10861.25 58747.72 39434.79 0.3878016 1.286546 0.7682202 0.01151144
> coeftest(tf.arima1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.85809   0.11988   7.1580 8.187e-13 ***
ma1 -0.56970   0.19413  -2.9347 0.003339 **
xreg -0.76935   1.63272  -0.4712 0.637491
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MAPE yang dihasilkan dari fungsi transfer ARIMA (1,1,1) adalah 1,29% lebih kecil, jadi lebih baik dari MAPE fungsi transfer ARIMA(1,1,0). Selanjutnya dilakukan pengujian *coeftest* untuk fungsi transfer ARIMA(1,1,1) untuk melihat signifikansi dari komponen AR dan MA. Dari hasil pengujian di atas, ditemukan bahwa komponen AR1 signifikan dan komponen MA1 juga signifikan pada tingkat kepercayaan 99%, sebaliknya komponen fungsi transfer xreg tidak signifikan. Berdasarkan hasil penelusuran model dan uji yang telah dilakukan maka disimpulkan bahwa model fungsi transfer luas areal karet ARIMA (1,1,1) lebih baik dibandingkan model ARIMA fungsi transfer (1,1,0).

### Peramalan berbasis fungsi transfer

Berdasarkan model fungsi transfer dengan model noise ARIMA (1, 1, 1), dilakukan peramalan berbasis nilai aktual dimana luas areal karet diestimasi menggunakan data aktual harga karet sheet periode 2014-2019. Meskipun data aktual luas areal karet periode 2014-2019 telah ada, dilakukan peramalan luas areal karet untuk mengecek *performance/kelayakan* model fungsi transfer.

Tabel 27. Pengujian MAPE Model Fungsi Transfer ARIMA (1,1,1) Dengan Faktor Input Harga Karet Sheet Data Aktual tahun 2014 - 2019

```

Series: test.h[, "Areal"]
Regression with ARIMA(1,1,1) errors

Coefficients:
      ar1      ma1      xreg
    0.8581 -0.5697 -0.7694
s.e.  0.0000  0.0000  0.0000

sigma^2 estimated as 3.805e+09:  log likelihood=-53.81
AIC=109.63  AICC=110.96  BIC=109.24

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 7440.983 10149.37 8264.082 0.2044491 0.2268684 0.534982 0.2204719
    
```

Peramalan luas areal karet periode 2014-2019 menggunakan fungsi transfer ARIMA (1,1,1) dengan input harga karet dunia **data aktual** menghasilkan MAPE 0,23%.

Tujuan melakukan pemodelan fungsi transfer adalah untuk mendapatkan nilai ramalan periode ke depan, yakni luas areal karet tahun 2020-2024. Karena data series input harga karet dunia tersedia hingga tahun 2019, maka perlu dilakukan peramalan harga karet dunia terlebih dahulu atau dengan kata lain peramalan luas areal dilakukan berbasis nilai ramalan harga karet.

Oleh karenanya, untuk pengujian data testing terlebih dahulu dilakukan estimasi harga karet dunia periode 2014 - 2019 menggunakan model ARIMA (1,1,0) sebagaimana yang telah diperoleh dari tahap pencarian model tentatif untuk variabel input, sebagai variabel input harga karet sheet. Pemilihan variabel input harga karet sheet karena secara tidak langsung berpengaruh juga pada luas areal karet nasional. Selanjutnya dilakukan peramalan luas areal karet dengan fungsi transfer ARIMA (1, 1, 1) sebagai model terbaik berdasarkan tahapan pengepasan model dengan noise. Peramalan luas areal dengan fungsi transfer ARIMA (1,1,1) menggunakan **nilai ramalan** harga karet sheet yang telah diestimasi dengan ARIMA (1,1,0).

Tabel 28. Pengujian MAPE Model Fungsi Transfer ARIMA (1,1,1) Dengan Faktor Input Harga Karet Sheet Nilai Ramalan tahun 2014 – 2019.

```

Series: test.h[, "Areal"]
Regression with ARIMA(1,1,1) errors

Coefficients:
      ar1      ma1      xreg
    0.8581 -0.5697 -0.7694
s.e.  0.0000  0.0000  0.0000

sigma^2 estimated as 3.805e+09:  log likelihood=-52.87
AIC=107.73  AICC=109.07  BIC=107.34

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 6950.511 8438.594 6950.511 0.1910169 0.1910169 0.449947 0.1931367
    
```

Estimasi luas areal karet berbasis fungsi transfer dengan model fungsi transfer ARIMA (1,1,1) selama 6 tahun terakhir (2014-2019) menggunakan input harga karet sheet hasil **angka ramalan** ARIMA (1,1,0) menghasilkan MAPE 0,19%. Nilai MAPE yang kurang dari 1% ini untuk penyusunan estimasi sudah sangat baik, karena kesalahannya sangat kecil.

Setelah dilakukan peramalan luas areal karet baik menggunakan input (harga karet dunia) data aktual maupun ramalan, tahapan berikutnya adalah pengepasan model arima output. Pengepasan model ARIMA output dimaksudkan untuk membandingkan hasil ramalan luas areal baik berdasarkan data training (1970-2013) maupun data testing (2014-2019). Model ARIMA biasa, tanpa fungsi transfer, terbaik untuk variabel luas areal yang terpilih adalah ARIMA (1,1,0). Model ARIMA (1,1,0) tersebut merupakan ARIMA biasa dan bukan hasil dari fungsi transfer. Hasil ramalan model ARIMA tanpa fungsi transfer dan model ARIMA dengan fungsi transfer selanjutnya akan dibandingkan.

Tabel 29. Pengujian MAPE Model ARIMA (1,1,0) Tanpa Fungsi Tranfer Untuk Faktor Output (Luas Areal Karet) tahun 2014 – 2019.

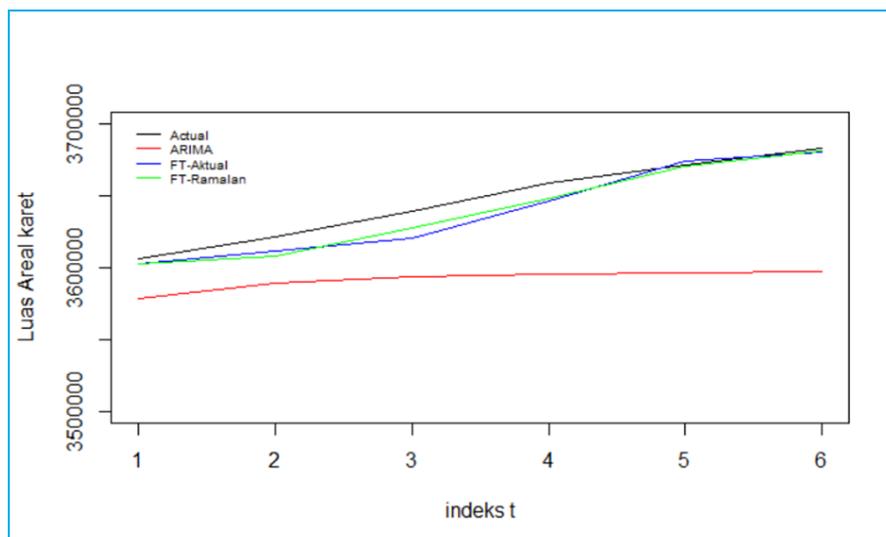
```
Series: train.h[, "Areal"]
ARIMA(1,1,0)
Coefficients:
      ar1
      0.4552
s.e.    0.1351

sigma^2 estimated as 3.864e+09:  log likelihood=-522.78
AIC=1049.56  AICC=1049.87  BIC=1053.03

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 15982.35 60698.41 42217.86 0.5447585 1.376808 0.8224365 -0.1098817
> ramalan.output <- forecast(model.output,6)
> accuracy(ramalan.output,test.h[, "Areal"])
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 15982.35 60698.41 42217.86 0.5447585 1.376808 0.8224365 -0.1098817
Test set      54956.80 59048.44 54956.80 1.5027068 1.502707 1.0706010      NA
```

Hasil peramalan luas areal karet dengan ARIMA biasa atau ARIMA tanpa fungsi transfer menunjukkan bahwa hasil ramalan jika menggunakan data training menghasilkan MAPE = 1,37%. Di sisi lain, MAPE yang dihasilkan dari data testing yaitu 1,50%. Artinya jika model yang digunakan hanya ARIMA tanpa fungsi transfer maka MAPE data testing cenderung lebih tinggi dibandingkan dengan MAPE fungsi transfer ARIMA (1,1,1).

Untuk membandingkan ketepatan model estimasi, dilakukan perbandingan hasil estimasi terhadap data aktual luas areal karet pada tahun 2014-2019 (data testing). Hasil ramalan yang dibandingkan yaitu ramalan dengan model ARIMA(1,1,0), model fungsi transfer ARIMA(1,1,1) dimana input harga karet sheet yang digunakan adalah data aktual maupun ramalan. Selanjutnya hasil ramalan dibandingkan dalam sebuah grafik seperti Gambar 15.



Gambar 15. Perbandingan Hasil Ramalan Luas areal Karet Tahun 2014-2019

Dari grafik di atas terlihat jika dibandingkan dengan data aktual luas areal karet 2014-2019 (warna hitam), maka hasil peramalan dengan model ARIMA tanpa fungsi transfer (warna merah) tidak mengikuti pola data aktualnya dan cenderung *underestimate*, karena hasil peramalan selalu lebih kecil dari nilai aktualnya / nilai sebenarnya. Sebaliknya, peramalan dengan fungsi transfer khususnya jika input harga karet dunia yang digunakan adalah data aktual maka hasil ramalan luas arealnya (warna biru) sangat menyerupai pola data luas areal sebenarnya selama 6 tahun terakhir. Jika input harga karet dunia yang digunakan adalah hasil ramalan, maka estimasi luas arealnya (warna hijau) juga menghasilkan angka estimasi yang sangat baik karena makin mendekati nilai luas areal sebenarnya (data actual), hal ini juga terlihat di grafik data hasil ramalan dan data actual hampir berimpit pada 2 tahun terakhir.

Hasil identifikasi model untuk meramalkan luas areal karet dengan ke-3 cara diatas menghasilkan nilai MAPE masing-masing sebagai berikut:

Tabel 30. Perbandingan MAPE untuk Arima biasa dan Arima Fungsi Transfer

No	Model	MAPE
1	ARIMA (1,1,0) (tanpa Fungsi transfer)	1,37% (data training) dan 1,50% (data testing)
2	Fungsi Transfer dengan ARIMA (1,1,1) input data aktual	0,23%
3	Fungsi Transfer dengan ARIMA(1,1,1) input data ramalan	0,19%

Berdasarkan nilai MAPE yang dihasilkan maka model fungsi transfer ARIMA(1,1,1) dengan input data aktual lebih bagus jika dibandingkan dengan model fungsi transfer ARIMA (1,1,1) dengan input data ramalan maupun model ARIMA(1,1,0) tanpa fungsi transfer. Namun demikian, karena tujuan penyusunan model adalah untuk mendapatkan angka ramalan

tahun 2020 -2024, maka digunakan model Fungsi Transfer berbasis input data ramalan, yakni dengan terlebih dahulu dilakukan estimasi harga karet sheet lima tahun ke depan menggunakan model ARIMA (1,1,0). Hasil estimasi harga karet dunia tersebut kemudian menjadi input untuk meramalkan luas areal karet lima tahun ke depan menggunakan model fungsi transfer ARIMA (1,1,1) dan sebagai variabel input harga karet sheet. Hasil ramalan luas areal karet disajikan pada Tabel 26 dan Tabel 27.

Selain mencari model terbaik untuk meramalkan luas areal karet, akan diestimasi juga luas areal karet lima tahun ke depan (2020-2024) menggunakan fungsi transfer ARIMA (1,1,0) dengan menggunakan seluruh data. Setelah dilakukan run model Fungsi Transfer dengan ARIMA (1,1,0) untuk seluruh data (tidak lagi menggunakan data training), maka dihasilkan MAPE untuk seluruh data sebesar 1,16%, nilai ini juga relative kecil, sehingga estimasi lima tahun kedepan cukup akurat.

Tabel 31. Hasil Peramalan Luas Areal Karet Model Fungsi Transfer Arima (1,1,1) Tahun 2020 – 2024

```
Series: dataestimasi[, "Areal"]
Regression with ARIMA(1,1,1) errors

Coefficients:
      ar1      ma1      xreg
    0.8594 -0.5683 -0.6315
s.e.  0.1106  0.1811  1.4559

sigma^2 estimated as 3.32e+09:  log likelihood=-592.92
AIC=1193.83  AICC=1194.76  BIC=1201.32

Training set error measures:
Training set      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
9560.19  55220.97  35752.41  0.3405945  1.160396  0.7515264  0.01371162
> ramalan.tf.final <- forecast(tf.arima.final,h=5, xreg=ramalan.input2$mean)
> ramalan.tf.final$mean
Time Series:
Start = 50
End = 54
Frequency = 1
[1] 3694629 3703988 3712086 3719032 3725004
```

Tabel 32. Hasil Peramalan Luas Areal Karet Tahun 2020-2024 Model Fungsi Transfer ARIMA (1,1,1)

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	3.683.482	
2020	3.694.629	0,303
2021	3.703.988	0,253
2022	3.712.086	0,219
2023	3.719.032	0,187
2024	3.725.004	0,161
Rata Rata Pertumbuhan		0,205

Tahun 2019 : Angka Sementara (Ditjenbun)  
Tahun 2020 - 2024 : Estimasi Berdasarkan Model Arima

Pada tahun 2019 Angka Sementara untuk luas areal karet sebesar 3.683.482 hektar (Ditjen Perkebunan, 2019). Hasil peramalan dengan fungsi transfer ARIMA (1,1,1) menunjukkan bahwa selama lima tahun ke depan luas areal karet akan meningkat rata-rata 0,205% per tahun. Tahun 2020 luas areal karet diestimasi sebesar 3.694.629 hektar atau naik 0,30% dibandingkan tahun 2019. Hasil estimasi tersebut hampir sama dibandingkan Angka Estimasi Ditjenbun untuk luas areal karet tahun 2020 adalah sebesar 3.694.716 hektar atau lebih tinggi 87 hektar atau 0,002% . Luas areal karet diramalkan terus meningkat hingga tahun 2024, dengan pertumbuhan yang semakin melambat, hingga pada tahun 2024 luas arealnya diestimasi sebesar 3.725.004 hektar.

#### **D. Model VAR (*Vector Auto Regressive*)**

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen. Model VAR berlaku pada saat nilai setiap variabel dalam sebuah system tidak hanya bergantung pada lag-nya sendiri, namun juga pada nilai lag variabel lain.

Tahapan yang dilakukan dalam penyusunan model VAR adalah sebagai berikut: persiapan data, pembagian data training dan testing, pemilihan lag dan type, pengajuan asumsi, ramalan data training, testing, penghitungan MAPE, dan plot, pemilihan model terbaik, dan pengepasan model untuk seluruh data dan peramalannya. Disamping itu dibahas Interpretasi Impulse Response Function (IRF) dan Variance Decomposition.

Variabel yang digunakan untuk estimasi model VAR adalah luas areal (areal) dalam satuan hektar, harga karet sheet (HRsheet) dalam satuan Rp/kg, harga karet dunia (HKaretworld) dalam satuan US\$/kg, volume ekspor karet (volekkaret) dalam satuan Ton, dan volume impor (volimkaret) dalam satuan Ton. Data luas areal karet, luas areal, volume ekspor karet, volume impor karet, dan harga karet sheet diperoleh dari publikasi Ditjen Perkebunan, sementara data variabel harga karet dunia diperoleh dari World Bank. Series masing-masing variabel berbeda karena keterbatasan ketersediaan data. Series data luas areal karet adalah dari tahun 1970-2019, series data harga sheet adalah dari tahun 1971-2019, series data harga karet dunia adalah dari tahun 1970-2019, sementara series data volume ekspor dan impor karet adalah dari tahun 1970-2019. Format data yang digunakan bisa dalam bentuk excell, CSV, ataupun SPSS.

#### **Pembagian Data Training dan Testing**

Series data yang digunakan adalah series tahun 1971 – 2019 akan dibagi menjadi 2 set data yakni set data training (tahun 1971-2013) atau 43 titik dan set data testing (2014-2019) atau 6 titik.

#### **Pemilihan Lag (p) dan Type**

Dalam permodelan VAR karet ini digunakan lima variabel, yaitu luas areal karet (areal), harga sheet (HRsheet), harga karet dunia (HKaretWorld), volume ekspor karet

(volekkaret), dan volume impor (volimkaret). Selain komposisi variabel tersebut, komponen konstanta dan trend juga perlu dilakukan pengujian untuk mengetahui signifikan atau tidak keberadaannya. Jika keduanya signifikan, maka komponen tersebut harus dimasukkan ke dalam model VAR dengan type “both”. Jika hanya konstanta yang signifikan, maka trend perlu dikeluarkan dari model VAR dengan model VAR type “const”. Jika hanya trend yang signifikan maka konstanta dikeluarkan dari model menggunakan model VAR type “trend”, dan jika keduanya tidak signifikan, maka type yang digunakan model VAR adalah “none”.

Keberadaan konstanta dan trend dapat dideteksi dari plot data awal, namun terkadang hal tersebut sulit dilakukan. Untuk mengantisipasi hal tersebut, sebaiknya dilakukan uji coba/trial and error menggunakan model VAR dari lag  $p=1$  s.d 5 dengan type “both” untuk mengetahui signifikan atau tidak keberadaannya. Untuk panjang lag maksimum bisa dilakukan *trial error* sampai tidak memungkinkan untuk dilakukan permodelan.

Untuk data karet ini, setelah dilakukan running model VAR dengan lag  $p=1$  type “both”, diperoleh informasi komponen konstanta tidak signifikan, tetapi trend signifikan, jumlah variabel yang signifikan ada 3 variabel dari total 5 variabel dalam system. Selanjutnya dilakukan uji VAR(2) type both, model ini menghasilkan konstanta dan trend yang signifikan, dan ada 5 variabel yang signifikan dari total 10 variabel dalam model. Untuk model VAR(3) type both, model ini menghasilkan konstanta dan trend yang signifikan, dan ada 3 variabel yang signifikan dari total 15 variabel dalam model. Untuk model VAR(4) type both, model ini menghasilkan konstanta dan trend yang signifikan, dan ada 5 variabel yang signifikan dari total 20 variabel dalam model. Selanjutnya model VAR(5) type both, model ini menghasilkan konstanta dan trend yang signifikan, dan hanya ada 1 variabel yang signifikan dari total 25 variabel dalam model. Oleh karena jumlah variabel yang signifikan semakin sedikit sehingga penelusuran berhenti di VAR(5) type both (Tabel 28)

Selanjutnya dilakukan juga pemilihan model dengan model VAR(1), VAR(2), VAR (3), VAR(4), VAR(5) untuk type constant. Jumlah Variabel yang banyak signifikan untuk type constant ini adalah model VAR(2) dan VAR(4).

Pemilihan lag  $p$  ditentukan dengan melihat banyaknya variabel yang signifikan dalam lag tersebut sekaligus memastikan harus ada variabel/peubah yang signifikan pada lag terpilih dimaksud. Setelah dilakukan run model ternyata ada 4 kandidat model VAR terbaik yaitu VAR (2) type “both” , VAR(3) type “both”, VAR(2) type “constant” dan VAR(4) type “constant“. Untuk menguji keempat model VAR yang terbaik maka dilakukan pengujian MAPE untuk data training dan MAPE data testing.

Tabel 33. Hasil Pengujian Model VAR pada Beberapa Tingkat p dan Type

Lag (p)	Type	Sig. Type	Var. Syg	Jumlah Variabel	Adj-R <sup>2</sup>
p=1	both	Const tidak Syg, trend Syg	3	5	0,9855
p=2	both	Const dan Trend Syg	5	10	0,9882
p=3	both	Const dan Trend Syg	3	15	0,9860
p=4	both	Const dan Trend Syg	5	20	0,9874
p=5	both	Const dan Trend Syg	1	25	0,9868
p=1	constant	Const tidak Syg	2	5	0,9846
p=2	constant	Const tidak Syg	5	10	0,9854
p=3	constant	Const tidak Syg	2	15	0,9826
p=4	constant	Const tidak Syg	5	20	0,9844
p=5	constant	Const tidak Syg	2	25	0,9787

Hasil pengujian besaran MAPE untuk 4 kandidat model VAR terbaik seperti pada Tabel 29, menunjukkan bahwa untuk data training yang memiliki MAPE terkecil adalah VAR(4) type Constant, sedangkan untuk MAPE data testing yang paling kecil adalah VAR(2) type Constant. Oleh karena tujuan permalan adalah untuk menghasilkan angka yang akurat maka model terbaik yang terpilih adalah model VAR(4) type constant, karena menghasilkan MAPE data training terkecil, meskipun di MAPE data testing bukan yang paling kecil.

Tabel 34. Perbandingan MAPE Pengujian Model VAR

VAR (p)	Type	MAPE training	MAPE Testing
p=2	both	1,04	8,61
p=3	both	1,07	15,66
p=2	constant	1,25	3,63
p=4	constant	0,89	5,29

Untuk model VAR yang terbaik adalah Model VAR (p=4) dengan konstanta. Hasil model VAR(2) type constant, untuk mengestimasi Luas areal (t) karet menunjukkan bahwa koefisien constanta menunjukkan hasil yang tidak signifikan pada tingkat kepercayaan 95%. Untuk variabel lain yang signifikan untuk mengestimasi luas areal (t) antara lain luas areal lag1 (99%), volume ekspor karet lag 1 (90%), volume impor karet lag 1 (95%), harga karet dunia lag 2 (95%), dan volume kespore karet lag 2 (90%). Ketiga variabel tersebut signifikan dengan tingkat kepercayaan antara 90% - 99%.

Model VAR (2) type both ini menghasilkan nilai Adjusted R Square = 98,54%, artinya keragaman luas areal dipengaruhi oleh variabel-variabel penjelasnya sebesar 98,54%. Nilai F hitung = 270,2, sehingga nilai p-value untuk model luas areal ini sangat kecil atau jauh lebih kecil dari 0,05, sehingga model layak untuk digunakan.

Tabel 35. Output Model VAR(2) type=constant

```

VAR Estimation Results:
=====
Endogenous variables: Areal, HRSheet, HKaretworld, volekkaret, volimkaret
Deterministic variables: const
Sample size: 41
Log Likelihood: -1784.335
Roots of the characteristic polynomial:
1.013 0.9647 0.6144 0.6144 0.5829 0.5829 0.5628 0.5628 0.3077 0.3077
Call:
VAR(y = karet[5:47, c(2, 6, 7, 10, 11)], p = 2, type = "const")

Estimation results for equation Areal:
=====
Areal = Areal.l1 + HRSheet.l1 + HKaretworld.l1 + volekkaret.l1 + volimkaret.l1 + Areal.l2 + HRSheet.l2 + HKaretworld.l2 + volekkaret.l2 + volimkaret.l2 + const

      Estimate Std. Error t value Pr(>|t|)
Areal.l1      8.281e-01  2.021e-01  4.098 0.000291 ***
HRSheet.l1    -2.002e+00  6.963e+00  -0.288 0.775664
HKaretworld.l1 -2.610e+04  2.698e+04  -0.967 0.341104
Volekkaret.l1  1.837e-01  1.022e-01  1.799 0.082162 .
Volimkaret.l1 -5.007e+00  2.335e+00  -2.145 0.040188 *
Areal.l2      2.052e-01  2.128e-01  0.964 0.342559
HRSheet.l2    -7.513e+00  6.533e+00  -1.150 0.259176
HKaretworld.l2  6.653e+04  2.762e+04  2.409 0.022350 *
Volekkaret.l2 -1.882e-01  9.963e-02  -1.889 0.068657 .
Volimkaret.l2 -3.045e+00  2.113e+00  -1.441 0.159883
const         -4.087e+04  8.562e+04  -0.477 0.636540
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56100 on 30 degrees of freedom
Multiple R-Squared: 0.989, Adjusted R-squared: 0.9854
F-statistic: 270.2 on 10 and 30 DF, p-value: < 2.2e-16

```

### Pengujian Asumsi VAR(2) type 'constant'

Selanjutnya dilakukan pengujian asumsi non autokorelasi, normalitas, dan homoskedastisitas pada sisaan model VAR terbaik. Untuk data karet akan dilakukan pengujian sisaan pada dua model terbaik VAR (2) type 'constant'.

Pemeriksaan autokorelasi residual model menggunakan fungsi "serial.test" yang di dalamnya dilakukan pengujian Portmanteau-and Breusch-Godfrey test. Nilai p-value lebih besar dari 0,05 maka sisaan saling bebas atau asumsi non autokorelasi terpenuhi. Pengujian Jarque-Bera tests untuk menguji kenormalan, hasil pengujian menunjukkan Nilai p-value lebih kecil dari 0,05, namun karena jumlah data yang digunakan cukup banyak, maka series tersebut dapat dianggap normal. Pemeriksaan heteroskedastisitas model menggunakan fungsi "arch.test" yang di dalamnya dilakukan pengujian ARCH-LM tests. Nilai p-value lebih besar dari 0,05 maka ragam sisaan homogen atau asumsi homoskedastisitas terpenuhi.

Tabel 36. Ouput Pengujian Asumsi VAR(2) type=constant

```

Portmanteau Test (asymptotic)
data: Residuals of VAR object varhsheet2
Chi-squared = 366.35, df = 350, p-value = 0.2631
$JB

      JB-Test (multivariate)
data: Residuals of VAR object varhsheet2
Chi-squared = 143.97, df = 10, p-value < 2.2e-16

      Skewness only (multivariate)
data: Residuals of VAR object varhsheet2
Chi-squared = 34.022, df = 5, p-value = 2.357e-06

      Kurtosis only (multivariate)
data: Residuals of VAR object varhsheet2
Chi-squared = 109.95, df = 5, p-value < 2.2e-16

      ARCH (multivariate)
data: Residuals of VAR object varhsheet2
Chi-squared = 525, df = 1350, p-value = 1
    
```

### Ramalan Data Training, Testing, Penghitungan MAPE, dan Plot

Selanjutnya dilakukan peramalan data, baik untuk data training maupun untuk data testing sekaligus dilakukan penghitungan MAPE. MAPE (Mean Absolute Percentage Error) merupakan rata-rata dari keseluruhan persentase kesalahan (selisih) antara data aktual dengan data hasil peramalan. Untuk menguji suatu model lebih baik dengan model yang lain, maka dilakukan pengujian model dengan membandingkan Nilai MAPE baik untuk data training maupun data testing. Data Testing hasil ramalan luas areal dengan VAR(p) nilai  $p=2$  type=constant, menghasilkan MAPE =3,63%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data testing adalah 3,63%. Data training hasil ramalan luas areal dengan nilai  $p=2$  type=constant menghasilkan MAPE =1,24%. Model VAR ini menunjukkan Ketika menggunakan data training sangat baik, terlihat dari MAPE yang kecil yaitu hanya sebesar 1,25%, namun Ketika digunakan untuk melakukan estimasi maka MAPE melonjak menjadi 3,63%, artinya kemampuan dalam meramalkan kurang baik.

Tabel 37. Pengujian Nilai MAPE untuk Model VAR(2) type=constant

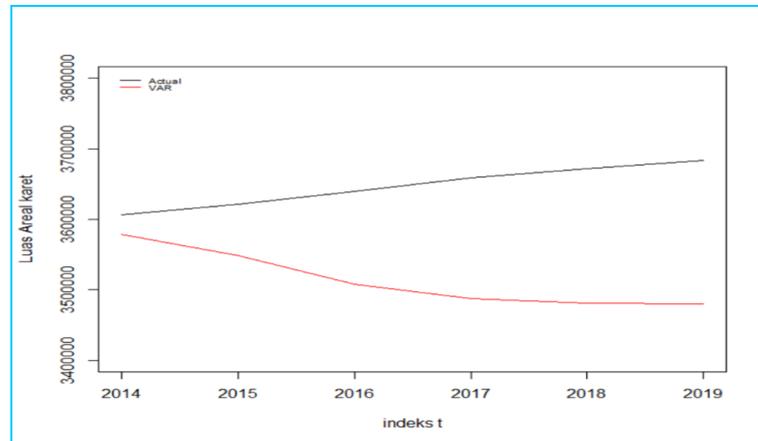
MAPE DATA Testing					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.777	2.394	4.146	3.633	5.070	5.537

MAPE DATA Training					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02895	0.41221	0.91816	1.24658	1.94552	3.68952

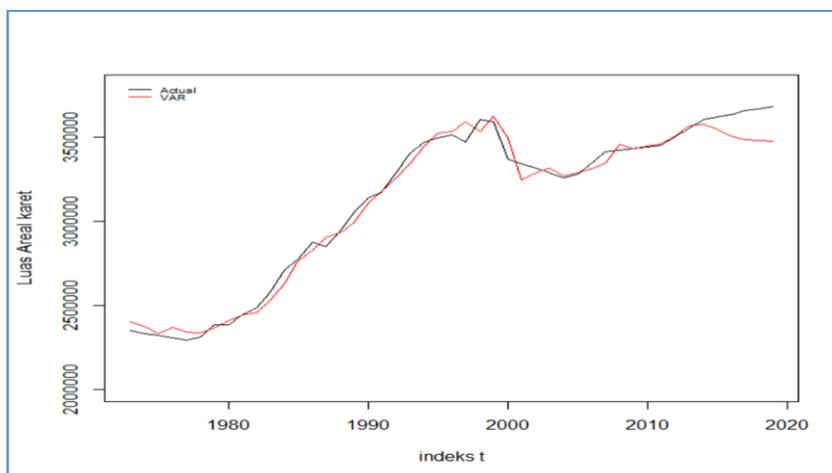
Dari grafik di bawah dapat dilihat bahwa pergerakan ramalan pada data testing tidak mengikuti pergerakan data asli/aktual. Sehingga mungkin model VAR (2) type “constant” kemampuan dalam meramlakan perlu dipertimbangkan. Hasil peramalan dengan model VAR (2) ini cenderung lebih rendah dari data actual. Pada tahun 2014 data actual sedikit lebih

tinggi dari data estimasi dengan VAR (2), pada tahun 2015 beda antara data actual dan data estimasi dengan VAR(2) ini semakin lebar, begitu juga tahun 2016 dan 2017, semakin jauh perbedaan antara data actual dan hasil estimasi. Untuk tahun 2018 dan 2019 perbedaan antara data estimasi dan data actual semakin lebar lagi. Untuk data testing ini rata-rata penyimpangannya adalah sebesar 3,63%. Dari segi besaran MAPE sebenarnya model ini sudah cukup baik karena MAPE dibawah 5%.



Gambar 16. Plot Ramalan dan Aktual Data Testing VAR (2) Type “constant”

Jika plot antara data testing dan data training digabungkan maka bentuk plotnya seperti Gambar 16. Untuk data tahun 1971 - 2014 atau data training plot sangat baik, karena antara data actual dan estimasi dengan model VAR (2) hampir selalu berimpit plotnya, sehingga MAPE akan kecil. MAPE hasil pengujian untuk data training adalah sebesar 1,25%, suatu nilai yang cukup kecil karena rata-rata penyimpangan hanya 1,25%, artinya model cukup akurat. Namun plot tahun 2014 – 2019 menunjukkan data actual dengan data estimasi dengan Model VAR (2) ini mulai agak renggang, dimana hasil estimasi rata-rata lebih rendah dari data actual. Hasil MAPE data testing ini, menunjukkan nilai yang lebih besar dari MAPE training yaitu sebesar 3,63%.



Gambar 17. Plot Ramalan dan Aktual Data Training dan Testing VAR (2) Type “constant”

## Pengepasan Model untuk Seluruh Data dan Peramalannya

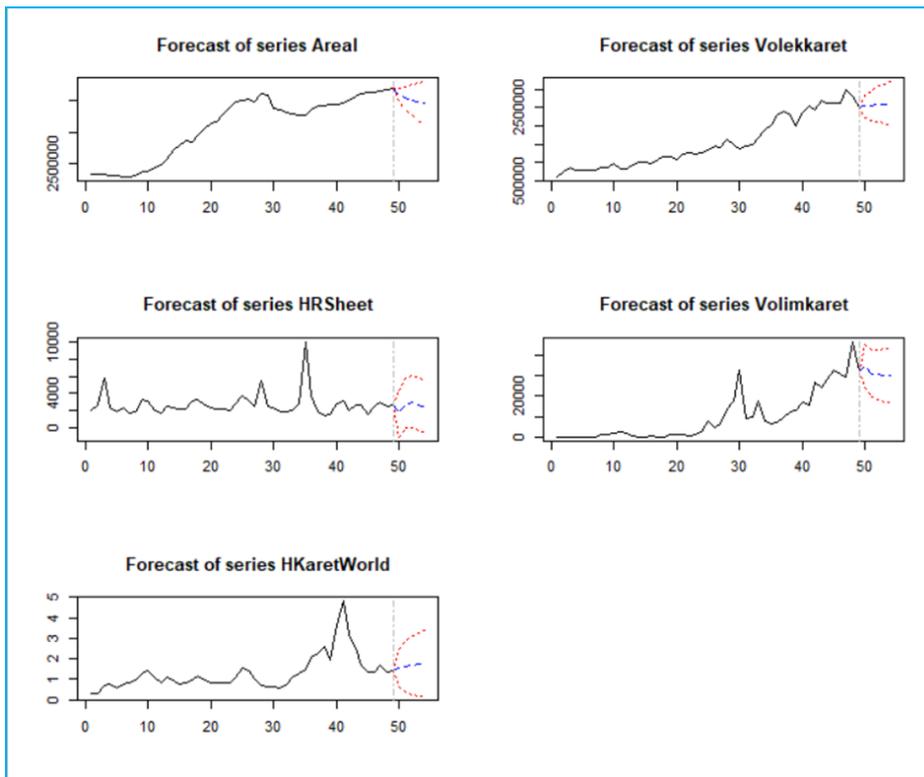
Model VAR terbaik dari hasil perbandingan nilai MAPE adalah model VAR (2) type “constant”, sehingga dapat dilanjutkan pengepasan model pada keseluruhan data. Setelah dilakukan pengepasan model, maka dilakukan estimasi luas areal karet untuk 5 tahun kedepan yaitu tahun 2020 -2024.

Peramalan dengan model VAR (2) type “constant” menggunakan keseluruhan data menunjukkan luas areal karet tahun 2020 sebesar 3,58 juta ha, turun pada tahun 2021 menjadi 3,54 juta ha, terus turun sehingga pada tahun 2024 luas areal karet diperkirakan mencapai 3,45 juta ha.

Tabel 38. Hasil Peramalan Luas areal Karet Tahun 2020-2024, Model VAR(2) Type=both

[1] 3584557 3539474 3509092 3483642 3459054
---

Keunggulan model VAR ini pada saat dilakukan permalan maka semua variabel dalam system akan ikut diramalkan. Hasil peramalan bukan hanya target variabel, tetapi semua variabel yang masuk dalam sistem. Hasil permalan 5 tahun ke depan untuk variabel luas areal menunjukkan luas areal akan terus menurunt secara linier. Sebaliknya harga karet sheet dalam negeri cenderung konstan meskipun ada fluktuasi, sementara harga karet dunia diperkirakan akan perlahan terus meningkat. Volume ekspor karet cenderung konstan, sebaliknya volume impor karet juga diperkirakan akan cenderung turun.



Gambar 18. Hasil Estimasi untuk Semua Vraiabel dalam Sistem Model VAR(2) Constant

Untuk model VAR (2) type “constant”, pergerakan hasil ramalan luas areal karet cenderung mengalami penurunan. Dari plot juga terlihat bahwa data cenderung turun. Jadi berdasarkan nilai MAPE, maka model terbaik yang terpilih untuk peramalan data luas areal karet adalah model VAR (2) type “constant” karena memberikan nilai MAPE training dan testing yang lebih kecil sekaligus memberikan plot ramalan yang cenderung menurun, meskipun tidak sesuai dengan pergerakan data historis sebelumnya.

Jika dibandingkan dengan tahun 2019 maka luas areal karet hasil estimasi tahun 2020 akan turun sebesar 2,69%, sehingga menjadi sekitar 3,58 juta hektar. Hal ini didukung dari realitas luas areal karet yang hanya sedikit saja mengalami peningkatan, karena harga karet dunia yang masih cenderung turun. Penurunan luas areal karet hasil estimasi selama 5 tahun kedepan adalah sekitar 0,89%. Meskipun terjadi penurunan hal ini bisa saja terjadi karena beberapa tahun belakangan ini harga karet alam dunia cenderung turun dan stabil di harga rendah. Beberapa petani karet telah mengkonversi lahannya ke tanaman yang lebih menguntungkan seperti kelapa sawit atau tanaman perkebunan lainnya.

Tabel 39. Hasil Estimasi Luas areal Karet Nasional dengan Model VAR(2), Type= constant

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	3.683.482	
2020	3.584.557	-2,686
2021	3.539.474	-1,258
2022	3.509.092	-0,858
2023	3.483.642	-0,725
2024	3.459.054	-0,706
Rata Rata Pertumbuhan		-0,887

Tahun 2019 : Angka Sementara (Ditjenbun)

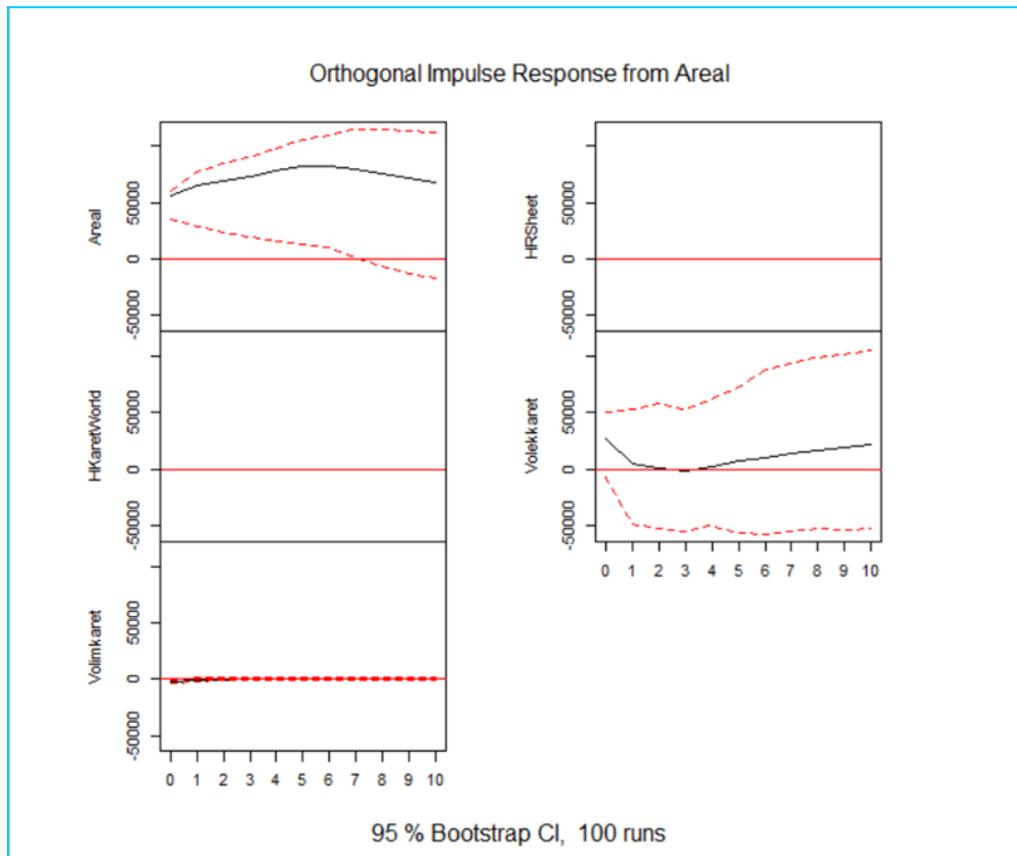
Tahun 2020 - 2024 : Estimasi Berdasarkan Model Arima

### Interpretasi Impulse Response Function (IRF) dan Variance Decomposition

Setelah diperoleh model terbaik, untuk permodelan VAR dapat dilakukan interpretasi lebih mendalam terkait Impulse Response Function (IRF) dan Variance Decomposition. Berikut interpretasi ketiga hal di atas untuk model terbaik VAR (2) type “both”. Impulse Response Function akan menjelaskan bagaimana perubahan atau gejolak yang terjadi pada suatu variabel/peubah di tahun tertentu akan berdampak pada variabel/peubah lain di tahun tertentu dan tahun-tahun setelahnya.

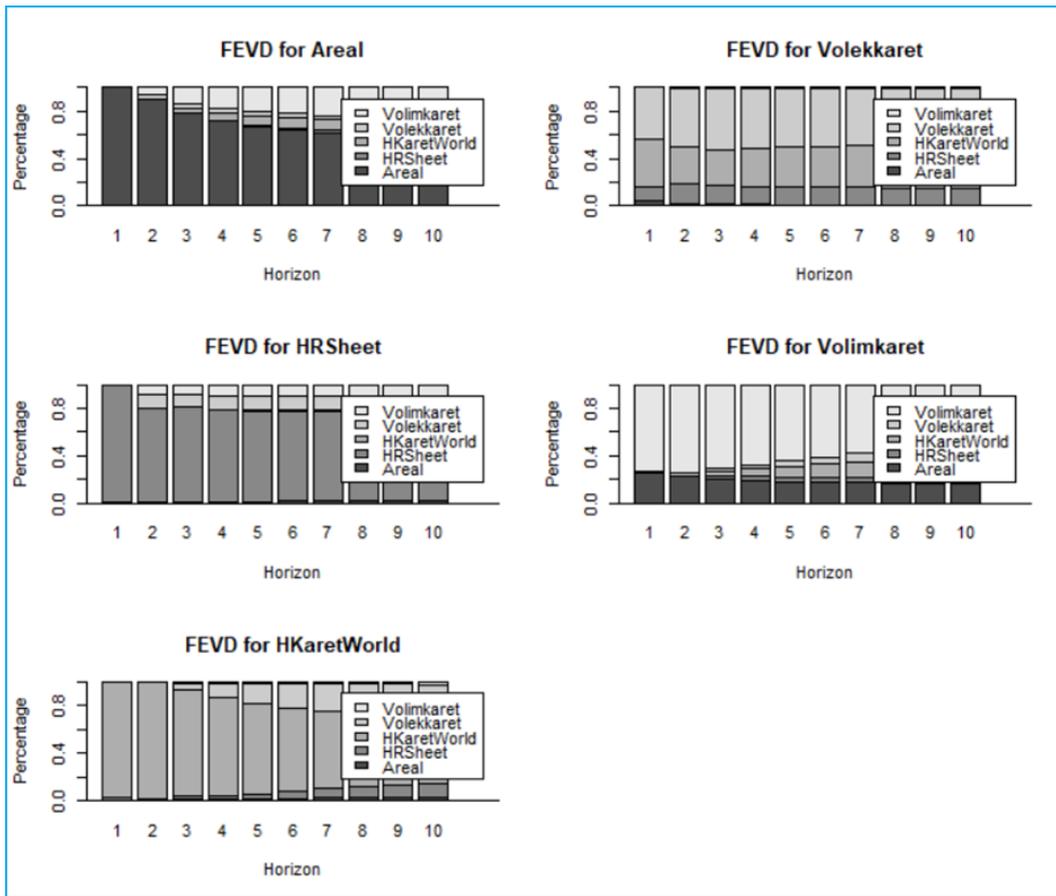
Dari grafik Impulse Response Function luas areal model terbaik VAR (2) type “constant”, dapat dilihat bahwa jika terjadi perubahan pada luas areal di tahun tertentu maka akan berdampak pada luas areal itu sendiri sampai 7 tahun ke depan. Dampak tersebut akan hilang setelah 7 tahun. Sedangkan dampak perubahan luas areal tidak berdampak pada harga sheet, luas areal, harga karet dunia, dan volume impor karet, namun masih sedikit berdampak

pada volume ekspor kedepan. Perubahan luas areal berdampak pada volume ekspor sampai dengan satu tahun dan akan hilang dampaknya setelah melewati tahun pertama tersebut.



Gambar 19. Impuls Respon Beberapa Variabel Terhadap Luas areal Model VAR(2) Constant

**Variance Decomposition** atau dekomposisi keragaman digunakan untuk melihat variabel apa saja yang mempengaruhi komposisi keragaman suatu variabel/peubah. Dari grafik dekomposisi keragaman model terbaik VAR (2) type “constant” di atas dapat dilihat bahwa komposisi luas areal pada tahun pertama dipengaruhi sepenuhnya oleh luas areal itu sendiri. Pada tahun kedua, komposisi luas areal 90% dipengaruhi oleh luas areal itu sendiri, 5% dipengaruhi oleh volume ekspor, dan 5% dipengaruhi oleh impor karet. Pada tahun ketiga, komposisi luas areal dipengaruhi 85% oleh luas areal itu sendiri, 10% dipengaruhi oleh volume ekspor, dan 5% dipengaruhi oleh harga karet dunia dan volume impor karet. Semakin bertambahnya tahun, pengaruh luas areal karet terhadap keragaman luas areal karet sendiri semakin berkurang diikuti makin tingginya pengaruh volume ekspor, volume impor, harga karet dunia dan harga karet sheet local.



Gambar 20. Dekomposisi Keragaman Beberapa Variabel Untuk Model VAR(2) Both

## E. Pemilihan Model Terbaik Estimasi Luas Areal Karet Nasional

Salah satu dasar penentuan model terbaik dapat dilakukan dengan membandingkan nilai MAPE untuk data testing dan training yaitu dengan memilih nilai MAPE yang paling kecil, terutama untuk data testing. Selain MAPE yang terkecil, pola pergerakan ramalan juga harus diperhatikan. Pilihlah plot yang paling berhimpit/bersesuaian dengan data asli/aktual atau dengan kata lain performa hasil ramalan seiring dengan data historisnya.

Berdasarkan data historis yang ada luas areal karet nasional berfluktuasi, luas areal tahun 2015 sebesar 3,62 juta ha atau turun 0,41%. Pada tahun 2016 dan 2017 luas areal karet nasional meningkat masing-masing sebesar 0,50% dan 0,55%, sehingga luas areal karet tahun 2017 menjadi sebesar 3,66 juta ha. Pada tahun 2018 luas areal karet nasional kembali naik sebesar 0,34%, kemudian pada tahun 2019 kembali naik sebesar 0,33%. Rata-rata pertumbuhan luas areal karet nasional selama 5 tahun terakhir atau tahun 2015 – 2019 sebesar 0,42%.

Tabel 40. Luas areal Karet Nasional Tahun 2015 – 2020

Tahun	Luas Areal Karet (Ha)	
	Indonesia	Pertumbuhan (%)
2015	3.621.103	0,41
2016	3.639.049	0,50
2017	3.659.090	0,55
2018	3.671.387	0,34
2019*)	3.683.482	0,33
2020 **)	3.694.716	0,30
Rata-rata pertumbuhan (%)		
2015 - 2019		0,42

Sumber : Direktorat Jenderal Perkebunan, diolah Pusdatin

Keterangan : 2019 Angka Sementara

2020 Angka Estimasi

Untuk menyusun angka estimasi luas areal karet telah dilakukan uji coba dengan 4 (empat) model. Model yang pertama adalah model time series atau ARIMA, model terbaik untuk ARIMA adalah pada orde ARIMA (1,1,0). Untuk model estimasi luas areal karet nasional dengan ARIMA (1,1,0) menghasilkan MAPE untuk data training sebesar 1,34% dan MAPE untuk data testing sebesar 0,69%. Hal ini menunjukkan bahwa kemampuan untuk melakukan estimasi dengan model ARIMA ini rata-rata akan mengalami kesalahan sekitar 0,69% lebih tinggi atau 0,69% lebih rendah. Hasil estimasi dengan model ARIMA pertumbuhan 5 tahun kedepan relatif lambat, yaitu hanya 0,031%/tahun. Hal ini berbeda dengan data 5 tahun ke belakang (2014 – 2019), dimana pertumbuhan luas areal mencapai 0,42% per tahun.

Model yang kedua adalah menggunakan regresi linier sebagai variabel bebas adalah waktu atau tahun. Untuk model regresi linier sederhana menghasilkan MAPE yang cukup besar yaitu 5,38% untuk data training dan 5,74% untuk data testing. Model regresi linier kurang tepat untuk mengestimasi luas areal karet karena menghasilkan MAPE yang cukup besar. Untuk model regresi kuadratik dengan variabel bebas waktu (tahun) menghasilkan MAPE yang lebih baik, yaitu sebesar 6,61% untuk data training dan 4,25% untuk data testing. MAPE untuk model regresi kuadratik lebih kecil untuk data testing, namun jika dilihat dari hasil estimasi angka pertumbuhan angka estimasi kurang realistis karena selama 5 tahun ke depan rata-rata pertumbuhan luas areal karet turun 0,39% per tahun, bertentangan dengan data historisnya.

Metode estimasi yang ketiga adalah dengan model fungsi transfer, untuk melakukan estimasi luas areal karet dengan variabel bebas adalah harga karet sheet. Untuk model fungsi transfer ini menghasilkan MAPE data training 0,23%, sementara untuk MAPE data testing sebesar 0,19%. Model fungsi transfer ini menghasilkan MAPE yang lebih kecil dibandingkan model ARIMA atau Model Regresi, sehingga model fungsi transfer lebih akurat dalam melakukan estimasi. Hasil estimasi juga menunjukkan angka yang lebih realistis, dengan

angka estimasi tahun 2020 sebesar 3,694 juta hektar, atau naik 0,30%. Disamping itu untuk estimasi 5 tahun kedepan angka pertumbuhan sebesar 0,205%/tahun, sementara angka pertumbuhan 5 tahun sebelumnya sebesar 0,42%/tahun, pertumbuhan ini mendekati pertumbuhan data historisnya (5 tahun kebelakang).

Tabel 41. Perbandingan MAPE Model Arima, Regresi, Fungsi Transfer dan VAR

No	Urain Model	MAPE		Hasil Estimasi Produksi Karet (Ton)					Pertumbuhan (%) Tahun 2020 - 2024
		Training	Testing	2020	2021	2022	2023	2024	
1	ARIMA (1,1,0)	1,34	0,69	3.689.094	3.691.698	3.692.906	3.693.467	3.693.727	0,031
2	Regresi Linier	5,38	5,74	3.976.231	4.010.475	4.044.719	4.078.964	4.113.208	0,850
	Regresi Kuadratik	6,61	4,25	3.467.337	3.456.242	3.443.583	3.429.362	3.413.577	-0,390
3	Arima (1,1,1) xreg=HRSheet	0,23	0,19	3.694.629	3.703.988	3.712.086	3.719.032	3.725.004	0,205
4	Var (2) p=2 type=constant	1,25	3,63	3.584.557	3.539.474	3.509.092	3.483.642	3.459.054	-0,887
Angka Sementara Tahun 2019 :		3.683.482 hektar							
Rata-rata Pertumbuhan 2015 - 2019 :		0,42%							

Untuk model estimasi yang terakhir adalah dengan model VAR (Vector Auto Regressive). Untuk model VAR ini menggunakan 5 variabel yaitu luas areal, harga karet sheet, harga karet dunia, volume ekspor dan volume impor karet. Model yang terbaik untuk Model VAR adalah nilai p=2 dan type="constant", p=2 artinya menggunakan variabel bebas sampai lag-2 dan ada konstanta. Estimasi luas areal karet dengan menggunakan model VAR ini menghasilkan ketelitian yang cukup tinggi yaitu MAPE untuk data training 1,25% dan MAPE untuk data testing 3,63%. MAPE untuk data testing ini model VAR lebih besar dibandingkan dengan model Fungsi Transfer atau Model Arima. Jika dibandingkan angka pertumbuhan luas areal karet antara hasil estimasi 5 tahun kedepan dengan rata-rata pertumbuhan -0,887% per tahun, hasil ini berlawanan dengan angka pertumbuhan 5 tahun terakhir yaitu sebesar 0,42% per tahun. Angka hasil estimasi untuk luas areal karet nasional tahun 2020 sebesar 3,58 juta hektar, sementara untuk angka sementara tahun 2019 sebesar 3,68 juta hektar atau turun sebesar 2,69%.

Berdasarkan Tabel 41 diatas, untuk data training dan data yang paling baik adalah yang memiliki MAPE terkecil, sehingga Model Fungsi Transfer ARIMA (1,1,1) dengan faktor input harga karet sheet merupakan model yang terbaik untuk menyusun angka Estimasi Luas Areal Karet nasional. Disamping faktor MAPE, hasil estimasi 5 tahun kedepan (2020 – 2024) menunjukkan pertumbuhan rata-rata luas areal karet nasional 0,21% per tahun, paling mendekati dari data aktual 5 tahun kebelakang yaitu pertumbuhan luas areal karet tahun 2015 – 2019 sebesar 0,42% per tahun. Berdasarkan MAPE dan angka pertumbuhan luas areal maka model Fungsi Transfer menjadi model terbaik untuk meramalkan luas areal karet nasional.

## KESIMPULAN

Untuk meningkatkan akurasi dalam penyusunan angka estimasi, maka dilakukan pengembangan metode estimasi luas areal karet nasional. Metode estimasi data perkebunan selama ini menggunakan model *Single Smoothing Exponential (SSE)* atau menggunakan *Double Smoothing Exponential (DSE)*. Meskipun dua metode tersebut dapat menghasilkan angka estimasi yang cukup baik, namun masih perlu melakukan pengembangan model alternatif yang diharapkan lebih akurat.

Untuk analisis ini data dibagi menjadi 2 kelompok, yaitu data training tahun 1967 – 2013, dan data testing tahun 2014 – 2019. Data training untuk penyusunan model, sedangkan data testing untuk uji coba model dalam melakukan estimasi 6 tahun kedepan. Untuk estimasi luas areal karet alternatif model pertama adalah Model ARIMA. Model ARIMA terbaik adalah ARIMA (1,1,0), menghasilkan MAPE untuk data training 1,34%, dan MAPE data testing 0,69%. Model kedua adalah model Regresi Linier sederhana dan Model Regresi Kuadratik. Model regresi linier menghasilkan MAPE data training sebesar 5,38% dan MAPE data testing 5,74%. Model regresi kuadratik menghasilkan MAPE data training sebesar 6,61%, MAPE data testing 4,25%. Untuk model yang ketiga dengan menggunakan Fungsi Transfer dengan variabel input harga karet sheet, menghasilkan MAPE untuk data training sebesar 0,23% dan MAPE data testing 0,19%. Untuk model yang keempat model VAR(2) type 'constant' ada pengaruh konstanta, menghasilkan MAPE data training 1,25% dan data MAPE data testing 3,63%.

Berdasarkan perbandingan besarnya MAPE baik data testing maupun data training dan hasil estimasi luas areal 5 tahun kedepan, maka model terbaik yang terpilih adalah model Fungsi Tranfer ARIMA(1,1,1) dengan factor input harga karet sheet karena menghasilkan akurasi yang cukup tinggi sehingga MAPE rata-rata data testing sebesar 0,19%. Hasil estimasi luas areal karet nasional untuk model Fungsi Tranfer ARIMA(1,1,1) dengan factor input harga karet sheet untuk tahun 2020 sebesar 3.694.629 hektar, tahun 2021 sebesar 3.703.988 hektar, tahun 2022 sebesar 3.712.086 hektar, tahun 2023 sebesar 3.719.032 hektar, dan tahun 2024 sebesar 3.725.004 hektar. Laju pertumbuhan estimasi luas areal karet nasional selama 5 tahun kedepan (2020 – 2024) rata-rata 0,21% per tahun.

## DAFTAR PUSTAKA

- Anonim. 2013. Pedoman Pelaksanaan Pengelolaan Data Komoditas Perkebunan (PDKP). Dirjen Perkebunan – Kementerian Pertanian.
- Enders, W. 2010. Applied Econometric Time Series. USA: University of Alabama. Wiley, Third Edition.
- Fitriani, D.R, Darsyah, M.Y., & Wasono, R. 2013. Peramalan Fungsi Transfer pada Harga Emas Pasar Komoditi. Seminar Nasional Pendidikan Sains dan Teknologi, Fakultas MIPA, Universitas Muhammadiyah Semarang.
- Guha, B and Bandyopadhyay, G. 2016. Gold Price Forecasting Using ARIMA Model. Journal of Advanced Management Science Vol. 4, No. 2, March 2016
- Gujarati, D.N. and D.C. Porter, 2010. Dasar-dasar Ekonometrika. Jakarta: Penerbit Salemba Empat, Buku 2, Edisi 5.
- Kementerian Pertanian. 2019. Statistik Perkebunan Indonesia 2018-2020 (Karet). Jakarta: Direktorat Jenderal Perkebunan.
- M. Firdaus 2020. Aplikasi Ekonometrika dengan E-Views, Stata dan R. Institut Pertanian Bogor (IPB) Press.
- Montgomery DC, Johnson LA & Gardiner JS. 1990. Forecasting and Time Series Analysis. Singapore:Mc-Graw Hill.
- Myers R. 1994. Classical And Modern Regression with Applications. Boston: PWS – KENT Publishing Company.
- Ryan TP. 1997. Modern Regression Methods. New York,USA: John Wiley & Sons, INC



# KAJIAN MODEL PERAMALAN LUAS AREAL KELAPA SAWIT (*CRUDE PALM OIL/CPO*) DI INDONESIA:

## PENDEKATAN MODEL ARIMA, REGRESI, FUNGSI TRANSFER, DAN VAR

Lasmiyati

*Subdirektorat Statistik Tanaman Perkebunan-Badan Pusat Statistik  
Jalan Dr. Sutomo no. 6-8 Gd 4 Lt 4, Pasar Baru-Jakarta Pusat, Indonesia*

### ABSTRAK

*Crude Palm Oil (CPO)* atau minyak kelapa sawit merupakan bahan baku dari berbagai barang kebutuhan sehari-hari bagi masyarakat di dunia. Kebutuhan dunia akan CPO yang cukup tinggi sungguh sangat menguntungkan bagi Indonesia sebagai negara beriklim tropis yang cocok untuk pertumbuhan kelapa sawit. Guna merumuskan kebijakan terkait luas areal, sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Penelitian ini akan mengkaji empat metode yaitu metode *Auto-Regressive Integrated Moving Average (ARIMA)*, regresi, fungsi transfer dan *Vector Auto Regression (VAR)* dengan menggunakan R Studio guna melakukan pemodelan luas areal kelapa sawit di Indonesia. Metode yang dipilih guna melakukan peramalan data luas areal kelapa sawit adalah metode yang terbaik yang dikaji dari nilai *Mean Absolute Percentage Error (MAPE)* terkecil. Berdasarkan dari hasil nilai MAPE disimpulkan bahwa fungsi transfer adalah metode yang terbaik dengan MAPE 4,76% untuk luas areal sawit kelapa sawit Indonesia.

### ABSTRACT

*Crude Palm Oil (CPO) or palm oil is the raw material for various daily necessities for people in the world. The world's demand for CPO which is quite high is very beneficial for Indonesia as a tropical country that is suitable for the growth of oil palm. In order to formulate policies related to area size, the availability of up-to-date data, and even forecasts for several periods in the future, is highly demanded. This study will examine four methods, namely the Auto-Regressive Integrated Moving Average (ARIMA) method, regression, transfer function and Vector Auto Regression (VAR) using R Studio to model the area of oil palm in Indonesia. The method chosen for forecasting oil palm area data is the best method which is assessed from the smallest MAPE (Mean Absolute Percentage Error) value. Based on the results of the MAPE value, it is concluded that the transfer function is the best method with a MAPE of 4.76% for the total area of Indonesian oil palm oil.*

## PENDAHULUAN

*Crude Palm Oil* (CPO) atau minyak kelapa sawit merupakan bahan baku dari berbagai barang kebutuhan sehari-hari bagi masyarakat di dunia. Kebutuhan dunia akan CPO yang cukup tinggi, sungguh sangat menguntungkan bagi Indonesia sebagai negara beriklim tropis yang cocok untuk pertumbuhan kelapa sawit. Penggunaannya sebagai bahan baku pada berbagai produk, misalnya sabun, kosmetik, minyak goreng, mentega, dll menjadikannya terus dibutuhkan ditengah polemiknya sebagai salah satu kandidat perusak lingkungan. Oleh karena itu, menjadi sangat penting untuk mempelajari estimasi data luas areal kelapa sawit sebagai dasar penentu berbagai kebijakan terkait.

Guna merumuskan kebijakan terkait luas areal kelapa, sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Makalah ini akan mencoba membandingkan beberapa metode statistik yang telah dikembangkan oleh para ahli, diantaranya metode *Auto-Regressive Integrated Moving Average* (ARIMA), Regresi, Fungsi Transfer dan *Vector Auto Regression* (VAR) dalam melakukan pemodelan dan peramalan luas areal kelapa sawit di Indonesia.

Oleh karena itu, tujuan dari disusunnya kegiatan ini adalah:

- c. Melakukan analisis dan peramalan data luas areal kelapa sawit menggunakan model ARIMA, Regresi, Fungsi Transfer dan VAR di Indonesia.
- d. Membandingkan metode tersebut dalam memperoleh ramalan data luas areal kelapa komoditas sawit.
- e. Menentukan metode terbaik dalam meramal data luas areal kelapa komoditas sawit di Indonesia.

## BAHAN DAN METODE

Data yang digunakan dalam melakukan pemodelan luas areal kelapa sawit adalah data *series* luas areal kelapa sawit dalam satuan hektar di Indonesia tahun 1967 s.d. 2018, produksi sawit (CPO) dalam satuan ton di Indonesia tahun 1967 s.d. 2018, harga CPO dunia dalam satuan US\$ tahun 1967 s.d. 2018, dan volume ekspor serta volume impor CPO dalam satuan ton di Indonesia tahun 1981-2018. Peubah yang diasumsikan mempengaruhi besaran luas areal kelapa sawit untuk model regresi adalah peubah waktu. Peubah yang diasumsikan mempengaruhi besaran luas areal kelapa sawit untuk Fungsi Transfer adalah harga CPO dunia. Sedangkan untuk pemodelan VAR luas areal sawit, peubah yang diasumsikan mempengaruhi adalah harga CPO dunia, produksi CPO Indonesia, volume ekspor CPO, dan volume impor CPO. Harga CPO dunia diperkirakan berpengaruh karena kenaikan harga akan mendorong petani meningkatkan luas areal. Peningkatan volume ekspor dan penurunan impor diperkirakan akan membangkitkan minat petani untuk mengusahakan kelapa sawit.

Peramalan data luas areal kelapa sawit dilakukan melalui pengujian dengan beberapa metode, yakni ARIMA, Regresi, Fungsi Transfer, dan VAR menggunakan R Studio.

Penjelasan metode yang digunakan dalam melakukan peramalan luas areal kelapa sawit di Indonesia adalah sebagai berikut:

- e. ARIMA

ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampainya. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

$Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t

$Y_{t-p}$  = data *time series* pada kurun waktu ke (t-P)

$\mu$  = suatu konstanta

$\theta_1 \theta_2 \dots \theta_p$  = parameter-parameter model

$\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

#### f. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih variabel.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

dengan

$y$  = variabel respon/dependen/terikat

$x_1, x_2, \dots, x_k$  = variabel prediktor/independen/bebas

$\beta_0$  = konstanta/intercept

$\beta_1, \beta_2, \dots, \beta_k$  = koefisien regresi variabel bebas

$\varepsilon$  = error/residu/sisaan

#### g. Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau  $Y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $Y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $X_t$ ) dengan deret output tersebut.

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t$$

Dimana:

- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$

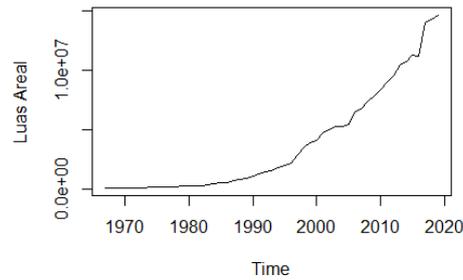
#### h. VAR

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen, karena pada kenyataannya suatu

variabel dapat bertindak sebagai endogen atau dependen dari suatu variabel yang lainnya (Gujarati, Damodar. N dan Porter, Dawn. C. 2009).

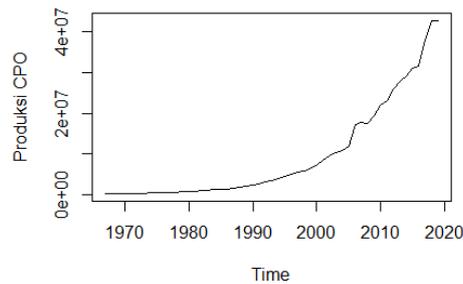
## HASIL DAN PEMBAHASAN

### h. Eksplorasi Data



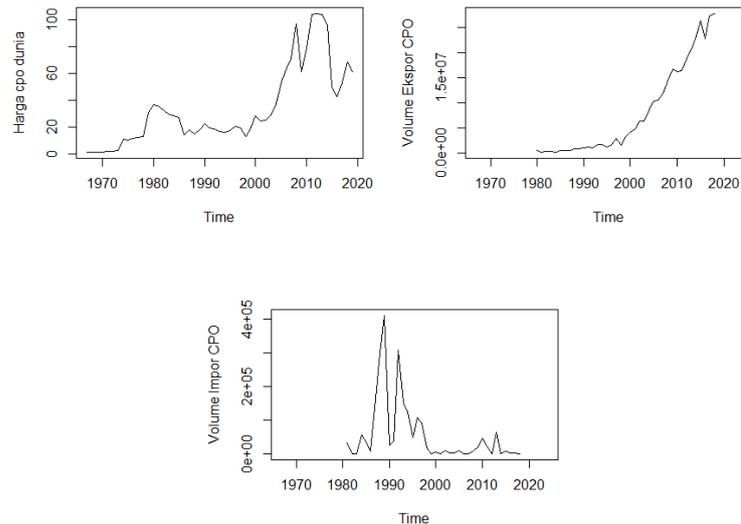
**Gambar 1.** Perkembangan Luas Areal Sawit (Ha), 1967-2018

Luas areal sawit di Indonesia selama periode 1967 – 2018 juga cenderung mengalami trend peningkatan dari tahun ke tahun dengan rata-rata pertumbuhan sebesar 10,18% per tahun (Gambar 1). Pertumbuhan tersebut sedikit lebih rendah dibandingkan dengan pertumbuhan produksi sawit (CPO).



**Gambar 2.** Perkembangan Produksi Sawit (CPO) (Ton), 1967- 2018

Produksi sawit (CPO) di Indonesia selama periode 1967 – 2018 cenderung mengalami trend peningkatan dari tahun ke tahun dengan rata-rata pertumbuhan sebesar 11,49% per tahun (Gambar 2).



**Gambar 3.** Perkembangan Harga CPO dunia (US\$) 1967-2018, Volume Eskpor Karet (Ton) 1980-2018, dan Volume Impor Karet (Ton) 1981-2018

Pergerakan harga CPO dunia selama periode 1967 – 2018 sangat berfluktuasi dari tahun ke tahun dengan kecenderungan meningkat (Gambar 3). Rata-rata pertumbuhan harga CPO dunia sebesar 14,01% per tahun. Pergerakan volume ekspor CPO selama periode 1980 – 2018 cenderung menunjukkan trend peningkatan, sedangkan pergerakan volume impor karet selama periode 1981 – 2018 menunjukkan pergerakan yang cenderung tidak stabil.

#### i. ARIMA

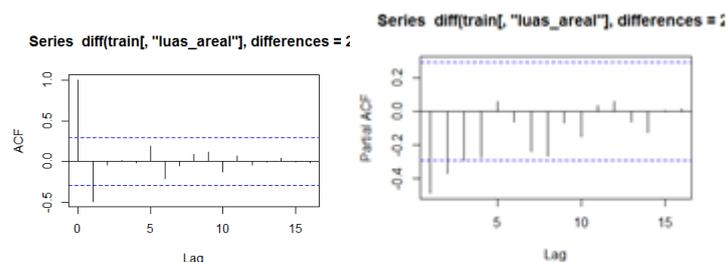
Dalam melakukan pemodelan luas areal kelapa sawit menggunakan ARIMA, data yang digunakan adalah periode 1967-2018. Periode data tersebut kemudian dipisahkan menjadi set data *training* dan *testing*. Panjang *series* data pada data *training* adalah tahun 1967-2012, sementara data *testing* adalah periode tahun 2013-2018. Data *training* digunakan untuk melakukan penyusunan model, sementara data *testing* digunakan untuk validasi model.

Uji kestasioneran data dilakukan menggunakan uji formal statistik *Augmented Dickey-Fuller* maupun secara visual dengan melihat plot ACF dan PACF. Tabel 1 menunjukkan bahwa data luas areal kelapa sawit stasioner pada *differencing* 2. Berdasarkan plot ACF dan PACF pada *differencing* 2 terlihat data sudah stasioner atau dengan kata lain rata-rata dan variannya konstan (Kirchgassner, Gebhard dan Wolters, Jurgen. 2007) (Gambar 1).

**Tabel 1.** Hasil Uji ADF

Kondisi Data	Hasil Uji ADF Luas Areal Sawit
Asli	Value of test-statistic is: 0.9624 16.017 14.4593 Critical values for test statistics: 1pct 5pct 10pct <b>Blm stasioner</b> tau3 -4.04 -3.45 -3.15 phi2 6.50 4.88 4.16 phi3 8.73 6.49 5.47
Diff 1	Value of test-statistic is: -2.5446 Critical values for test statistics: 1pct 5pct 10pct <b>Blm stasioner</b> tau1 -2.6 -1.95 -1.61
Diff 2	Value of test-statistic is: -8.4603 Critical values for test statistics:

	1pct	5pct	10pct	Stasioner
tau1	-2.6	-1.95	-1.61	



**Gambar 1.** Plot ACF dan Partial ACF data *differencing* luas areal kelapa sawit (CPO)

Selanjutnya dilakukan pendugaan model ARIMA dengan hasil seperti tertera dalam tabel 2.

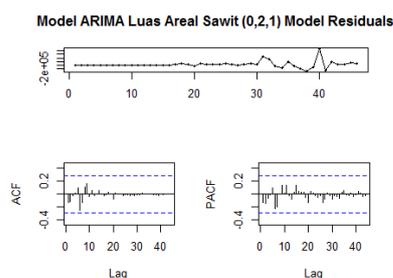
**Tabel 2.** Pendugaan Model ARIMA

---

```
Series: train[, "luas_areal"]
ARIMA(0,2,1)
sigma^2 estimated as 3.708e+10: log likelihood=-597.73
AIC=1199.47 AICc=1199.76 BIC=1203.04
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 43872.06 186187.6 96737.22 2.399218 4.757331 0.4598172 -0.1301828
```

---

Dari hasil *running* program model auto ARIMA menggunakan R Studio pada luas areal kelapa sawit diperoleh model ARIMA (0,2,1) dengan hasil pemeriksaan *residual* yang sudah cukup stasioner pada rata-rata sebagaimana terlihat pada Gambar 2. Selain itu, hasil pemeriksaan kenormalan *residual* menggunakan Ljung Box juga menunjukkan bahwa *residual* sudah secara signifikan mendekati distribusi normal pada lag 5,10,15,20,25,30,35, dan 40. Hal tersebut terlihat dari nilai p yang signifikan atau di atas taraf 5 persen (Tabel 3).



**Gambar 2.** Grafik Pemeriksaan Model *Residual* ARIMA pada Data *Training*

**Tabel 3.** Hasil Uji L-Jung test *Residual* ARIMA pada Data *Training*

Lag	Box-Ljung test Luas Areal Sawit	Hasil
5	X-squared=2.1904,df= 5,p-value = 0.8222	Sign normal
10	X-squared=8.9751,df= 10,p-value = 0.5345	Sign normal
15	X-squared=9.3678,df= 15,p-value = 0.8575	Sign normal
20	X-squared=10.095,df= 20,p-value = 0.9664	Sign normal
25	X-squared=10.204,df= 25,p-value = 0.9961	Sign normal
30	X-squared=10.414,df= 30,p-value = 0.9997	Sign normal
35	X-squared=10.454,df = 35,p-value = 1	Sign normal
40	X-squared=10.806,df = 40,p-value = 1	Sign normal

45	X-squared=10.878,df = 45,p-value = 1	Sign normal
----	--------------------------------------	-------------

Selanjutnya dilakukan peramalan data *testing* diikuti penghitungan nilai MAPE. Tabel 4 menunjukkan perbandingan nilai MAPE data *training* dan data *testing*, terlihat bahwa nilai MAPE luas areal kelapa cukup bagus yaitu 4,76% dan 5,7%.

**Tabel 4.** Perbandingan MAPE Data *Training* dan Data *Testing*

Luas Areal	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	43872.06	186187.6	96737.22	2.399218	<b>4.757331</b>	0.4598172	-0.1301828
Test set	594838.77	1027952.0	753592.73	4.281559	<b>5.698820</b>	3.5820221	NA

Hasil pengepasan model ARIMA (0,2,1) untuk luas areal pada keseluruhan data *training* dan *testing* diperoleh model sebagai tertera pada Tabel 5 dengan nilai MAPE 5,13% untuk luas areal kelapa sawit.

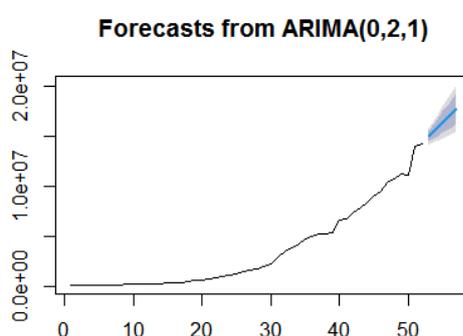
**Tabel 5.** Hasil Pengepasan Model ARIMA untuk Keseluruhan Data

Call: arima(x = total[, "luas_areal"], order = c(0, 2, 1))							
Coefficients:							
	ma1						
	-0.8484						
	s.e. 0.0576						
sigma^2 estimated as 1.678e+11: log likelihood = -717.74, aic = 1439.48							
Training set error measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	88326.91	401707.7	172325	3.120533	<b>5.134172</b>	0.6129376	-0.3720598

Hasil peramalan 5 tahun ke depan dengan menggunakan model ARIMA (0,2,1) untuk luas areal kelapa sawit tertera pada Tabel 6. dengan plot yang terlihat pada gambar 3. Dari plot terlihat bahwa nilai ramalan cenderung mengalami kenaikan pada 5 tahun mendatang.

**Tabel 6.** Peramalan Luas Areal ARIMA (0,2,1) 2019-2023

Point Forecast	Luas areal
53	15033414
54	15740479
55	16447543
56	17154607
57	17861672



**Gambar 3.** Plot peramalan Luas Areal ARIMA (0,2,1)

## j. Regresi

Peubah bebas yang digunakan dalam melakukan pemodelan data luas areal kelapa sawit adalah peubah waktu yang dalam hal ini adalah tahun. Panjang *series* data pada data *training*

adalah tahun 1967-2012, sementara data *testing* adalah periode tahun 2013-2018. Pendugaan model regresi awal dilakukan dengan menggunakan model regresi linier sederhana, namun karena nilai MAPE yang sangat tinggi (MAPE *training* 247,6, MAPE *testing*= 36,51), akhirnya dilakukan pendugaan model regresi kuadratik.

**Tabel 7.** Pendugaan Model Regresi Kuadratik dengan Peubah Waktu Data *Training*

---

```
Call:
lm(formula = luas_areal ~ t + t2, data = train)
Residuals:
    Min       1Q   Median       3Q      Max
-533219 -158130  76102  194078  307453
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  615753.3   108898.5    5.654 1.16e-06 ***
t            -142436.7    11193.3   -12.725 3.56e-16 ***
t2             7468.1      240.5    31.050 < 2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 257000 on 43 degrees of freedom
Multiple R-squared:  0.9923,    Adjusted R-squared:  0.992
F-statistic: 2783 on 2 and 43 DF,  p-value: < 2.2e-16
```

---

Dapat dilihat bahwa peubah waktu dan peubah waktu dalam kuadrat memiliki koefisien yang signifikan. *R-Squared* sebesar 99,2% menunjukkan bahwa peubah waktu tersebut dapat menjelaskan 99,2% keragaman dari luas areal kelapa sawit di Indonesai. Selain itu, hasil uji normalitas menunjukkan bahwa residual berdistribusi normal dengan tingkat kesalahan 5% dan memenuhi asumsi homoskedastisitas, meskipun hasil uji autokorekasi menunjukkan masih signifikan terjadinya masalah tersebut. Namun hal tersebut sementara diabaikan karena tujuan utama adalah mencari model untuk peramalan.

**Tabel 8.** Hasil Pengujian Asumsi Klasik Model Regresi Kuadratik Data *Training*

---

```
Luas Areal Sawit
Jarque-Bera test for normality
data:  reg2$residuals
JB = 4.6028, p-value = 0.055
Durbin-watson test
data:  reg2
DW = 0.45415, p-value = 2.099e-12
alternative hypothesis: true autocorrelation is greater than 0
studentized Breusch-Pagan test
data:  reg2
BP = 0.089847, df = 2, p-value = 0.9561
```

---

Sedangkan hasil penghitungan MAPE menunjukkan bahwa nilai MAPE data *testing* lumayan baik sekitar 6,47% untuk model luas areal kelapa sawit.

**Tabel 9.** MAPE Data Training dan Testing Model Regresi Kuadratik

---

Luas Areal Sawit		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Training		0.2232	4.8949	18.1719	<b>53.4285</b>	79.0556	481.9535
Testing		2.443	3.164	4.494	<b>6.466</b>	9.548	13.415

---

Hasil pengepasan model menggunakan keseluruhan data *training* dan *testing* menunjukkan bahwa keseluruhan koefisien signifikan dengan *R-Squared* sebesar 99,14% atau variabel peubah waktu dapat menjelaskan 99,14% keragaman luas areal kelapa sawit.

**Tabel 10.** Pengepasan dan MAPE Pendugaan Model Regresi Kuadratik dengan Peubah Waktu

---

```
Call:
lm(formula = luas_areal ~ t + t2, data = total)
Residuals:
    Min       1Q   Median       3Q      Max
-890047 -207584  12155  235270 1316926
```

---

---

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	780609.3	152741.2	5.111	5.29e-06 ***
t	-170425.9	13850.3	-12.305	< 2e-16 ***
t2	8189.0	262.6	31.180	< 2e-16 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 381400 on 49 degrees of freedom

Multiple R-squared: 0.9918, Adjusted R-squared: **0.9914**

F-statistic: 2947 on 2 and 49 DF, p-value: < 2.2e-16

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0827	4.2431	11.3695	<b>60.2228</b>	69.8164	637.7602

---

Hasil peramalan 5 tahun ke depan dengan menggunakan model regresi kuadratik tertera pada Tabel 11 dengan nilai ramalan cenderung mengalami kenaikan pada 5 tahun mendatang.

**Tabel 11.** Peramalan Luas Areal kelapa Sawit dengan Regresi Kuadratik 2019-2023

---

	Luas Areal kelapa Sawit
1	14061499
2	14750917
3	15456713
4	16178888
5	16917440

---

### i. Fungsi Transfer

Peubah yang diasumsikan dapat mempengaruhi gejala luas areal kelapa sawit adalah harga CPO dunia sehingga akan digunakan untuk melakukan analisis dengan model fungsi transfer. Panjang *series* data pada data training adalah tahun 1967-2012, sementara data *testing* adalah periode tahun 2013-2018. Dari pola data dan uji *Augmented Dickey-Fuller* (Tabel 12 dan Gambar 4) harga CPO dunia terlihat bahwa data belum stasioner sehingga akan dilakukan *differencing* terlebih dahulu sebelum digunakan dalam permodelan. Hasil pengujian deret *differencing* (Tabel 13) menunjukkan bahwa data *differencing* secara nyata sudah stasioner (nilai uji *mutlak* > nilai kritis *mutlak*), oleh karena itu bisa dilanjutkan dengan identifikasi model ARIMA untuk harga CPO dunia sebagai peubah input.

**Tabel 12.** Uji *Augmented Dickey-Fuller* Harga CPO Dunia

---

Residual standard error: 9.532 on 40 degrees of freedom

Multiple R-squared: 0.08813, Adjusted R-squared: 0.01974

F-statistic: 1.289 on 3 and 40 DF, p-value: **0.2915**

Value of test-statistic is: **-0.3169** 2.0409 1.351

Critical values for test statistics:

	1pct	5pct	10pct
tau3	<b>-4.15</b>	-3.50	-3.18
phi2	7.02	5.13	4.31
phi3	9.31	6.73	5.61

---

**Tabel 13.** Uji *Augmented Dickey-Fuller* untuk *differencing* Harga CPO dunia

---

Residual standard error: 9.5 on 39 degrees of freedom

Multiple R-squared: 0.6205, Adjusted R-squared: 0.5914

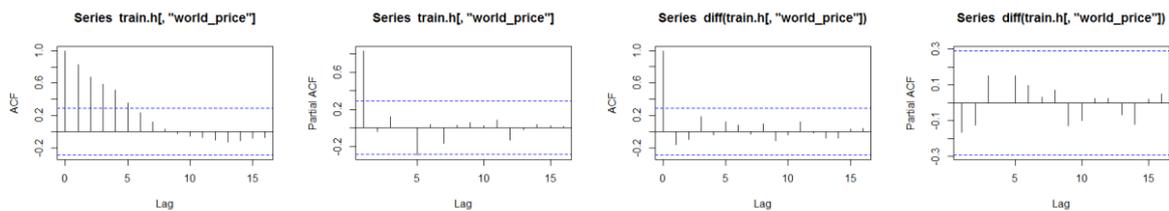
F-statistic: 21.26 on 3 and 39 DF, p-value: **2.524e-08**

Value of test-statistic is: **-5.7704** 11.1347 16.6727

Critical values for test statistics:

	1pct	5pct	10pct
tau3	<b>-4.15</b>	-3.50	-3.18
phi2	7.02	5.13	4.31
phi3	9.31	6.73	5.61

---



**Gambar 4.** ACF dan PACF Data Asli dan *Differencing* Harga CPO Dunia

Pendugaan model ARIMA dilakukan dengan memperhatikan hasil uji ADF dan pola dari ACF dan PACF yang signifikan. Model ARIMA deret input Harga CPO dunia data *training* ini akan digunakan untuk tahap *prewhitening*. Dengan menggunakan auto ARIMA untuk data training diperoleh model ARIMA (0,1,0) dengan MAPE 36,96 %. ARIMA (0,1,0) akan menghasilkan ramalan yang sama untuk periode ke depan sehingga dilakukan *trial error* dengan menggunakan ordo AR dan MA yang lain hingga diperoleh model ARIMA (1,1,0) dengan MAPE 19,04 %.

**Tabel 14.** Pendugaan Model ARIMA Deret Input Harga CPO Dunia Data *Training*

---

```

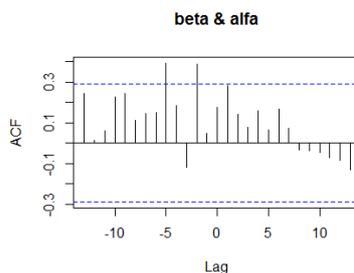
arima(x = train.h[, "world_price"], order = c(1, 1, 0))
Coefficients:
      ar1
    -0.0943
s.e.    0.1468
sigma^2 estimated as 93.14:  log likelihood = -165.87,  aic = 335.75
Training set error measures:

```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	2.464373	9.5452	5.770456	5.503966	<b>19.0398</b>	0.9842841	-0.07858562

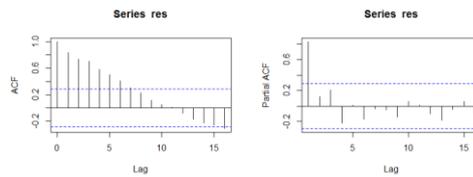
---

Selanjutnya dilakukan *prewhitening* dengan terlebih dahulu menentukan nilai alfa dan beta. Alfa merupakan nilai *residual* dari model arima variabel input yang dalam hal ini adalah harga CPO dunia, sedangkan beta merupakan nilai *residual* dari model arima output yang dalam hal ini adalah luas areal kelapa sawit. Korelasi nilai beta dan alfa akan menentukan nilai r, b, dan s yang akan digunakan dalam permodelan fungsi transfer.



**Gambar 5.** *Cross Corelation* dari Beta dan Alfa Data Training Luas Areal Kelapa Sawit

Berdasarkan grafik *cross corelation* beta dan alfa pada Gambar 5. dapat dilihat bahwa pengaruh dari input harga dunia CPO kepada luas areal kelapa sawit Indonesia terjadi pada saat yang sama tanpa jeda, sehingga nilai koefisien b dan s adalah 0, begitu juga halnya dengan nilai r atau jeda pengaruh variabel output sendiri adalah 0. Selanjutnya dilakukan pengepasan model awal data *training* dengan membuat permodelan fungsi transfer awal menggunakan koefisien b, s, dan r untuk mendapatkan nilai *residual*. Plot ACF dan PACF *residual* data *training* dapat dilihat pada Gambar 5 sedangkan hasil model hasil pendugaan residual dapat dilihat pada Tabel 15.



**Gambar 5.** Plot ACF dan PACF *Residual* Pendugaan Fungsi Transfer Data *Training*

**Tabel 15.** Hasil Pendugaan Model ARIMA Residual Data *Training*

---

```

Series: res
ARIMA(0,2,1)
Coefficients:
    ma1
    -1.0000
s.e.      0.0701
sigma^2 estimated as 7.779e+11: log likelihood=-666.19
AIC=1336.38  AICc=1336.67  BIC=1339.95
Training set error measures:
Training set  ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
25827.89  852755.8  530806.5  10.1576  84.45881  1.000694  -0.1947163
z test of coefficients:
Estimate Std. Error z value Pr(>|z|)
ma1 -0.999999  0.070079  -14.27 < 2.2e-16 ***
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

---

Diperoleh model ARIMA *residual* (0,2,1) untuk luas areal kelapa sawit. Nilai MA1 model ARIMA *residual* luas areal signifikan sehingga model ini dapat digunakan untuk permodelan fungsi transfer lebih lanjut meskipun nilai MAPE cukup tinggi. Selanjutnya dilakukan pendugaan model fungsi transfer dengan menggunakan model ARIMA *residual* (0,2,1) untuk output luas areal kelapa sawit dan input harga CPO dunia pada data *training* (Tabel 16).

**Tabel 16.** Pendugaan Fungsi Transfer Arima *Residual* (0,2,1) Output Luas Areal Kelapa Sawit dengan Input Harga CPO Dunia Data *Training*

---

```

Series: train.h[, "luas_areal"]
Regression with ARIMA(0,2,1) errors
Coefficients:
    ma1      xreg
    -0.7440  -218.4769
s.e.      0.0855  2803.3972
sigma^2 estimated as 3.796e+10: log likelihood=-597.73
AIC=1201.46  AICc=1202.06  BIC=1206.81
Training set error measures:
Training set  ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
43942.39  186178  96549.42  2.411322  4.750652  0.4589245  -0.1323079
z test of coefficients:
Estimate Std. Error z value Pr(>|z|)
ma1 -0.743992  0.085508  -8.7009  <2e-16 ***
xreg -218.476920  2803.397200  -0.0779  0.9379
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

---

Nilai MAPE pada tabel 16 menunjukkan bahwa model ARIMA *residual* (0,2,1) cukup bagus yaitu 4,75%. Berikutnya dilakukan peramalan dengan data *testing* menggunakan input nilai aktual dan nilai ramalan, hasil dapat dilihat pada Tabel 17.

**Tabel 17.** Peramalan Data *Testing* menggunakan Input Nilai Aktual dan Ramalan

---

```

Input Aktual Series: test.h[, "luas_areal"]
Regression with ARIMA(0,2,1) errors
Coefficients:
      ma1      xreg
    -0.744  -218.4769
s.e.    0.000    0.0000
sigma^2 estimated as 3.796e+10: log likelihood=-62.46
AIC=126.92  AICc=128.92  BIC=126.31
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 248975.7 1087874 646079.4 1.70976 4.762921 0.8118709 -0.5431961
Input Ramalan Series: test.h[, "luas_areal"]
Regression with ARIMA(0,2,1) errors
Coefficients:
      ma1      xreg
    -0.744  -218.4769
s.e.    0.000    0.0000
sigma^2 estimated as 3.796e+10: log likelihood=-62.46
AIC=126.91  AICc=128.91  BIC=126.3
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 247450.8 1086527 647707 1.699798 4.777491 0.8139161 -0.5464482

```

---

Selanjutnya dilakukan pendugaan ulang model input untuk keseluruhan data *training* dan *testing*, hasil dapat dilihat pada Tabel 18. Terlihat nilai MAPE masih cukup bagus yaitu 5,97% untuk model luas area kelapa sawit. Sedangkan ramalan untuk lima tahun ke depan dapat dilihat pada Tabel 19, terlihat bahwa nilai ramalan cenderung mengalami kenaikan.

**Tabel 18.** Pendugaan Ulang Model Fungsi Transfer Keseluruhan Data

---

```

Series: dataestimasi[, "luas_areal"]
Regression with ARIMA(0,2,1) errors
Coefficients:
      ma1      xreg
    -0.8456  4011.858
s.e.    0.0587  4708.426

sigma^2 estimated as 1.724e+11: log likelihood=-717.38
AIC=1440.76  AICc=1441.28  BIC=1446.49
Error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 87301.48 398893.3 179308.4 2.805767 5.971479 0.6377769 -0.4049617

```

---

**Tabel 19.** Peramalan Fungsi Transfer (0,0,0) (0,2,1) Luas Areal Kelapa Sawit 2019-2023

---

Point	Forecas	Luas Areal
53		15040169
54		15750526
55		16460680
56		17170822
57		17880963

---

#### j. Model VAR

Pemodelan dengan menggunakan model VAR dilakukan dengan menggunakan peubah luas areal, produksi sawit (CPO), harga CPO dunia, volume ekspor CPO, dan volume impor CPO. Hal ini diasumsikan bahwa apabila terjadi gejolak produksi sawit (CPO, harga CPO dunia, volume ekspor karet, dan volume impor karet akan mempengaruhi pergerakan luas areal. Panjang *series* data pada data *training* adalah tahun 1981-2012, sementara data *testing* adalah periode tahun 2013-2018. Panjang *lag* (p) dalam model VAR ( $lag=p$ ) ditentukan dengan melakukan *trial error* menggunakan nilai p dari 1 s.d 6 lalu memperhatikan hasil signifikansi lag baik dengan maupun tanpa menggunakan konstanta dan *trend*. Dari hasil *trial error* diperoleh nilai p=2 dengan *type* menggunakan konstanta dan *trend*. Penggunaan konstanta dan *trend* dilakukan karena dari sekian banyak *trial error* diperoleh informasi

bahwa kedua hal tersebut seringkali signifikan. Hasil pendugaan model VAR (2) dapat dilihat pada tabel 20. *R-squared* sebesar 99,45% menunjukkan bahwa peubah produksi sawit (CPO), harga CPO dunia, volume ekspor CPO, dan volume impor CPO dengan konstanta dan *trend* mampu menjelaskan 99,45% keragaman luas areal kelapa sawit.

**Tabel 20.** Pendugaan Model VAR (2) Data *Training*

---

VAR Estimation Results:  
=====

Endogenous variables: luas\_areal, prod\_cpo\_ind, world\_price, ekspor\_cpo, impor\_cpo  
Deterministic variables: both  
Sample size: 30  
Log Likelihood: -1724.098  
Roots of the characteristic polynomial:  
1.001 0.8011 0.8011 0.6375 0.6375 0.5892 0.5892 0.545 0.4969 0.4969  
Call:  
VAR(y = sawit[15:46, c(2, 3, 4, 5, 6)], p = 2, type = "both")

Estimation results for equation luas\_areal:  
=====

luas\_areal = luas\_areal.l1 + prod\_cpo\_ind.l1 + world\_price.l1 + ekspor\_cpo.l1 +  
impor\_cpo.l1 + luas\_areal.l2 + prod\_cpo\_ind.l2 + world\_price.l2 + ekspor\_cpo.l2 +  
impor\_cpo.l2 + const + trend

	Estimate	Std. Error	t value	Pr(> t )
<b>luas_areal.l1</b>	<b>1.042e+00</b>	<b>3.201e-01</b>	<b>3.256</b>	<b>0.00439 **</b>
prod_cpo_ind.l1	-1.324e-01	8.210e-02	-1.612	0.12426
world_price.l1	5.285e+03	6.509e+03	0.812	0.42736
ekspor_cpo.l1	4.577e-02	7.862e-02	0.582	0.56766
impor_cpo.l1	-2.913e-01	5.116e-01	-0.569	0.57618
luas_areal.l2	-3.199e-01	3.588e-01	-0.892	0.38433
prod_cpo_ind.l2	6.359e-02	8.281e-02	0.768	0.45247
world_price.l2	-2.069e+03	7.998e+03	-0.259	0.79881
ekspor_cpo.l2	6.577e-02	7.543e-02	0.872	0.39477
impor_cpo.l2	-4.182e-01	4.868e-01	-0.859	0.40154
const	-2.356e+05	4.383e+05	-0.538	0.59741
<b>trend</b>	<b>7.956e+04</b>	<b>3.684e+04</b>	<b>2.159</b>	<b>0.04456 *</b>

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
Residual standard error: 212900 on 18 degrees of freedom  
Multiple R-Squared: 0.9966, **Adjusted R-squared: 0.9945**  
F-statistic: 479.1 on 11 and 18 DF, p-value: < 2.2e-16

---

**Tabel 21.** Uji Asumsi Klasik Residual Model VAR (2) Data *Training* dan Keseluruhan Data

---

<b>Portmanteau Test (asymptotic) Training</b>	Chi-squared=110.94, df=108, p-value=0.8592	Chi-squared=313.08, df=350, p-value=0.9225
<b>JB-Test (multivariate)</b>	Chi-squared=22.14, df=10, p-value=0.01441	Chi-squared=35.623, df= 0, p-value=9.768e-05
<b>Skewness only (multivariate)</b>	Chi-squared=11.219, df=5, p-value=0.04721	Chi-squared=10.69, df=5, p-value=0.05788
<b>Kurtosis only (multivariate)</b>	Chi-squared=10.922, df=5, p-value=0.05296	Chi-squared=24.933, df=5, p-value=0.0001435
<b>ARCH (multivariate) Training</b>	Chi-squared=375, df=1125, p-value=1	Chi-squared=465, df=1125, p-value=1

---

Dari hasil pengujian asumsi non autokorelasi, normalitas, dan homoskedastisitas, diperoleh informasi bahwa *residual* yang diperoleh dari model VAR (2) data *training* memenuhi ketiga asumsi tersebut, dibuktikan dengan nilai *p-value* yang tidak signifikan (di atas 0,05). Selanjutnya dilakukan ramalan pada data *testing* dan pengepasan model pada seluruh data. Hasil pengujian normalitas, non autokorelasi, serta homoskedastisitas dengan model VAR (2) pada data keseluruhan menunjukkan bahwa *residual* yang diperoleh memenuhi ketiga asumsi tersebut yang ditunjukkan dengan nilai *p-value* yang tidak signifikan atau lebih dari 0,05 (Tabel 21).

**Tabel 22.** MAPE Data *Training*, *Testing*, dan Pengepasan Keseluruhan Data Model VAR (2)

Luas Areal Sawit	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Training	0.008161	1.039968	4.011846	<b>5.376829</b>	6.229610	28.904990
Testing	0.1865	2.6908	5.6857	<b>5.7781</b>	8.4778	12.0098
Keseluruhan Data	0.4531	2.1377	4.4105	<b>6.1427</b>	8.2975	29.3595

Dari hasil pehitungan MAPE data *training*, *testing*, dan pengepasan keseluruhan data diperoleh informasi MAPE data *testing* sudah cukup baik, yaitu sebesar 5,77% dan 5,37% untuk data *training*. Sedangkan ramalan untuk lima tahun ke depan dapat dilihat pada Tabel 23, terlihat bahwa nilai ramalan cenderung mengalami kenaikan.

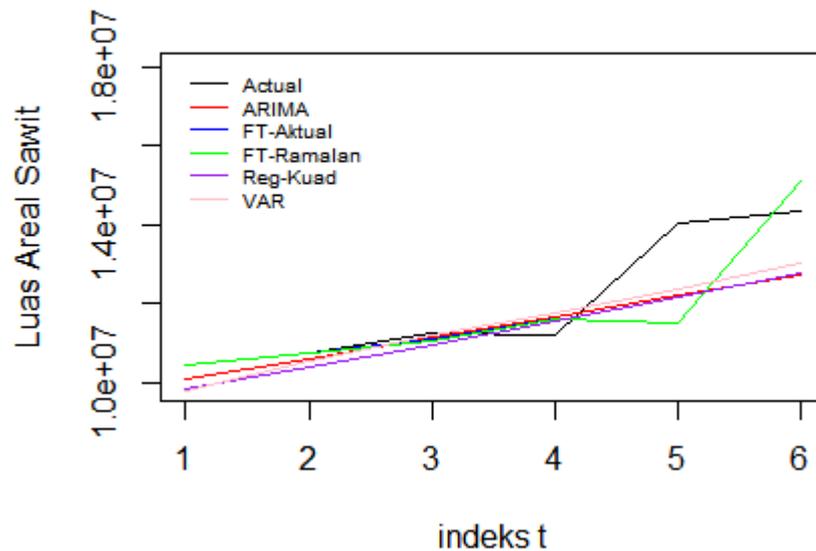
**Tabel 23.** Peramalan VAR (4) Luas Areal kelapa Sawit (CPO) 2019-2023

Fcst	luas areal produksi sawit (CPO)	CI
[1,]	15677985	
[2,]	16347415	
[3,]	17555578	
[4,]	19098145	
[5,]	20371548	

**Tabel 24.** Rekapitulasi MAPE *Training*, *Testing*, dan Pengepasan Keseluruhan Data Model ARIMA, Regresi, Fungsi Transfer, dan VAR

LUAS AREAL KELAPA SAWIT								
Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	
ARIMA								
Training set	43872.06	186187.6	96737.22	2.399218	<b>4.757331</b>	0.4598172	-0.1301828	
Test set	594838.77	1027952.0	753592.73	4.281559	<b>5.698820</b>	3.5820221	NA	
Training set	88326.91	401707.7	172325	3.120533	<b>5.134172</b>	0.6129376	-0.3720598	
Regresi Kuadratik		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
Traning set		0.2232	4.8949	18.1719	<b>53.4285</b>	79.0556	481.9535	
Test set		2.443	3.164	4.494	<b>6.466</b>	9.548	13.415	
Total set		0.0827	4.2431	11.3695	<b>60.2228</b>	69.8164	637.7602	
Fungsi Transfer		ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set		43942.39	186178	96549.42	2.411322	<b>4.750652</b>	0.4589245	-0.1323079
Test (input aktual)		248975.7	1087874	646079.4	1.70976	<b>4.762921</b>	0.8118709	-0.5431961
Test (input ramalan)		247450.8	1086527	647707	1.699798	<b>4.777491</b>	0.8139161	-0.5464482
Total set		87301.48	398893.3	179308.4	2.805767	<b>5.971479</b>	0.6377769	-0.4049617
VAR		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	
Training set		0.008161	1.039968	4.011846	<b>5.376829</b>	6.229610	28.904990	
Testing set		0.1865	2.6908	5.6857	<b>5.7781</b>	8.4778	12.0098	
Total set		0.4531	2.1377	4.4105	<b>6.1427</b>	8.2975	29.3595	

Dari hasil rekapitulasi nilai MAPE *training*, *testing*, dan pengepasan keseluruhan data model ARIMA, Regresi, Fungsi Transfer, dan VAR, terlihat bahwa MAPE data *testing* model fungsi transfer memberikan nilai yang paling kecil yaitu 4,76% untuk peramalan luas areal kelapa sawit.



**Gambar 6.** Perbandingan Plot Aktual dan Ramalan Data *Testing* menggunakan Metode ARIMA, Regresi Kuadratik, Fungsi Transfer, dan VAR Luas Areal Kelapa Sawit Indonesia.

## KESIMPULAN DAN SARAN

### Kesimpulan

- Hasil analisis yang didapat dari empat model adalah sebagai berikut:  
Peramalan Luas Areal Sawit
  - Model ARIMA terbaik adalah ARIMA (0,2,1) dengan MAPE data testing 5,69%.
  - Model regresi terbaik adalah regresi kuadratik dengan MAPE 6,47% .
  - Model Fungsi Transfer terbaik dengan peubah input harga CPO dunia adalah ARIMA (0,0,0)(0,2,1) dengan model input ARIMA (1,1,0).
  - Model VAR terbaik dengan memasukkan peubah produksi sawit (CPO), harga CPO dunia, volume ekspor CPO dan volume impor CPO adalah VAR(2) type “both”.
- Dari keempat model yang diujicobakan, model terbaik adalah fungsi transfer dengan nilai MAPE data testing terkecil yaitu 4,76% untuk luas areal sawit.
- Hasil ramalan fungsi transfer luas areal sawit untuk lima tahun ke depan adalah 15040169 Ha, 15750526 Ha, 16460680 Ha, 17170822 Ha, 17880963Ha.

### Saran

- Perlu dilakukan kajian mendalam dengan metode peramalan lainnya.
- Untuk model regresi, perlu diujicobakan menggunakan variabel bebas/peubah input selain waktu.

- Untuk model fungsi transfer dan VAR perlu diujicobakan dengan menggunakan variabel lain yang sekiranya lebih berpengaruh, baik secara teoritis maupun praktis.
- Perlu dikaji metode peramalan untuk data non parametrik, sehingga tidak terhalang oleh tidak terpenuhinya berbagai asumsi.

## DAFTAR PUSTAKA

- Direktorat Jenderal Perkebunan. 2019. Statistik Perkebunan Indonesia 2018-2020 Jakarta: Direktorat Jenderal Perkebunan, Kementerian Pertanian. Statistik Perkebunan Indonesia 2018-2020.
- Gujarati, Damodar. N dan Porter, Dawn. C. 2009. Basic Econometrics. Boston: Douglas Reiner.
- Kirchgassner, Gebhard dan Wolters, Jurgen. 2007. Introduction to Modern Time Series Analysis. Berlin: Springer Berlin Heidelberg Newyork.
- Wei, William WS. 2006. Time Series Analyis. Phladelphia: Department of Statistics The Fox School of Business and Management Temple University.
- Wooldridge, Jeffrey M. 2015. Introductory Econometrics:A Modern Approach. Boston:Cegage Learning.

# KAJIAN METODE ESTIMASI LUAS PANEN TEBU INDONESIA

Efi Respati - *Statisticians*

*Center for Agricultural Data and Information System - Ministry of Agriculture*

*Jln. Harsono RM 3 Gd D Lt 4, Ragunan-Jakarta Selatan, Indonesia*

*E-mail: [efi@pertanian.go.id](mailto:efi@pertanian.go.id)*

## ABSTRAK

Publikasi resmi data luas panen tebu di Indonesia dirilis oleh pemerintah setiap tahun dengan lag 2 tahun. Perumusan kebijakan produksi, impor, distribusi pasokan serta pengendalian harga, sangat memerlukan data terkini bahkan data hasil peramalan beberapa periode ke depan. Angka estimasi disusun oleh Direktorat Jenderal Perkebunan dengan model univariate. Kajian estimasi luas panen tebu Indonesia dilakukan dengan pendekatan metode ARIMA, Regresi, Fungsi Transfer dan model Vector Autoregression (VAR) dengan software pengolah R-Studio. Hasil uji statistic menunjukkan bahwa Fungsi Transfer merupakan metode terbaik untuk estimasi luas panen tebu dengan nilai MAPE sebesar 3.34% dan data hasil estimasi tahun 2020-2024 yang cukup realistis.

*Kata kunci: gula, ARIMA, regresi, fungsi transfer, var, MAPE*

## ABSTRACT

*The official publication of sugarcane harvested area data in Indonesia is released by the government annually, and has 2 years lag. Formulation of production, import, supply distribution and price control policies, requires the current data and even the forecast data for the next few periods. Estimated figures have been compiled by the Directorate General of Plantation, resulted from the univariate model. This paper compares the forecasting of sugarcane harvested area in Indonesia by the ARIMA, Regression, Transfer Function and Vector Autoregression (VAR) model approaches with the R-Studio software. The best method is Fungsi Transfer with MAPE = 3.34% and the sugarcane harvested area estimated data is more realistic.*

*Keywords: projection model, arima, regression, transfer function, var*

## PENDAHULUAN

Kegiatan pengumpulan tebu/gula untuk perkebunan rakyat di Indonesia saat ini mengacu pada Buku Pedoman Pelaksanaan Pengelolaan Data Komoditas Perkebunan yang diterbitkan oleh Direktorat Jenderal Perkebunan Tahun 2013 (Edisi Kedua). Berdasarkan buku pedoman tersebut, komoditas tebu/gula merupakan tanaman semusim, sehingga pengumpulan datanya dilakukan semesteran, namun demikian rilis data resmi dilakukan tahunan. Pengumpulan data tebu/gula untuk perkebunan besar, baik Perkebunan Besar Negara maupun Perkebunan Besar Swasta dilakukan oleh Badan Pusat Statistik melalui pelaporan oleh Pabrik Gula secara online. Rilis angka tahunan dilakukan secara bersama-sama antara Kementerian Pertanian dan BPS serta sinkronisasi akhir dengan Perusahaan Gula.

Ketersediaan data tebu/gula yang dirilis secara resmi oleh Direktorat Jenderal Perkebunan saat ini adalah data dengan lag t-2, yakni pada tahun 2020 baru tersedia angka tetap tahun 2018. Disamping angka tetap, Direktorat Jenderal Perkebunan merilis juga Angka Sementara dengan lag t-1, karena belum semua daerah/perusahaan gula melaporkan realisasi produksi pada periode tersebut. Pada periode yang sama, Direktorat Jenderal Perkebunan juga merilis Angka Estimasi tahun ke-t, yang diperoleh dengan menggunakan metode peramalan univariate.

Makalah ini membahas kajian metode ARIMA, Regresi, Fungsi Transfer dan VAR yang bisa diterapkan dalam memodelkan luas panen tebu Indonesia, serta menentukan metode yang paling akurat dalam melakukan peramalan produksi gula 5 (lima) tahun kedepan.

## METODOLOGI

### 4. Sumber Data

Data tebu/gula yang digunakan dalam makalah ini adalah sbb.:

No	Variabel	Periode data	Sumber data	Level data
1	Luas panen tebu	1989-2019	Ditjen Perkebunan	Nasional
2	Harga Internasional Gula	1989-2019	World Bank	Internasional
3	Ekspor Molases	1989-2019	BPS	Nasional
4	Impor GKP	1989-2019	BPS	Nasional

### 5. Software

Software yang digunakan dalam menyusun makalah ini adalah software *R-Studio*. Keunggulan software R-Studio adalah software yang *open source*, sehingga tidak memerlukan biaya untuk pembelian maupun perpanjangan lisensi.

### 6. Tinjauan Literatur

Metode statistik yang dikembangkan oleh para ahli untuk melakukan peramalan data sangat beragam, baik peubah tunggal maupun peubah ganda, diantaranya metode ARIMA (*Auto Regressive Integrated Moving Average*), Regresi, Fungsi Transfer dan VAR (*Vector Autoregresion*).

#### c. *Autoregressive Integrated Moving average (ARIMA)*

*Autoregressive Integrated Moving average (ARIMA)* atau biasa disebut juga dengan metode time series Box Jenkins, sangat sesuai digunakan untuk melakukan peramalan jangka pendek, sementara untuk peramalan jangka panjang kurang baik ketepatannya. Metode ARIMA merupakan metode yang hanya menggunakan peubah dependen dan mengabaikan peubah independen sewaktu melakukan peramalan.

Metode ARIMA dibagi kedalam tiga kelompok model, yaitu *autoregressive model (AR)*, *moving average model (MA)* dan model campuran yang memiliki karakteristik kedua model di atas yaitu *autoregressive integrated moving average (ARIMA)*.

#### 1) *Autoregressive Model (AR)*

AR adalah suatu model yang menjelaskan pergerakan suatu peubah melalui peubah itu sendiri di masa lalu.

Model *autoregressive* orde ke-*p* dapat ditulis sebagai berikut:

ARIMA (*p*,0,0)

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t \dots\dots\dots(1)$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $Y_{t-p}$  = data *time series* pada kurun waktu ke (t-P)
- $\mu$  = suatu konstanta
- $\theta_1 \dots \theta_p$  = parameter autoregressive ke-p
- $\varepsilon_t$  = nilai kesalahan pada waktu ke t

### 2) *Moving Average Model (MA)*

MA adalah suatu model yang melihat pergerakan peubahnya melalui sisaannya di masa lalu.

Bentuk model MA dengan ordo q atau MA (q) atau model ARIMA (0,d,g) dapat ditulis sebagai berikut:

$$Y_t = \mu - \phi_1\varepsilon_{t-1} - \phi_2\varepsilon_{t-2} - \dots - \phi_q\varepsilon_{t-q} + \varepsilon_t \dots\dots\dots(2)$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $\phi_1 \dots \phi_q$  = parameter-parameter moving average
- $\varepsilon_{t-q}$  = nilai kesalahan pada waktu ke (t-q)

### 3) *Autoregressive Intergrated Moving Everage (ARIMA)*

ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampainya. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t \dots\dots\dots(3)$$

dimana:

- $Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t
- $Y_{t-p}$  = data *time series* pada kurun waktu ke (t-P)
- $\mu$  = suatu konstanta
- $\theta_1 \theta_q \phi_1 \phi_n$  = parameter-parameter model
- $\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

### d. Model Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih peubah. Pada analisis regresi, dibedakan menjadi dua, yaitu peubah respon atau biasa juga disebut peubah bergantung (*dependent variable*) dan peubah *explanatory* atau biasa disebut penduga (*predictor variable*) atau disebut juga peubah bebas (*independent peubah*). Model regresi digunakan untuk meramalkan atau memprediksi nilai masa depan dari peubah respon berdasarkan nilai-nilai tertentu yang diberikan oleh peubah prediktor.

Asumsi yang mendasari penggunaan analisis regresi linier adalah:

1. Sisaan mengikuti fungsi distribusi normal;
2. Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas);

3. Tidak terdapat autokorelasi antara sisaan untuk setiap data pengamatan; dan
4. Tidak terdapat multikolinearitas antara peubah respon.

- **Analisis Regresi Linier**

**b. Analisis Regresi Linier Sederhana**

Model regresi linier sederhana melibatkan satu peubah prediktor dan diformulasikan dalam model matematis sebagai berikut:

$$y = \beta_0 + \beta_1 x + \varepsilon \dots\dots\dots(4)$$

dimana:

- $y$  = peubah respon/dependen/terikat
- $x$  = peubah prediktor/independen/bebas
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi peubah bebas
- $\varepsilon$  = error/residu/sisaan

Parameter model atau koefisien regresi yaitu  $\beta_0$  dan  $\beta_1$  masing-masing memiliki interpretasi sebagai intercept dan slope dari suatu garis lurus.  $\beta_1$  mengukur perubahan rata-rata dari peubah respon  $y$  untuk setiap perubahan dari peubah prediktor  $x$ . Parameter ini biasanya tidak diketahui dan harus diprediksi dari data sampel.  $\varepsilon$  merupakan kesalahan statistik, didefinisikan sebagai sisa atau perbedaan hasil antara nilai data pengamatan peubah respon terhadap nilai peubah respon hasil prediksi.

**Menaksir Parameter  $\beta_0$  dan  $\beta_1$**

Misalkan kita mempunyai n pasangan observasi, katakan  $(y_1, x_1), (y_2, x_2) \dots (y_n, x_n)$ , data ini dapat digunakan untuk memprediksi parameter  $\beta_0$  dan  $\beta_1$  sehingga jumlah kuadrat dari deviasi antara observasi-observasi dan garis regresi menjadi minimum. Kita gunakan persamaan diatas untuk menaksir nilai parameter tersebut,

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \dots\dots\dots(5)$$

dengan meminimumkan fungsi kuadrat tersebut maka akan diperoleh taksiran untuk  $\beta_0$  dan  $\beta_1$  sebagai berikut.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Sedangkan varian untuk intercept dan slope adalah:

$$v(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \quad v(\hat{\beta}_0) = \sigma^2 \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}} \right]$$

$$\sigma^2 = \frac{\left( \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \right) - \hat{\beta}_1 \left( \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i\right)}{n} \right)}{n - p}$$

dimana n = jumlah pengamatan

p = jumlah parameter (untuk regresi linier sederhana, p = 2)

- **Analisis Regresi Linier Berganda**

Jika terdapat k peubah prediktor, maka disebut dengan regresi linier berganda dengan model matematis sebagai berikut:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \dots\dots\dots(6)$$

dengan

- y** = peubah respon/dependen/terikat
- $x_1, x_2, \dots, x_k$  = peubah prediktor/independen/bebas
- $\beta_0$**  = konstanta/intercept
- $\beta_1, \beta_2, \dots, \beta_k$  = koefisien regresi peubah bebas
- $\varepsilon$**  = error/residu/sisaan

**e. Metode Fungsi Transfer**

Dalam bidang pertanian, produksi komoditas pertanian sangat dipengaruhi oleh peubah lainnya seperti serangan OPT, penggunaan saprodi, gejolak harga komoditas tersebut atau komoditas lainnya, dan lainnya. Apabila peramalan produksi hanya didasarkan pada besarnya produksi komoditas yang bersangkutan saja tanpa memperhatikan faktor-faktor yang mempengaruhinya, maka informasi untuk pembuatan perencanaan menjadi tidak lengkap, sehingga tujuan peramalan menjadi tidak tercapai secara utuh. Salah satu upaya menganalisis data deret waktu multivariat agar diperoleh hasil yang dapat memberikan informasi yang lengkap dan simultan, adalah dengan menggunakan model Fungsi Transfer.

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau Yt) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri (Yt) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau Xt) dengan deret output tersebut. Model fungsi transfer merupakan fungsi dinamis yang pengaruhnya tidak hanya pada hubungan linier antara deret input dengan deret output pada waktu ke-t, tetapi juga pada waktu t+1, t+2, ..., t+k. Hubungan seperti ini pada fungsi transfer dapat menimbulkan delay (waktu senjang) antara peubah input dan peubah output.

Tujuan pemodelan fungsi transfer adalah untuk menetapkan model yang sederhana, yang menghubungkan deret output (Yi) dengan deret input (Xi) dan gangguan/noise(ni). Wei (1994) juga menjelaskan bahwa di dalam fungsi transfer terdapat rangkaian output yang mungkin

dipengaruhi oleh rangkaian *multiple input*. Pada kasus *single input* peubah, dapat menggunakan metode korelasi silang yang dianjurkan oleh Box and Jenkins (1976). Teknik ini juga dapat digunakan ketika terdapat *single input* peubah yang lebih dari satu selama antar variable *input* tidak berkorelasi silang. Jika beberapa atau semua peubah input berkorelasi silang maka teknik prewhitening atau metode korelasi silang tidak dapat digunakan secara langsung. Alasan utama bagi perlunya suatu perencanaan atau peramalan adalah adanya tenggang waktu pengambilan keputusan yang dapat berkisar dari beberapa hari atausampai beberapa tahun. Pada analisis fungsi transfer untuk peramalan deret berkala univariate, terdapat deret berkala output yang diperkirakan dipengaruhi oleh deret berkala input dan input-input lain yang digabungkan dalam satu kelompok yang disebut gangguan (noise). Deret input mempengaruhi deret output melalui sebuah fungsi transfer yang mendistribusikan pengaruhnya secara dinamis melalui beberapa periode waktu yang akan datang dengan persentase tertentu yang disebut sebagai bobot respons impuls atau bobot fungsi transfer.

Model umum Fungsi Transfer:

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \quad \dots\dots\dots(7)$$

Dimana:

- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang lag  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$
- $q \rightarrow$  ordo MA bagi noise  $N_t$

**f. Vector Autoregression (VAR)**

*Vector Autoregression* (VAR) dikemukakan pertama kali oleh Christopher Sims (1980). Sims mengembangkan model ekonometri dengan mengabaikan pengujian asumsi secara apriori. VAR dikembangkan oleh Sims sebagai kritik atas metode simultan. Jumlah peubah yang besar dan klasifikasi endogen dan eksogen pada metode simultan merupakan dasar dari kritik tersebut. Penggunaan pendekatan struktural atas pemodelan persamaan simultan biasanya menerapkan teori ekonomi di dalam usahanya untuk mendeskripsikan hubungan antar peubah yang ingin diuji. Akan tetapi sering ditemukan bahwa teori ekonomi saja ternyata tidak cukup kaya di dalam menyediakan spesifikasi yang ketat dan tepat atas hubungan dinamis antar peubah (Gujarati, 2010). Model VAR merupakan jalan keluar atas permasalahan ini melalui pendekatan non-struktural atau tidak mendasarkan pada teori ekonomi tertentu (*atheoretical*). Metode VAR memperlakukan seluruh peubah secara simetris tanpa memperlakukan peubah dependen dan independen atau dengan kata lain model ini memperlakukan seluruh peubah sebagai peubah endogen., karena pada kenyataannya suatu peubah dapat bertindak sebagai endogen atau dependen dari suatu peubah yang lainnya (Gujarati, 2010).

Kelebihan dalam penggunaan metode VAR (Gujarati, 2010):

- e. Kemudahan dalam penggunaan, tidak perlu mengkhawatirkan tentang penentuan peubah endogen dan peubah eksogen.
- f. Kemudahan dalam estimasi, metode *Ordinary Least Square* (OLS) dapat diaplikasikan pada tiap persamaan secara terpisah.
- g. *Forecast* atau peramalan yang dihasilkan pada beberapa kasus ditemukan lebih baik daripada yang dihasilkan oleh model persamaan simultan yang kompleks.
- h. Perangkat estimasi yang digunakan adalah *Impulse Respon Function* (IRF) untuk melacak respon dari peubah dependen dalam sistem VAR terhadap *shock* dari *error term* dan *Variance*

*Decomposition* yang memberikan informasi mengenai pentingnya masing-masing *error term* dalam mempengaruhi peubah-peubah dalam VAR.

Di sisi lain, terdapat beberapa kritik terhadap model VAR menyangkut permasalahan berikut (Gujarati, 2010) :

- 6) Model VAR merupakan model yang *atheoretic* atau tidak berdasarkan teori, hal ini tidak seperti pada persamaan simultan. Pada persamaan simultan, pemilihan peubah yang akan dimasukkan dalam persamaan memegang peranan penting dalam mengidentifikasi model.
- 7) Pada model VAR penekanannya terletak pada *forecasting* atau peramalan sehingga model ini kurang cocok digunakan dalam menganalisis kebijakan.
- 8) Permasalahan yang besar dalam model VAR adalah pada pemilihan *lag length* atau panjang lag yang tepat. Karena semakin panjang lag, maka akan menambah jumlah parameter yang akan bermasalah pada *degrees of freedom*.
- 9) Peubah yang tergabung pada model VAR harus stasioner. Apabila tidak stasioner, perlu dilakukan transformasi bentuk data, misalnya melalui *first difference*.
- 10) Sering ditemui kesulitan dalam menginterpretasi tiap koefisien pada estimasi model VAR, sehingga sebagian besar peneliti melakukan interpretasi pada estimasi fungsi *impulse respon*.

### Estimasi Model VAR

Enders (2004) menjelaskan, ketika peneliti tidak memiliki kepastian untuk menentukan bahwa suatu peubah adalah eksogen, maka suatu perluasan analisis fungsi perpindahan alami akan memperlakukan masing-masing peubah secara simetris. Sebagai contoh, pada kasus-kasus peubah yang membiarkan alur waktu atau *time path*  $\{s_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{y_t\}$  dan membiarkan *time path*  $\{y_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{s_t\}$ .

Di dalam sistem *bivariate*, hubungan tersebut dapat digambarkan seperti pada persamaan (1.4) di bawah ini:

$$\begin{aligned} s_t &= b_{10} - b_{12}y_t + \gamma_{11}s_{t-1} + \gamma_{12}y_{t-1} + \varepsilon_{s_t} \\ y_t &= b_{20} - b_{21}s_t + \gamma_{21}s_{t-1} + \gamma_{22}y_{t-1} + \varepsilon_{y_t} \end{aligned} \quad \dots \dots \dots (8)$$

Dengan mengasumsikan bahwa kedua peubah  $s_t$  dan  $y_t$  adalah stasioner:  $\varepsilon_{s_t}$  dan  $\varepsilon_{y_t}$  adalah *disturbances* yang memiliki rata-rata nol dan matriks kovarians terbatas atau bersifat *white noise* dengan standar deviasi yang berurutan  $\sigma_s$  dan  $\sigma_y$ ; serta  $\{\varepsilon_{s_t}\}$  dan  $\{\varepsilon_{y_t}\}$  adalah *disturbances* yang independen dengan rata-rata nol dan kovarian terbatas (*uncorrelated white-noise disturbances*). Kedua persamaan di atas merupakan orde pertama VAR, karena panjang *lag* nya hanya satu. Agar persamaan (4.1) lebih mudah dipahami dan digunakan sebagai alat analisis maka ditransformasikan dengan menggunakan matriks aljabar, dan hasilnya dapat dituliskan secara bersama seperti pada persamaan di bawah ini:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} s_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix}$$

Atau dengan bentuk lain:

$$Bx_t = \Gamma_0 + \Gamma_1 x_{t-1} + \varepsilon_t \dots\dots\dots(9)$$

Dimana:

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \quad x_t = \begin{bmatrix} s_t \\ y_t \end{bmatrix} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix}$$

Dengan melakukan pengalihan antara persamaan (4.2) dengan  $B^{-1}$  atau invers matriks B, maka akan dapat ditentukan model VAR dalam bentuk standar, seperti dituliskan pada persamaan di bawah ini:

$$x_t = A_0 + A_1 x_{t-1} + \ell_t \dots\dots\dots(10)$$

$$A_0 = B^{-1} \Gamma_0$$

dimana  $A_1 = B^{-1} \Gamma_1$

$$\ell_t = B^{-1} \varepsilon_t$$

Untuk tujuan notasi, maka  $\{a_{i0}\}$  dapat didefinisikan sebagai elemen ke-i dari vektor  $A_0$ ;  $\{a_{ij}\}$  sebagai elemen dalam baris ke-i dan baris ke-j dari matriks  $A_1$ ; dan  $\{e_{it}\}$  sebagai elemen ke-i dari vektor  $e_t$ . Dengan menggunakan notasi baru yang telah dijelaskan sebelumnya, maka persamaan (4.3) dapat ditulis menjadi:

$$\begin{aligned} s_t &= a_{10} + a_{11}s_{t-1} + a_{12}y_{t-1} + e_{1t} \\ y_t &= a_{20} + a_{21}s_{t-1} + a_{22}y_{t-1} + e_{2t} \dots\dots\dots(11) \end{aligned}$$

**Fungsi Impulse Response**

Fungsi *impulse response* menggambarkan tingkat laju dari *shock* peubah yang satu terhadap peubah yang lainnya pada suatu rentang periode tertentu. Sehingga dapat dilihat lamanya pengaruh dari *shock* suatu peubah terhadap peubah lain sampai pengaruhnya hilang atau kembali ke titik keseimbangan. Analisis fungsi *impulse respon* dapat dituliskan dalam bentuk *Vector Moving Avarage (VMA)* dari bentuk standar VAR pada persamaan (4.4).

$$\begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} - \\ s \\ - \\ y \end{bmatrix} + \sum_{i=1}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \end{bmatrix} \dots\dots\dots(12)$$

dimana  $s_t$  dan  $y_t$  memiliki hubungan dengan  $e_{1t}$  dan  $e_{2t}$  secara berurutan. Selanjutnya dengan menggunakan operasi aljabar matriks maka *vector error* dapat ditentukan sebagai berikut :

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \frac{1}{1-b_{12}b_{21}} \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{s_{t-i}} \\ \varepsilon_{y_{t-i}} \end{bmatrix} \dots\dots\dots(13)$$

dengan menggabungkan persamaan (4.5) dan (4.6) akan didapat:

$$\begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} - \\ s_t \\ - \\ y_t \end{bmatrix} + \frac{1}{1-b_{12}b_{21}} \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{s_{t-i}} \\ \varepsilon_{y_{t-i}} \end{bmatrix} \dots\dots\dots(14)$$

Persamaan (4.7) dapat disederhanakan dengan mendefinisikan matriks 2x2  $\Phi_i$  dengan elemen  $\Phi_{jk}$  (i) seperti persamaan berikut :

$$\Phi_i = A_1^i / (1 - b_{12}b_{21}) \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \dots\dots\dots(15)$$

sehingga diperoleh bentuk matriks persamaan fungsi *impulse respon*:

$$\begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} \bar{s} \\ \bar{y} \end{bmatrix} + \sum_{i=0}^n \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{st-1} \\ \varepsilon_{yt-1} \end{bmatrix} \dots\dots\dots(16)$$

dimana :

$\Phi_{ij}(i)$  = efek dari *structural shock* pada s dan y

$\Phi_{ij}(0)$  = *impact multipliers*

$\sum \Phi_{ij}(i)$  = *cumulative multipliers*

$\sum \Phi_{ij}(i)$  pada saat  $n \rightarrow \infty$  = *long run multipliers*

**Variance Decomposition**

*Variance decomposition* atau disebut juga *forecast error variance decomposition* merupakan perangkat pada model VAR yang akan memisahkan variasi dari sejumlah peubah yang diestimasi menjadi komponen-komponen *shock* atau menjadi peubah *innovation*, dengan asumsi bahwa peubah-peubah *innovation* tidak saling berkorelasi. Kemudian, *variance decomposition* akan memberikan informasi mengenai proporsi dari pergerakan pengaruh *shock* pada sebuah peubah terhadap *shock* peubah yang lain pada periode saat ini dan periode yang akan datang.

Bentuk VMA dari peubah x pada satu periode ke depan dapat dituliskan sbb.:

$$x_{t+1} = \bar{x} + \sum_{i=0}^{\infty} \phi_i \varepsilon_{t+1-i} \dots\dots\dots(17)$$

*Forecast error* pada satu periode kedepan adalah:

$$E_t x_{t+1} = \bar{x} + \sum_{i=1}^{\infty} \phi_i \varepsilon_{t+1-i} \dots\dots\dots(18)$$

*Forecast* satu periode ke depan dilambangkan dengan  $\Phi_0 \varepsilon_{t+1}$ . *Forecast error* pada periode n ke depan adalah:

$$x_{t+n} - E_t x_{t+n} = \bar{x} + \sum_{i=0}^{n-1} \phi_i \varepsilon_{t+n-i} \dots\dots\dots(19)$$

*Forecast error* pada n periode ke depan untuk peubah s adalah:

$$s_{t+n} - E_t s_{t+n} = \phi_{11}(0)\varepsilon_{st+n} + \phi_{11}(1)\varepsilon_{st+n-1} + \dots + \phi_{11}(n-1)\varepsilon_{yt+1} \\ + \phi_{12}(0)\varepsilon_{yt+n} + \phi_{12}(1)\varepsilon_{yt+n-1} + \dots + \phi_{12}(n-1)\varepsilon_{yt+1} \dots\dots\dots(20)$$

*Variance dari forecast error*  $s_{t+n}$  periode n ke depan adalah  $\sigma_s(n)^2$  dimana:

$$\sigma_s(n)^2 = \sigma_s^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2] + \sigma_y^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2] \dots\dots\dots(21)$$

Forecast error variance decomposition adalah proporsi dari  $\sigma_s(n)^2$  terhadap shock  $s$  dan shock  $y$ . Sehingga forecast error variance decomposition pada shock  $s$  adalah:

$$\sigma_s^2 [\phi_{11}(0)^2 + \phi_{11}(1)^2 + \dots + \phi_{11}(n-1)^2] / \sigma_s(n)^2 \dots\dots\dots(22)$$

sedangkan forecast error variance decomposition pada shock  $y$  adalah:

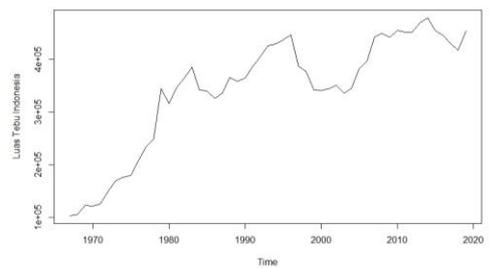
$$\sigma_y^2 [\phi_{12}(0)^2 + \phi_{12}(1)^2 + \dots + \phi_{12}(n-1)^2] / \sigma_y(n)^2 \dots\dots\dots(23)$$

## HASIL PEMBAHASAN

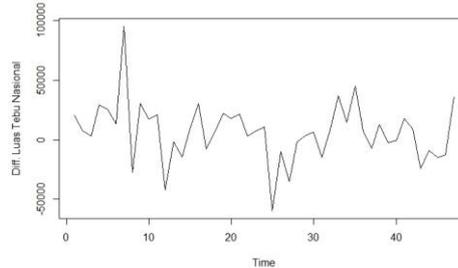
### 5. Estimasi Metode ARIMA

Eksplorasi data series luas panen tebu Indonesia tahun 1972-2019 untuk pemodelan ARIMA dilakukan dengan membuat plot data yang menunjukkan kecenderungan tren meningkat (Gambar 1). Hasil uji Augmented Dickey-Fuller untuk data luas panen tebu dengan pembedaan (*differencing*) tingkat 1 menunjukkan data sudah stasioner seperti tersaji pada Gambar 2 serta hasil Uji Augmented Dickey-Fuller menunjukkan nilai statistic uji sudah lebih kecil dari nilai kritis, sebagai berikut:

```
Value of test-statistic is: -5.5465
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```



Gambar 1. Plot Luas Panen Tebu, 1972-2019



Gambar 2. Plot Luas Panen Tebu dengan Differencing Tingkat 1, 1972-2019

Data series luas panen tebu untuk pemodelan dibagi menjadi series *data training* (1972 – 2013) dan series *data testing* (2014-2019). Series data *training* digunakan untuk mencari model terbaik dan akan divalidasi dengan menggunakan series data *testing*.

Penentuan ordo lag  $p$  (AR) dan  $q$  (MA) dilakukan melalui investigasi plot ACF dan PACF namun belum terlihat kriteria lag AR dan MA-nya, sehingga dilakukan dengan memanfaatkan script pada software R-Studio untuk menunjukkan 10 model tentative, dengan hasil seperti tersaji pada Tabel 2.

**Tabel 2. Sepuluh Model Tentative Luas Panen Tebu oleh R-Studio**

No	p	q	sbc
[1,]	0	1	820.5969
[2,]	1	1	820.881
[3,]	2	1	822.8913
[4,]	0	2	824.8961
[5,]	1	2	825.1658
[6,]	3	1	826.605
[7,]	4	1	827.9579
[8,]	2	2	828.4604
[9,]	0	3	829.0532
[10,]	1	3	829.1226

Uji signifikansi terhadap komponen AR dan MA pada ke-10 model tentative menunjukkan bahwa komponen AR maupun MA pada beberapa model tentatif menunjukkan signifikan baik pada tingkat kepercayaan 90% atau lebih tinggi. Model akhir yang dipilih didasarkan pada nilai MAPE pada data testing yang lebih kecil serta hasil estimasi yang logis, yakni model ARIMA (4,1,1).

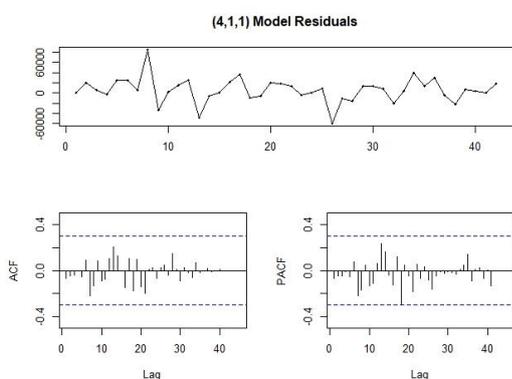
Uji signifikansi terhadap komponen MA1 dan MA2 adalah sbb.

```
Series: train[, "Luas_Total"]
ARIMA(4,1,1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.67874    0.39887  1.7017 0.08882 .
ar2  0.31329    0.18698  1.6756 0.09382 .
ar3 -0.06780    0.21317 -0.3181 0.75044 .
ar4 -0.17719    0.15835 -1.1190 0.26316 .
ma1 -0.67039    0.38375 -1.7470 0.08064 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pemeriksaan sisaan terhadap model ARIMA (4,1,1) menunjukkan pola terdistribusi normal serta pola ACF dan PACF sisaan yang tidak nyata, seperti tersaji pada Gambar 3. Hasil Uji Ljung-Box yang mengindikasikan autokorelasi sisaan tidak signifikan hingga 30 lag (Tabel 3).



**Gambar 3. Plot sisaan ARIMA (4,1,1) Peubah Luas Panen Tebu**

**Tabel 3. Hasil Uji Ljung-Box Arima (4,1,1)**

lags	statistic	df	p-value
5	0.5098289	5	0.991759
10	5.3555490	10	0.866198
15	10.1508035	15	0.810148
20	17.4698697	20	0.622286
25	21.6802043	25	0.654137
30	26.2581170	30	0.661894

Setelah asumsi sisaan telah memenuhi syarat, maka dilakukan peramalan untuk set data testing yang menghasilkan nilai MAPE sebesar 6,15% dan ramalan data testing sebagai berikut.

```

Time Series:
Start = 43
End = 48
Frequency = 1
[1] 466321.4 470427.4 471180.0 469989.8 469654.2 468275.0
Training set ME RMSE MAE MPE MAPE MASE ACF1
Test set -22886.139 30531.96 26815.02 -5.329401 6.151157 1.4271844 NA

```

Pengepasan seluruh data terhadap model ARIMA (4,1,1), maka diperoleh nilai MAPE yang cukup bagus sebesar 4,92% untuk series data *training*.

```

Series: olahtebu[, "Luas_Total"]
ARIMA(4,1,1)

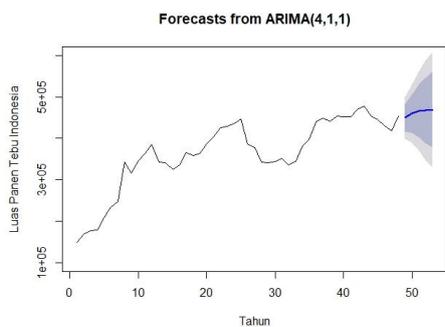
Coefficients:
      ar1      ar2      ar3      ar4      ma1
s.e.  0.6860  0.2781 -0.0632 -0.1734 -0.6740
      0.3714  0.1811  0.1990  0.1508  0.3574

sigma^2 estimated as 641517736:  log likelihood=-540.75
AIC=1093.5  AICc=1095.6  BIC=1104.6

Training set error measures:
Training set ME RMSE MAE MPE MAPE MASE ACF1
5453.407 23692.36 17186.83 1.737708 4.918569 0.9227419 -.05271031

```

Hasil peramalan produksi gula tahun 2020 – 2024 menggunakan metode ARIMA (4,1,1) serta plot hasil ramalannya tersaji pada Tabel 4 dan Gambar 4.



Gambar 4. Plot hasil ramalan luas panen tebu metode ARIMA (4,1,1)

Tabel 4. Hasil Estimasi Luas Panen Tebu Indonesia Metode ARIMA (4,1,1)

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
49	449222.2	416762.8	481681.6	399579.8	498864.5
50	459744.4	413564.2	505924.5	389117.9	530370.8
51	465808.2	403289.7	528326.7	370194.4	561422
52	466965.3	389005.8	544924.8	347736.5	586194.1
53	469476.6	378678.6	560274.5	330613.1	608340

## 6. Estimasi Metode Regresi

Estimasi luas panen tebu menggunakan metode regresi linear dan kuadratik dilakukan dengan peubah Tahun (t) sebagai peubah bebas.

### a. Regresi linear sederhana

Hasil regresi linear sederhana luas panen tebu dengan peubah tahun menunjukkan bahwa baik intercept maupun peubah tahun nyata pada taraf kepercayaan 100% dengan  $R^2=65,81\%$ .

Persamaan regresi linearnya adalah : **Produksi = -10620482 + 5508Tahun**

```
lm(formula = Luas_Total ~ Tahun, data = train)
```

```

Residuals:
Min      1Q  Median      3Q      Max

```

```

-91771 -43710 6308 36181 83308
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -10620482 1235611 -8.595 1.02e-10 ***
Tahun        5508      620 8.884 4.20e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 50450 on 41 degrees of freedom
Multiple R-squared: 0.6581, Adjusted R-squared: 0.6498
F-statistic: 78.92 on 1 and 41 DF, p-value: 4.201e-11

```

Hasil permalan luas panen tebu tahun 2020 – 2024 dengan menggunakan metode regresi linear tersebut adalah sbb.:

Tabel 5. Hasi Estimasi Luas Panen Metode Regresi Linear

No	Tahun	Luas Panen (Ha)	Pertumbuhan (%)
1	2019*)	453,238	
2	2020	505,678	11.57
3	2021	511,186	1.09
4	2022	516,694	1.08
5	2023	522,202	1.07
6	2024	527,710	1.05

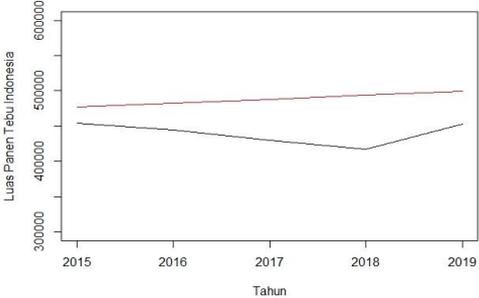
Nilai MAPE data training sebesar 13,99%  
 Min. 1st Qu. Median Mean 3rd Qu. Max.  
 0.6254 3.0109 10.6596 13.9898 16.4368 61.7116  
 Nilai MAPE data testing sebesar 11,10%  
 Min. 1st Qu. Median Mean 3rd Qu. Max.  
 5.094 8.480 10.171 11.095 13.468 18.261

Keterangan: \*) Angka Sementara, Ditjenbun

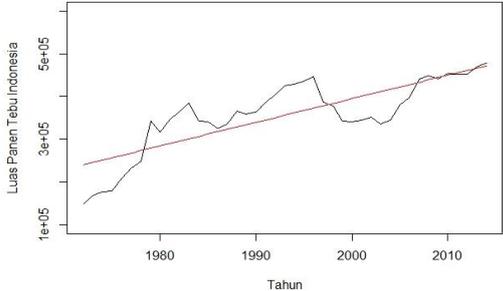
Uji yang dilakukan terhadap model regresi tersebut dan hasilnya adalah:

- Uji normalitas menggunakan Uji Jarque-Bera, diperoleh informasi bahwa data tidak menyebar normal, yakni nilai JB = 2.2275, p-value = 0.164.
- Terindikasi ada autokorelasi dengan ditunjukkan dari hasil uji Durbin-Watson sebesar 0.2436291, dan Statistic p-value = 0
- Uji Chi-Square menunjukkan Non-constant Variance Score Test  
 Variance formula: ~ fitted.values  
 Chisquare = 4.309379, Df = 1, p = 0.037903

Grafik pola garis regresi pada data training dan testing adalah sbb.:



Gambar 5. Plot hasil regresi dengan data testing



Gambar 5. Plot hasil regresi dengan data training

**c. Regresi kuadratik**

Hasil regresi kuadratik luas panen tebu dengan peubah bebas tahun menunjukkan bahwa baik intercept, peubah tahun, maupun tahun kuadrat nyata pada taraf kepercayaan 90%-100%, dengan R2 = 73,47%. Persamaan regresi kuadratik adalah :

$$\text{Produksi} = 191.853,67 + 12.623,83\text{Tahun} + -169,43\text{Tahun}^2$$

```
lm(formula = Luas_Total ~ t + t2, data = train)
Residuals:
    Min       1Q   Median       3Q      Max
-84641 -32768  12900  30784  74159
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 191853.67   19661.93    9.758 3.90e-12 ***
t            12623.83    2165.57    5.829 8.21e-07 ***
t2           -169.43     49.85   -3.399 0.00154 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 44990 on 40 degrees of freedom
Multiple R-squared:  0.7347, Adjusted R-squared:  0.7215
F-statistic: 55.39 on 2 and 40 DF, p-value: 2.98e-12
```

Hasil permalan produksi gula tahun 2020 – 2024 dengan menggunakan metode regresi kuadratik adalah sbb.:

Tabel 6. Hasil Estimasi Luas Panen Tebu Metode Regresi Kuadratik

No	Tahun	Luas Panen (Ha)	Pertumbuhan (%)
1	2019*)	453,238	
2	2020	403,620	-10.95
3	2021	399,470	-1.03
3	2022	394,982	-1.12
5	2023	390,154	-1.22
4	2024	384,988	-1.32

Nilai MAPE data training= 13,99%

Min. 1st Qu. Median Mean 3rd Qu.  
 Max.  
 0.6254 3.0109 10.6596 13.9898 16.4368  
 61.7116

Nilai MAPE data testing= 5,27%

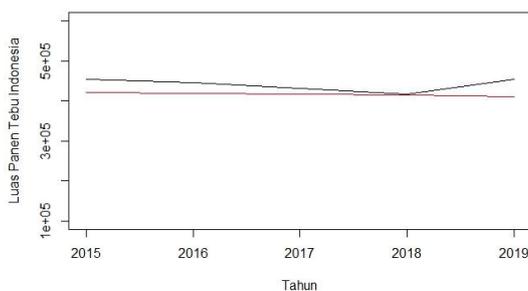
Min. 1st Qu. Median Mean 3rd Qu. Max.  
 0.850 3.147 5.796 5.270 7.217 9.343

Keterangan: \*) Angka Sementara, Ditjenbun

Uji yang dilakukan terhadap model regresi tersebut dan hasilnya adalah:

- Uji normalitas menggunakan Uji Jarque-Bera, diperoleh informasi bahwa data tidak menyebar normal, yakni nilai JB = 2.1348, p-value = 0.198
- Terindikasi ada autokorelasi dengan ditunjukkan dari hasil uji Durbin-Watson sebesar 0.3073562, dan statistic p-value = 0.0.
- Uji Chi-Square menunjukkan Non-constant variance score Test variance formula: ~ fitted.values  
 chisquare = 0.2780305, Df = 1, p = 0.59799

Grafik pola garis regresi pada data training dan testing adalah sbb.:



Gambar 5. Plot hasil regresi kuadratik dengan data testing

## 7. Estimasi Metode Fungsi Transfer

Dengan asumsi bahwa luas panen tebu Indonesia sangat dipengaruhi oleh harga gula di pasar global (Pelelangan Amerika Serikat), maka dilakukan pemodelan Fungsi Transfer dengan menyertakan peubah harga gula global tersebut.

Data deret input harga gula global dan data deret output luas panen tebu mempunyai tren meningkat dari waktu ke waktu atau tidak stasioner sehingga harus dilakukan pembedaan/*differencing* sebagai syarat untuk melakukan pemodelan Fungsi Transfer. Pembedaan tingkat 1 pada kedua data series tersebut telah stasioner berdasarkan hasil uji Augmented Dickey-Fuller, sbb.:

<p>Hasil uji Augmented Dickey_Fuller data series produksi gula:</p> <p>Value of test-statistic is: -4.3546 3.3441 3.601</p> <p>Critical values for test statistics:</p> <table border="1"> <tr> <td></td> <td>1pct</td> <td>5pct</td> <td>10pct</td> </tr> <tr> <td>tau3</td> <td>-4.15</td> <td>-3.50</td> <td>-3.18</td> </tr> <tr> <td>phi2</td> <td>7.02</td> <td>5.13</td> <td>4.31</td> </tr> <tr> <td>phi3</td> <td>9.31</td> <td>6.73</td> <td>5.61</td> </tr> </table>		1pct	5pct	10pct	tau3	-4.15	-3.50	-3.18	phi2	7.02	5.13	4.31	phi3	9.31	6.73	5.61	<p>Hasil uji Augmented Dickey_Fuller data series harga gula global:</p> <p>Value of test-statistic is: -5.0342</p> <p>Critical values for test statistics:</p> <table border="1"> <tr> <td></td> <td>1pct</td> <td>5pct</td> <td>10pct</td> </tr> <tr> <td>tau1</td> <td>-2.62</td> <td>-1.95</td> <td>-1.61</td> </tr> </table>		1pct	5pct	10pct	tau1	-2.62	-1.95	-1.61
	1pct	5pct	10pct																						
tau3	-4.15	-3.50	-3.18																						
phi2	7.02	5.13	4.31																						
phi3	9.31	6.73	5.61																						
	1pct	5pct	10pct																						
tau1	-2.62	-1.95	-1.61																						

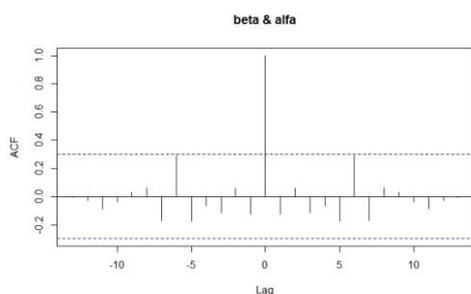
Penelusuran model deret input Harga Gula Global oleh R-Studio ditunjukkan bahwa **ARIMA (2,1,1)** merupakan model terbaik dengan komponen AR dan MA yang nyata, sbb.:

```

arima(x = train.h[, "Harga_US"], order = c(2, 1, 1))
Coefficients:
      ar1      ar2      ma1
    0.4329 -0.4655 -0.7285
s.e. 0.1707  0.1650  0.1448
sigma^2 estimated as 0.009365:  log likelihood = 37.89,  aic = -67.79
z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1  0.43286   0.17074  2.5351  0.011240 *
ar2 -0.46548   0.16503 -2.8206  0.004793 **
ma1 -0.72852   0.14478 -5.0321  4.852e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Kedua series data baik deret input (harga gula global) maupun deret output (luas panen tebu), mempunyai tren atau tidak stasioner, sehingga menghasilkan hubungan yang semu. Oleh karenanya, perlu dilakukan kajian analisis korelasi silang terhadap deret input Harga Gula Global dengan Luas Panen Tebu dengan hasil sbb.:



Gambar 6. Plot Korelasi Silang Deret Input Harga Gula Global - Luas Panen Tebu

Plot korelasi disamping menunjukkan nyata pada lag=0 atau nilai b=0 yang berarti bahwa tidak ada jeda pengaruh harga gula global terhadap luas panen tebu Indonesia, dan nilai s=0 atau tidak ada panjang pengaruh harga gula global terhadap luas panen tebu Indonesia. Nilai r=0 atau tidak ada pengaruh luas panen periode sebelumnya terhadap luas panen tebu periode ini karena adanya perubahan harga gula global. Hal ini mengingat data series luas panen tebu merupakan data tahunan.

Setelah dilakukan penelusuran model noise, diperoleh model **ARIMA (3,1,1)** sebagai model terbaik, dengan signifikansi masing-masing koefisien adalah sbb.:

```

Series: res
ARIMA(3,1,1)

Coefficients:
      ar1      ar2      ar3      ma1
-0.8470  0.1485  0.0859  0.7624
s.e.    0.3342  0.2005  0.1916  0.2919

sigma^2 estimated as 1.173e+09:  log likelihood=-496.18
AIC=1002.36  AICC=1004.02  BIC=1011.04

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1 -0.847045  0.334198 -2.5346 0.011259 *
ar2  0.148507  0.200480  0.7408 0.458839
ar3  0.085929  0.191573  0.4485 0.653760
ma1  0.762421  0.291901  2.6119 0.009004 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Peramalan berbasis Fungsi Transfer dengan derajat  $(r,b,s) = (0,0,0)$  dan model noise ARIMA  $(3,1,1)$  dilakukan terhadap data aktual (series data training), diperoleh nilai **MAPE = 5,24%**

```

Series: test.h[, "Prod_Total"]
Regression with ARIMA(3,1,1) errors

Coefficients:
      ar1      ar2      ar3      ma1      xreg
-0.0567  0.2837  0.1531  0.1054 -19337.6
s.e.    0.0000  0.0000  0.0000  0.0000  0.0

sigma^2 estimated as 657796741:  log likelihood=-53.87
AIC=109.74  AICC=111.74  BIC=109.12

Training set error measures:
      ME  RMSE  MAE  MPE  MAPE  MASE  ACF1
Trainingset-14030.17148424 115083.4 -0.6479129 5.241765 0.9556407 0.1674493

```

Untuk melakukan peramalan luas panen tebu Indonesia tahun 2020-2024, maka menggunakan metode Fungsi Transfer dengan sebelumnya dilakukan peramalan deret input harga gula global yang telah diperoleh pada penelusuran model ARIMA sebelumnya, yakni ARIMA  $(2,1,1)$ . Model ini menghasilkan nilai **MAPE data testing** sebesar **3,34%** sbb.:

```

Series: test.h[, "Luas_Total"]
Regression with ARIMA(3,1,1) errors

Coefficients:
      ar1      ar2      ar3      ma1      xreg
-0.0567  0.2837  0.1531  0.1054 -19337.6
s.e.    0.0000  0.0000  0.0000  0.0000  0.0

sigma^2 estimated as 657796741:  log likelihood=-45.95
AIC=93.91  AICC=95.91  BIC=93.29

Training set error measures:
      ME  RMSE  MAE  MPE  MAPE  MASE  ACF1
Trainingset 2143.931 20523.92 14801.33 0.3984613 3.340167 0.8193717 -0.04094452

```

Hasil ramalan luas panen tebu tahun 2020 – 2024 menggunakan model Fungsi Transfer dengan derajat  $(r,b,s) = (0,0,0)$  dan model noise ARIMA  $(3,1,1)$  adalah sebagai berikut:

```

Time Series:
Start = 49
End = 53
Frequency = 1
[1] 449437.9 456665.7 460917.1 461721.2 464008.4

```

## 8. Estimasi Vector Auto Regression (VAR)

Model VAR merupakan alat analisis yang sangat berguna dalam memahami adanya hubungan timbal balik (interrelationship) antara peubah ekonomi maupun dalam pembentukan ekonomi yang berstruktur. Model VAR akan dibangun dengan mengikutsertakan peubah luas panen tebu, harga gula global, ekspor molasses dan impor gula Indonesia. Model VAR dapat digunakan untuk menjelaskan fenomena semua peubah yang masuk ke dalam sistem, namun dalam sub bab ini lebih dibahas fenomena terhadap peubah luas panen tebu Indonesia, dan menggunakan hasil model VAR untuk peramalan luas panen tebu tahun 2020-2024.

Hasil penelusuran model diperoleh bahwa VAR dengan lag=1 dengan tidak mengikutsertakan tren merupakan model terbaik, dengan signifikansi sebagai berikut:

```
VAR Estimation Results:
=====
Endogenous variables: Luas_Total, Harga_EU, Harga_US, Ekspor_Mol, Impor_Gula
Deterministic variables: const
Sample size: 47
Log Likelihood: -1709.599
Roots of the characteristic polynomial:
0.9846 0.8974 0.6107 0.6107 0.1442
Call:
VAR(y = olahtebu[1:48, c(5, 10, 11, 12, 13)], p = 1, type = "const")

Estimation results for equation Luas_Total:
=====
Luas_Total = Luas_Total.l1 + Harga_EU.l1 + Harga_US.l1 + Ekspor_Mol.l1 +
Impor_Gula.l1 + const

Estimate Std. Error t value Pr(>|t|)
Luas_Total.l1 6.829e-01 9.731e-02 7.018 1.56e-08 ***
Harga_EU.l1 8.065e+04 3.728e+04 2.164 0.03637 *
Harga_US.l1 1.039e+04 3.472e+04 0.299 0.76636
Ekspor_Mol.l1 3.678e-02 1.999e-02 1.840 0.07301 .
Impor_Gula.l1 8.290e-03 3.891e-03 2.130 0.03919 *
const 5.306e+04 1.802e+04 2.944 0.00531 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22800 on 41 degrees of freedom
Multiple R-squared: 0.9265, Adjusted R-squared: 0.9176
F-statistic: 103.4 on 5 and 41 DF, p-value: < 2.2e-16
```

Uji normalitas terhadap sisaan, homokedastisitas dan non autokorelasi sudah terpenuhi sebagai berikut:

```
Portmanteau Test (asymptotic)
data: Residuals of VAR object varhsheet1
Chi-squared = 282.44, df = 375, p-value = 0.9999

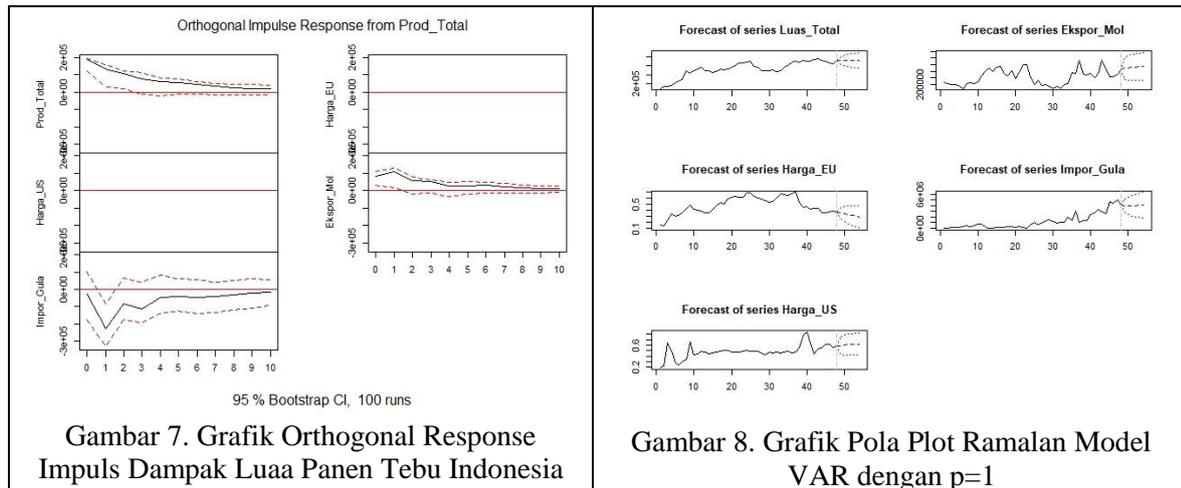
skewness only (multivariate)
data: Residuals of VAR object varhsheet1
Chi-squared = 38.263, df = 5, p-value = 3.341e-07

$Kurtosis
Kurtosis only (multivariate)
data: Residuals of VAR object varhsheet1
Chi-squared = 45.057, df = 5, p-value = 1.412e-08

ARCH (multivariate)
data: Residuals of VAR object varhsheet1
Chi-squared = 630, df = 1125, p-value = 1
```

Hasil grafik orthogonal impuls respons menunjukkan bahwa apabila ada perubahan luas panen tebu Indonesia pada tahun tertentu, maka luas panen tebu akan merespon hingga 3 tahun ke depan, ekspor

molasses akan terpengaruh hingga 1 tahun ke depan. Perubahan luas panen tebu Indonesia tidak berdampak pada harga gula global maupun impor gula Indonesia.



Nilai MAPE data ser testing sebesar 4,13%, sementara nilai MAPE data set traning sebsar 9,26% sebagai berikut:

MAPE DATA TRAINING  
 Min. 1st Qu. Median Mean 3rd Qu. Max.  
 0.5089 3.3332 6.2377 9.6029 9.2570 146.4256

MAPE DATA TESTING  
 Min. 1st Qu. Median Mean 3rd Qu. Max.  
 0.6466 1.2726 3.7680 4.1298 5.9326 9.5012

Model VAR dengan p=2 dengan mempertimbangkan tren digunakan untuk melakukan estimasi luas panen tebu Indonesia dengan hasil sebagai berikut:

Tabel 7. Hasil Ramalan Luas Panen Tebu Metode VAR

[1] 455634.9 457107.5 457695.7 457662.1 457250.8 456644.4

### SIMPULAN DAN SARAN

Ke-empat metode yang digunakan untuk memodelkan dan meramalkan data luas panen tebu di Indonesia dapat ditabulasikan sebagai berikut:

Tabel 8. Perbandingan MAPE dan Hasil Ramalan dengan Empat Metode

No	Model	MAPE (%)		Hasil ramalan produksi gula (Ton)					Pertumbuhan 2020-2024 (%)
		Training	Testing	2020	2021	2022	2023	2024	
1	Arima (4,1,1)	5.06	6.15	449,222	459,744	465,808	466,965	469,477	1.11
2	Regresi								
	a. Linear Sederhana	13.99	11.10	505,678	511,186	516,694	522,202	527,710	1.07
	b. Kuadratik	13.99	5.27	403,620	399,470	394,982	390,154	384,988	-1.17
3	Fungsi Transfer	5.24	3.34	449,438	456,666	460,917	461,721	464,008	0.80
4	VAR	9.60	4.13	455,635	457,108	457,696	457,662	457,251	0.09

Berdasarkan hasil analisis ke-4 metode dapat disimpulkan bahwa Model Fungsi Transfer merupakan model terbaik karena mempunyai nilai MAPE yang terkecil dengan hasil ramalan luas panen yang cukup realistis.

Berdasarkan simpulan diatas, maka beberapa saran yang diajukan adalah sbb.;

- Model Fungsi Transfer dapat disarankan untuk digunakan dalam memodelkan data luas panen tebu Indonesia dan meramalkan beberapa tahun ke depan.
- Model VAR dapat dipertimbangkan sebagai model alternatif terbaik karena mempunyai nilai MAPE yang tidak terlalu berbeda dengan model Fungsi Transfer.

### **DAFTAR PUSTAKA**

- Enders, W. 2010. Applied Econometric Time Series. USA: University of Alabama. Wiley, Third Edition.
- Guha, B and Bandyopadhyay, G. 2016. Gold Price Forecasting Using ARIMA Model. Journal of Advanced Management Science Vol. 4, No. 2, March 2016
- Gujarati, D.N. and D.C. Porter, 2010. Dasar-dasar Ekonometrika. Jakarta: Penerbit Salemba Empat, Buku 2, Edisi 5.



# KAJIAN PENGEMBANGAN METODE ESTIMASI LUAS AREAL KAKAO

Yuliawati Rohmah

*Center for Agricultural Data and Information System-Ministry of Agriculture*

*Jalan Harsono RM No. 3 Gd D Lt 4, Ragunan-Jakarta Selatan, Indonesia*

*E-mail: yuliawati.rohmah@gmail.com*

## ABSTRAK

Kakao merupakan salah satu komoditas perkebunan strategis di Indonesia. Selama ini data luas areal kakao yang dirilis oleh Direktorat Jenderal Perkebunan bersama BPS berupa data tahunan yang disajikan untuk kondisi 2 tahun yang lalu merupakan Angka Tetap (ATAP), satu tahun yang lalu merupakan Angka Sementara (ASEM) dan untuk tahun yang berjalan merupakan Angka Estimasi (AESTI). Selama ini metode yang digunakan untuk menyusun AESTI adalah Metode *Single Exponential Smoothing* (SES) dan *Double Exponential Smoothing* (DES). Metode untuk menghasilkan AESTI data perkebunan perlu dikaji kembali agar didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik. Sehingga kajian ini bertujuan untuk melakukan pemodelan, membandingkan efektivitas dari pemodelan tersebut dan menentukan metode terbaik dalam mengestimasi luas areal kakao. Metode yang diterapkan adalah metode *Autoregressive Integrated Moving Average* (ARIMA), regresi, fungsi transfer dan *Vector Auto Regression* (VAR) dengan menggunakan *software program RStudio*. Pemilihan model terbaik dilakukan dengan membandingkan besaran nilai *Mean Absolute Percentage Error* (MAPE) dan kerealistisan hasil estimasi dibandingkan dengan data series sebelumnya. Berdasarkan dari hasil estimasi dan nilai MAPE disimpulkan bahwa Fungsi Transfer adalah model terbaik untuk estimasi luas areal kakao.

Kata kunci: ARIMA, regresi, fungsi transfer, VAR, luas areal kakao

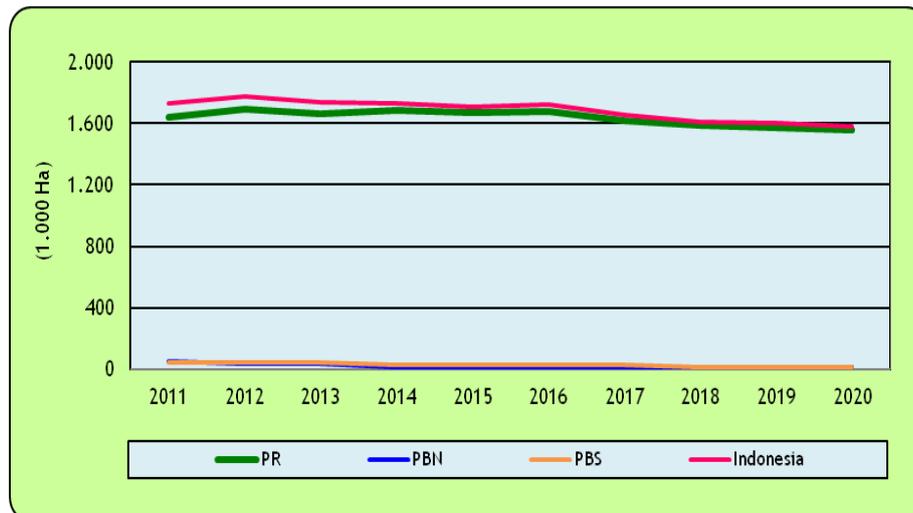
## ABSTRACT

*Cocoa is one of the strategic estate crop commodities in Indonesia. All this time the cocoa area data released by the Directorate General of Estate Crop together with BPS annually with a lag n-2 year for Fixed Number (ATAP), a lag n-1 year for Provisional Number (ASEM) and the current year for Estimated Number (AESTI). All this time, the methods used to compile AESTI are the Single Exponential Smoothing (SES) and Double Exponential Smoothing (DES). The method for producing AESTI data needs to be reviewed in order to obtain a method that is more accurate, more objective and statistically better. The purpose of this paper are modeling, comparing the effectiveness of the models and determine the best method in estimating cocoa area. The method applied are Autoregressive Integrated Moving Average (ARIMA), regression, transfer function and Vector Auto Regression (VAR) using the RStudio program. The best method will be selected by comparing the value of the Mean Absolute Percentage Error (MAPE) and the reliability of the estimation results to the previous data series. Based on the estimation results and MAPE value, it is concluded that Transfer Function is the best model for estimating cocoa area.*

*Keywords: ARIMA, regression, transfer function, VAR, cocoa area*

## PENDAHULUAN

Kakao merupakan salah satu komoditas ekspor perkebunan strategis yang ditetapkan pemerintah. Berdasarkan data FAO selama tahun 2014-2018, Indonesia menempati posisi ketiga sebagai negara dengan luas areal kakao terbesar di dunia dengan kontribusi sebesar 15,34% terhadap total luas areal dunia dengan luas mencapai 1,7 juta hasetelah Pantai Gading dan Ghana dengan masing-masing kontribusi mencapai 31,85% dan 15,37%. Perkebunan kakao di Indonesia didominasi oleh Perkebunan Rakyat (PR) sebesar 97,06% dan sisanya dimiliki oleh Perkebunan Besar Nasional (PBN) sebanyak 1,29% dan Perkebunan Besar Swasta (PBS) mencapai 1,65%. Daerah sentra luas areal kakao di Indonesia pada tahun 2018 pada urutan 3 teratas yang memberikan kontribusi diatas 10% terhadap total luas areal nasional adalah Provinsi Sulawesi Tengah (17,61%), Sulawesi Tenggara (15,82%) dan Sulawesi Selatan (13,54%). Perkembangan luas areal kakao Indonesia dalam 10 tahun terakhir berfluktuatif dan cenderung menurun dengan rata-rata pertumbuhan sebesar negatif 0,39% per tahun (Gambar 1).



Gambar 1. Perkembangan Luas areal Kakao Indonesia Tahun 2011-2020

Saat ini, rilis resmi data luas areal kakao dikeluarkan oleh Direktorat Jenderal Perkebunan bersama BPS berupa data tahunan yang disajikan untuk kondisi 2 tahun yang lalu (*lag n-2*) merupakan Angka Tetap (ATAP), satu tahun yang lalu (*lag n-1*) merupakan Angka Sementara (ASEM) dan untuk tahun yang berjalan merupakan Angka Estimasi (AESTI). Data statistik perkebunan yang diperoleh merupakan hasil sinkronisasi dan validasi yang dilaksanakan secara berjenjang dari tingkat kabupaten sampai tingkat nasional dengan mengacu pada Pedoman Pengelolaan Data Komoditas Perkebunan (PDKP) yang diterbitkan oleh Direktorat Jenderal Perkebunan. Selama ini metode yang digunakan untuk menyusun AESTI adalah Metode Pemulusan Eksponensial Tunggal atau *Single Exponential Smoothing* (SES) dan Pemulusan Eksponensial Ganda atau *Double Exponential Smoothing* (DES). Pemilihan model terbaik dilakukan dengan membandingkan besaran nilai *Mean Absolute Percentage Error* (MAPE) dan kerealistisan hasil estimasi dibandingkan dengan data series sebelumnya. Metode untuk menghasilkan AESTI data perkebunan perlu dikaji kembali agar

didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik dibandingkan metode yang telah dilakukan selama ini.

Berdasarkan hal di atas, maka kajian ini bertujuan untuk:

- f. Melakukan analisa dan estimasi data luas areal kakao menggunakan model *Auto Regressive Integrated Moving Average* (ARIMA), Regresi, Fungsi Transfer dan *Vector Auto Regression* (VAR).
- g. Membandingkan metode tersebut dalam memperoleh model estimasi data luas areal komoditas kakao yang memiliki tingkat akurasi tertinggi.
- h. Menentukan metode terbaik dalam meramal data luas areal komoditas kakao.

## BAHAN DAN METODE

Data yang digunakan dalam kajian ini adalah data sekunder *time series* tahunan. Variabel, satuan, level, periode dan sumber data yang digunakan dalam kajian ini adalah sebagai berikut:

Tabel 1. Jenis Variabel, Periode dan Sumber Data

No	Variabel Data	Satuan	Level	Periode	Sumber
1	Luas areal	Ha	Nasional	1980-2019	Ditjen Perkebunan
2	Produksi	Ton	Nasional	1980-2019	Ditjen Perkebunan
3	Harga dunia	US\$/Kg	Internasional	1980-2019	World Bank
4	Volume ekspor	Ton	Nasional	1980-2019	BPS
5	Volume impor	Ton	Nasional	1980-2019	BPS

Data yang digunakan dalam metode ARIMA dan regresi adalah luas areal, sedangkan variabel produksi digunakan pada metode fungsi transfer sebagai variabel yang diduga dapat mempengaruhi luas areal kakao nasional. Adapun pada metode VAR, data yang digunakan adalah luas areal, produksi, harga dunia, volume ekspor dan volume impor karena kakao merupakan komoditas ekspor andalan. Pertimbangan lain dalam pemilihan dan penggunaan variabel data dalam model adalah ketersediaan series data dalam jangka waktu yang cukup panjang.

Pada proses pengolahan dan analisa, data *time series* dibagi menjadi dua bagian yakni data *training* untuk penyusunan model periode tahun 1980 - 2013 dan sisanya sebagai data *testing* untuk validasi model periode tahun 2014 - 2019. Kemudian dari hasil data *training* disusun model dan dilakukan estimasi sesuai periode data *testing*, setelah itu dilakukan evaluasi kesesuaian ramalannya. Model terbaik dipilih dari berbagai alternatif metode estimasi yang dicoba dengan melihat nilai *Mean Absolute Percentage Error* (MAPE) dan kesesuaian hasil ramalan dengan historis data aktualnya. Model estimasi terbaik yang terpilih kemudian dilakukan untuk estimasi 5 tahun ke depan yakni tahun 2020 – 2024 dengan menggabungkan seluruh data (*training* dan *testing*). Metode estimasi luas areal kakao

nasional yang digunakan dalam kajian ini terdiri dari ARIMA, Regresi, Fungsi Transfer dan VAR menggunakan *software* program *RStudio*.

## 1. ARIMA

Metode ARIMA merupakan metode yang hanya menggunakan variabel dependen dan mengabaikan variabel independen sewaktu melakukan estimasi. ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampaunya. Bentuk modelnya adalah:

$$y_t = \mu + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \dots + \theta_p y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

$y_t$	= data <i>time series</i> sebagai variable dependen pada waktu ke-t
$y_{t-p}$	= data <i>time series</i> pada kurun waktu ke (t-p)
$\mu$	= suatu konstanta
$\theta_1 \theta_q \phi_1 \phi_n$	= parameter-parameter model
$\varepsilon_{t-q}$	= nilai sisaan pada waktu ke-(t-q)

## 2. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih variabel.

### a. Regresi Linier Sederhana

Model regresi linier sederhana melibatkan satu variabel prediktor dan diformulasikan dalam model matematis sebagai berikut:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

dimana:

$y_t$	= variabel respon/dependen/terikat
$x_t$	= variabel prediktor/independen/bebas
$\beta_0$	= konstanta/intercept
$\beta_1$	= koefisien regresi variabel bebas
$\varepsilon_t$	= error/residu/sisaan

### b. Regresi Kuadratik

Model regresi kuadratik merupakan nilai variabel tak bebas dengan bentuk naik atau turun secara linier atau terjadi secara parabola dengan model matematis berikut ini:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t$$

dimana:

$y_t$	= variabel respon/dependen/terikat
$x_t$	= variabel prediktor/independen/bebas

- $\beta_0$  = konstanta/intercept
- $\beta_1, \beta_2$  = koefisien regresi variabel bebas
- $\varepsilon_t$  = error/residu/sisaan

### 3. Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret *output* atau  $y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $x_t$ ) dengan deret *output* tersebut.

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t$$

dimana :

- $b$  = panjang jeda pengaruh  $x_t$  terhadap  $y_t$
- $r$  = panjang *lag*  $y$  periode sebelumnya yang masih mempengaruhi  $y_t$
- $s$  = panjang *lag*  $x$  periode sebelumnya yang masih mempengaruhi  $y_t$
- $p$  = ordo AR bagi *noise*  $N_t$
- $q$  = ordo MA bagi *noise*  $N_t$

### 4. VAR

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen., karena pada kenyataannya suatu variabel dapat bertindak sebagai endogen atau dependen dari suatu variabel yang lainnya (Gujarati, 2010).

$$x_t = \beta_{10} + \sum_{i=1}^p \alpha_{1i} x_{t-i} + \sum_{i=1}^p \beta_{1i} y_{t-i} + \sum_{i=1}^p \gamma_{1i} z_{t-i} + e_{1t}$$

$$y_t = \beta_{20} + \sum_{i=1}^p \alpha_{2i} x_{t-i} + \sum_{i=1}^p \beta_{2i} y_{t-i} + \sum_{i=1}^p \gamma_{2i} z_{t-i} + e_{2t}$$

$$z_t = \beta_{30} + \sum_{i=1}^p \alpha_{3i} x_{t-i} + \sum_{i=1}^p \beta_{3i} y_{t-i} + \sum_{i=1}^p \gamma_{3i} z_{t-i} + e_{3t}$$

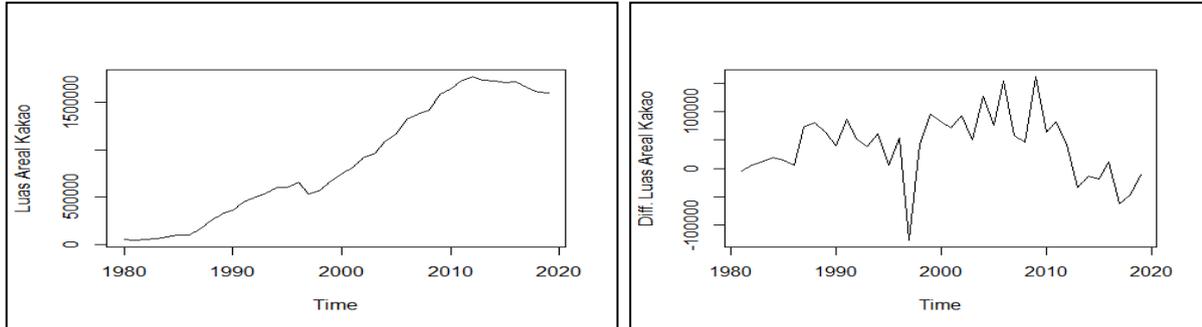
dimana .

- $x_t, y_t, z_t$  = variabel endogen
- $\beta_0$  = vektor konstanta  $n \times 1$ .
- $\alpha, \beta, \gamma$  = parameter dari  $x, y,$  dan  $z$
- $p$  = panjang *lag*
- $t$  = waktu
- $\varepsilon$  = vektor dari *shock* masing-masing variabel

## HASIL DAN PEMBAHASAN

### 1. ARIMA

#### Eksplorasi Data Luas areal Kakao Indonesia

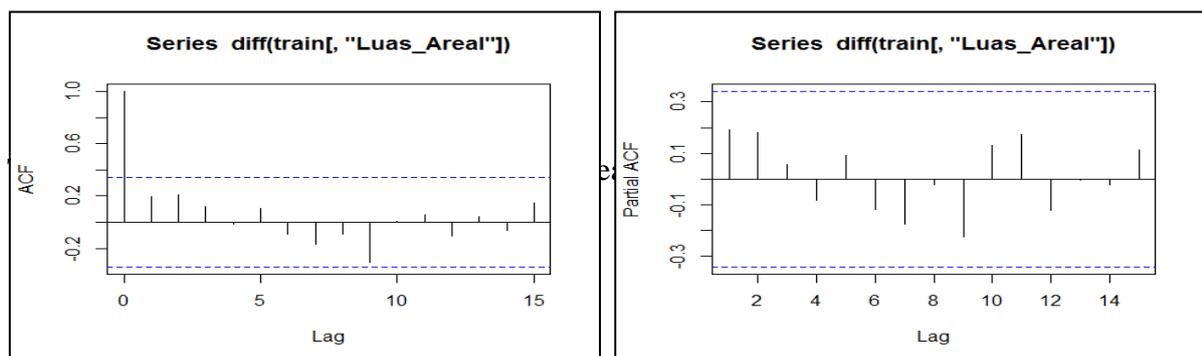


Gambar 2. Luas areal Kakao Tahun 1980-2019      Gambar 3. Luas areal Kakao *Differencing 1*

Luas areal kakao dalam periode 40 tahun terakhir berfluktuatif dengan rata-rata pertumbuhan sebesar 10,41% per tahun. Berdasarkan uji kestasioneran data menggunakan *Augmented Dickey Fuller* (ADF) memperlihatkan data luas areal kakao tidak stasioner, karena nilai *t-statistic* (-1,34) lebih besar dibandingkan *critical value* (-4,15; -3,50; -3,18) sehingga perlu dilakukan proses *differencing level 1*. Hal ini diperkuat dengan plot luas areal kakao berdasarkan sebaran datanya (Gambar 1) yang tidak konstan di sekitar rata-rata. Hasil *differencing level 1* luas areal kakao telah bersifat stasioner karena nilai *t-statistic* (-1,73) lebih kecil dibandingkan *critical value* pada alpha 10% (-1,61) dan sebaran datanya (Gambar 2) memiliki pola *single mean* atau konstan sekitar rata-rata bukan nol.

#### Model ARIMA Luas areal Kakao

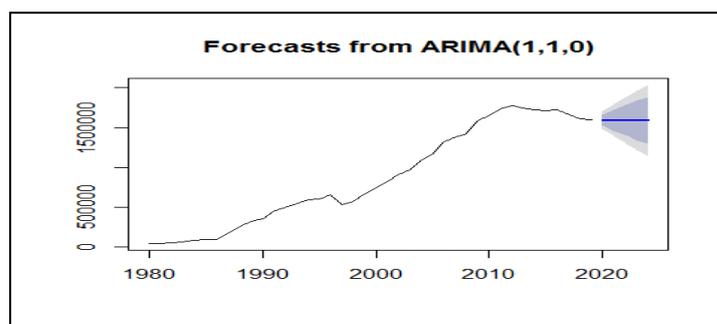
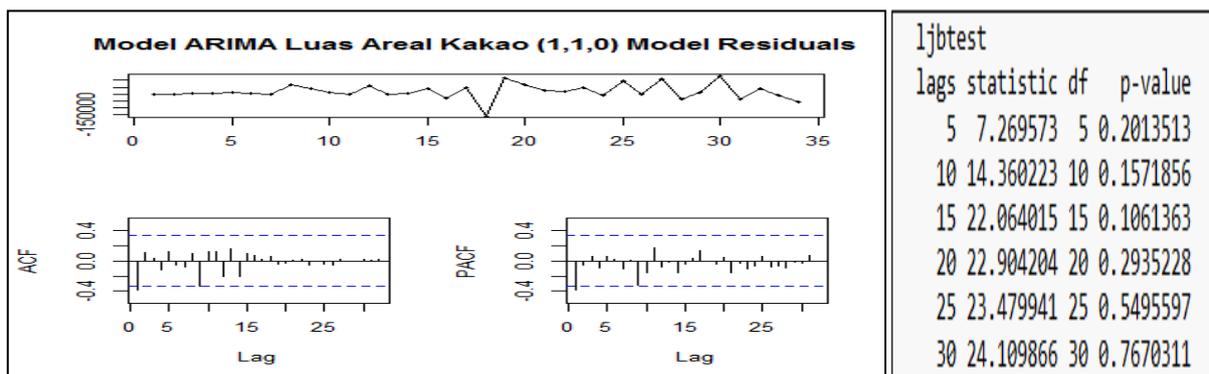
Tahap awal pada metode estimasi dengan ARIMA, setelah dipastikan data bersifat stasioner, maka dilakukan identifikasi model ARIMA berdasarkan hasil plot *Auto Correlation Function* (ACF) dan *Partial Auto Correlation Function* (PACF) seperti tampak pada Gambar 4 dan Gambar 5 dibawah ini. Dari hasil plot ACF bersifat *cut off* pada lag 1, sedangkan plot PACF, data sudah tidak memiliki pola khusus baik *tail off* ataupun *cut off* sehingga diperoleh dugaan awal untuk model ARIMA (1,1,0) atau ARIMA (0,1,1). Alternatif lain untuk mendapatkan model ARIMA dapat diperoleh dengan melakukan *overfitting* dari hasil *armaselect* dengan uji *Minimum Information Criterion* (Minic) seperti yang terdapat Tabel 2.



minic1	p	q	sbc	No	Model	MAPE (%)		Luas Areal (Ton)				
						Training	Testing	2020	2021	2022	2023	2024
[1,]	0	2	717.3321	1	ARIMA (0,1,2)	8,28	2,97	1.600.697	1.609.680	1.609.680	1.609.680	1.609.680
[2,]	0	0	718.2844	2	ARIMA (0,1,0)	16,36	15,16	1.600.647	1.600.647	1.600.647	1.600.647	1.600.647
[3,]	1	0	720.2641	3	ARIMA (1,1,0)	7,81	2,47	1.594.494	1.590.842	1.588.675	1.587.388	1.586.625
[4,]	1	2	720.4322	4	ARIMA (1,1,2)	7,87	7,00	1.574.951	1.555.230	1.537.896	1.522.661	1.509.270
[5,]	2	0	723.0150	5	ARIMA (2,1,0)	7,71	3,54	1.579.459	1.567.879	1.555.859	1.547.200	1.539.613
[6,]	2	2	723.6289	6	ARIMA (2,1,2)	7,92	5,24	1.570.459	1.545.291	1.525.106	1.509.522	1.497.968
[7,]	0	3	724.7791	7	ARIMA (0,1,3)	8,12	2,48	1.577.857	1.579.710	1.586.621	1.586.621	1.586.621
[8,]	3	2	726.8322	8	ARIMA (3,1,2)	7,86	6,85	1.572.838	1.554.686	1.537.295	1.521.809	1.508.708
[9,]	3	0	727.2303	9	ARIMA (3,1,0)	7,86	5,61	1.573.084	1.554.078	1.538.138	1.523.097	1.510.614
[10,]	1	3	727.5623	10	ARIMA (1,1,3)	7,86	7,16	1.574.139	1.551.258	1.532.814	1.516.828	1.502.972

Setelah model dan hasil estimasi diperoleh, maka tahap berikutnya adalah mengevaluasi hasil ramalan baik dengan menggunakan uji MAPE untuk data *training* dan data *testing* maupun dengan melihat kerealistisan hasil estimasi dengan data aktualnya. Model terbaik terpilih adalah ARIMA (1,1,0) karena memiliki nilai MAPE terkecil untuk data training dan data testing yakni sebesar 7,81% dan 2,47%. Hasil estimasinya juga dianggap paling mendekati data aktual (Tabel 2). Langkah selanjutnya berupa pemeriksaan sisaan, plot sisaan terdistribusi normal dan dan plot ACFserta PACF sisaan tidak nyata. Sedangkan dari hasil uji *Ljung-Box*, autokorelasi sisaan tidak signifikan pada 30 *lag* (Tabel 3). Hasil estimasi luas areal kakao dengan model ARIMA (1,1,0) untuk 5 tahun kedepan menurun tiap tahunnya dengan rata-rata sebesar -0,12% per tahun (Gambar 6).

Tabel 3. Hasil Uji Pemeriksaan Sisaan dan Hasil Uji *Ljung-Box* ARIMA (1,1,0)



Gambar 6. Hasil Estimasi Luas areal Kakao Model ARIMA (1,1,0) Tahun 2020-2024

## II. REGRESI

### a. Regresi Linier Sederhana

Dalam model regresi linier sederhana, variabel eksogen yang digunakan adalah tahun. Tahap awal untuk pembentukan model regresi adalah melakukan pendugaan model untuk mengetahui hubungan antara tahun dengan luas areal kakao. Hasil uji *coefficients* memperlihatkan signifikansi baik pada intersep maupun variabel tahun pada alpha 0,1%. Model regresi yang dihasilkan juga memiliki selang kepercayaan yang sangat baik dimana nilai  $R^2$  dan *adjusted R<sup>2</sup>* masing-masing sebesar 95% (Tabel 4). Persamaan model regresi linier luas areal kakao adalah sebagai berikut:

$$\text{Luas areal} = -110060447 + 55495\text{Tahun}$$

Tabel 4. Hasil Test Coefficients Model Regresi Linier Sederhana Luas areal Kakao

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-110060447	4388559	-25.08	<2e-16 ***
Tahun	55495	2198	25.25	<2e-16 ***
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 125700 on 32 degrees of freedom				
Multiple R-squared: 0.9522, Adjusted R-squared: 0.9507				
F-statistic: 637.4 on 1 and 32 DF, p-value: < 2.2e-16				

Dari evaluasi terhadap hasil estimasi dengan menggunakan uji MAPE atau nilai kesalahan rata-rata dalam meramalkan sebesar 48,82% untuk *training* dan 11,07% untuk *testing* (Tabel 5). Tahap selanjutnya melakukan serangkaian pengujian terhadap beberapa asumsi model regresi yakni sebaran data, autokorelasi dan homoskedastisitas. Seperti yang terdapat pada Tabel 6, berdasarkan hasil Uji Jarque-Bera diperoleh data menyebar normal, terindikasi ada autokorelasi dari hasil Uji Durbin-Watson dan homoskedastisitas terpenuhi dari hasil Uji Chi-Square. Luas areal kakao pada tahun 2020-2024 akan mengalami kenaikan yang sangat signifikan dengan rata-rata per tahun sebesar 2,62% berdasarkan hasil estimasi dengan menggunakan model regresi linier (Tabel 7).

Tabel 5. Nilai MAPE Data *Training* dan Data *Testing* Model Regresi Linier Luas areal Kakao

Data	MAPE (%)
Training	48,82
Testing	11,07

Tabel 6. Hasil Uji Asumsi Model Regresi Linier Luas areal Kakao

---

```

Uji Jarque-Bera
Jarque-Bera test for normality
data: reg1$residuals
JB = 0.5823, p-value = 0.6995

```

---

```

Uji Durbin-watson
lag Autocorrelation D-W Statistic p-value
 1          0.8503682    0.1805956    0
Alternative hypothesis: rho != 0

```

---

```

Uji Chi-Square
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.7172004, Df = 1, p = 0.39706

```

---

Tabel 7. Hasil Estimasi Luas areal Kakao Model Regresi Linier Tahun 2020-2024

Tahun	Luas areal Kakao (Ton)
2020	2.038.996
2021	2.094.491
2022	2.149.985
2023	2.205.480
2024	2.260.975
Rata-rata Pertumbuhan (%)	2,62

## b. Regresi Kuadratik

Variabel eksogen yang digunakan dalam metode regresi kuadratik adalah tahun. Berdasarkan hasil pendugaan model regresi kuadratik untuk luas areal kakao diperoleh signifikansi untuk tahun pada alpha 1% dan variabel tahun kuadrat pada alpha 0,1%, sedangkan untuk intersep tidak signifikan dari hasil uji *coefficient* dengan selang kepercayaan masing-masing sebesar 99% untuk *R-squared* dan 98% untuk *adjusted R-squared* (Tabel 8). Persamaan model regresi kuadratik luas areal kakao adalah sebagai berikut:

$$\text{Luas areal} = 10486,3 + 14365,4\text{Tahun} + 1175,1\text{Tahun}^2$$

Tabel 8. Hasil Test Coefficients Model Regresi Kuadratik Luas areal Kakao

---

```

Coefficients:
(Intercept) Estimate Std. Error t value Pr(>|t|)
t            14365.4    5152.8    2.788  0.00898 **
t2           1175.1     142.8    8.228  2.71e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 71600 on 31 degrees of freedom
Multiple R-squared:  0.985,    Adjusted R-squared:  0.984
F-statistic: 1017 on 2 and 31 DF,  p-value: < 2.2e-16

```

---

Tahap berikutnya yakni melakukan penghitungan MAPE data *training* dan data *testing*. Nilai MAPE data *testing* cukup besar yakni 32,28% (Tabel 9).

Tabel 9. Nilai MAPE Data *Training* dan *Testing* Model Regresi Kuadratik Luas areal Kakao

Data	MAPE (%)
Training	-
Testing	32,28

Tahap selanjutnya melakukan pengujian terhadap beberapa asumsi model regresi dengan Uji Jarque-Bera, Uji Durbin-Watson dan Uji Chi-Square. Berdasarkan serangkaian uji tersebut diperoleh hasil bahwa data tersebar normal, diduga terdapat autokorelasi dan heteroskedastisitas tidak terjadi (Tabel 10). Tahapan akhir yakni melakukan estimasi sementara luas areal kakao untuk tahun 2020-2024 dengan hasil estimasi seperti tertera pada Tabel 11 yang memperlihatkan kenaikan luas areal kakao yang sangat signifikan setiap tahunnya dengan rata-rata 4,21%.

Tabel 10. Hasil Uji Asumsi Model Regresi Kuadratik Luas areal Kakao

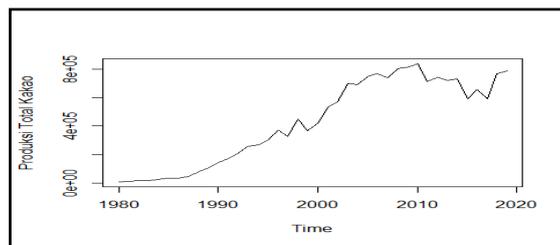
Uji Jarque-Bera			
Jarque-Bera test for normality			
data: reg2\$residuals			
JB = 1.7406, p-value = 0.231			
Uji Durbin-Watson			
lag	Autocorrelation	D-w Statistic	p-value
1	0.6961509	0.5191436	0
Alternative hypothesis: rho != 0			
Uji Chi-Square			
Non-constant Variance Score Test			
Variance formula: ~ fitted.values			
Chisquare = 0.341341, Df = 1, p = 0.55906			

Tabel 11. Hasil Estimasi Luas areal Kakao Model Regresi Kuadratik Tahun 2020-2024

Tahun	Luas areal Kakao (Ton)
2020	2.574.853
2021	2.686.753
2022	2.801.004
2023	2.917.606
2024	3.036.557
Rata-rata Pertumbuhan (%)	4,21

### III. FUNGSI TRANSFER

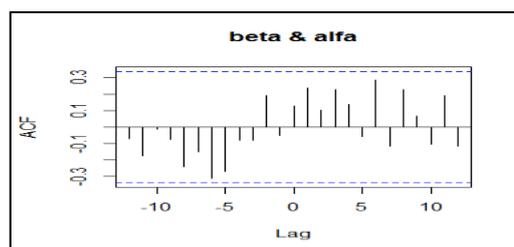
Pada metode fungsi transfer, peubah input yang digunakan adalah produksi (Gambar 7) dengan pertimbangan naik turunnya produksi di tahun sebelumnya akan sangat mempengaruhi petani untuk menanam/memelihara kakao atau mengganti komoditas lain untuk ditanam di areal kebunnya, sehingga produksi diduga dapat menentukan luas areal. Langkah pertama untuk proses analisa model adalah dengan mengidentifikasi model ARIMA peubah input berdasarkan hasil uji ADF dan plot ACF serta PACF. Berdasarkan hasil uji ADF diketahui data peubah input non stasioner dimana nilai *test-statistic* (-1,24) lebih besar dibanding *critical value* (-4,15; -3,50; -3,18), sehingga harus dilakukan proses *differencing*. Data produksi stasioner setelah di *differencing level 1* dengan nilai *critical value* lebih tinggi pada (-1,95) pada alpha 5% dibandingkan nilai *test-statistic* (-2,46).



Gambar 7. Produksi Kakao Tahun 1980-2019

Langkah kedua menduga model ARIMA peubah input baik dengan *autoarima* ataupun *armaselect*. Setelah melakukan *overfitting* dari berbagai kemungkinan model ARIMA peubah input, maka dipilih ARIMA (1,1,0) dengan pertimbangan nilai MAPE dan hasil *coefficients test* yang signifikan untuk ar1 pada alpha 10% dan *drift* pada alpha 0,01%. Langkah ketiga yakni melakukan *prewhitening* dan korelasi silang antara deret input dengan luas areal yang menghasilkan nilai r,s,b yakni (0,0,0) karena tidak ada yang nyata seperti tampak pada Gambar 8. Nilai b merupakan *lag* pertama kali dampak input berpengaruh terhadap output, s *lag* berikutnya setelah b dimana input berdampak terhadap output, dan r merupakan pengaruh output terhadap dirinya sendiri. Pada gambar 8 dapat dijelaskan bahwa nilai b dan s adalah 0 karena tidak ada yang nyata, sedangkan r dianggap 0 karena kakao merupakan tanaman tahunan.

Langkah keempat yaitu pengepasan model (r,s,b) = (0,0,0) yang menghasilkan nilai MAPE 17,56%. Identifikasi model *noise* atau residual dari peubah input merupakan langkah kelima yang dilakukan dengan model ARIMA seperti langkah kedua yang menghasilkan model ARIMA (2,1,5) sebagai model terpilih untuk residual. Langkah selanjutnya melakukan pengepasan model (r,s,b) = (0,0,0) dan *noise* (2,1,5) dengan nilai MAPE 7,61%.



Gambar 8. Plot Hasil *Prewhitening* dan Korelasi Silang Antara Deret Input dengan Output

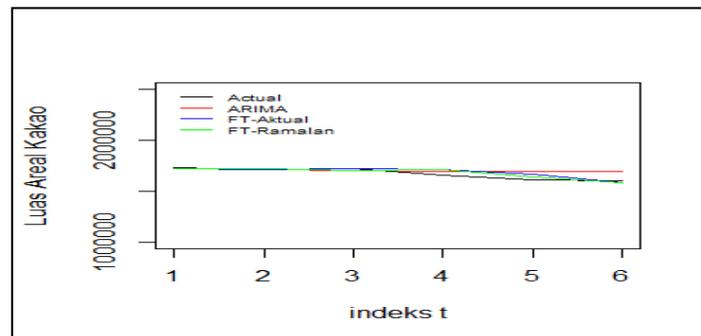
Langkah keenam melakukan serangkaian estimasi dan membandingkan nilai MAPE data *training* yang terdiri dari:

- d. Model fungsi transfer dengan data input nilai aktual.
- e. Model fungsi transfer dengan data input nilai estimasi.
- f. Model ARIMA output yang telah lebih dahulu dibahas di awal.

Nilai MAPE untuk model fungsi transfer yang menggunakan data input *testing* nilai aktual dan model fungsi transfer dengan data input nilai estimasi ARIMA (1,1,0) memiliki nilai MAPE yang relatif sama yakni 1,36% dan 1,43% (Tabel 12). Sedangkan dari plot hasil estimasi dengan fungsi transfer baik data input ramalan maupun data input aktual sangat mengikuti pola data aktual dibandingkan model ARIMA (Gambar 9).

Tabel 12. Nilai MAPE Data Training Model Fungsi Transfer Luas areal Kakao

No	Model Estimasi	MAPE (%)	
		Training	Testing
1	FT ARIMA (2,1,5) xreg=produksi aktual	7,61	1,36
2	FT ARIMA (2,1,5) xreg=produksi ARIMA (1,1,0)	7,61	1,43
3	ARIMA (1,1,0)	7,81	2,47



Gambar 9. Nilai MAPE Hasil Estimasi Model Fungsi Transfer

Langkah ketujuh yang merupakan langkah terakhir adalah menduga ulang model input produksi dengan model ARIMA (1,1,0) dan menduga ulang fungsi transfer ARIMA (2,1,5) untuk melakukan estimasi luas areal kakao 5 tahun kedepan yang menunjukkan pertumbuhan menurun sekitar 0,65% per tahun secara rata-rata (Tabel 13).

Tabel 13. Hasil Estimasi Luas areal Kakao Model Fungsi Transfer Tahun 2020-2024

Tahun	Luas areal Kakao (Ton)
2020	1.581.724
2021	1.568.544
2022	1.550.355
2023	1.541.735
2024	1.541.035
Rata-rata Pertumbuhan (%)	-0,65

#### IV. VAR

Pada model VAR, variabel yang dimasukkan yakni luas areal, produksi, harga kakao dunia, volume ekspor dan volume impor. Tahap awal dalam penentuan model VAR adalah melakukan penelusuran model dari *lag* atau  $p=1$  sampai dengan  $p=4$  dengan dan tanpa tren (*type both and const*) dengan hasil model terpilih yakni VAR (1) tipe=*const* dengan pertimbangan karena merupakan model VAR dengan variabel yang signifikan paling banyak.

Tabel 14. Model VAR Luas areal Kakao Terpilih

p=1, type=const								
Endogenous Variable	Luas Areal	Produksi	Harga Dunia	Voleks	Volimp	const	R <sup>2</sup>	R <sup>2</sup> Adjusted
Luas Areal	***						99%	99%
Produksi		**	*		0	**	98%	96%
Harga Dunia			***		*		81%	78%
Voleks		*	*	*		**	95%	94%
Volimp					***		95%	94%

12 variabel yg signifikan pada alpha = 10%

Tahap selanjutnya melakukan serangkaian pengujian terhadap model yakni normalitas sebaran, autokorelasi dan keragaman. Dari hasil Uji Chi-squared, Uji Jarque-Bera dan ARCH dapat disimpulkan asumsi normalitas dan non autokorelasi terpenuhi serta ragam homogen (Tabel 15).

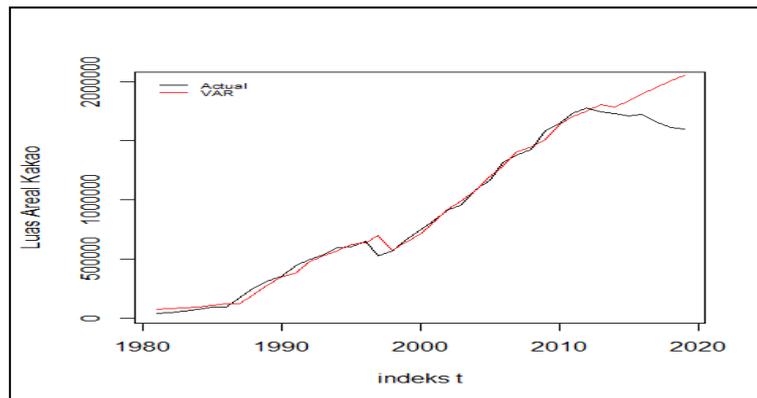
Tabel 15. Hasil Uji Asumsi Model VAR Luas areal Kakao

Portmanteau Test (asymptotic)	
data: Residuals of VAR object varhsheet1	Chi-squared = 341.62, df = 375, p-value = 0.8911
JB-Test (multivariate)	
data: Residuals of VAR object varhsheet1	Chi-squared = 48.054, df = 10, p-value = 6.066e-07
Skewness only (multivariate)	
data: Residuals of VAR object varhsheet1	Chi-squared = 19.021, df = 5, p-value = 0.001905
Kurtosis only (multivariate)	
data: Residuals of VAR object varhsheet1	Chi-squared = 29.033, df = 5, p-value = 2.284e-05
ARCH (multivariate)	
data: Residuals of VAR object varhsheet1	Chi-squared = 420, df = 1125, p-value = 1

Tahap berikutnya dalam model VAR adalah menghitung nilai MAPE dari data *training* dan data *testing* seperti yang tampak di Tabel 16. Untuk MAPE data *training* lebih besar 15,15% dibandingkan MAPE data *testing* yakni 12,78%. Pada Gambar 10 terlihat hasil plot estimasi data *training* dan data *testing* terhadap data aktual, dimana data *training* lebih mengikuti pola data aktual dibandingkan hasil estimasi data *testing*.

Tabel 16. Nilai MAPE Data Training dan Data Testing Model VAR Luas areal Kakao

Data	MAPE (%)
Training	15,15
Testing	12,78



Gambar 10. Plot Data Ramalan Model VAR(1) *type=const* Terhadap Data Aktual Luas areal Kakao Tahun 1980-2019

Tahap akhir dari serangkaian tahapan pada proses pemodelan dengan metode VAR berupa estimasi luas areal kakao untuk periode tahun 2020-2024 yang menduga akan terjadi kenaikan luas areal kakao secara kontinyu pada 5 tahun mendatang dengan rata-rata bertumbuh sebesar 3,02% per tahun seperti tampak pada Tabel 17 berikut.

Tabel 17. Hasil Estimasi Luas areal Kakao Model VAR (1) *type=const* Tahun 2020-2024

Tahun	Luas areal Kakao (Ton)
2020	1.785.255
2021	1.840.454
2022	1.898.126
2023	1.955.385
2024	2.010.733
Rata-rata Pertumbuhan (%)	3,02

### Model Estimasi Terbaik

Hasil analisa dari 4 model estimasi yang digunakan pada kajian ini dapat diringkas sebagai berikut:

Tabel 18. Ringkasan Hasil Analisa Model Estimasi Luas areal Kakao

No	Model	MAPE (%)		Hasil Estimasi Luas Areal Kakao (Ha)					Rata-rata Pertumbuhan (%)
		Training	Testing	2020	2021	2022	2023	2024	
1	ARIMA (1,1,0)	7,81	2,47	1.594.494	1.590.842	1.588.675	1.587.388	1.586.625	-0,12
2	Regresi Linier	48,82	11,07	2.038.996	2.094.491	2.149.985	2.205.480	2.260.975	2,62
3	Regresi Kuadratik	-	32,38	2.574.853	2.686.753	2.801.004	2.917.606	3.036.557	4,21
4	FT ARIMA (2,1,5) <i>xreg=produksi</i>	7,61	1,43	1.581.724	1.568.544	1.550.355	1.541.735	1.541.035	-0,65
5	VAR (1) <i>type=const</i>	12,78	15,15	1.785.255	1.840.454	1.898.126	1.955.385	2.010.733	3,02

Berdasarkan Uji MAPE dari Tabel 18 maka model terbaik dan terpilih untuk estimasi luas areal kakao adalah model FT ARIMA (2,1,5)  $x_{reg} = \text{produksi}$  karena memiliki nilai MAPE terkecil untuk data *training* 7,61% maupun data *testing* 1,43% dibandingkan model-model lainnya. Pertimbangan lain karena hasil estimasinya dianggap sangat mendekati data aktualnya. Luas areal kakao tahun 2020 diestimasi sebesar 1.581.724 ha, turun 1,19% dibandingkan tahun 2019 yang mencapai 1.600.647 ha. Hasil estimasi luas areal kakao yang dilakukan oleh Pusdatin mendekati estimasi Ditjen Perkebunan sebesar 1.582.406 ha atau hanya selisih 0,04%. Luas areal kakao menurun kembali sebesar 0,83% menjadi 1.568.544 ha di tahun 2021. Penurunan luas areal kakao terjadi sampai tahun 2024 sehingga rata-rata pertumbuhan luas areal kakao 5 tahun kedepan sebesar -0,65% (Tabel 19).

Tabel 19. Hasil Estimasi Model Fungsi Transfer Luas areal Kakao Tahun 2020-2024

Tahun	Produksi (Ton)		Pertumbuhan (%)
	Ditjen Perkebunan	Pusdatin	
2019*	1.600.647		
2020**	1.582.406	1.581.724	
2021		1.568.544	-0,83
2022		1.550.355	-1,16
2023		1.541.735	-0,56
2024		1.541.035	-0,05
Rata-rata Pertumbuhan (%)			-0,65

## KESIMPULAN

Dari keempat metode estimasi yang digunakan dalam kajian ini yaitu ARIMA, Regresi, Fungsi Transfer dan VAR, metode estimasi terbaik untuk estimasi luas areal kakao berdasarkan pertimbangan statistik dan kerealistisan hasil estimasi dengan historis data aktualnya adalah Model Fungsi Transfer.

## DAFTAR PUSTAKA

- Enders, Walter. 2010. *Applied Econometric Time Series*. USA: University of Alabama. Wiley, Third Edition.
- Gujarati, D.N. and D.C. Porter, 2010. *Dasar-dasar Ekonometrika*. Jakarta: Penerbit Salemba Empat, Buku 2, Edisi 5.
- Kementerian Pertanian. 2019. *Statistik Perkebunan Indonesia 2018-2020*. Jakarta: Direktorat Jenderal Perkebunan Kementerian Pertanian.
- Rohmah, Yuliawati. 2019. *Outlook Komoditas Perkebunan Kakao*. Jakarta: Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian.



**KAJIAN MODEL PERAMALAN LUAS AREAL KOPI DI INDONESIA:  
PENDEKATAN MODEL ARIMA, REGRESI, FUNGSI TRANSFER, DAN VAR**

Diah Mekita Sari

*Subdirektorat Statistik Tanaman Perkebunan-Badan Pusat Statistik  
Jalan Dr. Sutomo no. 6-8 Gd 4 Lt 4, Pasar Baru-Jakarta Pusat, Indonesia  
E-mail: mekita.sari@bps.go.id*

**ABSTRAK**

Kopi merupakan salah satu komoditas perkebunan yang memegang peranan cukup penting dalam perekonomian Indonesia. Kontribusi penting dari komoditas kopi bagi perekonomian nasional tercermin pada kinerja perdagangan dan peningkatan nilai tambahnya. Saat ini rilis resmi data luas dan produksi kopi oleh Badan Pusat Statistik (BPS) adalah data tahunan dengan *lag*  $n-1$  tahun. Guna merumuskan kebijakan produksi, luas areal, pasokan dan distribusi komoditas kopi sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Pada Tahun 2020, Pusdatin bersama BPS dan Dirjenbun mengkaji empat metode yaitu ARIMA, regresi, fungsi transfer, dan VAR (*Vector Auto Regression*) dengan menggunakan *software* RStudio guna melakukan pemodelan data kopi di Indonesia. Metode yang dipilih guna melakukan peramalan data luas areal kopi adalah metode yang terbaik dikaji dari nilai MAPE (*Mean Absolute Percentage Error*) terkecil dan kerealistikan hasil peramalan. Berdasarkan dari hasil perbandingan MAPE disimpulkan model terbaik adalah model Fungsi Transfer dengan peubah input volume ekspor kopi yang memiliki MAPE terkecil yaitu 0,55%.

**ABSTRACT**

*Coffee is a commodity that plays an important role in the Indonesian economy. The important contribution of the coffee commodity to the national economy is reflected in trade performance and increased value added. Currently, the official release of coffee production and area data by the BPS is annual data with a lag of  $n-1$  year. In order to formulate a production policy, the area size, supply and distribution of coffee, the availability of up-to-date data, and even forecasts for the next few periods, is urgently required. In 2020, Pusdatin together with BPS and Dirjenbun will examine four methods, namely ARIMA, regression, transfer function, and VAR (*Vector Auto Regression*) using RStudio software to model coffee data in Indonesia. The method chosen to forecast coffee area data is the best method studied from the smallest MAPE (*Mean Absolute Percentage Error*) value and the reliability of the forecasting results. Based on the results of the MAPE comparison, it can be concluded that the best model is the Transfer Function model with the coffee export volume as input variable having the smallest MAPE, namely 0.55%.*

## PENDAHULUAN

Tanaman kopi termasuk dalam golongan famili *Rubiaceae* yang mempunyai 500 macam genus dan lebih dari 6000 spesies. Tanaman kopi biasanya tumbuh berupa semak atau pohon kecil yang dapat mencapai 5 meter ketika tidak berbuah. Daunnya berwarna hijau gelap dan mengkilat, biasanya panjangnya 10-15 cm dan mempunyai lebar 6 cm. Kopi tumbuh di daerah tropis dan tumbuhan peralihan yang tumbuh di lereng gunung.

Kopi merupakan salah satu komoditas perkebunan yang memegang peranan cukup penting dalam perekonomian Indonesia. Kontribusi penting dari komoditas kopi bagi perekonomian nasional tercermin pada kinerja perdagangan dan peningkatan nilai tambahnya. Sebagai produk ekspor, komoditas kopi dapat memberikan kontribusi berupa penghasil devisa dan pendapatan negara, sumber pendapatan petani, penciptaan lapangan kerja, pendorong pertumbuhan sektor agribisnis dan agroindustri, pengembangan wilayah serta pelestarian lingkungan. Selain peluang ekspor yang semakin terbuka, pasar kopi di dalam negeri masih cukup besar.

Saat ini rilis resmi data produksi kopi oleh Badan Pusat Statistik (BPS) adalah data tahunan dengan *lag*  $n-1$  tahun. Guna merumuskan kebijakan produksi, pasokan dan distribusi komoditas kopi sangat dituntut ketersediaan data terkini, bahkan ramalan beberapa periode ke depan. Metode ramalan produksi kopi dapat digunakan dengan menggunakan beberapa metode statistik yang telah dikembangkan oleh para ahli.

Pada makalah ini akan dikaji metode ARIMA, Regresi, Fungsi Transfer dan VAR dalam melakukan pemodelan dan peramalan produksi dan luas areal kopi Indonesia.

Oleh karenanya, tujuan dari disusunnya makalah ini adalah:

- a. Melakukan analisis dan peramalan data produksi dan luas areal kopi menggunakan model ARIMA, Regresi, Fungsi Transfer dan VAR di Indonesia
- b. Membandingkan empat metode pada poin a dalam memperoleh ramalan data produksi dan luas areal komoditas kopi.
- c. Menentukan metode terbaik dalam meramal data produksi dan luas areal komoditas kopi di Indonesia

## BAHAN DAN METODE

Data yang digunakan dalam melakukan pemodelan produksi dan luas areal kopi Indonesia adalah data series tahunan tahun 1980 sampai 2019. Peubah yang diasumsikan mempengaruhi besaran produksi dan luas areal kopi untuk Fungsi Transfer adalah volume impor kopi dan untuk pemodelan VAR adalah produksi, luas areal, volume ekspor, volume impor, harga kopi dunia, dan harga pupuk urea dunia. Peramalan data produksi kopi, dilakukan uji coba beberapa metode yakni metode ARIMA, Regresi, Fungsi Transfer, dan VAR (*Vector Autoregression*), menggunakan *software* Rstudio.

Metode yang digunakan dalam melakukan peramalan data produksi dan luas areal kopi di Indonesia adalah sbb.

a. ARIMA (*Autoregressive Integrated Moving Average*)

ARIMA merupakan model dari fungsi linier nilai lampau beserta nilai sekarang dan sisaan lampainya. Bentuk modelnya adalah

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t$$

dimana:

$Y_t$  = data *time series* sebagai variable dependen pada waktu ke-t

$Y_{t-p}$  = data *time series* pada kurun waktu ke (t-P)

$\mu$  = suatu konstanta

$\theta_1 \theta_2 \dots \theta_p$  = parameter-parameter model

$\varepsilon_{t-q}$  = nilai sisaan pada waktu ke-(t-q)

b. Regresi

Regresi adalah suatu teknik analisis statistik yang digunakan untuk menyelidiki hubungan atau pengaruh antara dua atau lebih variabel.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

dengan

$y$  = variabel respon/dependen/terikat

$x_1, x_2, \dots, x_k$  = variabel prediktor/independen/bebas

$\beta_0$  = konstanta/intercept

$\beta_1, \beta_2, \dots, \beta_k$  = koefisien regresi variabel bebas

$\varepsilon$  = error/residu/sisaan

c. Fungsi Transfer

Model fungsi transfer adalah suatu model yang menggambarkan nilai dari prediksi masa depan dari suatu deret berkala (disebut deret output atau  $Y_t$ ) didasarkan pada nilai-nilai masa lalu dari deret itu sendiri ( $Y_t$ ) dan didasarkan pula pada satu atau lebih deret berkala yang berhubungan (disebut deret input atau  $X_t$ ) dengan deret output tersebut.

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t$$

Dimana:

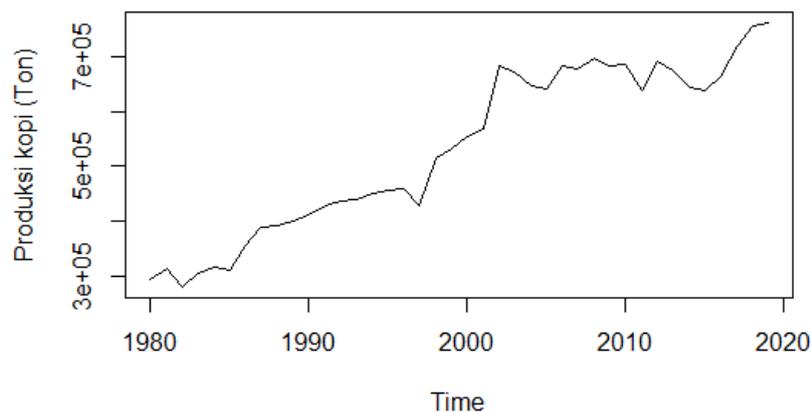
- $b \rightarrow$  panjang jeda pengaruh  $X_t$  terhadap  $Y_t$
- $r \rightarrow$  panjang *lag*  $Y$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $s \rightarrow$  panjang jeda  $X$  periode sebelumnya yang masih mempengaruhi  $Y_t$
- $p \rightarrow$  ordo AR bagi noise  $N_t$

d. VAR (*Vector*)

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen., karena pada kenyataannya suatu variabel dapat bertindak sebagai endogen atau dependen dari suatu variabel yang lainnya (Gujarati, 2010).

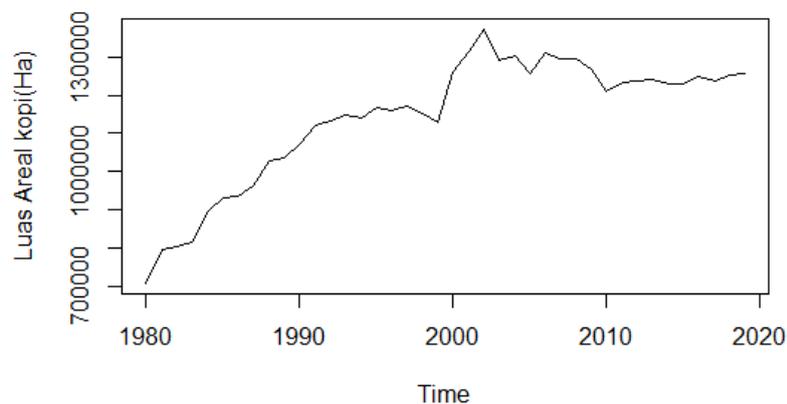
## HASIL DAN PEMBAHASAN

### a. Eksplorasi Data



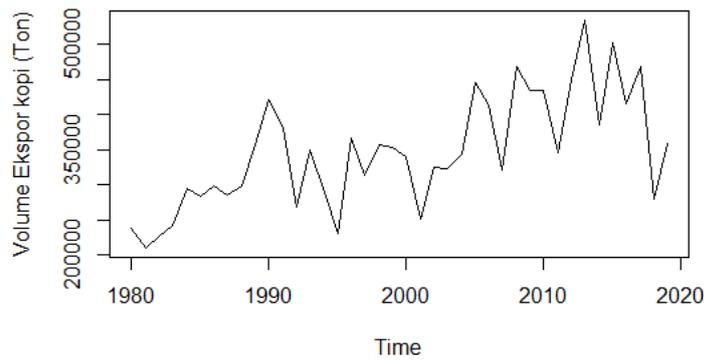
**Gambar 1.** Perkembangan Produksi Kopi Indonesia tahun 1980-2019

Produksi tahunan kopi di Indonesia periode 1980-2019 berfluktuasi setiap tahunnya disebabkan berbagai macam faktor namun cenderung meningkat dengan rata-rata pertumbuhan sebesar 2,64% (Gambar 1)



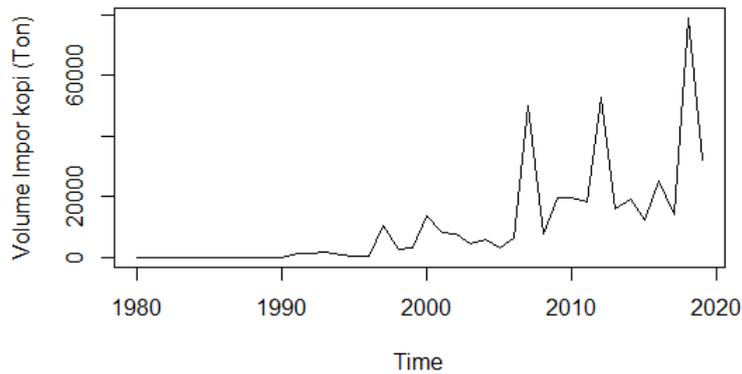
**Gambar 2.** Perkembangan Luas Areal Kopi Indonesia tahun 1980-2019

Luas areal kopi cukup berfluktuasi dengan kecenderungan meningkat hingga tahun 2002 dan menurun setelahnya (Gambar 2). Berkurangnya luas areal kopi disebabkan oleh banyak faktor, di antaranya alih fungsi lahan atau penggantian jenis tanaman.



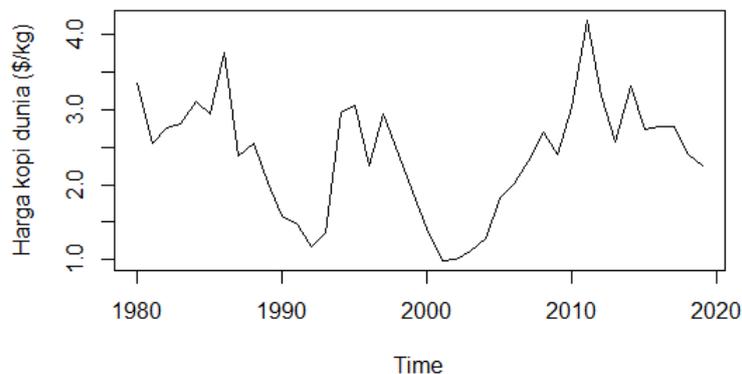
**Gambar 3.** Perkembangan Volume Ekspor Kopi Indonesia tahun 1980-2019

Volume Ekspor Kopi Indonesia periode 1980-2019 sangat berfluktuasi setiap tahunnya (Gambar 3). Bahkan sulit untuk memprediksi arah kecenderungan volume ekspor kopi disebabkan oleh fluktuasi yang sangat besar.



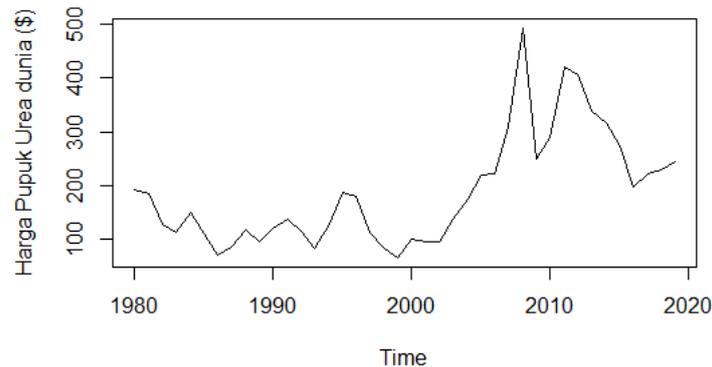
**Gambar 4.** Perkembangan Volume Impor Kopi Indonesia tahun 1980-2019

Volume Impor Kopi tahunan Indonesia periode Januari 1980 – Agustus 2019 berfluktuasi setiap tahunnya dengan jarak kenaikan dan penurunan yang cukup besar pada beberapa periode (Gambar 4). Volume impor kopi sangat berfluktuasi di antaranya disebabkan oleh nilai tukar rupiah dan harga kopi dalam negeri.



**Gambar 5.** Perkembangan Harga Kopi Dunia tahun 1980-2019

Harga kopi dunia secara tahunan periode 1980-2019 sangat berfluktuasi setiap tahunnya disebabkan berbagai macam faktor (Gambar 5). Harga kopi dunia diperkirakan turut memengaruhi minat petani dalam menanam dan merawat tanaman kopinya.



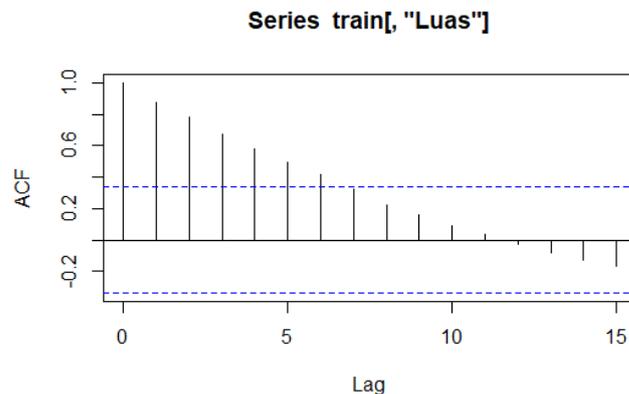
**Gambar 6.** Perkembangan Harga Pupuk Urea Dunia tahun 1980-2019

Harga pupuk urea dunia secara tahunan periode 1980-2019 terlihat berfluktuasi setiap tahunnyadan mencapai angka tertinggi pada tahun 2008 (Gambar 6). Pada penelitian Thamrin(2014), disebutkan bahwa produksi kopi salah satunya dipengaruhi oleh penggunaan pupuk urea.

#### **b. ARIMA Peramalan Luas Areal**

Dalam melakukan pemodelan luas areal kopi menggunakan model *Autoregressive Integrated Moving Average* (ARIMA), data yang digunakan adalah periode tahun 1980 hingga 2019. Periode data tersebut kemudian dipisahkan menjadi data set *training* dan *testing*. Panjang series data pada data set *training* adalah tahun 1980 hingga 2013, sementara dataset *testing* adalah periode tahun 2014-2019. Dataset *training* digunakan untuk melakukan penyusunan model, sementara dataset *testing* digunakan untuk validasi model.

Uji kestasioneran data seperti yang disyaratkan apabila melakukan pemodelan ARIMA dilakukan secara visual menggunakan hasil plot data maupun uji formal statistik. Gambar 7 menunjukkan plot *Auto Correlation Function* (ACF) dari data *training*. Tampak bahwa plot ACF data *training* memiliki pola tertentu sehingga diperkirakan data yang digunakan tidak stasioner. Hal ini diperkuat oleh hasil uji formal statistik yaitu dengan uji *Augmented Dickey-Fuller* yang mengindikasikan bahwa data luas areal kopi adalah tidak stasioner, terlihat dari nilai statistik uji (-1,5055) yang lebih besar dari nilai kritis taraf 5% yakni -3,50 (Tabel 1).



**Gambar 7.** Plot *Auto Correlation Function* (ACF) Data *Training* Luas areal Kopi

**Tabel 1.** Hasil Uji *Augmented Dickey-Fuller* Luas Areal Kopi

---

```

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

                Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.344e+05 6.494e+04 2.069 0.0462 *
z.lag.1     -1.056e-01 7.017e-02 -1.505 0.1414
tt          2.591e+01 1.015e+03 0.026 0.9798
z.diff.lag  -9.477e-02 1.615e-01 -0.587 0.5611

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36550 on 34 degrees of freedom
Multiple R-squared: 0.1646, Adjusted R-squared: 0.09084
F-statistic: 2.232 on 3 and 34 DF, p-value: 0.1023

Value of test-statistic is: -1.5055 3.4529 3.3478

Critical values for test statistics:
      1pct 5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61

```

---

Untuk mengatasi masalah ketidak-stasioneran, dilihat dari pola grafiknya maka metode yang dilakukan adalah melakukan *differencing*. Hasil pengujian *Augmented Dickey-Fuller test* pada Tabel 2, diperoleh hasil bahwa data *differencing* luas areal kopi telah stasioner. Hal tersebut diketahui dari nilai nilai statistik uji (-3,3606) yang lebih besar dari nilai kritis taraf 5% yakni -1,95.

**Tabel 2.** Hasil Uji *Augmented Dickey-Fuller Difference* Luas Areal Kopi

---

```

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

                Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1     -0.7209    0.2145 -3.361 0.00189 **
z.diff.lag  -0.1952    0.1556 -1.255 0.21786

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40250 on 35 degrees of freedom
Multiple R-squared: 0.4708, Adjusted R-squared: 0.4405
F-statistic: 15.57 on 2 and 35 DF, p-value: 1.458e-05

Value of test-statistic is: -3.3606

Critical values for test statistics:
      1pct 5pct 10pct
tau1 -2.62 -1.95 -1.61

```

---

Setelah mendapatkan data luas areal kopi yang stasioner, maka langkah selanjutnya adalah menentukan ordo AR dan MA menggunakan fungsi auto arima yang terdapat pada RStudio. Dari hasil pendugaan auto arima pada Tabel 3, maka model ARIMA yang terpilih adalah ARIMA (0,2,1) yang berarti model ARIMA yang optimal untuk mengestimasi data yang diolah adalah ARIMA dengan ordo *Auto Regressive* (AR) = 0, ordo *differencing* = 2, dan ordo *Moving Average* (MA) = 1.

**Tabel 3.** Hasil Fungsi Auto ARIMA Luas Areal Kopi

---

```

Series: train[, "Luas"]
      ARIMA(0,2,1)

Coefficients:
      ma1
      -0.8787
      s.e.   0.0854

sigma^2 estimated as 1.924e+09:  log likelihood=-387.68
      AIC=779.37   AICC=779.78   BIC=782.3

      Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -11228.82 41888.41 31071.14 -1.07353 2.744212 0.9274112 -0.1222689

```

---

Namun model ARIMA yang dipilih oleh Fungsi Autoarima dirasa kurang tepat karena pada model terpilih, ordo *differencing* = 2 sementara data luas areal kopi menurut perhitungan uji ADF telah stasioner pada ordo *differencing* = 1. Oleh karena itu, dilakukan *overfitting* menggunakan Fungsi Armaselect untuk mencoba ordo lain agar dapat menemukan model ARIMA terbaik yang dapat digunakan untuk melakukan peramalan luas areal kopi Indonesia.

**Tabel 4.** Hasil Pencarian Ordo ARIMA Optimum menggunakan Fungsi Armaselect (*diff*)

---

	p	q	sb	bc
[1,]	5	5	666.	3740
[2,]	4	5	694.	2403
[3,]	3	5	697.	0058
[4,]	0	5	698.	3290
[5,]	1	5	699.	2379
[6,]	2	5	701.	9494
[7,]	0	0	703.	8389
[8,]	1	0	705.	1157
[9,]	2	0	709.	2199
[10,]	4	0	711.	2626

---

Dari hasil pencarian ordo ARIMA optimum menggunakan fungsi armaselect (*diff*), maka diperoleh model ARIMA terbaik adalah ARIMA(5,1,5). Namun karena model ARIMA dengan ordo AR maupun MA = 5 dinilai terlalu kompleks, maka opsi selanjutnya adalah model ARIMA(0,1,0), ARIMA(1,1,0) dan ARIMA(2,1,0). Setelah dilakukan pengolahan data menggunakan ARIMA(0,1,0) dan ARIMA(1,1,0) didapatkan hasil bahwa nilai ramalan yang terbentuk bernilai konstan setelah 2 periode, sehingga model ARIMA yang diuji selanjutnya adalah ARIMA(2,1,0).

**Tabel 5.** Hasil pengujian ARIMA(2,1,0) untuk Luas Areal Kopi

---

```

Series: serieskopi[, "Luas"]
      ARIMA(2,1,0)

Coefficients:
      ar1      ar2
      0.0895  0.2091
      s.e.   0.1631  0.1620

sigma^2 estimated as 1.737e+09:  log likelihood=-469.24
      AIC=944.48   AICC=945.16   BIC=949.47

      Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 9671.395 40088.7 26704.68 0.966594 2.390681 0.8963879 -0.051304
Test set    -20735.126 22969.3 20735.13 -1.6754524 1.675452 0.6960096 NA

```

---

Hasil uji statistik terhadap model ARIMA(2,1,0) menunjukkan bahwa nilai MAPE data *training* sebesar 2,39% dan untuk data *testing* sebesar 1,68%. Nilai tersebut dinilai telah cukup baik untuk melakukan peramalan dan hasil uji Ljung-Box menunjukkan bahwa tidak terdapat autokorelasi pada residual model (Tabel 6).

**Tabel 6.** Hasil Uji Ljung-Box test ARIMA(2,1,0) untuk Luas Areal Kopi

	lags	statistic	df	<i>p-value</i>
	5	3.084407	5	0.6869744
	10	9.025686	10	0.5296677
	15	11.439987	15	0.7208268
	20	15.321439	20	0.7577288
	25	17.763369	25	0.8522693
	30	20.944625	30	0.8896114

Setelah diperoleh model ARIMA yang optimal, maka model ARIMA tersebut digunakan untuk melakukan peramalan luas areal kopi 5 tahun ke depan (2020-2024). Hasil peramalan luas areal kopi menggunakan ARIMA(2,1,0) ditampilkan pada Tabel 7 berikut.

**Tabel 7.** Hasil Peramalan Luas areal Kopi Indonesia menggunakan ARIMA(2,1,0)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	1261473	1208055	1314891	1179777	1343169
2021	1262870	1183874	1341865	1142057	1383683
2022	1263714	1158303	1369126	1102501	1424927
2023	1264082	1136523	1391641	1068997	1459166
2024	1264291	1116601	1411982	1038418	1490164

### c. Regresi untuk Peramalan Luas Areal

Peubah bebas yang digunakan dalam melakukan pemodelan data luas areal kopi di Indonesia adalah tahun. Data yang digunakan adalah periode tahun 1980 hingga 2019. Periode data tersebut kemudian dipisahkan menjadi data set *training* dan *testing*. Panjang series data pada data set *training* adalah tahun 1980 hingga 2013, sementara dataset *testing* adalah periode tahun 2014-2019. Dataset *training* digunakan untuk melakukan penyusunan model, sementara dataset *testing* digunakan untuk validasi model.

Dari hasil pendugaan parameter (Tabel 8) menunjukkan model cukup bagus, terlihat dari nilai *p-value*  $2,653 \times 10^{-12}$  atau  $H_0$  ditolak, artinya koefisien signifikan (berbeda dari nol) model yang jauh lebih kecil dari 0,05 dan variabel tahun signifikan secara uji t dengan *p-value*  $2,653 \times 10^{-12}$ . Nilai  $R^2$  model bernilai 78,78% yang artinya peubah tahun dapat menjelaskan 78,78% keragaman dari luas areal kopi di Indonesia.

**Tabel 8.** Penyusunan Model Regresi Linier Luas Areal Kopi Indonesia dengan Peubah Bebas Tahun

```

Call:
lm(formula = Luas ~ Tahun, data = train)

Coefficients:
(Intercept) -30025039      2858144 -10.51 6.70e-12 ***
Tahun          15601         1432  10.90 2.65e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81890 on 32 degrees of freedom
Multiple R-squared:  0.7878,    Adjusted R-squared:  0.7811
F-statistic: 118.8 on 1 and 32 DF,  p-value: 2.653e-12

```

Model luas areal kopi Indonesia diformulasikan sebagai berikut:

$$y_t = -30025039 + 15601T$$

Evaluasi nilai kesalahan model regresi yang terbentuk dilakukan dengan menghitung nilai MAPE atau nilai rata-rata persentase kesalahan peramalan. Nilai MAPE untuk dataset *training* dan *testing* masing-masing sebesar 6,08% dan 15,46% seperti tersaji pada Tabel 9.

**Tabel 9.** Nilai MAPE Model Regresi Linier Luas Areal Kopi Indonesia

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<b>MAPE Data Training</b>	0.0229	2.9839	4.9281	<b>6.081</b>	9.3082	22.3422
<b>MAPE Data Testing</b>	13.45	14.56	15.59	<b>15.46</b>	16.47	17.17

Setelah diperoleh model regresi yang akan digunakan, maka perlu dilakukan pengujian asumsi regresi linier, yakni normalitas, non-autokorelasi, dan homoskedas. Hasil uji asumsi normalitas menggunakan *Jarque Bera test* menunjukkan bahwa asumsi normalitas terpenuhi karena  $p\text{-value} > 0,05$  (Tabel 10). Uji asumsi selanjutnya yakni non-autokorelasi menggunakan *Durbin Watson test* juga menunjukkan bahwa asumsi terpenuhi, karena  $p\text{-value} < 0,05$  (Tabel 11) dan asumsi terakhir yakni homoskedas tidak terpenuhi karena uji NCV test menunjukkan bahwa  $p\text{-value} > 0,05$  (Tabel 12).

**Tabel 10.** Hasil Uji Asumsi Normalitas Regresi Linier Luas Areal Kopi menggunakan *Jarque Bera Test*

Jarque-Bera test for normality	
data:	reg1\$residuals
JB =	1.2213, $p\text{-value} = 0.3805$

**Tabel 11.** Hasil Uji Asumsi Non-Autokorelasi Regresi Linier Luas Areal Kopi menggunakan *Durbin Watson Test*

lag	Autocorrelation	D-w Statistic	$p\text{-value}$
1	0.76042	0.2731582	0
Alternative hypothesis: rho != 0			

**Tabel 12.** Hasil Uji Asumsi Normalitas Homoskedas Regresi Linier Luas Areal Kopi menggunakan *NCV test*

Non-constant Variance Score Test	
Variance formula:	~ fitted.values
Chisquare =	0.5677072, Df = 1, $p = 0.45117$

Setelah diperoleh model regresi linier yang optimal, maka model regresi tersebut digunakan untuk melakukan peramalan produksi kopi 5 tahun ke depan (2020-2024). Hasil peramalan produksi kopi Indonesia menggunakan regresi linier dengan peubah bebas yaitu tahun ditampilkan pada Tabel 13 berikut.

**Tabel 13.** Hasil Peramalan Luas Areal Kopi menggunakan Regresi Linier

2020	2021	2022	2023	2024
<b>1.489.579</b>	<b>1.505.180</b>	<b>1.520.782</b>	<b>1.536.383</b>	<b>1.551.984</b>

Selain melakukan peramalan menggunakan model regresi linier, peramalan model regresi juga dilakukan menggunakan model regresi kuadratik. Regresi kuadratik dilakukan menggunakan peubah bebas yakni tahun dan tahun kuadrat. Hasil pendugaan parameter

(Tabel 33) menunjukkan model cukup bagus, terlihat dari nilai  $p$ -value  $2,2 \times 10^{-16}$  atau  $H_0$  ditolak, artinya koefisien signifikan (berbeda dari nol) model yang jauh lebih kecil dari 0,05 dan variabel tahun dan tahun kuadrat signifikan secara uji t dengan  $p$ -value  $2 \times 10^{-16}$  dan  $1,04 \times 10^{-11}$ . Nilai  $R^2$  pada model regresi kuadrat sebesar 95,02% yang artinya peubah tahun dan tahun kuadrat dapat menjelaskan 95,02% keragaman dari luas areal kopi di Indonesia.

**Tabel 14.** Penyusunan Model Regresi Linier Luas Areal Kopi Indonesia dengan Peubah Bebas Tahun dan Tahun Kuadrat

---

```

Call:
lm(formula = Luas ~ t + t2, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-108113  -17744   1782    20215  109574

Coefficients:
(Intercept) 721915.40  18964.40  38.07 < 2e-16 ***
t            42528.49  2659.55  15.99 < 2e-16 ***
t2           -815.98    77.89  -10.48 1.04e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 39050 on 31 degrees of freedom
Multiple R-squared:  0.9532, Adjusted R-squared:  0.9502
F-statistic: 316 on 2 and 31 DF, p-value: < 2.2e-16

```

---

Model luas areal kopi Indonesia diformulasikan sebagai berikut:

$$y_i = 721915,40 + 42528,49T - 815,98T^2$$

Evaluasi nilai kesalahan model regresi kuadrat yang terbentuk dilakukan dengan menghitung nilai MAPE atau nilai rata-rata persentase kesalahan peramalan. Nilai MAPE untuk dataset *training* dan *testing* masing-masing sebesar 6,08% dan 4,64% seperti tersaji pada Tabel 34.

**Tabel 15.** Nilai MAPE Model Regresi Kuadrat Luas Areal Kopi Indonesia

---

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
MAPE Data <i>Training</i>	0.0229	2.9839	4.9281	<b>6.0815</b>	9.3082	22.3422
MAPE Data <i>Testing</i>	0.4778	2.1832	4.4844	<b>4.6411</b>	6.7882	9.4278

---

Setelah diperoleh model regresi kuadrat yang akan digunakan, maka perlu dilakukan pengujian asumsi regresi, yakni normalitas, non-autokorelasi, dan homoskedas. Hasil uji asumsi normalitas menggunakan *Jarque Bera test* menunjukkan bahwa asumsi normalitas tidak terpenuhi karena  $p$ -value  $< 0,05$  (Tabel 35). Uji asumsi selanjutnya yakni non-autokorelasi menggunakan Durbin Watson test juga menunjukkan bahwa asumsi terpenuhi, karena  $p$ -value  $< 0,05$  (Tabel 36) dan asumsi terakhir yakni homoskedas juga terpenuhi karena uji NCV test menunjukkan bahwa  $p$ -value  $< 0,05$  (Tabel 37).

**Tabel 16.** Hasil Uji Asumsi Normalitas Regresi Kuadrat Luas Areal K

opi menggunakan *Jarque Bera Test*

---

```

Jarque-Bera test for normality

data: reg2$residuals
JB = 6.2093, p-value = 0.034

```

---

**Tabel 17.** Hasil Uji Asumsi Non-Autokorelasi Regresi Kuadratik Luas Areal Kopi menggunakan *Durbin Watson Test*

<i>lag</i>	Autocorrelation	D-W Statistic	<i>p-value</i>
1	0.4673259	1.060411	0.002
Alternative hypothesis: rho != 0			

**Tabel 18.** Hasil Uji Asumsi Normalitas Homoskedas Regresi Kuadratik Luas Areal Kopi menggunakan *NCV test*

Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 4.999498, Df = 1, p = 0.025355

Setelah diperoleh model regresi kuadratik yang optimal, maka model regresi tersebut digunakan untuk melakukan peramalan luas areal kopi 5 tahun ke depan (2020-2024). Hasil peramalan luas areal kopi Indonesia menggunakan regresi kuadratik dengan peubah bebas yaitu tahun dan tahun kuadrat ditampilkan pada Tabel 19 berikut.

**Tabel 19.** Hasil Peramalan Luas Areal Kopi menggunakan Regresi Kuadratik

2020	2021	2022	2023	2024
1.117.494	1.093.928	1.068.731	1.041.902	1.013.440

**k. Fungsi Transfer untuk Peramalan Luas Areal Kopi**

Peubah yang diasumsikan dapat mempengaruhi gejala produksi kopi adalah volume ekspor kopi, dan peubah ini akan digunakan untuk melakukan analisis menggunakan model Fungsi Transfer. Dari pola data volume ekspor dan uji *Augmented Dickey-Fuller* (Tabel 20) terlihat data volume ekspor belum stasioner pada taraf uji 5%. Oleh karena itu, series data volume ekspor dilakukan *differencing* terlebih dahulu sebelum melakukan pemodelan (Tabel 21).

**Tabel 20.** Hasil Uji Augmented Dickey-Fuller Series Data Volume Ekspor Kopi

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.203e+05 6.655e+04 3.310 0.00257 **
z.lag.1     -8.798e-01 2.576e-01 -3.415 0.00197 **
tt          4.803e+03 1.673e+03 2.871 0.00772 **
z.diff.lag  9.786e-02 1.997e-01 0.490 0.62788

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56130 on 28 degrees of freedom
Multiple R-squared: 0.3776, Adjusted R-squared: 0.3109
F-statistic: 5.663 on 3 and 28 DF, p-value: 0.003668

Value of test-statistic is: -3.4148 4.4517 5.9084

Critical values for test statistics:
      1pct 5pct 10pct
tau3  -4.15 -3.50 -3.18
phi2   7.02  5.13  4.31
phi3   9.31  6.73  5.61
```

**Tabel 21.** Hasil Uji *Augmented Dickey-Fuller* Series Data Differencing Volume Ekspor Kopi

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression none

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1      -1.8592     0.3163  -5.879 2.22e-06 ***
z.diff.lag    0.3844     0.1920   2.002  0.0548 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 62820 on 29 degrees of freedom
Multiple R-squared:  0.6922,    Adjusted R-squared:  0.671
F-statistic: 32.61 on 2 and 29 DF,  p-value: 3.794e-08

Value of test-statistic is: -5.8786

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Pencarian ordo model ARIMA input untuk volume ekspor dilakukan dengan fungsi auto arima. Dari hasil fungsi auto arima, diperoleh ordo ARIMA terbaik untuk volume ekspor adalah ARIMA(0,1,1) sebagaimana ditampilkan pada Tabel 22.

**Tabel 22.** Hasil Fungsi Auto ARIMA Volume Ekspor Kopi

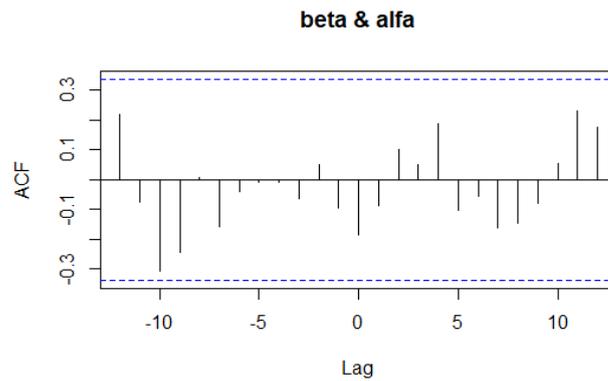
```
Series: train.h[, "Ekspor"]
ARIMA(0,1,1) with drift

Coefficients:
      ma1      drift
-0.7944  6587.680
s.e.    0.2290  2637.829

sigma^2 estimated as 3.294e+09:  log likelihood=-407.89
AIC=821.79  AICC=822.61  BIC=826.28

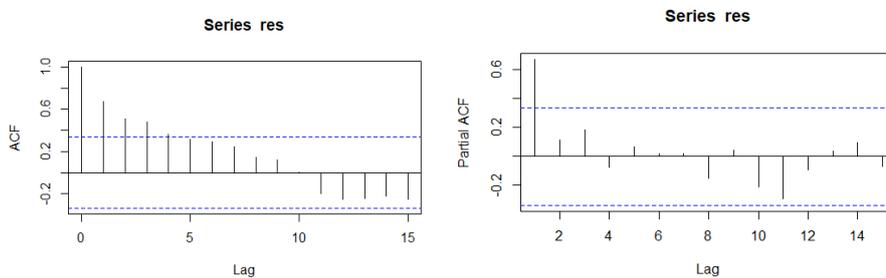
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 804.4137 54799.44 40991.07 -2.139735 12.43589 0.7689033
0.1059043
```

Dari hasil pencarian ordo arima optimum volume ekspor kopi menggunakan fungsi auto arima, diputuskan model arima input volume ekspor kopi adalah ARIMA (0,1,1). Tahap selanjutnya adalah melakukan *prewhitening* dan analisis korelasi silang antara residual model arima ekspor kopi dan luas areal kopi menggunakan ARIMA(0,1,1). Dari hasil plot ccf (*cross correlation function*) antara alfa (residual ekspor kopi) dan beta (residual luas kopi), diperoleh hasil bahwa tidak terdapat *lag* dari pengaruh yang diberikan oleh ekspor kopi terhadap luas areal kopi atau *lag* = 0. Hal tersebut dapat disimpulkan dari tidak adanya *bar* atau garis yang melewati ambang batas pada nilai *lag* positif (Gambar 8).

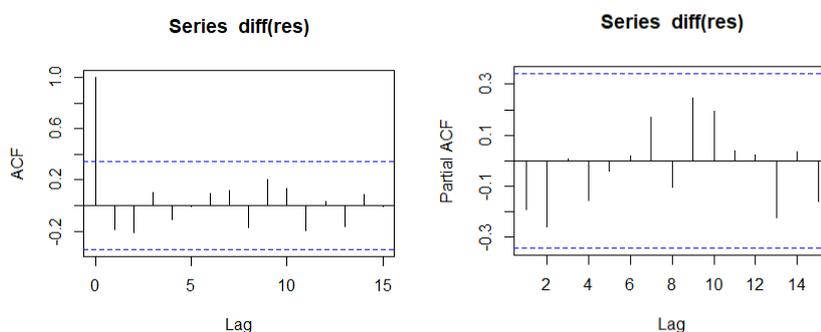


**Gambar 8.** Plot ACF Alfa (Residual Ekspor Kopi) dan Beta (Residual Luas Areal Kopi)

Karena tidak adanya *bar* atau garis yang melewati ambang batas pada nilai *lag* positif, maka model ARIMA untuk pengepasan model fungsi transfer adalah ARIMA(0,0,0). Kemudian residual model ARIMA(0,0,0) tersebut diteliti stasioneritasnya. Dari plot ACF dan PCF residual, diperoleh hasil bahwa residual model tidak stasioner (Gambar 9) sehingga perlu dilakukan *differencing*. Plot ACF dan PACF residual yang telah dilakukan *differencing* dapat dilihat pada Gambar 10.



**Gambar 9.** Plot ACF dan PACF Residual model ARIMA(0,0,0)



**Gambar 10.** Plot ACF dan PACF Residual Model ARIMA(0,0,0) setelah *Differencing*

Setelah data residual stasioner, maka dilakukan pencarian model ARIMA terbaik untuk residual tersebut atau dapat disebut dengan *noise*. Hasil dari fungsi auto arima menunjukkan bahwa model ARIMA(0,1,0) merupakan model terbaik (Tabel 23). Namun karena ordo ARIMA(0,1,0) dinilai kurang baik untuk peramalan karena akan menghasilkan peramalan

yang konstan, maka diperlukan referensi ordo ARIMA lain menggunakan fungsi *arma select* sebagaimana tampak pada Tabel 24.

**Tabel 23.** Hasil Fungsi Auto ARIMA Noise

---

```

Series: res
ARIMA(0,1,0)

sigma^2 estimated as 1.031e+10: log likelihood=-427.25
AIC=856.5 AICC=856.63 BIC=857.99

Training set error measures:
ME RMSE MAE MPE MAPE MASE ACF1
Training set 4329.337 100014.7 82533.53 11.44701 1192.781 0.9706869 -
0.1909202

```

---

**Tabel 24.** Hasil Pencarian Ordo ARIMA Optimum Noise menggunakan Fungsi Armaselect (diff)

---

```

p q sbc
[1,] 0 0 761.7994
[2,] 1 0 763.5312
[3,] 2 0 765.1619
[4,] 3 0 769.7195
[5,] 0 1 771.6555
[6,] 0 2 772.3672
[7,] 4 0 773.1920
[8,] 1 2 774.3536
[9,] 1 1 775.0690
[10,] 2 2 775.9740

```

---

Setelah dilakukan uji signifikansi model ARIMA dan ARIMA-xreg pada pilihan ordo arima di atas, diperoleh ordo ARIMA noise yang signifikan adalah ARIMA(1,1,2) sebagaimana tampak pada Tabel 25 berikut.

**Tabel 25.** Hasil Uji Signifikansi ARIMA-xreg (1,1,2)

---

```

z test of coefficients:

Estimate Std. Error z value Pr(>|z|)
ar1 -0.754003 0.146275 -5.1547 2.540e-07 ***
ma1 1.149843 0.276449 4.1593 3.192e-05 ***
ma2 0.639461 0.275959 2.3172 0.02049 *
xreg -0.111435 0.075255 -1.4808 0.13867
---
```

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

---

Oleh karena telah didapatkan model input serta noise yang terbaik untuk peramalan, maka langkah selanjutnya adalah melakukan peramalan. Peramalan dilakukan menggunakan dua jenis data input, yakni nilai aktual dan data ramalan. Dari hasil peramalan menggunakan nilai aktual sebagai input, diperoleh nilai MAPE sebesar 0,55% (Tabel 26) sementara peramalan menggunakan data ramalan sebagai input memiliki MAPE sebesar 0,61% (Tabel 27).

**Tabel 26.** Pengukuran Tingkat Kesalahan Model Fungsi Transfer dengan Nilai Aktual sebagai Input

---

```

Series: test.h[, "Luas"]
Regression with ARIMA(1,1,2) errors

```

---

---

AIC=108.56	AICC=109.89	BIC=108.17					
<i>Training set error measures:</i>							
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	
Training set	3688.406	8447.412	6796.583	0.295796	<b>0.5455912</b>	0.754388	-
			0.1151604				

---

**Tabel 27.** Pengukuran Tingkat Kesalahan Model Fungsi Transfer dengan Data Ramalan sebagai Input

---

Series: test.h[, "Luas"]							
Regression with ARIMA(1,1,2) errors							
AIC=109.32		AICC=110.65		BIC=108.93			
<i>Training set error measures:</i>							
ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	
Training set	4190.422	9112.714	7588.456	0.3336436	<b>0.6081011</b>	0.842282	-
				0.6025041			

---

Setelah diperoleh model fungsi transfer yang optimal, maka model tersebut digunakan untuk melakukan peramalan luas areal kopi 5 tahun ke depan (2020-2024). Hasil peramalan luas areal kopi Indonesia menggunakan fungsi transfer dengan volume ekspor kopi sebagai peubah ditampilkan pada Tabel 28 berikut.

**Tabel 28.** Hasil Peramalan Luas Areal Kopi menggunakan Fungsi Transfer

---

2020	2021	2022	2023	2024
<b>1.267.674</b>	<b>1.271.336</b>	<b>1.268.722</b>	<b>1.270.663</b>	<b>1.269.222</b>

---

### **I. Model Vector Auto Regression (VAR) untuk Peramalan Luas Areal Kopi**

Pemodelan dengan menggunakan model VAR dilakukan dengan memasukkan peubah luas areal kopi, produksi kopi, volume ekspor kopi, volume impor kopi, harga kopi dunia, dan harga pupuk urea dunia. Hal ini diasumsikan bahwa apabila terjadi gejolak pada variabel-variabel tersebut, akan mempengaruhi gejolak produksi kopi ataupun sebaliknya.

Pada tahap pertama, perlu dilakukan pemeriksaan apakah terdapat pengaruh trend dalam model. Hasil pengujian signifikansi trend dilakukan pada  $lag=1$  dan  $lag=2$ . Hasil pengujian trend pada  $lag=1$  menunjukkan bahwa *trend* berpengaruh signifikan pada persamaan produksi kopi (signifikan 5%) dan persamaan harga pupuk urea dunia (signifikan 10%) sementara hasil pengujian *trend* pada  $lag=2$  menunjukkan bahwa *trend* berpengaruh signifikan hanya pada persamaan harga pupuk urea dunia (signifikan 5%). Dari hasil *running* model menggunakan  $lag$  1 dan 2, diperoleh informasi komponen *trend* tidak terlalu berpengaruh signifikan sehingga *trend* dikeluarkan dalam model untuk *running* model VAR (p) selanjutnya.

Langkah selanjutnya adalah menentukan  $lag$  (p) yang akan digunakan dalam model VAR. Penentuan  $lag$  (p) dilakukan dengan melakukan pengujian VAR(1) hingga VAR(4) dengan pertimbangan kecukupan series data. Dari hasil *running* menggunakan  $p=1$  s.d  $p=4$ , diperoleh informasi bahwa variabel yang signifikan cukup banyak pada  $p=2$ , sehingga model VAR yang digunakan selanjutnya adalah VAR (2). Hal tersebut diperkuat juga dengan nilai *Adjusted R-squared*: 0,9495 atau dapat dikatakan bahwa 94.95% keragaman luas areal kopi

dapat dijelaskan oleh kelima variabel/peubah penjelas termasuk konstanta. Hasil pengujian model VAR(2) dapat dilihat dalam Tabel 29.

**Tabel 29.** Hasil Estimasi VAR Luas Areal Kopi dengan Peubah Produksi, Volume Ekspor, Volume Impor, Harga Kopi Dunia, dan Harga Pupuk Urea Dunia pada Lag=2 tanpa *Trend*

---

VAR Estimation Results:						
=====						
Endogenous variables: Produksi, Luas, Ekspor, Impor, world.Price, Urea.Price						
Deterministic variables: const						
Sample size: 38						
Log Likelihood: -1945.481						
Roots of the characteristic polynomial:						
1.003	0.8959	0.7582	0.7582	0.7209	0.7209	0.5579 0.5579 0.4862 0.4862 0.4756 0.4524
Call:						
VAR(y = kopi[, c(2, 3, 4, 5, 6, 7)], p = 2, type = "const")						
Estimation results for equation Luas:						
=====						
Luas = Produksi.l1 + Luas.l1 + Ekspor.l1 + Impor.l1 + world.Price.l1 + Urea.Price.l1 +						
Produksi.l2 + Luas.l2 + Ekspor.l2 + Impor.l2 + World.Price.l2 + Urea.Price.l2 + const						
	Estimate	Std. Error	t value	Pr(> t )		
Produksi.l1	-2.849e-01	2.249e-01	-1.267	0.21691		
Luas.l1	6.835e-01	1.732e-01	3.946	0.00057 ***		
Ekspor.l1	1.366e-01	1.178e-01	1.159	0.25742		
Impor.l1	2.545e-01	5.542e-01	0.459	0.65011		
World.Price.l1	-1.596e+04	1.227e+04	-1.300	0.20542		
Urea.Price.l1	6.099e+01	1.293e+02	0.472	0.64129		
Produksi.l2	3.576e-01	2.384e-01	1.500	0.14618		
Luas.l2	1.506e-01	1.674e-01	0.900	0.37687		
Ekspor.l2	1.509e-01	1.145e-01	1.317	0.19971		
Impor.l2	2.576e-03	7.535e-01	0.003	0.99730		
World.Price.l2	2.444e+04	1.274e+04	1.918	0.06663 .		
Urea.Price.l2	-3.967e+02	1.075e+02	-3.689	0.00110 **		
const	1.086e+05	8.629e+04	1.258	0.21988		
---						
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						
Residual standard error: 32260 on 25 degrees of freedom						
Multiple R-Squared: 0.9659, <b>Adjusted R-squared: 0.9495</b>						
F-statistic: 58.99 on 12 and 25 DF, <i>p-value</i> : 2.85e-15						
Covariance matrix of residuals:						
	Produksi	Luas	Ekspor	Impor	World.Price	Urea.Price
Produksi	979994490	317789702	411294269	1.296e+08	-5.205e+03	-334182.9
Luas	317789702	1040816640	-156407446	2.345e+07	-3.192e+03	3268.2
Ekspor	411294269	-156407446	3268060856	-3.409e+08	-1.227e+04	-246755.0
Impor	129603587	23454205	-340925128	1.698e+08	3.305e+02	-21601.1
World.Price	-5205	-3192	-12274	3.305e+02	3.187e-01	15.6
Urea.Price	-334183	3268	-246755	-2.160e+04	1.560e+01	3041.6
Correlation matrix of residuals:						
	Produksi	Luas	Ekspor	Impor	World.Price	Urea.Price
Produksi	1.0000	0.314660	0.22982	0.31773	-0.29448	-0.193562
Luas	0.3147	1.000000	-0.08481	0.05579	-0.17525	0.001837
Ekspor	0.2298	-0.084806	1.00000	-0.45768	-0.38030	-0.078265
Impor	0.3177	0.055793	-0.45768	1.00000	0.04493	-0.030059
World.Price	-0.2945	-0.175247	-0.38030	0.04493	1.00000	0.501086
Urea.Price	-0.1936	0.001837	-0.07827	-0.03006	0.50109	1.000000

---

Setelah terpilih *lag* model VAR yang optimal, maka perlu dilakukan pengujian asumsi non-autokorelasi, normalitas, dan homoskedas pada sisaan model VAR (2). Dari hasil pengujian non autokorelasi, diperoleh informasi bahwa model tersebut terbebas dari autokorelasi dengan tingkat kesalahan 5% (Tabel 30). Residual memenuhi asumsi kenormalan dengan tingkat kesalahan 5% melalui *normality test* (*Jarque Bera Test*, pengujian skewness, dan kurtosis) sebagaimana tampak pada Tabel 31, sedangkan asumsi homoskedastisitas atau kehomogenan ragam signifikan terpenuhi dengan *p-value* =1 (Tabel 32).

**Tabel 30.** Hasil Uji Asumsi Non-Autokorelasi model VAR(2) menggunakan *Serial Test*

---

Portmanteau Test (asymptotic)
data: Residuals of VAR object varhsheet2 Chi-squared = 430.43, df = 504, <b><i>p-value</i> = 0.9922</b>

---

**Tabel 31.** Hasil Uji Asumsi Normalitas Regresi Linier Produksi Kopi menggunakan *Normality Test*

---

\$JB JB-Test (multivariate) data: Residuals of VAR object varhsheet2 Chi-squared = 12.244, df = 12, <b><i>p-value</i> = 0.4263</b>
\$Skewness Skewness only (multivariate) data: Residuals of VAR object varhsheet2 Chi-squared = 9.3156, df = 6, <b><i>p-value</i> = 0.1566</b>
\$Kurtosis Kurtosis only (multivariate) data: Residuals of VAR object varhsheet2 Chi-squared = 2.928, df = 6, <b><i>p-value</i> = 0.8178</b>

---

**Tabel 32.** Hasil Uji Asumsi Homoskedas Regresi Linier Produksi Kopi menggunakan *ARCH Test*

---

ARCH (multivariate)
data: Residuals of VAR object varhsheet2 Chi-squared = 693, df = 2205, <b><i>p-value</i> = 1</b>

---

Sebelum melakukan peramalan, perlu dilakukan evaluasi nilai kesalahan model VAR yang terbentuk dengan menghitung nilai MAPE atau nilai rata-rata persentase kesalahan peramalan. Nilai MAPE VAR(2) untuk dataset *training* dan *testing* masing-masing sebesar 4,29% dan 1,41% seperti tersaji pada Tabel 33.

**Tabel 33.** Nilai MAPE Model VAR(2) Luas Areal Kopi Indonesia

---

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<b>MAPE Data <i>Training</i></b>	0.02455	1.40016	2.52837	<b>4.29215</b>	6.44759	15.44307
<b>MAPE Data <i>Testing</i></b>	0.1376	0.9089	1.4483	<b>1.4017</b>	2.0703	2.3690

---

Setelah diperoleh model VAR yang optimal yakni VAR(2), maka model tersebut digunakan untuk melakukan peramalan luas areal kopi 5 tahun ke depan (2020-2024). Hasil peramalan luas areal kopi Indonesia menggunakan metode VAR dengan peubah produksi kopi, volume ekspor kopi, volume impor kopi, harga kopi dunia, dan harga pupuk urea dunia ditampilkan pada Tabel 34 berikut.

**Tabel 34.** Hasil Peramalan Luas Areal Kopi menggunakan VAR(2)

---

2020	2021	2022	2023	2024
<b>1.329.924</b>	<b>1.416.176</b>	<b>1.514.353</b>	<b>1.500.741</b>	<b>1.442.753</b>

---

## KESIMPULAN DAN SARAN

### Kesimpulan

- Hasil analisis yang didapat dari empat model adalah sebagai berikut:
  - Model ARIMA terbaik adalah ARIMA (2,1,0) dengan MAPE 1,68%
  - Model regresi linier terbaik menghasilkan MAPE 15,46% dengan nilai probabilitas  $< 0,05$  dan  $R^2$  sebesar 78,11%.
  - Model regresi kuadratik terbaik memiliki MAPE sebesar 4,64% dengan nilai probabilitas  $< 0,05$  dan  $R^2$  sebesar 95,02%.
  - Model Fungsi Transfer terbaik dengan peubah input volume ekspor kopi adalah ARIMA (0,0,0)(1,1,2) dengan model input ARIMA (0,1,1). Model Fungsi Transfer tersebut memiliki MAPE sebesar 0,55%.
  - Model VAR terbaik dengan memasukkan peubah produksi kopi, volume ekspor kopi, volume impor kopi, harga kopi dunia, dan harga urea dunia adalah VAR(2) tanpa *trend* (hanya konstanta). MAPE yang dihasilkan adalah 1,41%.
- Dari keempat model yang dicobakan, model terbaik adalah model Fungsi Transfer dengan melihat MAPE terkecil yaitu 0,55%
- Hasil ramalan luas areal kopi model fungsi transfer dengan peubah input volume ekspor kopi dengan ARIMA (0,0,0)(1,1,2) dan model input ARIMA (0,1,1) untuk 2020 sampai 2024 adalah 1.267.674 Ha, 1.271.336 Ha, 1.268.722 Ha, 1.270.663 Ha, dan 1.269.222 Ha.

### Saran

- Perlu dilakukan kajian mendalam dengan metode peramalan lainnya.
- Untuk model regresi, perlu diujicobakan menggunakan variabel bebas/peubah input selain waktu.
- Untuk model fungsi transfer dan VAR perlu diujicobakan dengan menggunakan variabel lain yang sekiranya lebih berpengaruh, baik secara teoritis maupun praktis.
- Perlu dikaji metode peramalan untuk data non parametrik, sehingga tidak terhalang oleh tidak terpenuhinya berbagai asumsi.

## DAFTAR PUSTAKA

- Draper, N. R, dan Smith, H. 1992. Analisis Regresi Terapan, Jakarta : PT Gramedia Pustaka Utama
- Direktorat Jenderal Perkebunan. 2019. Statistik Perkebunan Indonesia 2018-2020 Jakarta: Direktorat Jenderal Perkebunan, Kementerian Pertanian.Statistik Perkebunan Indonesia 2018-2020.
- Gujarati, Damodar. N dan Porter, Dawn. C. 2009. Basic Econometrics. Boston: Douglas Reiner.
- Thamrin, Syahrini. 2014. Faktor-faktor yang Mempengaruhi Produksi Usahatani Kopi Arabika di Kabupaten Enrekang Sulawesi Selatan. Pangkajene Kepulauan: Politeknik Pertanian Negeri Pangkajene Kepulauan
- Malian, Husni & Sudi Mardianto dan Mewa Ariani. 2004. Faktor-faktor yang Mempengaruhi Produksi, Konsumsi dan Harga Beras serta Inflasi Bahan Makanan. Jurnal Agro Ekonomi, Volume 22 No. 2: 119-146
- Siagian, Viktor dan Muchamad Yusron. 2015. Analisis Faktor-Faktor yang Mempengaruhi Luas Lahan Garapan Usahatani Padi Sawah di Provinsi Banten. Lampung: Politeknik Negeri Lampung

# KAJIAN METODE ESTIMASI LUAS AREAL TEH INDONESIA

Diah Indarti  
Fungsional Statistisi, Pusdatin-Kementan

## ABSTRAK

Komoditas teh memiliki peranan penting dalam perekonomian nasional yaitu sebagai sumber pendapatan petani, penyerapan tenaga kerja, sumber devisa Negara, mendorong agroindustri pengembangan wilayah dan pelestarian lingkungan. Pada umumnya tanaman teh nasional dikembangkan di Indonesia sejak jaman Belanda. Khusus untuk teh rakyat mulai dikembangkan sekitar Tahun 1980-an sehingga kondisi tanaman pada umumnya merupakan tanaman tua/rusak dengan produktivitas yang sudah menurun dan sudah saatnya dilakukan perbaikan budidaya melalui rehabilitasi dan intensifikasi tanaman (Kementerian Pertanian, 2013).

Pada Tahun 2020 ini Pusat Data dan Sistem Informasi Pertanian bersama dengan Direktorat Jenderal Perkebunan dan Badan Pusat Statistik mengkaji beberapa metode peramalan untuk menghasilkan angka estimasi yang nilainya cenderung mendekati angka aktualnya. Tujuan dari makalah ini melakukan pemodelan, membandingkan efektivitas dari pemodelan tersebut dan menentukan metode terbaik dalam meramalkan dengan ARIMA, Regresi, Fungsi Transfer dan VAR pada luas areal teh di Indonesia.

**Kata Kunci :** ARIMA, regresi, fungsi transfer, VAR, estimasi luas areal

## PENDAHULUAN

### Latar Belakang

Estimasi/peramalan adalah ilmu yang mempelajari tentang prediksi peristiwa masa depan dengan pengambilan data historis dan memproyeksikannya ke masa depan dengan menggunakan model matematis (Heizer et al.,2011). Estimasi tidak akan menggambarkan kejadian atau peristiwa yang sebenarnya, tetapi dalam model matematis peramalan dapat dibuat sedemikian sehingga galat (error) peramalan (kesalahan peramalan) yang diperoleh adalah sekecil-kecilnya. Galat biasanya diukur dalam *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Root Mean Square Absolute Percentage Error* (RMSAPE), dan lain-lain.

Metode analisis yang digunakan untuk mencari pengaruh data historis terhadap data yang akan datang agar memperoleh hasil estimasi adalah analisis deret waktu. Data yang digunakan dalam analisis ini harus memiliki interval yang sama, misalkan dalam harian, bulanan, atau tahunan. Analisis deret waktu dapat digunakan untuk menganalisis data yang terdiri satu variabel (*univariate*) dan data yang terdiri lebih dari satu variabel (*multivariate*). Kelemahan dari analisis deret waktu univariat adalah tidak diperhi-tungkannya pengaruh variabel-variabel lain di luar model yang mungkin saja memiliki pengaruh signifikan terhadap model karena pada analisis deret waktu univariat hanya dianalisis satu variabel terhadap data

historisnya. Sedangkan pada analisis deret waktu multivariat dapat dianalisis hubungan dinamis antar variabel dan meramalkan data secara simultan (Tsay, 2014).

Pada tahun 2020 Pusdatin membahas dan mengkaji beberapa metode peramalan guna menghasilkan angka yang nilainya semakin mendekati nilai sebenarnya (aktual). Pemodelan yang diterapkan adalah pemodelan ARIMA, regresi, fungsi transfer dan *Vector Auto Regression* (VAR) dengan menggunakan *software Program R Studio*.

Makalah ini mengulas hasil penerapan metode tersebut untuk meramal angka luas areal teh khususnya di Indonesia serta membahas kelebihan dan kekurangan masing-masing metode yang diterapkan.

## **Tujuan**

Tujuan dari makalah ini adalah:

- a. Melakukan analisis dan pemodelan ARIMA, Regresi, Fungsi Transfer dan VAR untuk meramalkan luas areal teh di Indonesia.
- b. Membandingkan efektivitas metode tersebut dalam hasil peramalan luas areal teh.
- c. Menentukan metode terbaik dalam mengestimasi data luas areal komoditas teh di Indonesia.

## **METODOLOGI**

### **Sumber Data**

Data yang digunakan dalam kajian ini adalah data sekunder yang diperoleh dari BPS dan Direktorat Jenderal Perkebunan berupa data tahunan dari 1980 sampai dengan 2019. Datanya terdiri dari data luas areal teh, luas areal teh, harga konsumen teh, produksi dan volume impor (Badan Pusat Statistik) di Indonesia. Data tahun 1980 sampai dengan 2013 digunakan untuk membangun model sedangkan data tahun 2014 sampai dengan 2019 digunakan untuk validasi model.

### **Metode**

Tahapan yang dilakukan dalam kajian ini adalah:

1. Melakukan eksplorasi data untuk mengetahui pola data.
2. Bagi data menjadi dua bagian, untuk penyusunan model (*training*) dan untuk validasi model (*testing*).
3. Susun model dengan berbekal data training.
4. Lakukan peramalan dengan masing-masing metode sesuai dengan periode data *training* dan *testing*.
5. Memilih model terbaik dari keempat metode yang diterapkan. Semakin kecil nilai MAPE dan pola model mendekati garis aktual, menunjukkan data hasil peramalan semakin mendekati nilai aktual. Semakin kecil nilai AIC maka model semakin tepat.
6. Melakukan pendugaan ulang dengan model terbaik pada seluruh data (gabungan data *training* dan *testing*).
7. Menggunakan model terbaik untuk peramalan sesuai periode yang diinginkan. Dalam hal ini hingga tahun 2024.

### a. Metode ARIMA

Langkah-langkahnya adalah sebagai berikut:

- 1) Memeriksa kestasioneran data luas areal teh. Data dikatakan stasioner jika berfluktuasi di sekitar rata-rata untuk seluruh periode waktu. Hal ini dapat dilakukan dengan melihat plot data, plot ACF dan PACF serta uji *Augmented Dickey Fuller*.
- 2) Identifikasi ordo AR dan MA pada deret yang stasioner dari hasil amatan plot ACF dan PACF. Model AR ( $p$ ) dicirikan dengan pola *tail off* pada plot ACF dan pola *cut off* setelah lag ke- $p$  pada plot PACF. Model MA ( $q$ ) dicirikan dengan pola *cut off* setelah lag ke- $q$  pada plot ACF dan pola *tail off* pada plot PACF. Apabila ACF dan PACF keduanya berpola *tail off*, maka dapat dikatakan model yang digunakan adalah ARMA ( $p,q$ ). Bila terdapat proses differencing karena proses penstasioneran, maka model yang terbentuk adalah ARIMA ( $p,d,q$ ) dengan  $d$  merupakan banyaknya proses *differencing* dari data yang tidak stasioner.
- 3) Pendugaan model tentatif. Dari data yang sudah stasioner dilakukan penentuan besarnya ordo  $p$  dan  $q$ . Nilai ordo  $p$  ditentukan dari banyaknya koefisien pada plot PACF pertama yang berbeda nyata. Sedangkan nilai ordo  $q$  ditentukan dari banyaknya koefisien pada plot ACF pertama yang berbeda nyata. Perlu diperhatikan juga perilaku musiman ( $S$ ). Sehingga model yang diperoleh dapat ditulis dengan ARIMA ( $p,d,q$ )( $P,D,Q$ ) <sup>$S$</sup> . Penentuan komponen  $P,D,Q$  musiman sama halnya dengan penentuan  $p,d,q$  non musiman hanya diterapkan pada lag berjeda  $S$ .
- 4) *Overfitting*, merupakan penambahan komponen model untuk evaluasi bila ada komponen sistematis yang terlewat dan tercakup dalam sisaan dengan tujuan untuk memperoleh model terbaik. Bila penambahan komponen di model-model *overfitting* tidak nyata, maka kembali ke model awal.
- 5) Peramalan dengan menggunakan model ARIMA terbaik. Perhitungan dilakukan secara rekursif, yaitu menghitung peramalan satu periode kemudian dua periode, dan seterusnya sampai  $n$  periode ke depan.
- 6) Setelah nilai ramalan diperoleh, langkah berikutnya adalah evaluasi hasil ramalan tersebut pada dua hal: a) *magnitude* (nilai data) dan b) *direction* (pergerakan data). Bila pergerakan hasil model baik untuk *training* maupun *testing* dapat mengikuti pergerakan data aktual dengan baik, nilai data MAPE (*Mean Absolute Percentage Error*) yang diperoleh pun cukup kecil, maka model tersebut layak digunakan untuk peramalan.

### b. Metode Regresi

Langkah-langkahnya adalah sebagai berikut:

- 1) Penentuan parameter untuk pembentukan model.
- 2) Pendugaan model.
- 3) Pengujian hipotesis dengan Analisis Varians.
- 4) Peramalan dengan model yang terbentuk.

- 7) Evaluasi secara *magnitude* (MAPE) dan *direction* (Korelasi). Bila hasil keduanya bagus, maka model tersebut layak digunakan untuk peramalan.

**Estimasi dengan Model Regresi Linear Sederhana dan Regresi Kuadrat**

Model regresi linear sederhana dan regresi kuadrat menggunakan peubah prediktor waktu (*t*) sebagai variabel independen. Model regresi linear sederhana untuk mengestimasi luas areal teh dituliskan sebagai berikut:

$$y = \beta_0 + \beta_1 t + \varepsilon \dots\dots\dots(1)$$

dimana:

- y* = luas areal teh
- t* = tahun
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi variabel tahun
- $\varepsilon$  = error

Parameter  $\beta_1$  mengukur efek tahun terhadap rata-rata perubahan luas areal teh. Parameter ini diprediksi dari data sampel.  $\varepsilon$  menggambarkan kesalahan statistik, didefinisikan sebagai sisa atau perbedaan hasil antara nilai data pengamatan variabel dependen terhadap nilai hasil prediksi.

Model regresi kuadrat yang digunakan untuk meramalkan luas areal teh dituliskan sebagai berikut:

$$y = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon \dots\dots\dots(2)$$

dimana:

- y* = luas areal teh
- t* = tahun
- t*<sup>2</sup> = kuadrat dari tahun
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi variabel tahun
- $\beta_2$  = koefisien regresi variabel kuadrat tahun
- $\varepsilon$  = error

Dampak perubahan tahun terhadap rata-rata perubahan luas areal teh ditunjukkan oleh parameter  $\beta_1$ . Parameter  $\beta_2$  menunjukkan efek perubahan kuadrat tahun terhadap rata-rata perubahan luas areal teh.

Estimasi dengan model regresi linear sederhana dan regresi kuadrat berdasarkan pada beberapa asumsi antara lain:

**- Sisaan mengikuti fungsi distribusi normal**

Uji normalitas dilakukan dengan *Jarque-Bera (JB) test for normality*. Hipotesis yang mendasari yaitu:

Hipotesis: .....(3)

$H_0$  : Sisaan terdistribusi normal

$H_1$  : *Otherwise*

Jika nilai p-value hasil JB test lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi normalitas terpenuhi.

**- Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas)**

Asumsi homoskedastisitas diuji menggunakan *Non-constant Variance (NCV) Score Test* dengan hipotesis sebagai berikut:

Hipotesis: .....(4)

$H_0$  : Homoskedastisitas

$H_1$  : Heterokedastisitas

Jika nilai p-value pada NCV test lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi homoskedastisitas terpenuhi.

**- Tidak terdapat autokorelasi antar sisaan untuk setiap data pengamatan**

Ada tidaknya autokorelasi antar sisaan pada data amatan diuji berdasarkan nilai autokorelasi *D-W Statistic* dengan hipotesis sebagai berikut:

Hipotesis: .....(5)

$H_0$  : Tidak ada autokorelasi antar sisaan

$H_1$  : *Otherwise*

Jika nilai p-value *autocorrelation D-W Statistic* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi tidak adanya autokorelasi antar sisaan telah terpenuhi

**c. Metode Fungsi Transfer**

Model fungsi transfer menggambarkan ramalan luas areal teh yang didasarkan pada nilai-nilai masa lalu dari luas areal itu sendiri, serta didasarkan pada produksi teh

(variabel input) dan gangguan/noise. Model fungsi transfer untuk mengestimasi luas areal teh dituliskan sebagai berikut:

$$y_t = v(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \quad \dots(6)$$

dimana:

$y_t$  = luas areal teh tahun ke-t

$x_t$  = produksi teh tahun ke-t

$b$  = panjang jeda pengaruh produksi teh terhadap luas areal teh

$r$  = panjang lag luas areal teh periode sebelumnya yang masih mempengaruhi luas areal teh tahun-t

$s$  = panjang jeda produksi teh periode sebelumnya yang masih mempengaruhi luas areal teh tahun-t

$p$  = ordo AR bagi noise  $N_t$

$q$  = ordo MA bagi noise  $N_t$

Langkah-langkah dalam melakukan peramalan dengan menggunakan model Fungsi Transfer adalah sebagai berikut:

- 1) Memeriksa kestasioneran deret input luas tanaman menghasilkan dan deret output luas areal teh.
- 2) Membuat model ARIMA untuk deret input.
- 3) Melakukan *prewhitening* deret input.
- 4) Melakukan *prewhitening* deret output berdasarkan hasil *prewhitening* deret input.
- 5) Menghitung korelasi silang deret input dengan deret output untuk melihat hubungannya.
- 6) Menetapkan nilai  $b$ ,  $r$  dan  $s$  dengan melihat plot korelasi silang deret input dan output dengan ketentuan sebagai berikut:
  - a) Nilai  $b$  ditentukan berdasarkan *lag* positif yang pertama kali nyata pada plot korelasi silang.
  - b) Nilai  $s$  dilihat dari lamanya deret input mempengaruhi deret output setelah nyata yang pertama (nilai  $b$ ).
  - c) Nilai  $r$  menunjukkan lamanya deret output berhubungan dengan nilai terdahulu dari deret output itu sendiri. Nilai  $r$  dilihat dari plot *ACF* deret output atau ditentukan berdasarkan pola *lag* ( $b+s$ ), jika memiliki pola eksponensial maka  $r = 1$  dan memiliki pola gelombang sinus maka  $r = 2$ .
- 7) Membuat model ARIMA deret sisaan dengan mengamati plot *ACF* dan *PACF* deret sisaan.
- 8) Pendugaan model fungsi transfer.
- 9) Diagnosa model jika terdapat parameter yang tidak nyata, maka kembali ke langkah 6.
- 10) Peramalan dengan model yang diperoleh.

**d. Metode VAR**

Langkah-langkahnya adalah sebagai berikut:

- 1) Menguji kestasioneran data.
- 2) Menentukan panjang lag optimal. Jumlah lag ditentukan dengan menggunakan *Adjusted-R<sup>2</sup>* atau *AIC (Akaike Information Criterion)* terkecil. Bisa juga dengan mengambil Minimum Information Criterion terkecil.
- 3) Dugaan parameter.
- 4) Memeriksa korelasi silang antar sisaan.
- 5) Diagnosis model satu per satu untuk model keseluruhan dan sisaan.
- 6) Mengkaji respon impuls, yang merupakan informasi untuk pengaruh perubahan *shock* atau guncangan suatu peubah terhadap peramalan peubah lain. Dapat dilihat melalui plot respon impuls.
- 7) Mengkaji dekomposisi ragam, yang berarti proporsi keragaman galat suatu peubah yang dijelaskan oleh galat masing-masing peubah dan galat peubah lain.
- 8) Melakukan peramalan.

Berikut adalah persamaan yang dituliskan pada model VAR :

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} s_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix} \text{ atau dengan bentuk lain:}$$

$$Bx_t = \Gamma_0 + \Gamma_1 X_{t-1} + \varepsilon_t \dots \dots \dots (7)$$

dimana:

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \quad x_t = \begin{bmatrix} s_t \\ y_t \end{bmatrix} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix}$$

Dengan melakukan pengalihan antara persamaan (7) dengan B<sup>-1</sup> atau invers matriks B, maka akan dapat ditentukan model VAR dalam bentuk standar, seperti dituliskan pada persamaan di bawah ini:

$$x_t = A_0 + A_1 x_{t-1} + \ell_t \dots \dots \dots (8)$$

dimana:

$$A_0 = B^{-1} \Gamma_0$$

$$A_1 = B^{-1} \Gamma_1$$

$$\ell_t = B^{-1} \varepsilon_t$$

Pada kajian ini  $X_t$  merupakan matriks yang dibentuk dari luas areal teh, produksi, harga konsumen dan volume ekspor dan volume impor

Sebagaimana model regresi, model VAR juga didasarkan pada beberapa asumsi antara lain:

**- Sisaan mengikuti fungsi distribusi normal**

Uji normalitas pada model VAR didasarkan pada nilai *Jarque-Bera (JB) test (multivariate)*, *Skewness only (multivariate)*, dan *Kurtosis only (multivariate)*. Hipotesis yang mendasari yaitu:

Hipotesis: .....(9)

$H_0$  : Sisaan terdistribusi normal

$H_1$  : *Otherwise*

Jika nilai p-value dari *JB test*, *Skewness only* dan *Kurtosis only* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi normalitas terpenuhi.

**- Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas)**

Asumsi homoskedastisitas pada model VAR didasarkan pada nilai *ARCH (multivariate)* dengan hipotesis sebagai berikut:

Hipotesis: .....(10)

$H_0$  : Homoskedastisitas

$H_1$  : Heterokedastisitas

Jika nilai p-value pada *ARCH (multivariate)* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi homoskedastisitas terpenuhi.

**- Tidak terdapat autokorelasi antar sisaan untuk setiap data pengamatan**

*Pormanteau Test (asymptotic)* digunakan untuk menguji ada tidaknya autokorelasi antar sisaan pada data amatan. Hipotesis yang digunakan sebagai berikut:

Hipotesis: .....(11)

$H_0$  : Tidak ada autokorelasi antar sisaan

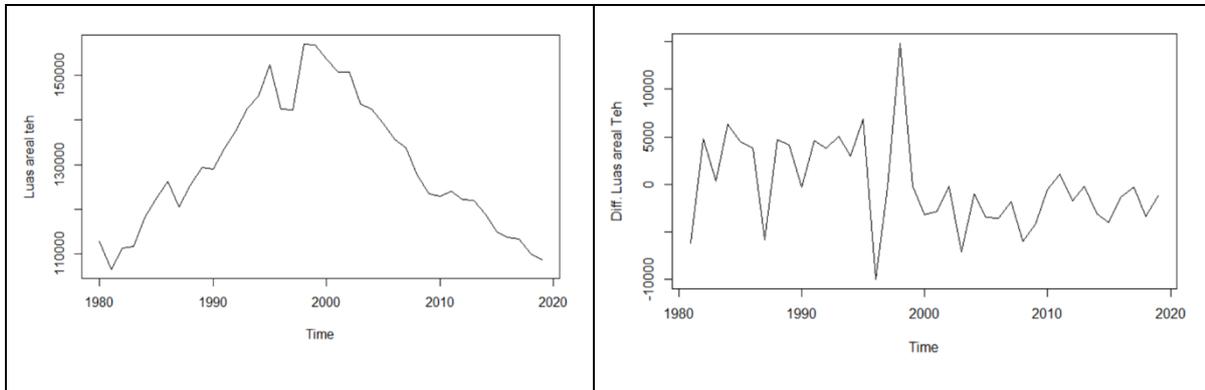
$H_1$  : *Otherwise*

Jika nilai p-value hasil uji *Pormanteau Test* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi tidak adanya autokorelasi antar sisaan telah terpenuhi.

## HASIL DAN PEMBAHASAN

### MODEL ARIMA

#### Eksplorasi Data



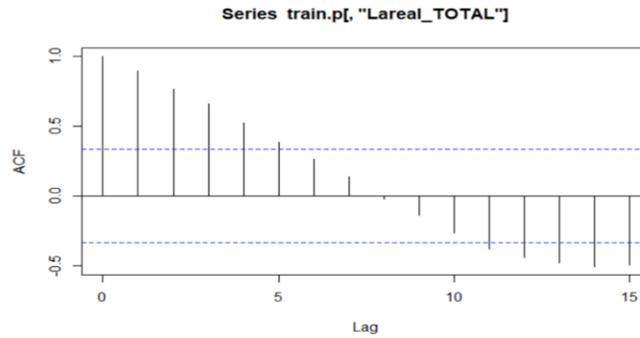
Gambar 1. Plot Luas Areal Teh, 1980-2019

Gambar 2. Plot Luas ArealTeh *Differencing*

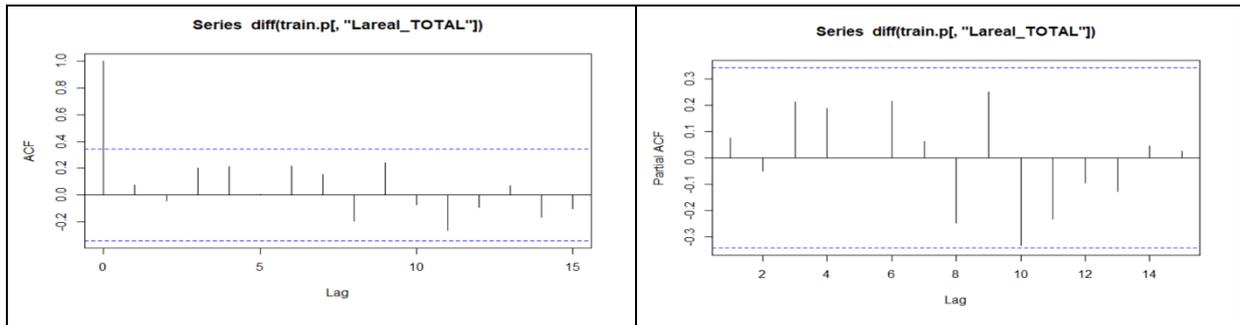
Gambar 1 adalah plot luas areal teh yang belum stationer, kemudian dilakukan *differencing* untuk mendapatkan plot yang stasioner (Gambar 2). Stasioneritas berarti bahwa tidak terdapat perubahan yang drastis pada data. Fluktuasi data berada disekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan variansi dari fluktuasi tersebut (Makridakis, 19951). Data time series dikatakan stasioner jika rata-rata dan variansinya konstan, tidak ada unsur trend dalam data, dan tidak ada unsur musiman. Apabila data tidak stasioner, maka perlu dilakukan modifikasi untuk menghasilkan data yang stasioner. Salah satu cara yang umum dipakai adalah metode pembedaan (*differencing*). Untuk menentukan apakah series stasioner, nonstasioner dapat dibantu dengan melihat plot dari series atau bentuk *difference*-nya. Proses *differencing* dapat dilakukan untuk beberapa periode sampai data stasioner, yaitu dengan cara mengurangkan suatu data dengan data sebelumnya.

#### Uji Kestasioneran Data

Data deret waktu memerlukan transformasi dan pembedaan untuk mencapai kestasioneran data. Transformasi diperlukan agar deret waktu stasioner dalam ragam. Sedangkan pembedaan diperlukan agar deret waktu stasioner dalam rataan. Kestasioneran data dapat diketahui pengujian secara visual dan uji statistik. Uji secara visual dapat dilihat dari plot data ACF (Gambar 3-5). Agar tidak terjadi unsur subjektivitas, maka diperlukan uji statistik, yaitu menggunakan Uji *Augmented Dickey Fuller*. Hasil uji *Augmented Dickey Fuller* (Tabel 1) mengindikasikan bahwa data luas areal teh konstan di sekitar rataan bukan nol atau berpola *Single Mean*. Hal ini terlihat dari nilai  $p$  yang lebih besar dari taraf nyatanya ( $\alpha=0,05$ ). Oleh karena itu, data bisa dapat langsung digunakan untuk pembentukan model.



Gambar 3. Plot *Auto Correlation Function* (ACF) Data Training Luas Areal Teh



Gambar 4. Plot ACF Luas Areal Teh *Differencing*

Gambar 5. Plot ACF Luas Areal Teh *Differencing*

**Tabel 1.** Hasil Uji *Augmented Dickey-Fuller* Luas Areal Teh

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
Min 1Q Median 3Q Max
-9246.7 -1791.4 259.2 1693.5 14978.1
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.279e+04 6.514e+03 1.963 0.05783 .
z.lag.1 -6.520e-02 4.853e-02 -1.344 0.18798
tt -2.048e+02 6.698e+01 -3.058 0.00432 **
z.diff.lag -3.981e-02 1.588e-01 -0.251 0.80353
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 4187 on 34 degrees of freedom
Multiple R-squared: 0.258, Adjusted R-squared: 0.1925
F-statistic: 3.94 on 3 and 34 DF, p-value: 0.01627
Value of test-statistic is: -1.3436 3.7664 5.6449
Critical values for test statistics:
1pct 5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2 7.02 5.13 4.31
phi3 9.31 6.73 5.61
```

Berdasarkan Gambar 1 dan Tabel 1, terlihat jelas bahwa plot ACF data *training* memiliki pola tertentu sehingga diperkirakan data yang digunakan tidak stasioner. Hal ini diperkuat oleh hasil uji formal statistik yaitu dengan uji *Augmented Dickey-Fuller* yang mengindikasikan bahwa data luas areal teh adalah tidak stasioner, terlihat dari nilai statistik uji (-1,3436) yang lebih besar dari nilai kritis taraf 5% yakni -4,15. Maka dilakukan stasioneritas dengan melakukan differencing dan dilakukan pengujian ADF kembali.

Dari hasil *Augmented Dickey-Fuller test* pada Tabel 2, menunjukkan bahwa luas areal teh telah stasioner. Hal tersebut diketahui dari nilai statistik uji (-4,0196) yang lebih kecil dibandingkan nilai kritis taraf 5% yakni -2,62.

**Tabel 2.** Hasil Uji Augmented Dickey-Fuller *Differencing* Luas Areal Teh

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
Residuals:
  Min    1Q  Median    3Q   Max
-10886.1 -3051.9 -600.5  3095.2 14507.5
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1  -0.88445   0.22004  -4.020 0.000295 ***
z.diff.lag  0.03504   0.16295   0.215 0.830985
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 4665 on 35 degrees of freedom
Multiple R-squared:  0.4342,    Adjusted R-squared:  0.4019
F-statistic: 13.43 on 2 and 35 DF,  p-value: 4.695e-05
Value of test-statistic is: -4.0196
Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.62 -1.95 -1.61
```

Data luas areal teh telah stasioner maka langkah selanjutnya adalah menentukan ordo AR dan MA menggunakan fungsi auto arima. Dari hasil pendugaan auto arima pada Tabel 3, maka model ARIMA yang terpilih adalah ARIMA (0,2,1) yang berarti model ARIMA yang optimal untuk mengestimasi data yang diolah adalah ARIMA dengan ordo *Auto Regressive* (AR) = 0, ordo *differencing* = 2, dan ordo *Moving Average* (MA) = 1.

**Tabel 3.** Hasil Fungsi Auto ARIMA Luas Areal Teh

```
Series: teh[, "Lareal_TOTAL"]
ARIMA(0,2,1)
Coefficients:
```

```

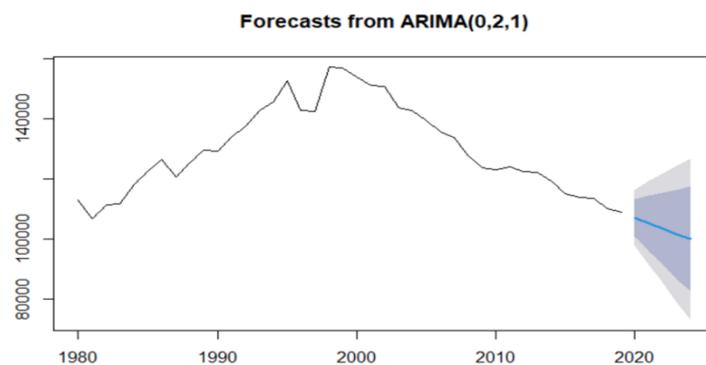
ma1
-0.8574
s.e. 0.0827
sigma^2 estimated as 22155556: log likelihood=-375.43
AIC=754.87 AICc=755.21 BIC=758.14
Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -234.657 4527.018 3239.901 -0.1211805 2.431637 0.905156 -0.01104558

```

Setelah diperoleh model ARIMA yang optimal, maka model ARIMA (0,2,1) tersebut digunakan untuk melakukan peramalan luas areal teh 5 tahun ke depan (2020-2024). Hasil peramalan luas areal teh menggunakan ARIMA(0,2,1) ditampilkan pada Tabel 4.

**Tabel 4.** Hasil Peramalan Produksi Teh dengan ARIMA (0,2,1)

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2020	106950.70	100918.48	112982.9	97725.21 116176.2
2021	105151.41	95992.02	114310.8	91143.33 119159.5
2022	103352.11	91352.11	115352.1	84999.69 121704.5
2023	101552.82	86781.74	116323.9	78962.40 124143.2
2024	99753.52	82205.62	117301.4	72916.32 126590.7



Gambar 6. Plot Estimasi Data Luas Areal Teh

## MODEL REGRESI

### Model Regresi Linier

Peubah bebas yang digunakan dalam melakukan pemodelan data luas areal teh di Indonesia adalah tahun. Data yang digunakan adalah periode tahun 1980 hingga 2019. Periode data tersebut kemudian dipisahkan menjadi data set *training* dan *testing*. Panjang series data pada data set *training* adalah tahun 1980 hingga 2013, sementara dataset *testing*

adalah periode tahun 2014-2019. Dataset *training* digunakan untuk melakukan penyusunan model, sementara dataset *testing* digunakan untuk validasi model.

Dari hasil pendugaan parameter (Tabel 5) menunjukkan model belum cukup bagus, terlihat dari nilai *p-value* 0,0752 atau  $H_0$  diterima, artinya koefisien belum signifikan karena model masih lebih besar dari 0,05 dan variabel tahun signifikan secara uji t dengan *p-value* 0,0414. Nilai  $R^2$  model bernilai 96,27% yang artinya peubah tahun dapat menjelaskan 96,27% keragaman dari luas areal teh di Indonesia.

**Tabel 5.** Pendugaan Model Regresi Linier

```
Call:
lm(formula = Lareal_TOTAL ~ Tahun, data = train)

Residuals:
    Min       1Q   Median       3Q      Max
-18944.1 -11668.5  -929.5   9721.2  23477.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -854500.0   464642.4  -1.839   0.0752 .
Tahun         494.5     232.7    2.125   0.0414 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13310 on 32 degrees of freedom
Multiple R-squared:  0.1237, Adjusted R-squared:  0.09627
F-statistic: 4.515 on 1 and 32 DF, p-value: 0.04141
```

### MAPE Training

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.2044	3.8665	8.2569	8.3441	12.9100	17.4750

### MAPE Testing

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
18.99	24.02	25.78	26.16	29.41	32.36

### Uji Kenormalan

```
Jarque-Bera test for normality
data: reg1$residuals
JB = 2.0002, p-value = 0.1825
```

## Model Regresi Kuadratis

Pendugaan model regresi linear kuadratis dilakukan dengan memperhatikan signifikansi dari model dan parameter-parameternya. Jika dibandingkan hasil penelusuran model antara model regresi linier dengan kuadratis maka dapat disimpulkan bahwa hasil regresi kuadratis lebih memenuhi kaidah-kaidah statistik yang baik. Terlihat bahwa nilai probabilitas ( $<0,0001$ ) model yang jauh lebih kecil dari 0,05. Dengan  $R^2$  sebesar 83,7% .

**Tabel 6.** Pendugaan Model Regresi Kuadratis

```
Call:
lm(formula = Lareal_TOTAL ~ t + t2, data = train)
```

```

Residuals:
  Min      1Q  Median      3Q      Max
-8424.3 -3823.8 -186.3  1587.3 12058.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 100641.59   2746.17   36.65 < 2e-16 ***
t            4997.94    385.12   12.98 4.54e-14 ***
t2          -136.47     11.28  -12.10 2.82e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 5655 on 31 degrees of freedom
Multiple R-squared:  0.8468, Adjusted R-squared:  0.837
F-statistic: 85.71 on 2 and 31 DF,  p-value: 2.342e-13

```

### MAPE Testing

```

Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 5.116  6.420  10.789  11.063  14.423  19.085

```

### MAPE Training

```

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.02186 1.00868  2.64470  3.17313  4.85407 10.69956

```

### Uji Kenormalan

Jarque-Bera test for normality

```

data: reg2$residuals
JB = 2.0187, p-value = 0.1765

```

## MODEL FUNGSI TRANSFER

Pada model Fungsi Transfer peubah input yang dapat mempengaruhi luas areal teh adalah produksi teh. Dari hasil uji *Augmented Dickey-Fuller* (Tabel 7) terlihat bahwa data produksi teh belum stasioner pada taraf uji 5%. Maka akan dilakukan *differencing* terlebih dahulu pada data produksi teh. Hasil uji ADF yang telah dilakukan *differencing* satu kali menunjukkan bahwa nilai *test-statistic* yaitu -6.201 lebih kecil dari *critical values* menunjukkan bahwa  $H_0$  ditolak yang berarti data luas tanaman menghasilkan teh telah stasioner (Tabel 8).

**Tabel 7.** Hasil uji *Augmented Dickey Fuller* pada Produksi Teh

```

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
  Min    1Q  Median    3Q   Max
-24831 -4018  1086  6259 19264
Coefficients:
            Estimate Std. Error t value Pr(>|t|)

```

```

(Intercept) 29927.8442 17953.4100 1.667 0.107
z.lag.1 -0.1888 0.1471 -1.283 0.210
tt -36.8840 307.3167 -0.120 0.905
z.diff.lag -0.2790 0.1886 -1.480 0.150
Residual standard error: 10270 on 28 degrees of freedom
Multiple R-squared: 0.2287, Adjusted R-squared: 0.1461
F-statistic: 2.768 on 3 and 28 DF, p-value: 0.06025
Value of test-statistic is: -1.2828 1.7133 2.2157
Critical values for test statistics:
 1pct 5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2 7.02 5.13 4.31
phi3 9.31 6.73 5.61

```

**Tabel 8.** Hasil uji *Augmented Dickey Fuller* pada Luas Tanaman Teh (*Differencing 1*)

```

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
Residuals:
  Min   1Q Median   3Q   Max
-20382 -3900 3095 11196 19340
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1  -1.7074   0.2753  -6.201 9.19e-07 ***
z.diff.lag  0.3113   0.1701   1.829  0.0776 .
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 10030 on 29 degrees of freedom
Multiple R-squared: 0.7018, Adjusted R-squared: 0.6813
F-statistic: 34.13 on 2 and 29 DF, p-value: 2.398e-08
Value of test-statistic is: -6.201
Critical values for test statistics:
 1pct 5pct 10pct
tau1 -2.62 -1.95 -1.61

```

Pencarian ordo model ARIMA input untuk produksi teh dilakukan dengan fungsi auto arima. Dari hasil fungsi auto arima, diperoleh ordo ARIMA terbaik untuk produksi teh adalah ARIMA(0,1,1) sebagaimana ditampilkan pada Tabel 9.

**Tabel 9.** Auto Model ARIMA (0,1,1) Produksi  
Series: train.h[, "Produksi\_TOTAL"]  
ARIMA(0,1,1)

Coefficients:

ma1

-0.3751

s.e. 0.1419

sigma<sup>2</sup> estimated as 104899113: log likelihood=-351.12

AIC=706.25 AICc=706.65 BIC=709.24

Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 1897.353 9936.225 7702.993 1.14934 5.490948 0.9136643 -0.05934335

### Penentuan nilai $b$ , $r$ dan $s$

Jika hasil korelasi silang nyata, maka dilanjutkan dengan mencari nilai  $b$ ,  $r$  dan  $s$  dengan melihat plot korelasi silang deret input dan output. Plot korelasi silang antara produksi teh dan luas areal menunjukkan nilai yang signifikan pada  $lag$  ke-  $b$ . Nilai  $s$  dilihat pada banyaknya  $lag$  korelasi silang yang signifikan setelah  $lag$  ke- $b$  yang diartikan dengan lamanya nilai luas tanaman menghasilkan mempengaruhi produksi teh setelah nyata yang pertama, dari diperoleh nilai  $s=0$ . Sedangkan nilai  $r$  dilihat dari plot ACF deret output yang signifikan setelah signifikan yang pertama dan diperoleh  $r=1$ .

### Membuat Model ARIMA Deret Sisaan

Pendugaan model ARIMA deret sisaan dilakukan dengan mengamati plot ACF dan PACF deret sisaan yang memperlihatkan bahwa masih terdapat  $lag$  yang nyata. Model yang baik adalah model yang tidak mempunyai  $lag$  yang nyata baik pada plot ACF dan PACF, di pengujian korelasi serta di korelasi silang deret sisaan. Bila pada plot ACF dan PACF masih terlihat pola sistematis, maka perlu dilakukan *overfitting*.

### Peramalan Berbasis Fungsi Transfer

Model fungsi transfer didapat dari kombinasi model deret input awal dengan model ARIMA deret sisaan. Model deret input adalah data produksi, sementara model ARIMA sisaan menggunakan data luas areal.

Hasil pengujian dari model ARIMA (0,2,1) menghasilkan MAPE untuk data testing adalah 3,48 % dan MAPE untuk data training adalah 2,63% (Tabel 10).

**Tabel 10.** Evaluasi *Magnitude* dan *Direction* Model ARIMA (0,2,1) Data Produksi

Data	N	<i>Magnitude</i> (MAPE)
Testing	6	3.48
Training	34	2.63

<sup>a</sup> Probabilitas < 0,05

Setelah diperoleh model fungsi transfer yang optimal, maka model tersebut digunakan untuk melakukan peramalan luas areal teh 5 tahun ke depan (2020-2024). Hasil peramalan

luas areal teh Indonesia menggunakan fungsi transfer dengan produksi teh sebagai peubah ditampilkan pada Tabel 11 berikut.

**Tabel 11.** Hasil Peramalan Luas Areal Teh menggunakan Fungsi Transfer

2020	2021	2022	2023	2024
107.060	105.321	103.582	101.842	100.103

### MODEL VECTOR AUTO REGRESSION (VAR)

Model VAR untuk peramalan luas areal teh memasukkan peubah produksi, luas areal, harga konsumen, volume ekspor dan volume impor teh. Hal ini diasumsikan jika terjadi gejolak pada variabel-variabel tersebut akan mempengaruhi gejolak luas areal teh atau sebaliknya.

Pada tahap pertama, perlu dilakukan pemeriksaan apakah terdapat pengaruh trend dalam model. Hasil pengujian signifikansi trend dilakukan pada  $lag=1$  dan  $lag=2$ . Hasil pengujian trend pada  $lag=1$  menunjukkan bahwa *trend* berpengaruh signifikan pada persamaan luas areal teh (signifikan 5%). Dari hasil *running* model menggunakan  $lag$  1 dan 2, diperoleh informasi komponen *trend* tidak terlalu berpengaruh signifikan sehingga *trend* dikeluarkan dalam model untuk *running* model VAR (p) selanjutnya.

Langkah selanjutnya adalah menentukan  $lag$  (p) yang akan digunakan dalam model VAR. Penentuan  $lag$  (p) dilakukan dengan melakukan pengujian VAR(1) hingga VAR(4) dengan pertimbangan kecukupan series data. Dari hasil *running* menggunakan  $p=1$  sampai dengan  $p=4$ , diperoleh informasi bahwa variabel yang signifikan cukup banyak pada  $p=2$ , sehingga model VAR yang digunakan selanjutnya adalah VAR (2) *type both*. Hal tersebut diperkuat juga dengan nilai *Adjusted R-squared*: 0,9016 atau dapat dikatakan bahwa 90,16% keragaman luas areal teh dapat dijelaskan oleh kelima variabel/peubah penjelas termasuk konstanta. Hasil pengujian model VAR(2) dapat dilihat dalam Tabel 12.

**Tabel 12.** Hasil Estimasi VAR Luas Areal Teh dengan Peubah Produksi, Luas Areal, Harga Konsumen, Volume Ekspor, Volume Impor pada VAR (2) type “both”.

VAR Estimation Results:

=====

Endogenous variables: Produksi\_TOTAL, Lareal\_TOTAL, harga\_kons, vol\_ekspor, vol\_impor

Deterministic variables: both

Sample size: 33

Log Likelihood: -1614.578

Roots of the characteristic polynomial:

0.9744 0.6301 0.5618 0.5618 0.01709

Call:

VAR(y = teh[1:34, c(2, 3, 5, 7, 8)], p = 1, type = "both")

Estimation results for equation Lareal\_TOTAL:

=====

Lareal\_TOTAL = Produksi\_TOTAL.l1 + Lareal\_TOTAL.l1 + harga\_kons.l1 + vol\_ekspor.l1 + vol\_impor.l1 + Produksi\_TOTAL.l2 + Lareal\_TOTAL.l2 + harga\_kons.l2 + vol\_ekspor.l2 + vol\_impor.l2 + const + trend

	Estimate	Std. Error	t value	Pr(> t )
Produksi_TOTAL.l1	1.705e-01	1.287e-01	1.325	0.2001
Lareal_TOTAL.l1	5.927e-01	1.869e-01	3.171	0.0048 **
harga_kons.l1	1.949e-01	3.301e-01	0.590	0.5615
vol_ekspor.l1	-1.247e-01	8.167e-02	-1.527	0.1424
vol_impor.l1	1.711e-01	3.836e-01	0.446	0.6603
Produksi_TOTAL.l2	2.349e-01	1.285e-01	1.827	0.0827 .
Lareal_TOTAL.l2	-4.511e-02	1.867e-01	-0.242	0.8116
harga_kons.l2	-3.983e-01	3.295e-01	-1.209	0.2409
vol_ekspor.l2	7.246e-02	7.867e-02	0.921	0.3680
vol_impor.l2	2.795e-02	4.882e-01	0.057	0.9549
const	1.588e+04	1.303e+04	1.218	0.2373
trend	-3.638e+02	5.139e+02	-0.708	0.4872

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
 Residual standard error: 4105 on 20 degrees of freedom  
 Multiple R-Squared: 0.9365, Adjusted R-squared: 0.9016  
 F-statistic: 26.83 on 11 and 20 DF, p-value: 1.433e-09

Setelah terpilih *lag* model VAR yang optimal, maka perlu dilakukan pengujian asumsi non-autokorelasi, normalitas, dan homokedastitas pada sisaan model VAR (2) *type both*. Dari hasil pengujian non autokorelasi, diperoleh informasi bahwa model tersebut terbebas dari autokorelasi dengan tingkat kesalahan 5% (Tabel 12). Residual memenuhi asumsi kenormalan dengan tingkat kesalahan 5% melalui *normality test* (*Jarque Bera Test*, pengujian skewness, dan kurtosis) sebagaimana tampak pada Tabel 13, sedangkan asumsi homoskedastisitas atau kehomogenan ragam signifikan terpenuhi dengan *p-value* =1 (Tabel 15).

**Tabel 13.** Hasil Uji Asumsi Non-Autokorelasi model VAR(2) *type both* menggunakan *Serial Test*

Portmanteau Test (asymptotic)  
 data: Residuals of VAR object varhsheet.b2  
 Chi-squared = 278.8, df = 350, p-value = 0.998

**Tabel 14.** Hasil Uji Asumsi Normalitas Luas Areal Teh menggunakan *Normality Test*

\$JB  
 JB-Test (multivariate)  
 data: Residuals of VAR object varhsheet.b2  
 Chi-squared = 44.63, df = 10, p-value = 2.536e-06  
 \$Skewness  
 Skewness only (multivariate)

data: Residuals of VAR object varhsheet.b2  
 Chi-squared = 19.02, df = 5, p-value = 0.001906  
 \$Kurtosis

Kurtosis only (multivariate)

data: Residuals of VAR object varhsheet.b2  
 Chi-squared = 25.61, df = 5, p-value = 0.0001062

**Tabel 15.** Hasil Uji Asumsi Homokedasitas Luas Areal Teh menggunakan ARCH Test

ARCH (multivariate)

data: Residuals of VAR object varhsheet.b2  
 Chi-squared = 405, df = 1125, p-value = 1

Sebelum melakukan peramalan, perlu dilakukan evaluasi nilai kesalahan model VAR yang terbentuk dengan menghitung nilai MAPE atau nilai rata-rata persentase kesalahan peramalan. Nilai MAPE VAR(2) type both untuk dataset *training* dan *testing* masing-masing sebesar 19,12% dan 10,76% seperti tersaji pada Tabel 16.

**Tabel 16.** Nilai MAPE Model VAR(2) type both Luas Areal Teh Indonesia

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<b>Data Training</b>	12.37	17.91	19.54	19.12	21.01	24.79
<b>Data Testing</b>	<b>1.933</b>	<b>5.860</b>	<b>10.768</b>	<b>10.812</b>	<b>15.559</b>	<b>22.274</b>

Setelah diperoleh model VAR yang optimal yakni VAR(2) *type both*, maka model tersebut digunakan untuk melakukan peramalan luas areal teh 5 tahun ke depan (2020-2024). Hasil peramalan luas areal teh Indonesia menggunakan metode VAR dengan peubah produksi teh, harga konsumen teh, volume ekspor teh, volume impor teh ditampilkan pada Tabel 17 berikut.

**Tabel 17.** Hasil Peramalan Luas Areal Teh menggunakan VAR(2) *type both*

	2020	2021	2022	2023	2024
	<b>120.137</b>	<b>117.565</b>	<b>114.724</b>	<b>111.772</b>	<b>108.770</b>

## KESIMPULAN DAN SARAN

Kesimpulan dari penyusunan angka estimasi luas areal teh dilakukan dengan penelusuran 4 model terbaik yaitu ARIMA, Regresi, Fungsi Transfer dan VAR.

1. Model yang pertama adalah model time series atau ARIMA, model terbaik untuk ARIMA adalah pada orde ARIMA (0,2,1). Untuk model estimasi luas areal teh nasional dengan ARIMA (0,2,1) menghasilkan MAPE untuk data training sebesar 2,43%.
2. Model yang kedua adalah menggunakan regresi linier sebagai variabel bebas adalah waktu atau tahun. Untuk model regresi linier sederhana menghasilkan MAPE yang cukup besar yaitu 8,34% untuk data training dan 26,16% untuk data testing. Untuk model regresi kuadratik dengan variabel bebas waktu (tahun) menghasilkan MAPE yang lebih baik, yaitu sebesar 3,17% untuk data training dan 11,06% untuk data testing.
3. Model ketiga adalah menggunakan fungsi transfer, dengan data input nya adalah produksi. Model ARIMA data input menggunakan ARIMA (0,2,1). Hasil dari pengepasan fungsi transfer untuk MAPE data training 2,63% dan MAPE untuk data testing 3,48%.
4. Model keempat adalah *Vector Auto Regressive* (VAR) dengan menggunakan 5 variabel yaitu produksi, luas areal, harga konsumen, volume ekspor dan volume impor. Untuk Model VAR adalah nilai  $p=2$  dan  $type="both"$ ,  $p=2$  artinya menggunakan variabel bebas sampai lag-2 dan ada konstanta dan faktor trend. MAPE yang dihasilkan dari model VAR (2)  $type="both"$  adalah MAPE untuk data training 19,12% dan MAPE untuk data testing 10,81%.
5. Pada model regresi, variabel bebas/peubah sebaiknya dapat menggunakan variabel lain yang berpengaruh selain variabel waktu.
6. Model fungsi transfer dan VAR sebaiknya dapat diujicobakan dengan menggunakan variabel lain yang dapat berpengaruh secara teori.
7. Sebaiknya dapat dilakukan kajian lebih mendalam terkait dengan metode estimasi dengan menggunakan software lain

## DAFTAR PUSTAKA

- Budiar, Aindra. 2013. Peramalan Produksi TBS Kelapa Sawit Dengan Model Arima dan Model Fungsi Transfer Input Ganda. Institut Pertanian Bogor, Bogor
- Gujarati, D.N. and D.C. Porter, 2010. Dasar-dasar Ekonometrika. Jakarta: Penerbit Salemba Empat, Buku 2, Edisi 5.
- Heizer, J., Render, B. & Munson, C., 2011. Operations Management Sustainability and Supply Chain Management. Boston: Pearson.
- Kementerian Pertanian. 2013. Pedoman Teknis Pengembangan Tanaman Teh Tahun 2014. Jakarta: Direktorat Jenderal Perkebunan Kementerian Pertanian

Kementerian Pertanian. 2019. Statistik Perkebunan Indonesia 2018-2020 (Teh). Jakarta: Direktorat Jenderal Perkebunan.

Mardiyanto, Atqo. 2000. Kajian Peramalan Dengan Model Struktural dan Non Struktural (VAR dan ARIMA). Institut Pertanian Bogor, Bogor

Makridakis S, Wheelwright SC, McGee VE. 1999. Metode dan Aplikasi Peramalan. Suminto H, penerjemah. Binarupa Aksara, Jakarta.

M. Firdaus 2020. Aplikasi Ekonometrika dengan E-Views, Stata dan R. Institut Pertanian Bogor (IPB) Press.

Tsay, R.S., 2014. Multivariate Time Series Analysis With R and Financial Applications. New Jersey: John Wiley & Sons.



# ESTIMASI LUAS AREAL LADA DI INDONESIA

## *Estimation of Pepper Area in Indonesia*

Roydatul Zikria<sup>1\*</sup>

<sup>1</sup>*Pusat Data dan Sistem Informasi Pertanian, Kementerian Pertanian  
Jl. Harsono RM No. 3 Gedung D Lantai 4, Jakarta Selatan, Indonesia*

\*Korespondensi penulis. E-mail: [roydatul\\_z@pertanian.go.id](mailto:roydatul_z@pertanian.go.id)

### ABSTRACT

Indonesia pepper area has increased for the last five years. Estimation of pepper area in the next few years uses ARIMA model, regression, transfer function and VAR. Regarding estimation of those models, transfer function is selected as the best model to estimate pepper area. In detail, the model is transfer function ARIMA (1,1,0) using export price as an input variable. Transfer model produces MAPE for training data and testing data which are respectively 4.92 and 1.83. Estimation of pepper production for the next five years (2020-2024) increases. The estimation in 2020 is 188,517 ha while in 2024 is estimated 188.993 ha of area.

**Keywords:** *estimation, ARIMA, regression, transfer function, VAR*

### ABSTRAK

Luas areal lada di Indonesia cenderung meningkat selama lima tahun terakhir. Untuk mengestimasi luas areal lada beberapa tahun ke depan, penelitian ini menggunakan model ARIMA, regresi, fungsi transfer dan VAR. Hasil estimasi dari keempat model tersebut dibandingkan dan terpilih fungsi transfer sebagai model estimasi terbaik. Model fungsi transfer yang terpilih adalah fungsi transfer ARIMA (1,1,0) dengan variabel input harga ekspor. Hasil estimasi dengan fungsi transfer menghasilkan MAPE data training sebesar 4,92 dan MAPE data testing 1,83. Luas areal lada lima tahun ke depan (2020-2024) diramalkan naik. Tahun 2020 luas areal lada diramalkan sebesar 188.517 ha kemudian meningkat di tahun 2024 menjadi 188.993 ha.

**Kata kunci:** *estimasi, ARIMA, regresi, fungsi transfer, VAR*

### PENDAHULUAN

Lada merupakan salah satu komoditas yang berperan penting dalam perekonomian Indonesia. Usahatani lada mampu menyumbang negara, menyediakan lapangan pekerjaan serta menjadi bahan baku industri maupun konsumsi langsung (Kemala, 2006). Pengembangan usahatani lada di Indonesia sebagian besar berada di Luar Jawa seperti Kepulauan Bangka Belitung, Lampung, Sumatera Selatan, Sulawesi Selatan dan Kalimantan Timur. Berdasarkan data Kementerian Pertanian (2019) luasan lada yang berasal dari kelima provinsi tersebut sebesar 65% dari total luas areal nasional. Jenis lada yang paling banyak diusahakan di Indonesia adalah lada hitam yang berasal dari Lampung atau dikenal dengan nama *Lampung black pepper* dan lada putih atau *Muntok white pepper* yang dihasilkan di Kepulauan Bangka Belitung (Nurdjannah, 2006).

Pengusahaan lada di Indonesia didominasi oleh perkebunan rakyat. Pengusahaan lada oleh perkebunan rakyat umumnya masih dilakukan secara tradisional (Damanik, 2001). Menurut Kardinan et al. (2018) budidaya lada secara tradisional dilakukan petani dimana penggunaan bahan kimia sebagai asupan tambahan menjadi andalan dan merupakan garansi keberhasilan petani. Oleh karena itu, ketika pupuk dan pestisida sulit diperoleh maka asupan tanaman lada berkurang yang mengakibatkan produksinya turun.

Luas areal lada di Indonesia selama sepuluh tahun terakhir menunjukkan adanya trend peningkatan seiring dengan peningkatan produksinya. Tahun 2010 luas areal lada sebesar 179.318 ha dan meningkat menjadi 188.041 ha pada tahun 2019 atau terjadi penambahan luas 0,18% setiap tahunnya (Kementerian Pertanian, 2019). Peningkatan tersebut mengindikasikan bahwa luas areal lada beberapa tahun ke depan masih memungkinkan bertambah jika didukung dengan kebijakan yang tepat. Estimasi luas areal lada beberapa tahun ke depan sangat diperlukan sebagai bahan penentu kebijakan maupun *Early Warning System (EWS)* mengingat data Angka Tetap (ATAP) Perkebunan memiliki lag yang cukup jauh dibandingkan data tahun berjalan.

Estimasi luas areal pada tahun berjalan maupun beberapa periode ke depan sangat penting untuk penentuan kebijakan di subsektor perkebunan. Direktorat Jenderal Perkebunan telah mengupayakan penyediaan data yang lebih *real time* dengan melakukan penyusunan Angka Estimasi (AESTI) pada tahun berjalan. Namun estimasi luas areal komoditas perkebunan untuk lima tahun ke depan masih belum tersedia. Estimasi Ditjen Perkebunan hanya dilakukan untuk satu tahun ke depan menggunakan model *univariate* seperti *Double Exponential Smoothing (DES)*. Salah satu kelemahan dari model *univariate* yaitu variabel yang digunakan hanya satu misalnya luas areal. Akibatnya hasil analisis hanya mampu memberikan gambaran terhadap satu variabel saja tanpa adanya intervensi dari variabel lain. Padahal luas areal komoditas perkebunan tidak terlepas dari pengaruh variabel-variabel lain seperti harga ekspor lada, harga pupuk, produksi serta variabel lainnya. Selain itu, untuk menentukan kebijakan subsektor perkebunan seperti ekstensifikasi atau perluasan tanam, diperlukan informasi variabel input lain yang diduga turut berpengaruh terhadap luas areal komoditas perkebunan. Oleh karena itu diperlukan model yang mampu menyajikan analisis mendalam dalam mengestimasi luas areal dengan melibatkan variabel input lain, misalnya model *multivariate*.

Penelitian ini menganalisis hasil estimasi luas areal lada di Indonesia dengan model *univariate* maupun *multivariate*. Terdapat empat model yang digunakan dalam mengestimasi luas areal lada antara lain *Autoregressive Integrated Moving Average (ARIMA)*, regresi, fungsi transfer dan *Vector Autoregression (VAR)*. Model ARIMA menghasilkan estimasi luas areal lada tanpa ada pengaruh dari variabel lain. Model regresi mengestimasi luas areal yang dipengaruhi oleh variabel waktu. Model fungsi transfer menghasilkan angka estimasi luas areal dengan memasukkan intervensi dari satu variabel pendukung yang dianggap paling berpengaruh terhadap luas areal. Model VAR mengestimasi luas areal dengan mempertimbangkan pengaruh dari beberapa variabel lain atau terdapat lebih dari satu variabel pendukung yang diduga berpengaruh terhadap luasannya. Hasil estimasi dari keempat model tersebut dibandingkan untuk selanjutnya ditentukan model terbaik untuk meramalkan luas areal lada di Indonesia beberapa tahun ke depan.

## METODE PENELITIAN

### Kerangka Pemikiran

Kerangka pemikiran dalam penelitian menggambarkan hasil estimasi luas areal lada dengan empat model yaitu ARIMA, regresi, fungsi transfer dan VAR. Model ARIMA umumnya digunakan untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang kurang baik ketepatannya. Estimasi dengan model ARIMA hanya menggunakan variabel dependen dan mengabaikan variabel independennya. Dengan kata lain, untuk mengestimasi luas areal lada beberapa tahun ke depan maka variabel yang digunakan hanya luas areal itu sendiri.

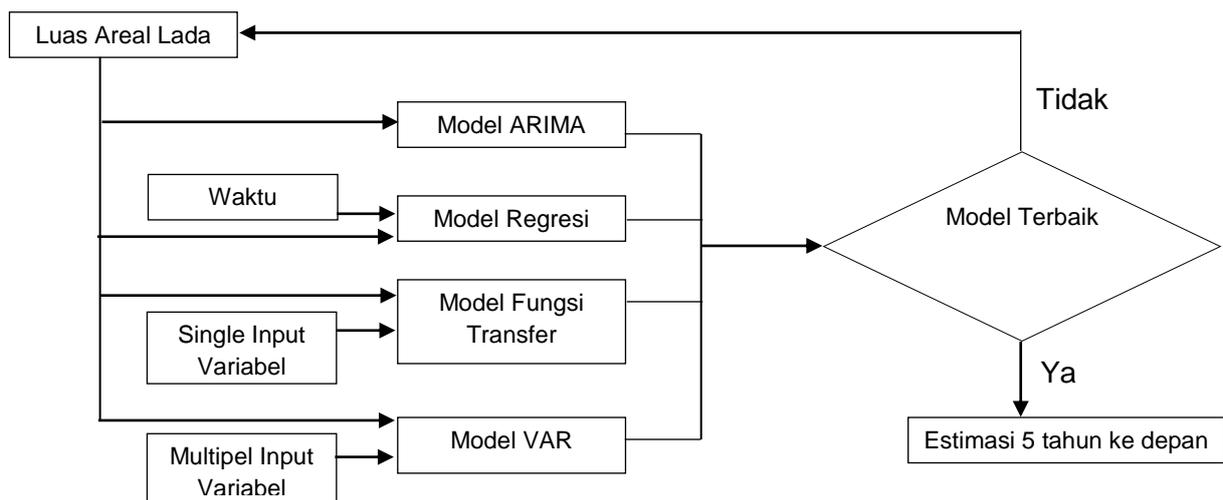
Estimasi dengan model regresi pada penelitian dibagi menjadi dua yaitu regresi linear sederhana dan regresi kuadratik. Kedua model regresi tersebut digunakan untuk menyelidiki pengaruh antara dua atau lebih variabel. Secara umum, estimasi dengan model regresi melibatkan variabel dependen dan variabel independen. Variabel dependen yang dimaksud pada analisis ini adalah luas areal lada sedangkan variabel independen yang dipilih yaitu waktu ( $t$ ). Selanjutnya dilakukan peramalan luas areal lada berdasarkan nilai-nilai tertentu yang diberikan oleh variabel independennya. Terdapat beberapa asumsi yang mendasari penggunaan model regresi linear sederhana maupun kuadratik antara lain sisaan mengikuti fungsi distribusi normal, varians sisaan konstan untuk setiap data amatan (homoskedastisitas) dan tidak terdapat autokorelasi antara sisaan untuk setiap data amatan.

Model fungsi transfer menggambarkan nilai ramalan masa depan dari suatu deret berkala (deret output) yang didasarkan pada nilai-nilai masa lalu dari deret itu sendiri serta didasarkan pula pada suatu deret berkala yang berhubungan (deret input). Model fungsi transfer merupakan fungsi dinamis yang pengaruhnya tidak hanya pada hubungan linear antara waktu ke- $t$  dengan deret/variabel input, tetapi juga terdapat hubungan antara variabel input dengan variabel output pada waktu ke- $t$ ,  $t+1$ , ...,  $t+k$ . Pada fungsi transfer terdapat rangkaian output yang mungkin dipengaruhi oleh rangkaian *multiple* input. Untuk kasus *single input* variabel pada fungsi transfer, dapat menggunakan metode korelasi silang. Penelitian ini menggunakan *single input* variabel yaitu harga ekspor untuk meramalkan luas areal lada sebagai variabel outputnya.

Model VAR menggunakan pendekatan non-struktural atau tidak mendasarkan pada teori ekonomi tertentu dalam melakukan peramalan. Model ini memperlakukan seluruh peubah secara simetris tanpa mempermasalahkan variabel dependen dan independen atau dengan kata lain model VAR memperlakukan seluruh variabel sebagai variabel dependen/endogen, karena pada kenyataannya suatu variabel dapat bertindak sebagai endogen atau dependen dari suatu peubah lain (Gujarati & Porter, 2010). Untuk meramalkan luas areal lada beberapa tahun ke depan, penelitian ini menggunakan beberapa variabel antara lain produksi lada, harga ekspor, dan harga pupuk.

Pembentukan model estimasi luas areal lada dilakukan dengan membagi series data aktual menjadi data training dan data testing. Data training digunakan untuk menentukan model

estimasi dan meramalkan data testing yang sebenarnya sudah tersedia data aktualnya. Hasil ramalan data testing tersebut kemudian dibandingkan dengan data aktual untuk dihitung tingkat kesalahan (*error*) hasil ramalan. Model terbaik untuk estimasi adalah model dengan tingkat *error* yang paling kecil, dalam hal ini ditunjukkan oleh nilai *Mean Absolute Percentage Error* (MAPE) terkecil. Selain MAPE, pemilihan model terbaik juga mempertimbangkan kelogisan hasil ramalan dengan historis data sebelumnya. Berdasarkan hasil identifikasi model ARIMA, regresi, fungsi transfer dan VAR, dipilih model terbaik untuk meramalkan luas areal lada di Indonesia selama lima tahun ke depan. Secara umum tahapan penelitian ini disajikan melalui kerangka konseptual penelitian pada Gambar 1.



Gambar 1. Kerangka Konseptual Penelitian

### Pengumpulan Data

Data yang digunakan pada penelitian ini bersumber dari Direktorat Jenderal Perkebunan, Kementerian Pertanian serta World Bank. Variabel yang digunakan antara lain luas areal lada, produksi lada, harga ekspor lada dan harga pupuk. Luas areal lada merupakan penjumlahan dari luas Tanaman Belum Menghasilkan (TBM), Tanaman Menghasilkan (TM) dan Tanaman Rusak/Tanaman Tidak Menghasilkan (TR/TTM). Produksi lada merupakan total produksi yang diusahakan oleh Perkebunan Rakyat (PR), Perkebunan Besar Negara (PBN) dan Perkebunan Besar Swasta (PBS). Harga ekspor dihitung dari nilai ekspor dibagi volume ekspor. Volume ekspor maupun nilai ekspor dihitung berdasarkan enam kode HS yaitu 09041110, 09041120, 09041190, 09041210, 09041220, 09041290. Pemilihan variabel harga ekspor dilakukan dengan pertimbangan keterbatasan ketersediaan series data harga lada di tingkat domestik (harga produsen maupun harga konsumen). Selain itu, digunakan juga harga pupuk di pasar dunia yang bersumber dari World Bank. Series data yang digunakan pada penelitian ini yaitu data tahun 1976-2019, dimana data tahun 2019 merupakan Angka

Sementara (ASEM) sedangkan sisanya merupakan Angka Tetap (ATAP). Berdasarkan series data tersebut, selanjutnya dilakukan pengelompokan data training untuk periode 1976-2013 dan data testing untuk periode 2014-2019, sehingga diperoleh total observasi sebanyak 44. Berdasarkan model terbaik yang terpilih, estimasi luas areal lada di Indonesia dilakukan untuk lima tahun ke depan yaitu 2020-2024.

**Analisis Data**

Secara empiris, penelitian ini membandingkan hasil estimasi luas areal lada dengan empat model yaitu ARIMA, regresi, fungsi transfer dan VAR. Tahapan penelitian dimulai dengan mencari model estimasi berdasarkan historis data training untuk meramalkan data testing. Selanjutnya hasil estimasi data testing dibandingkan dengan nilai aktual luas arealnya untuk mengetahui tingkat kesalahan berdasarkan nilai MAPE. Berdasarkan nilai MAPE yang dihasilkan oleh keempat model estimasi tersebut dipilih model ramalan dengan MAPE terkecil. Model dengan MAPE terkecil dipilih sebagai model terbaik untuk mengestimasi luas areal lada lima tahun ke depan. Model terbaik yang terpilih juga harus memenuhi asumsi statistik yang ditetapkan di masing-masing model. Pengolahan data untuk estimasi luas areal lada baik dengan model ARIMA, regresi, fungsi transfer maupun VAR dilakukan dengan *software* RStudio.

**Estimasi dengan Model ARIMA**

Model ARIMA dibagi ke dalam tiga kelompok model yaitu *Autoregressive Model (AR)*, *Moving Average Model (MA)* dan *Autoregressive Integrated Moving Average Model (ARIMA)*. Model AR menjelaskan pergerakan suatu peubah itu sendiri di masa lalu. Model AR ordo ke-*p* untuk mengestimasi luas areal lada atau dapat ditulis ARIMA (*p*, 0, 0) sebagai berikut:

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + \varepsilon_t \dots\dots\dots(1)$$

dimana:

$Y_t$  = luas areal lada pada waktu ke-*t*

$Y_{t-p}$  = luas areal lada pada kurun waktu ke (*t-p*)

$\mu$  = suatu konstanta

$\theta_1 \dots \theta_p$  = parameter autoregressive ke-*p*

$\varepsilon_t$  = nilai kesalahan pada waktu ke-*t*

Model MA menjelaskan pergerakan peubahnya melalui sisaannya di masa lalu. Model MA dengan ordo *q* untuk mengestimasi luas areal lada atau ARIMA (0,0,*q*) ditulis sebagai berikut:

$$Y_t = \mu - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t \dots \dots \dots (2)$$

dimana:

$Y_t$  = luas areal lada pada waktu ke- $t$

$\phi_1 \dots \phi_q$  = parameter-parameter moving average

$\varepsilon_{t-q}$  = nilai kesalahan pada waktu ke ( $t-q$ )

Model ARIMA merupakan model dari fungsi linear nilai lampau beserta nilai sekarang dan sisaan lampauya. Bentuk model ARIMA ( $p,d,q$ ) untuk mengestimasi luas areal lada ditulis sebagai berikut:

$$Y_t = \mu + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_q \varepsilon_{t-q} + \varepsilon_t \dots \dots \dots (3)$$

dimana:

$Y_t$  = luas areal lada pada waktu ke- $t$

$Y_{t-p}$  = luas areal lada pada kurun waktu ke ( $t-p$ )

$\mu$  = suatu konstanta

$\theta_1 \theta_2 \dots \theta_p \phi_1 \phi_2 \dots \phi_q$  = parameter-parameter model

$\varepsilon_{t-q}$  = nilai sisaan pada waktu ke- $(t-q)$

Penggunaan model ARIMA mensyaratkan series data yang stasioner. Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan *differencing*. *Differencing* yaitu menghitung perubahan atau selisih nilai observasi. Data yang telah dilakukan *differencing* perlu dicek kembali apakah telah stasioner atau belum. Pengecekan stasioneritas data dapat dilihat dari beberapa cara antara lain melihat sebaran data, menggunakan uji *Augmented Dickey-Fuller Test (ADF) Test Unit Root Test* dan melihat dari perilaku autokorelasi berdasarkan plot *Autocorrelation Function (ACF)* dan *Partial Autocorrelation Function (PACF)*.

Berdasarkan sebaran datanya, data yang telah stasioner menyebar secara acak dan tidak memiliki pola-pola tertentu baik pola musiman maupun *trend*. Pengecekan stasioneritas dengan uji ADF memiliki hipotesis sebagai berikut:

Hipotesis: .....(4)

$H_0$  : Data tidak stasioner

$H_1$  : Data stasioner

Jika nilai *test-statistic* pada uji ADF lebih kecil dari *critical value for test-statistic* baik pada taraf ( $\alpha$ ) 1%, 5% atau 10% maka  $H_0$  ditolak yang berarti data telah stasioner. Pengecekan stasioneritas dari perilaku *autokorelasi* dilihat dari plot ACF dan PACF. Jika pada kedua plot tersebut tidak banyak lag yang keluar dari *confidence interval* maka data telah stasioner.

Data yang telah stasioner selanjutnya dilakukan tahapan pendugaan model ARIMA menggunakan fungsi *auto.arima* atau *armaselect* yang tersedia pada software RStudio. Software tersebut akan memberikan rekomendasi model terbaik untuk mengestimasi luas areal lada. Berdasarkan model terbaik yang terpilih, kemudian dilakukan pemeriksaan sisaan menggunakan pengujian LJungBox. Jika autokorelasi sisaan tidak signifikan yang ditandai dengan nilai p-value yang lebih besar dari 5% atau 10%, maka model ARIMA tersebut sudah cukup baik untuk mengemas data luas areal lada.

Model ARIMA yang terpilih digunakan untuk mengestimasi data testing. Hasil ramalan data testing selanjutnya dibandingkan dengan data aktualnya untuk mengecek akurasi hasil ramalan. Akurasi hasil ramalan model ARIMA ditunjukkan oleh MAPE data training dan data testing. Jika model terpilih dirasa telah menghasilkan MAPE yang kecil, maka model tersebut dapat digunakan untuk mengestimasi luas areal lada untuk beberapa periode ke depan.

### Estimasi dengan Model Regresi Linear Sederhana dan Regresi Kuadratik

Model regresi linear sederhana dan regresi kuadratik pada penelitian ini menggunakan peubah prediktor waktu ( $t$ ) sebagai variabel independen. Model regresi linear sederhana untuk mengestimasi luas areal lada dituliskan sebagai berikut:

$$y = \beta_0 + \beta_1 t + \varepsilon \dots\dots\dots(5)$$

dimana:

$y$  = luas areal lada

$t$  = tahun

$\beta_0$  = konstanta/intercept

$\beta_1$  = koefisien regresi variabel tahun

$\varepsilon$  = error

Parameter  $\beta_1$  mengukur efek tahun terhadap rata-rata perubahan luas areal lada. Parameter ini diprediksi dari data sampel.  $\varepsilon$  menggambarkan kesalahan statistik, didefinisikan sebagai sisa atau perbedaan hasil antara nilai data pengamatan variabel dependen terhadap nilai hasil prediksi.

Model regresi kuadratik yang digunakan untuk meramalkan luas areal lada dituliskan sebagai berikut:

$$y = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon \dots\dots\dots(6)$$

dimana:

- $y$  = luas areal lada
- $t$  = tahun
- $t^2$  = kuadrat dari tahun
- $\beta_0$  = konstanta/intercept
- $\beta_1$  = koefisien regresi variabel tahun
- $\beta_2$  = koefisien regresi variabel kuadrat tahun
- $\varepsilon$  = error

Dampak perubahan tahun terhadap rata-rata perubahan luas areal lada ditunjukkan oleh parameter  $\beta_1$ . Parameter  $\beta_2$  menunjukkan efek perubahan kuadrat tahun terhadap rata-rata perubahan luas areal lada.

Estimasi dengan model regresi linear sederhana dan regresi kuadratik mendasarkan pada beberapa asumsi antara lain:

**- Sisaan mengikuti fungsi distribusi normal**

Uji normalitas dilakukan dengan *Jarque-Bera (JB) test for normality*. Hipotesis yang mendasari yaitu:

Hipotesis: .....(7)

$H_0$  : Sisaan terdistribusi normal

$H_1$  : *Otherwise*

Jika nilai p-value hasil JB test lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi normalitas terpenuhi.

**- Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas)**

Asumsi homoskedastisitas diuji menggunakan *Non-constant Variance (NCV) Score Test* dengan hipotesis sebagai berikut:

Hipotesis: .....(8)

$H_0$  : Homoskedastisitas

$H_1$  : Heterokedastisitas

Jika nilai p-value pada NCV test lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi homoskedastisitas terpenuhi.

**- Tidak terdapat autokorelasi antar sisaan untuk setiap data pengamatan**

Ada tidaknya autokorelasi antar sisaan pada data amatan diuji berdasarkan nilai autokorelasi *D-W Statistic* dengan hipotesis sebagai berikut:

Hipotesis: .....(9)

$H_0$  : Tidak ada autokorelasi antar sisaan

$H_1$  : *Otherwise*

Jika nilai p-value *autocorrelation D-W Statistic* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi tidak adanya autokorelasi antar sisaan telah terpenuhi.

**Estimasi dengan Model Fungsi Transfer**

Perkembangan luas areal lada di Indonesia tidak terlepas dari pengaruh variabel lain. Menurut Sitorus & Mauludi (1992) salah satu variabel yang berpengaruh terhadap luas areal lada di Indonesia adalah harga lada. Penelitian ini menggunakan harga ekspor lada sebagai variabel input dalam mengestimasi luas areal (variabel output) menggunakan model fungsi transfer. Harga ekspor dipilih dengan pertimbangan ketersediaan series data yang panjang, mengingat series data harga domestik yang tersedia cukup pendek sehingga tidak mendukung untuk pembentukan model estimasi. Harga ekspor dihitung dari nilai ekspor dibagi volume ekspor lada. Model fungsi transfer pada penelitian ini menggambarkan ramalan luas areal lada yang didasarkan pada nilai-nilai masa lalu dari luas areal itu sendiri, serta didasarkan pada harga ekspor (variabel input) dan gangguan/noise. Model fungsi transfer untuk mengestimasi luas areal lada dituliskan sebagai berikut:

$$y_t = \nu(B)x_t + N_t \quad \longrightarrow \quad y_t = \frac{\omega_s(B)}{\delta_r(B)} x_{t-b} + \frac{\theta_q(B)}{\phi_p(B)} \varepsilon_t \quad \dots(10)$$

dimana:

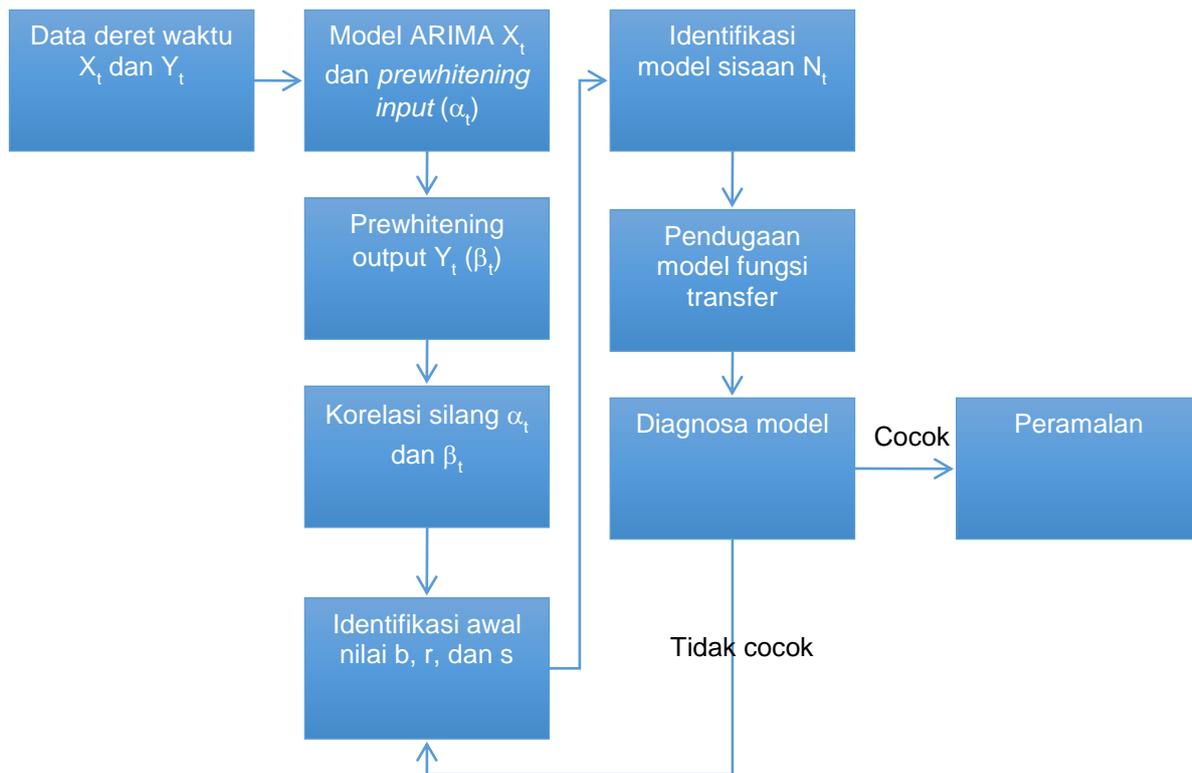
- $y_t$  = luas areal lada tahun ke-t
- $x_t$  = harga ekspor lada tahun ke-t
- $b$  = panjang jeda pengaruh harga ekspor terhadap luas areal lada
- $r$  = panjang lag luas areal lada periode sebelumnya yang masih mempengaruhi luas areal lada tahun-t
- $s$  = panjang jeda harga ekspor lada

periode sebelumnya yang masih  
mempengaruhi luas areal lada tahun-t

$p$  = ordo AR bagi noise  $N_t$

$q$  = ordo MA bagi noise  $N_t$

Pemodelan untuk estimasi luas areal lada dengan fungsi transfer dilakukan melalui beberapa langkah sebagaimana disajikan pada Gambar 2 berikut:



Gambar 2. Langkah-langkah Estimasi Luas Areal Lada dengan Fungsi Transfer

### Estimasi dengan Model VAR

Pemodelan dengan *Vector Autoregression (VAR)* memperlakukan seluruh peubah secara simetris tanpa mempermasalahkan apakah variabel tersebut merupakan variabel dependen maupun independen. Penelitian ini menggunakan luas areal lada sebagai variabel dependen. Luas areal lada tersebut diduga dipengaruhi oleh beberapa variabel, tidak seperti model fungsi transfer yang hanya menggunakan harga ekspor saja sebagai variabel inputnya. Selain harga ekspor (Sitorus & Mauludi, 1992), variabel input lain yang diduga juga berpengaruh terhadap luas areal lada di Indonesia antara lain produksi (Fatma et al., 2020; Nursalam, 2020), dan harga (Asnawi et al., 2017; Fatma et al., 2020; Hamdani et al., 2015; Nursalam, 2020; Sahara et al., 2004).

Harga ekspor digunakan sebagai pendekatan harga lada dengan pertimbangan keterbatasan series data harga lada baik di tingkat produsen maupun konsumen. Berdasarkan beberapa penelitian, produksi lada dipengaruhi oleh luas areal. Mengingat model VAR memperlakukan seluruh peubah secara simetris tanpa mempermasalahkan variabel dependen dan independen atau dengan kata lain seluruh variabel diperlakukan sebagai variabel dependen/endogen, maka produksi dipilih sebagai salah satu variabel untuk mengestimasi luas areal lada. Harga pupuk pada penelitian ini menggunakan harga pupuk dunia dikarenakan keterbatasan series data harga pupuk domestik. Harga pupuk yang digunakan adalah UREA karena harga pupuk NPK tidak tersedia. Meskipun umumnya pupuk yang digunakan untuk tanaman lada adalah NPK, UREA merupakan salah satu penyusun komposisi pupuk NPK sebagai unsur sumber hara N (Daras et al., 2012).

Model VAR termasuk kategori model sistem, dimana ketika tidak ada kepastian untuk menentukan bahwa suatu peubah adalah eksogen (independen) maka suatu perluasan analisis fungsi perpindahan alami akan memperlakukan masing-masing peubah secara simetris (Enders, 2004). Sebagai contoh, pada kasus-kasus peubah yang membiarkan alur waktu atau *time path*  $\{s_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{y_t\}$  dan membiarkan *time path*  $\{y_t\}$  dipengaruhi oleh nilai saat ini dan waktu sebelumnya dari  $\{s_t\}$ . Di dalam sistem *bivariate*, hubungan tersebut dapat digambarkan seperti pada persamaan berikut:

$$\begin{aligned} s_t &= b_{10} - b_{12}y_t + \gamma_{11}s_{t-1} + \gamma_{12}y_{t-1} + \varepsilon_{s_t} \\ y_t &= b_{20} - b_{21}s_t + \gamma_{21}s_{t-1} + \gamma_{22}y_{t-1} + \varepsilon_{y_t} \end{aligned} \quad \dots(11)$$

Dengan mengasumsikan bahwa kedua peubah  $s_t$  dan  $y_t$  adalah stasioner:  $\varepsilon_{s_t}$  dan  $\varepsilon_{y_t}$  adalah *disturbances* yang memiliki rata-rata nol dan matriks kovarians terbatas atau bersifat *white noise* dengan standar deviasi yang berurutan  $\sigma_s$  dan  $\sigma_y$ : serta  $\{\varepsilon_{s_t}\}$  dan  $\{\varepsilon_{y_t}\}$  adalah *disturbances* yang independen dengan rata-rata nol dan kovarian terbatas (*uncorrelated white-noise disturbances*). Kedua persamaan di atas merupakan orde pertama VAR, karena panjang *lag* nya hanya satu. Agar Persamaan (11) lebih mudah dipahami dan digunakan sebagai alat analisis maka ditransformasikan dengan menggunakan matriks aljabar, dan hasilnya dapat dituliskan secara bersama seperti pada persamaan di bawah ini:

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} s_t \\ y_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} s_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix} \text{ atau dengan bentuk lain:}$$

$$Bx_t = \Gamma_0 + \Gamma_1 X_{t-1} + \varepsilon_t \dots \dots \dots (12)$$

dimana:

$$B = \begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \quad x_t = \begin{bmatrix} s_t \\ y_t \end{bmatrix} \quad \Gamma_0 = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} \quad \Gamma_1 = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{s_t} \\ \varepsilon_{y_t} \end{bmatrix}$$

Dengan melakukan pengalihan antara persamaan (12) dengan  $B^{-1}$  atau invers matriks B, maka akan dapat ditentukan model VAR dalam bentuk standar, seperti dituliskan pada persamaan di bawah ini:

$$x_t = A_0 + A_1 x_{t-1} + \ell_t \dots \dots \dots (13)$$

dimana:

$$A_0 = B^{-1} \Gamma_0$$

$$A_1 = B^{-1} \Gamma_1$$

$$\ell_t = B^{-1} \varepsilon_t$$

Pada penelitian ini  $X_t$  merupakan matriks yang dibentuk dari luas areal lada, produksi, harga ekspor dan harga pupuk.

Sebagaimana model regresi, model VAR juga didasarkan pada beberapa asumsi antara lain:

**- Sisaan mengikuti fungsi distribusi normal**

Uji normalitas pada model VAR didasarkan pada nilai *Jarque-Bera (JB) test (multivariate)*, *Skewness only (multivariate)*, dan *Kurtosis only (multivariate)*. Hipotesis yang mendasari yaitu:

Hipotesis: .....(14)

$H_0$  : Sisaan terdistribusi normal

$H_1$  : *Otherwise*

Jika nilai p-value dari *JB test*, *Skewness only* dan *Kurtosis only* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi normalitas terpenuhi.

**- Varians sisaan konstan untuk setiap data pengamatan (homoskedastisitas)**

Asumsi homoskedastisitas pada model VAR didasarkan pada nilai *ARCH (multivariate)* dengan hipotesis sebagai berikut:

Hipotesis: .....(15)

$H_0$  : Homoskedastisitas

$H_1$  : Heterokedastisitas

Jika nilai p-value pada *ARCH (multivariate)* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi homoskedastisitas terpenuhi.

**- Tidak terdapat autokorelasi antar sisaan untuk setiap data pengamatan**

*Pormanteau Test (asymptotic)* digunakan untuk menguji ada tidaknya autokorelasi antar sisaan pada data amatan. Hipotesis yang digunakan sebagai berikut:

Hipotesis: .....(16)

$H_0$  : Tidak ada autokorelasi antar sisaan

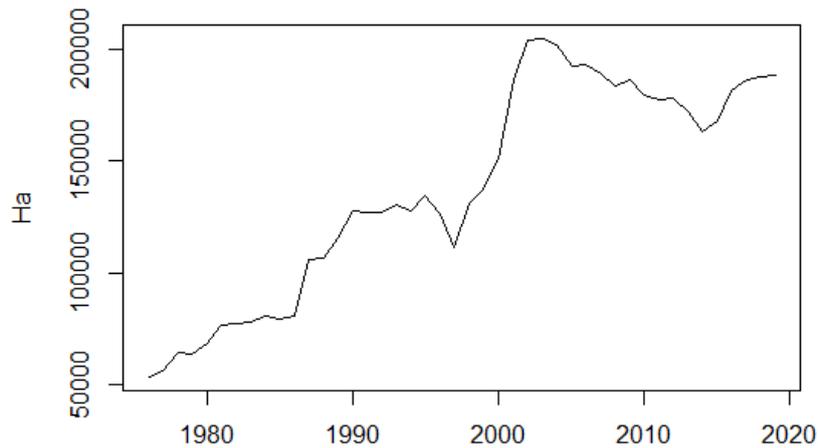
$H_1$  : *Otherwise*

Jika nilai p-value hasil uji *Pormanteau Test* lebih besar dari nilai  $\alpha$  maka  $H_0$  diterima, dengan kata lain asumsi tidak adanya autokorelasi antar sisaan telah terpenuhi.

## HASIL DAN PEMBAHASAN

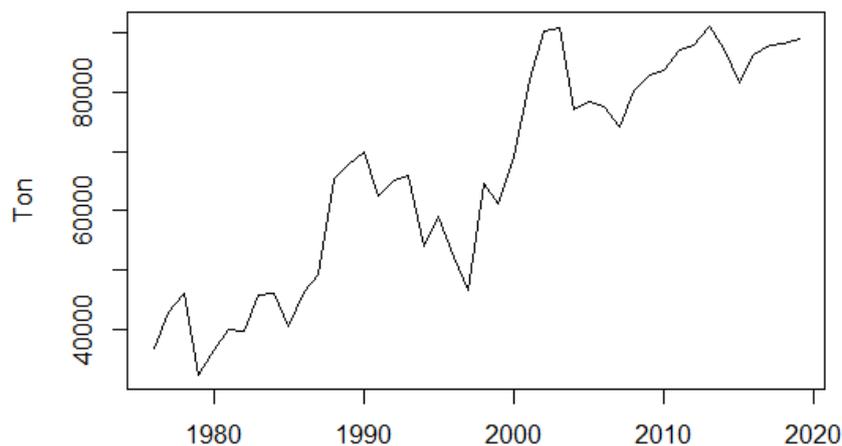
### Keragaan Lada di Indonesia

Lada merupakan salah satu komoditas rempah unggulan sub sektor perkebunan. Pengembangan usahatani lada di Indonesia sebagian besar berada di Luar Jawa seperti Kepulauan Bangka Belitung, Lampung, Sumatera Selatan, Sulawesi Selatan dan Kalimantan Timur. Berdasarkan data Kementerian Pertanian (2019) kelima provinsi tersebut berkontribusi sebesar 65% terhadap luas areal lada di Indonesia. Jenis lada yang paling banyak diproduksi di Indonesia adalah lada hitam yang berasal dari Lampung atau dikenal dengan nama *Lampung black pepper* dan lada putih atau *Muntok white pepper* yang dihasilkan di Kepulauan Bangka Belitung (Nurdjannah, 2006). Perkembangan luas areal lada dari tahun 1976 hingga tahun 2019 cenderung meningkat (Gambar 3). Pada tahun 1976 luas areal lada yang meliputi penjumlahan dari TBM, TM maupun TR/TTM sebesar 53.817 ha kemudian meningkat menjadi 188.041 ha pada tahun 2019. Selama sepuluh tahun terakhir (2010-2019) peningkatan luas areal lada tercatat sebesar 0,18% per tahun, dengan luas areal tertinggi dicapai pada tahun 2019.



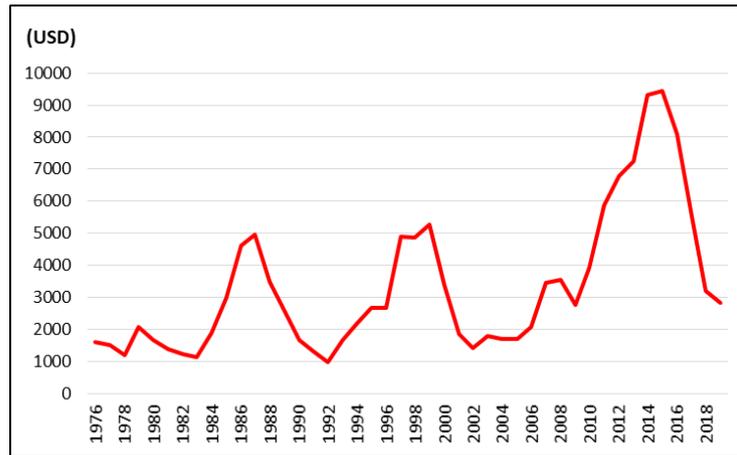
Gambar 3. Perkembangan Luas Areal Lada di Indonesia Tahun 1976-2019

Peningkatan produksi lada seiring dengan perkembangan luas areal yang terus bertambah dari tahun ke tahun (Gambar 4). Pada tahun 1976 produksi lada sebesar 36.854 ton kemudian mengalami peningkatan menjadi 88.949 ton pada tahun 2019. Dengan rata-rata pertumbuhan produksi sebesar 0,78% per tahun selama sepuluh tahun terakhir, upaya untuk meningkatkan produksi lada masih memungkinkan untuk dilakukan. Produksi lada di Indonesia sendiri didominasi Perkebunan Rakyat (PR) diikuti oleh Perkebunan Besar Swasta (PBS).



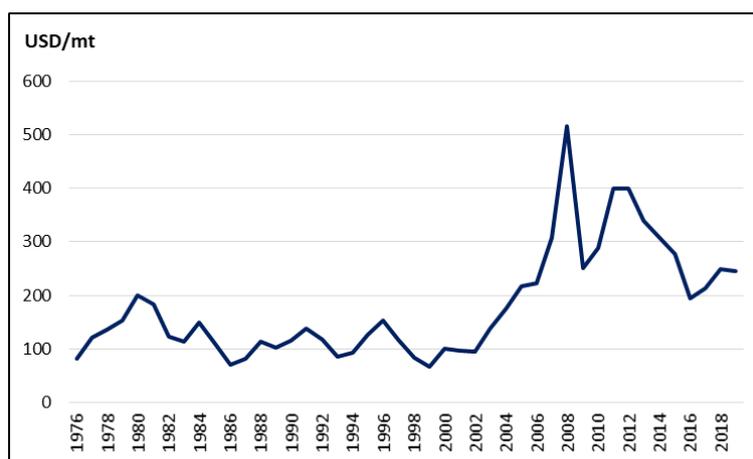
Gambar 4. Perkembangan Produksi Lada di Indonesia Tahun 1976-2019

Harga ekspor merupakan pendekatan terhadap harga lada domestik yang ketersediaan series datanya terbatas. Harga ekspor dihitung dari nilai ekspor dibagi volume ekspor. Harga ekspor lada selama periode 1976-2019 cenderung fluktuatif (Gambar 6). Harga ekspor lada tahun 1976 sebesar 1,6 ribu USD dan meningkat menjadi 2,8 ribu USD pada tahun 2019. Pada tahun 2015 harga lada mencapai nilai tertinggi yaitu 9,4 ribu USD.



Gambar 5. Perkembangan Harga Ekspor Lada di Indonesia Tahun 1976-2019

Harga pupuk dunia merupakan pendekatan terhadap harga pupuk domestik yang ketersediaannya terbatas. Harga pupuk dunia yang digunakan adalah harga UREA. Umumnya, jenis pupuk yang paling banyak digunakan pada tanaman lada adalah NPK. Menurut Daras et al. (2012) UREA merupakan salah satu penyusun komposisi pupuk NPK sebagai unsur sumber hara N. Oleh karena itu, harga UREA dipilih sebagai pendekatan harga pupuk. Harga pupuk selama periode 1976-2019 cenderung fluktuatif (Gambar 6). Harga pupuk pada tahun 1976 sebesar 82 USD/mt dan meningkat menjadi 245 USD/mt pada tahun 2019. Pada tahun 2008 harga pupuk mencapai nilai tertinggi yaitu 515 USD/mt.

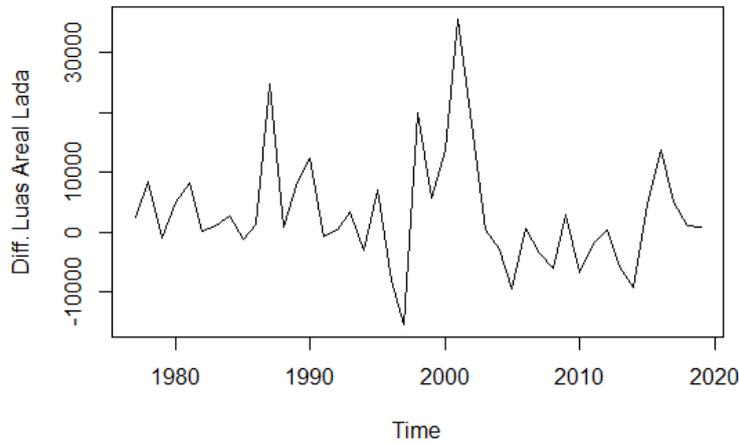


Gambar 6. Perkembangan Harga Pupuk di Dunia Tahun 1976-2019

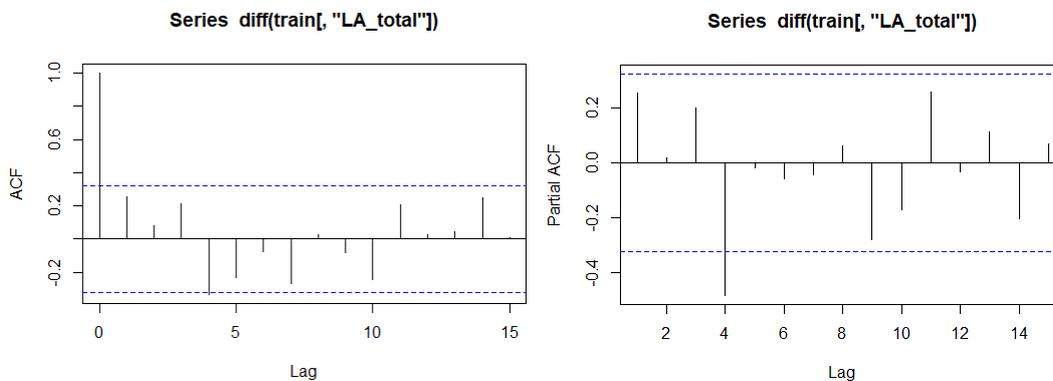
### Hasil Estimasi dengan Model ARIMA

Pemodelan ARIMA dilakukan untuk data yang telah stasioner. Karena berdasarkan plot data serta hasil uji ADF terindikasi bahwa data luas areal lada belum stasioner, sehingga dilakukan *differencing* satu kali. Setelah *differencing* data tersebut menjadi stasioner (Gambar 7) yang ditunjukkan dengan nilai *test-statistic* -3,57, dimana nilai tersebut lebih kecil dari

*critical value* baik untuk  $\alpha$  sebesar 1%, 5% maupun 10%. Stasioneritas data tersebut juga didukung dengan plot ACF dan PACF, dimana pada plot tersebut tidak banyak lag yang keluar dari *confidence interval* (Gambar 8).



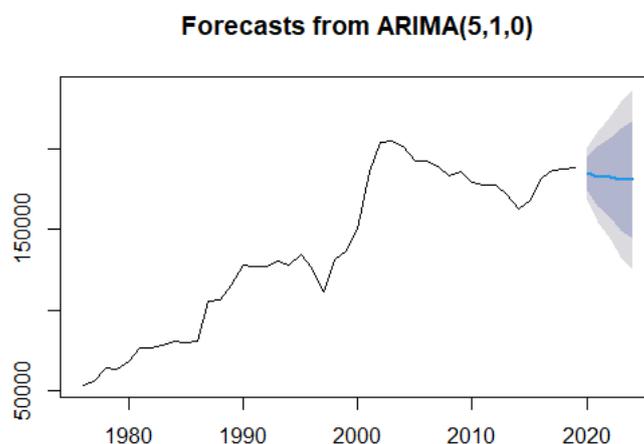
Gambar 7. Plot Luas Areal Lada Setelah *Differencing*



Gambar 8. Plot ACF dan PACF Setelah Dilakukan *Differencing* pada Data Luas Areal Lada

Dengan menggunakan software RStudio, diperoleh model terbaik untuk mengesimasi luas areal lada yaitu ARIMA (5,1,0). Model ini menghasilkan MAPE data training sebesar 4,60 dan MAPE data testing sebesar 6,46. Berdasarkan hasil pemeriksaan sisaan dengan uji LJungBox, ditemukan bahwa autokorelasi sisaan tidak signifikan pada seluruh lag yang ditandai dengan nilai p-value yang lebih besar dari 5%. Hal ini menunjukkan bahwa model ARIMA (5,1,0) sudah cukup baik untuk mengepas data luas areal lada. Selanjutnya dengan model ARIMA (5,1,0) dilakukan estimasi luas areal lada periode 2020-2024. Data Kementerian Pertanian (2019) menunjukkan bahwa pada tahun 2019 (ASEM) luas areal lada sebesar 188.041 ha. Hasil estimasi lada dengan ARIMA (5,1,0) menghasilkan ramalan luas areal pada tahun 2020 sebesar 184.691 ha kemudian turun menjadi 182.505 ha pada tahun

2021. Tahun 2022 estimasi luas areal lada sebesar 182.138 ha kemudian pada tahun 2023 turun menjadi 180.815 ha. Luas areal lada diramalkan sebesar 180.932 ha pada tahun 2024. Hasil estimasi luas areal lada disajikan pada Gambar 9.



Gambar 9. Hasil Estimasi Luas Areal Lada Tahun 2020-2024

### Hasil Estimasi dengan Model Regresi Linear Sederhana dan Regresi Kuadratik

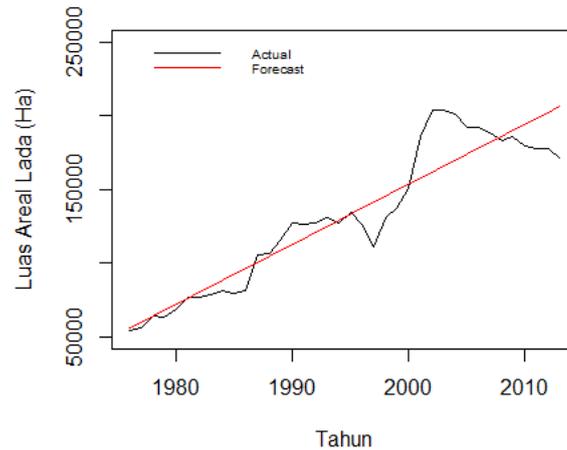
Model regresi linear sederhana untuk mengestimasi luas areal lada menghasilkan persamaan regresi sebagai berikut:

$$Luas = -8.010.799 + 4.082 t + \varepsilon \dots \dots \dots (17)$$

Secara parsial , intercept dan tahun masing-masing berkorelasi signifikan terhadap luas areal lada yang ditunjukkan oleh nilai *p-value* lebih kecil dari 1%. Dengan *p-value* model yang juga lebih kecil dari 1% menunjukkan bahwa secara umum model regresi linear tersebut telah signifikan untuk mengestimasi luas areal lada.  $R^2$  yang dihasilkan pada model ini sebesar 0,88 menunjukkan bahwa 88% luas areal lada di Indonesia diduga dipengaruhi oleh tahun, sedangkan 12% lainnya dipengaruhi oleh variabel lain yang tidak masuk dalam model. Model (17) menunjukkan bahwa penambahan tahun sebanyak satu unit satuan akan meningkatkan luas areal lada sebesar 4.082 ha. Dengan MAPE data training 8,68 dan MAPE data testing 23,72 model ini meramalkan luas areal lada memiliki kecenderungan meningkat sebagaimana disajikan pada Gambar 10. Luas areal lada tahun 2020 diramalkan sebesar 235.272 ha dan meningkat hingga mencapai 251.600 ha pada tahun 2024.

Berdasarkan hasil pengujian asumsi pada Hipotesis (7), (8) dan (9) ditemukan bahwa sisaan mengikuti fungsi distribusi normal yang ditunjukkan dengan nilai *p-value* hasil uji *JB test* sebesar 0,18. Asumsi homoskedastisitas pada model ini juga terpenuhi dimana *p-value* dari uji NCV sebesar 0,00 yang menunjukkan bahwa terdapat heteroskedastisitas. Selain itu terdapat autokorelasi antar sisaan yang ditunjukkan dengan nilai *p-value* dari *autocorrelation D-W Statistic* lebih kecil dari  $\alpha$ . Mengingat tidak semua asumsi terpenuhi, maka model regresi

linear sederhana tidak direkomendasikan untuk dipilih sebagai model terbaik dalam mengestimasi luas areal lada.



Gambar 10. Hasil Estimasi Luas Areal Lada dengan Model Regresi Linear Sederhana

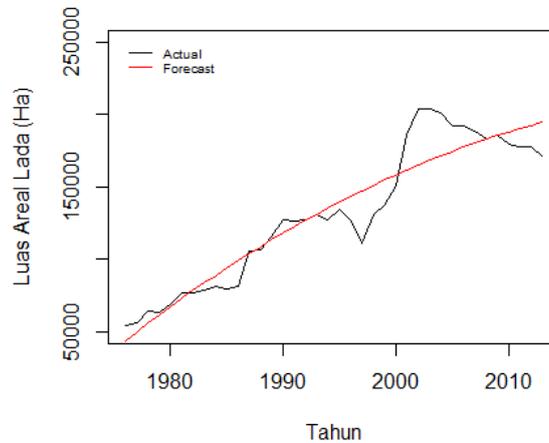
Model regresi kuadratik untuk mengestimasi luas areal lada menghasilkan persamaan regresi sebagai berikut:

$$Luas = 43.823 + 6.054 t - 53t^2 + \epsilon... (18)$$

Secara parsial, intercept, tahun dan kuadrat tahun masing-masing berkorelasi signifikan terhadap luas areal lada yang ditunjukkan oleh nilai *p-value* lebih kecil dari 10%. Dengan *p-value* model yang juga lebih kecil dari  $\alpha$  menunjukkan bahwa secara umum model regresi kuadratik telah signifikan untuk mengestimasi luas areal lada. *Adjusted-R<sup>2</sup>* yang dihasilkan pada model ini sebesar 0,89 menunjukkan bahwa 89% luas areal lada di Indonesia diduga dipengaruhi oleh tahun dan kuadrat tahun, sedangkan 11% lainnya dipengaruhi oleh variabel lain yang tidak masuk dalam model. Model (18) menunjukkan bahwa penambahan tahun sebanyak satu unit satuan akan meningkatkan luas areal lada sebesar 6.054 ha, sedangkan penambahan kuadrat tahun sebanyak satu unit satuan akan menurunkan luas areal lada 53 ha. Dengan MAPE data training 8,90 dan MAPE data testing 12,88 model ini meramalkan luas areal lada memiliki kecenderungan meningkat sebagaimana disajikan pada Gambar 11. Luas areal lada tahun 2020 diramalkan sebesar 211.632 ha dan meningkat hingga mencapai 214.531 ha pada tahun 2024.

Hasil pengujian asumsi pada Hipotesis (7), (8) dan (9) ditemukan bahwa sisaan mengikuti fungsi distribusi normal yang ditunjukkan dengan nilai *p-value* hasil uji *JB test* sebesar 0,31. Namun asumsi homoskedastisitas pada model ini tidak terpenuhi dimana *p-value* dari uji NCV sebesar 0,01 atau lebih kecil dari  $\alpha$ . Selain itu, terdapat asumsi lain yang dilanggar yaitu ditemukan autokorelasi antar sisaan yang ditunjukkan dengan nilai *p-value* dari *autocorrelation D-W Statistic* lebih kecil dari  $\alpha$ . Mengingat banyak asumsi yang terlanggar,

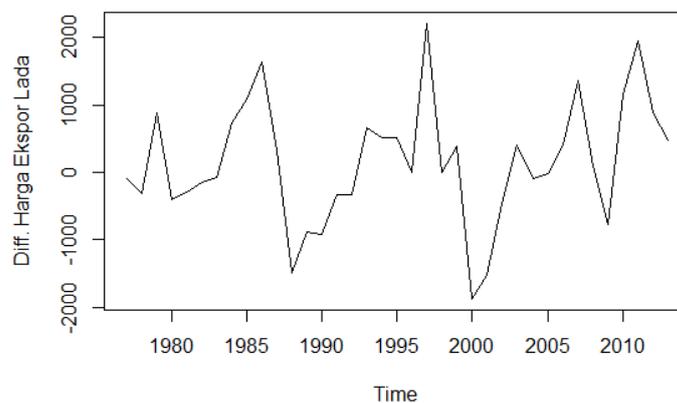
maka model regresi kuadratik juga tidak direkomendasikan untuk dipilih sebagai model terbaik dalam mengestimasi luas areal lada.



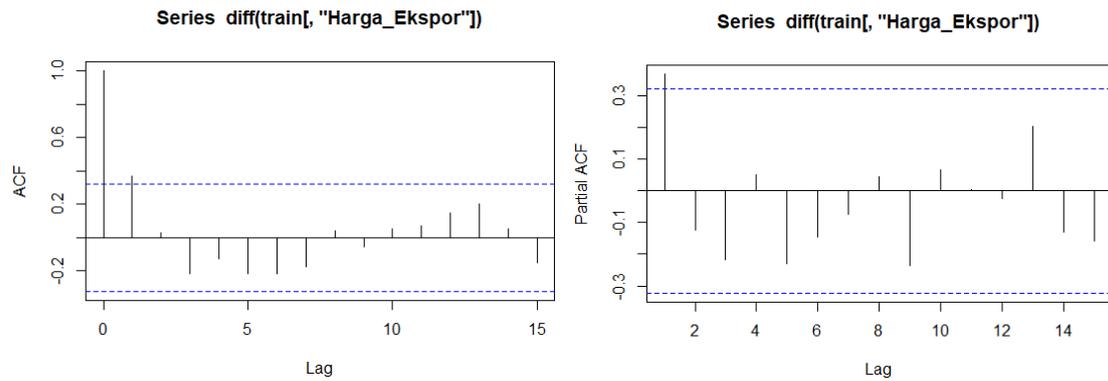
Gambar 11. Hasil Estimasi Luas Areal Lada dengan Model Regresi Kuadratik

### Hasil Estimasi dengan Model Fungsi Transfer

Estimasi dengan model fungsi transfer diawali dengan meramalkan nilai variabel input yaitu harga ekspor menggunakan model ARIMA. Dikarenakan data harga ekspor belum stasioner maka dilakukan *differencing* satu kali sehingga menghasilkan data yang telah stasioner (Gambar 12) yang ditunjukkan dengan nilai *test-statistic*  $-3,52$ , dimana nilai tersebut lebih kecil dari *critical value* baik untuk  $\alpha$  sebesar 1%, 5% maupun 10%. Stasioneritas harga ekspor lada juga didukung dengan plot ACF dan PACF, dimana pada plot tersebut tidak banyak lag yang keluar dari *confidence interval* (Gambar 13). Model ARIMA terbaik untuk mengestimasi harga ekspor lada adalah ARIMA (1,1,0) dengan AIC model sebesar 607,5.



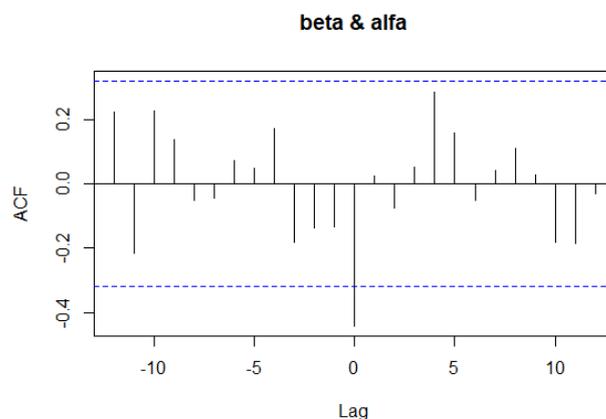
Gambar 12. Plot Harga Ekspor Lada Setelah *Differencing*



Gambar 13. Plot ACF dan PACF Setelah Dilakukan *Differencing* pada Data Harga Ekspor Lada

Tahapan selanjutnya dalam pemodelan fungsi transfer adalah prewhitening dan korelasi silang antar residual harga ekspor dengan residual luas areal lada yang menghasilkan plot ACF (Gambar 14). Berdasarkan plot ACF tersebut diperoleh nilai  $b=0$  dimana lag pertama kali signifikan pada lag 0. Interpretasi dari nilai  $b=0$  yaitu tidak ada jeda pengaruh dampak harga ekspor terhadap luas areal lada. Selain itu, diperoleh nilai  $s=0$  karena tidak ada tambahan lag yang signifikan setelah lag 0, yang berarti korelasi antara harga ekspor dengan luas areal terjadi di tahun yang sama. Nilai  $r$  diasumsikan 0 karena data luas areal maupun harga ekspor lada merupakan data tahunan yang tidak mengandung pola musiman.

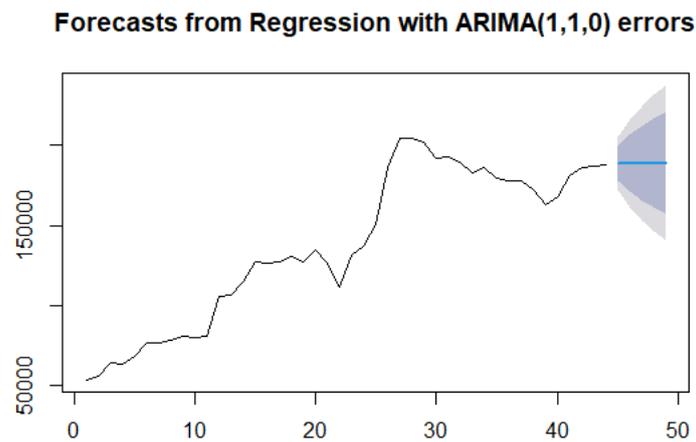
Setelah pengepasan model berdasarkan nilai  $r, s$ , dan  $b$ , dilakukan identifikasi model noise dengan memanfaatkan fungsi *auto.arima* pada RStudio. Model noise yang direkomendasikan yaitu ARIMA (1,1,0) dengan MAPE data training 4,93 dan MAPE data testing 1,83.



Gambar 14. Plot ACF Hasil Korelasi Silang Residual

Secara umum, model fungsi transfer terbaik untuk mengestimasi luas areal lada adalah fungsi transfer ARIMA (1,1,0) dimana variabel input (harga ekspor) terlebih dahulu

diestimasi dengan ARIMA (1,1,0). Estimasi luas areal lada lima tahun ke depan cenderung meningkat namun dengan pertumbuhan yang cukup kecil (Gambar 15). Hasil estimasi luas areal lada pada tahun 2020 sebesar 188.517 ha kemudian meningkat menjadi 188.773 ha pada tahun 2021. Tahun 2022 luas areal lada diramalkan naik menjadi 188.902 ha kemudian meningkat lagi menjadi 188.964 ha pada tahun 2023 dan 188.993 ha pada tahun 2024. Dengan nilai MAPE data testing yang cukup kecil, model fungsi transfer ini direkomendasikan sebagai salah satu model untuk mengestimasi luas areal lada beberapa periode ke depan.

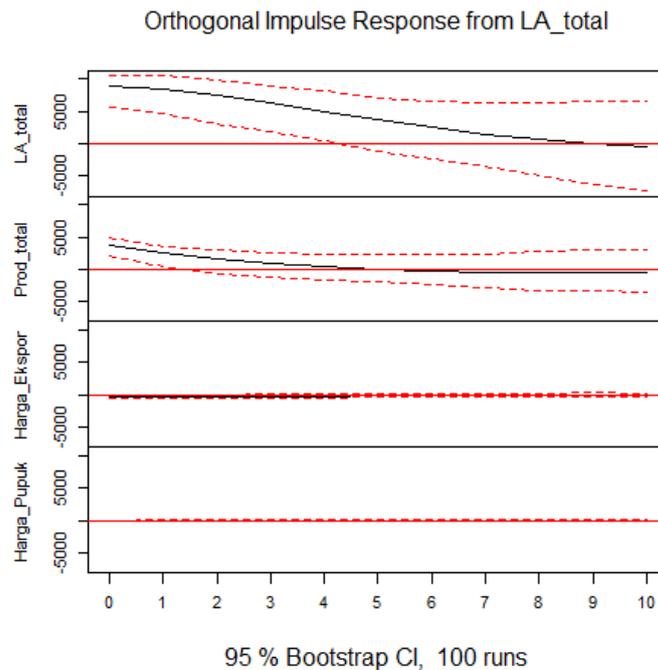


Gambar 15. Hasil Estimasi Luas Areal Lada dengan Model Fungsi Transfer

### Hasil Estimasi dengan Model VAR

Estimasi dengan model VAR diawali dengan pemilihan ordo- $p$  serta type model VAR yang paling baik. Model VAR terbaik ditandai dengan banyaknya variabel yang signifikan pada ordo yang telah ditentukan. Model VAR yang terpilih yaitu VAR dengan ordo-1 atau VAR(1) tanpa memasukkan konstanta maupun trend. Pada model VAR(1) beberapa variabel yang signifikan yaitu luas areal  $t-1$ , harga ekspor  $t-1$  dan harga pupuk  $t-1$ , sedangkan produksi  $t-1$  tidak signifikan.

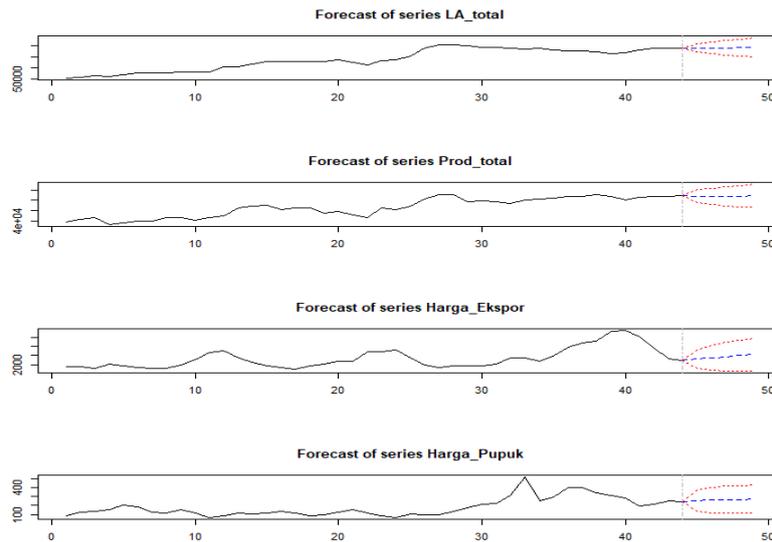
Salah satu kelebihan dari model VAR adalah dapat menampilkan *impulse respon* antar variabel. *Impulse respon* menggambarkan tingkat laju dari *shock* peubah yang satu terhadap peubah yang lainnya pada suatu rentang periode tertentu. Sehingga dapat dilihat lamanya pengaruh dari *shock* suatu peubah terhadap peubah lain sampai pengaruhnya hilang atau kembali ke titik keseimbangan. *Impulse respon* luas areal lada pada model VAR(1) ditunjukkan pada Gambar 16. Berdasarkan gambar tersebut terlihat bahwa perubahan luas areal berdampak pada luas areal itu sendiri sampai 4 tahun ke depan, kemudian dampaknya hilang setelah 4 tahun. Perubahan luas areal lada juga berdampak pada produksi, namun tidak berdampak pada harga ekspor dan harga pupuk.



Gambar 16. *Impulse Respon* Luas Areal pada Model VAR(1)

Model VAR(1) menghasilkan MAPE data training sebesar 4,88 dan MAPE data testing sebesar 23,69. Berdasarkan hasil pengujian asumsi pada Hipotesis (14), (15) dan (16) ditemukan bahwa sisaan tidak terdistribusi secara normal yang ditunjukkan dengan nilai *p-value* lebih kecil dari  $\alpha$  berdasarkan hasil uji *JB test*, *Skewness*, dan *Kurtosis*. Asumsi homoskedastisitas pada model ini terpenuhi dimana *p-value* pada ARCH sebesar 1. Selain itu, tidak terdapat autokorelasi antar sisaan yang ditunjukkan dengan nilai *p-value* dari *Portmanteu Test* sebesar 0,93. Mengingat fokus pemodelan pada penelitian ini adalah peramalan menggunakan series data yang cukup banyak, sehingga meskipun asumsi normalitas tidak terpenuhi, model VAR(1) dapat dipertimbangkan untuk digunakan dalam mengestimasi luas areal lada.

Berdasarkan model VAR(1), estimasi luas areal lada selama lima tahun ke depan diramalkan naik. Pada tahun 2020 luas areal lada diprediksi sebesar 188.611 ha dan terus meningkat hingga mencapai 191.413 ha pada tahun 2024. Selain ramalan luas areal lada, model VAR(1) ini juga meramalkan variabel input lain yaitu produksi, harga ekspor dan harga pupuk sebagaimana disajikan pada Gambar 17.



Gambar 17. Estimasi Variabel Input pada Model VAR(1)

### Pemilihan Model Estimasi Terbaik

Kriteria pemilihan model estimasi terbaik pada penelitian ini didasarkan pada nilai terkecil dari MAPE data testing serta kelogisan hasil ramalan jika dibandingkan dengan perkembangan luas areal lada beberapa tahun terakhir. Dari keempat model estimasi yang dibandingkan, diperoleh informasi bahwa model fungsi transfer memberikan MAPE data testing paling kecil dibandingkan model lain. Dengan luas areal lada pada tahun 2019 sebesar 188.041 ha (Kementerian Pertanian, 2019) dan rata-rata pertumbuhan luas areal selama lima tahun terakhir (2015-2019) sebesar 2,97%, maka luas areal lada pada periode 2020-2024 diramalkan naik 0,06%. Perbandingan hasil estimasi luas areal lada selama lima tahun ke depan dengan model ARIMA, regresi, fungsi transfer dan VAR disajikan pada Tabel 1.

Tabel 1. Perbandingan Hasil Estimasi Luas Areal Lada dengan Model ARIMA, Regresi, Fungsi Transfer dan VAR

No	Model	MAPE		Estimasi Luas Areal Lada (Ha)					Pertumb. (%)
		Train	Test	2020	2021	2022	2023	2024	
1	ARIMA(5,1,0)	4,60	6,46	184.691	182.505	182.138	180.815	180.932	-0,51
2	Regresi Linear	8,68	23,72	235.272	239.354	243.436	247.518	251.600	1,69
	Regresi Kuadratik	8,90	12,88	211.632	212.516	213.294	213.966	214.531	0,34
3	Fungsi Transfer ARIMA(1,1,0), xreg=harga ekspor	4,92	1,83	188.517	188.773	188.902	188.964	188.993	0,06
4	VAR(1) type=none	4,88	23,69	188.611	189.076	189.625	190.381	191.413	0,37

## KESIMPULAN DAN SARAN

### Kesimpulan

Model terbaik yang terpilih untuk mengestimasi luas areal lada adalah fungsi transfer ARIMA (1,1,0) dengan variabel inputnya harga ekspor. Pemilihan fungsi transfer sebagai model estimasi dengan pertimbangan MAPE data testing yang dihasilkan merupakan yang terkecil dibandingkan model lain. Hasil estimasi luas areal lada selama lima tahun ke depan naik sebesar 0,06% per tahun. Hal ini seiring dengan rata-rata pertumbuhan luas areal lada lima tahun terakhir yang juga mengalami peningkatan. Dengan kata lain, estimasi luas areal lada periode 2020-2024 masih logis jika dibandingkan dengan historis luas areal lada beberapa tahun ke belakang. Hasil estimasi dengan fungsi transfer menunjukkan pada tahun 2020 luas areal lada di Indonesia sebesar 188.517 ha kemudian meningkat hingga tahun 2024 menjadi 188.993 ha.

### Saran

Penelitian ini membatasi estimasi luas areal lada menggunakan empat model yaitu ARIMA, regresi, fungsi transfer dan VAR. Meskipun model fungsi transfer dipilih menjadi model terbaik untuk mengestimasi luas areal lada, akan lebih baik jika pada penelitian berikutnya mencoba model lain untuk mengestimasi luasan lada baik dengan model *univariate* atau *multivariate*. Selain itu, variabel lain selain yang telah digunakan pada penelitian ini dapat juga dicobakan untuk mengestimasi luas areal lada.

## DAFTAR PUSTAKA

- Asnawi, R., Zahara, & Arief, R.W. 2017. Pengaruh Pengelolaan Faktor Internal Usahatani Terhadap Produktivitas Lada di Provinsi Lampung. *Jurnal Littri*. 23(1): 1-10.
- Damanik, Sabarman. 2001. Analisis Penawaran dan Permintaan Lada Indonesia di Pasar Internasional. *Jurnal Litri*. 7(4):113-119.
- Daras, U., Sobari, I., & Towaha, J. Formulasi Pemupukan Berimbang pada Tanaman Lada di Bangka Belitung. *Buletin RISTRI*. 3(2): 185-192.
- Enders, W. 2004. *Applied Econometric Time Series*. USA: University of Alabama.
- Fatma, Hikmah, N., & Usman. 2020. Faktor-faktor yang Memengaruhi Produksi Lada di Desa Kongkomas Kecamatan Basidondo Kabupaten Tolitoli. *Jurnal Agrotech*. 10(1):35-40.
- Gujarati, D.N., & Porter, D.C. 2010. *Dasar-dasar Ekonometrika*. Jakarta: Salemba Empat, Buku 2, Edisi 5.
- Hamdani, Tety, E., & Eliza. 2015. Faktor-faktor yang Mempengaruhi Produksi Lada di Indonesia. *Jom Faperta*. 7(2):1-7.

- Kardinan, A., Laba, I.W., & Rismayani. 2018. Peningkatan Daya Saing Lada Melalui Budidaya Organik. *Perspektif*. 7(1):26-39.
- Kemala, Syarif. 2006. Strategi Pengembangan Sistem Agribisnis Lada untuk Meningkatkan Pendapatan Petani. *Perspektif*. 5(10): 48-54.
- Kementerian Pertanian. 2019. Statistik Perkebunan Indonesia 2018-2020 (Lada). Jakarta: Direktorat Jenderal Perkebunan Kementerian Pertanian.
- Nurdjannah, Nanan. 2006. Perbaikan Mutu Lada Dalam Rangka Meningkatkan Daya Saing di Pasar Dunia. *Perspektif*. 5(1): 13-25.
- Nursalam. 2020. Analisis Produksi dan Efisiensi Alokatif Usahatani Lada di Desa Ameroro Kecamatan Tinondo Kabupaten Kolaka Timur. *Agrimor*. 5(3): 57-59.
- Sahara, D., Yusuf, & Sahardi. 2004. Pengaruh Faktor Produksi pada Usahatani Lada di Sulawesi Tenggara (Kasus Integrasi Lada – Ternak di Kecamatan Landono, Kabupaten Kendari). *Jurnal Pengkajian dan Pengembangan Teknologi Pertanian*. 7(2):139-145.
- Sitorus, D.T., & Mauludi, L. 1992. Analisis Respon Penawaran Lada Putih di Indonesia. *Buletin Littro*. 7(2): 1-5.
- Zikria, Roydatul. 2019. Outlook Lada. Jakarta: Pusat Data dan Sistem Informasi Pertanian



# PENGEMBANGAN METODOLOGI ESTIMASI DATA LUAS AREAL CENGKEH 2020

Vera Junita Siagian, Fungsional Statistisi, Pusdatin Kementan

## ABSTRAK

Pada tahun 2020, Pusdatin bekerjasama dengan Ditjen Perkebunan dan BPS mengkaji 4 metode estimasi yang nantinya didapatkan metode yang lebih akurat, lebih objektif dan lebih baik secara statistik dari metode sebelumnya dan diharapkan hasil estimasinya cenderung mendekati angka aktualnya. Melalui kegiatan pengembangan metode estimasi data perkebunan tahun 2020, diharapkan akan dihasilkan metode yang paling sesuai untuk mengestimasi luas areal cengkeh nasional dengan membandingkan besaran MAPE (Mean Absolut Percentage Error) dan kerealistisan hasil permalan dengan data series sebelumnya

Metode yang diterapkan adalah ARIMA (*Autoregressive Integrated Moving Average*), Regresi, Fungsi Transfer dan VAR (*Vector Auto Reggression*) dengan menggunakan software R Studio

## ABSTRACT

*In 2020, Center for Agriculture Data and Informasi System (CADIS) together with the Directorate General of Plantation and the BPS-Statistic Indonesia examined 4 methods of estimation that will be obtained more accurately, more objectively and statistically better than the previous method and it is expected that the estimated results tend to be close to the actual number. Through the development of plantation data estimation method in 2020, it is expected that the most suitable method will be produced to estimate the national clove area by comparing the amount of MAPE (Mean Absolut Percentage Error) and the realism of the results of the permalan with the previous series data*

*. The methods applied are ARIMA (Autoregressive Integrated Moving Average), Regression, Transfer Function and VAR (Vector Auto Reggression) using R Studio software*

## V. PENDAHULUAN

### 1.1. Latar Belakang

Metode yang dilakukan selama ini untuk menyusun angka estimasi (AESTI) luas areal cengkeh nasional menurut Buku Pedoman Pelaksanaan Pengeolaan Data Komoditas Perkebunan (PDKP) adalah metode pemulusan eksponensial tunggal atau Single Exponential Smoothing (SES) dan pemulusan eksponensial ganda atau DSS (Double Exponential Smoothing). Untuk pemilihan model terbaik yaitu dengan membandingkan besaran MAPE (Mean Absolut Percentage Error) dan kerealistisan hasil permalan dibandingkan dengan data series sebelumnya. Menurut T. Hani Handoko (2011), *Exponential Smoothing* adalah suatu tipe teknik peramalan rata-rata bergerak yang melakukan penimbangan terhadap data masa lalu dengan cara

*eksponensial sehingga data paling akhir mempunyai bobot atau timbangan lebih besar dalam rata-rata bergerak.*

Pada tahun 2020 ini, Pusdatin mengkaji 4 (empat) metode estimasi guna menghasilkan angka yang nilainya semakin mendekati nilai sebenarnya. Ke empat metode yang di kaji yaitu ARIMA (*Autoregressive Integrated Moving Average*), Regresi, Fungsi Transfer dan VAR (*Vector Auto Regression*) dengan menggunakan software R Studio

## **1.2. Tujuan**

Tujuan dari makalah ini adalah:

- Melakukan analisis dan pemodelan dengan ARIMA, Regresi, Fungsi Transfer dan VAR untuk meramalkan luas areal cengkeh nasional
- Membandingkan efektivitas metode tersebut dalam hasil estimasi luas areal cengkeh
- Menentukan metode terbaik dalam meramal data luas areal cengkeh nasional

## **VI. Sumber data dan Struktur Data**

Data yang digunakan dalam kajian ini adalah data sekunder yang bersumber dari Direktorat Jenderal Perkebunan dan Badan Pusat Statistik (BPS). Untuk data luas areal dan luas areal cengkeh bersumber dari Direktorat Jenderal Perkebunan sementara untuk data harga produsen dan volume ekspor bersumber dari BPS dengan series data tahunan yaitu dari tahun 1983 sampai tahun 2019.

## **VII. Metodologi**

### **3.1. ARIMA**

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung flat (mendatar/konstan) untuk periode yang cukup panjang. Model Autoregresif Integrated Moving Average (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (time series) secara statistik berhubungan satu sama lain (dependent) [https://daps.bps.go.id/ile\\_artikel/77/arima.pdf](https://daps.bps.go.id/ile_artikel/77/arima.pdf)

### **3.2. Regresi**

Analisis regresi mempelajari bentuk hubungan antara satu atau lebih peubah/[variabel](#) bebas (X) dengan satu peubah tak bebas (Y). Pada regresi harus ada variabel yang ditentukan dan variabel yang menentukan atau dengan kata lain adanya ketergantungan variabel yang satu dengan variabel yang lainnya. Untuk menentukan bentuk hubungan (model) diperlukan pemisahan yang tegas antara variabel bebas yang diberi simbol X dan variabel tak bebas yang diberi simbol Y. Dalam bentuk yang paling

sederhana yaitu satu peubah bebas (X) dengan satu peubah tak bebas (Y) mempunyai persamaan:

$$Y=a+bX+\varepsilon$$

Dimana:

a : intersept

b: koefisien arah atau koefisien beta

### 3.3. Fungsi Transfer

Model fungsi transfer merupakan gabungan dari karakteristik analisis regresi berganda dengan karakteristik deret berkala ARIMA (*Autoregressive Integrated Moving Average*). Dalam model ini, selain mengandung keterkaitan dengan kejadian pada waktu-waktu sebelumnya, juga keterkaitan yang sifatnya kausal dengan variable lain sebagaimana pada model regresi. Model fungsi transfer dengan variabel input lebih dari dua deret berkala disebut dengan model fungsi transfer multivariat. Prosedur pembentukan model fungsi transfer multivariat melalui dua tahap, yaitu pembentukan model fungsi transfer tunggal dari masing-masing input, baru dilakukan pembentukan model fungsi transfer secara simultan dari semua variabel.

Beberapa hal yang berkaitan dengan model fungsi transfer antara lain deret berkala output, disebut  $Y_t$ , yang diperkirakan akan dipengaruhi oleh deret berkala input, disebut  $X_t$ , dan input-input lain yang digabungkan dalam satu kelompok yang disebut gangguan (noise)  $N_t$ . Model fungsi transfer bivariat ditulis dalam dua bentuk umum. Bentuk pertama adalah sebagai berikut: (Makridakis, dkk:1999:448)

$$Y_t = v(B)X_t + N_t$$

### 3.4. Model Var

VAR (*Vector Auto Regression*) digunakan untuk memproyeksikan sebuah sistem dengan variabel runtut waktu dan untuk menganalisis dampak dinamis dari faktor gangguan yang terdapat dalam sistem variabel tersebut. Pada dasarnya Analisis VAR sama dengan suatu model persamaan simultan, karena dalam Analisis VAR kita mempertimbangkan beberapa variabel endogen secara bersama-sama dalam suatu model. Sebenarnya analisis ini mirip dengan model persamaan simultan biasa. Hanya saja di dalam Analisis VAR masing-masing variabel selain diterangkan oleh nilainya di masa lampau, juga dipengaruhi oleh nilai masa lalu dari semua variabel endogen lainnya dalam model yang diamati. Selain itu, dalam analisis VAR biasanya tidak ada variabel eksogen di dalam model.

## VIII. Hasil Pembahasan

### 4.1. Penyusunan Model Menggunakan R Studio

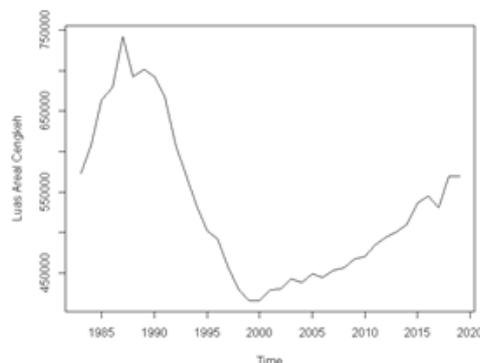
R Studio adalah bahasa pemrograman dan sistem perangkat lunak yang dirancang khusus untuk mengerjakan segala hal terkait komputasi statistik. Bahasa pemrograman ini pertama kali dikembangkan pada tahun 1993 oleh dua orang pakar statistik yaitu Ross Ihaka dan Robert Gentleman di Auckland University, New Zealand. Sampai saat

ini, bahasa pemrograman R terus berkembang secara pesat seiring dengan semakin populernya terminologi “**Big Data**” dan meningkatnya kebutuhan perusahaan akan *data scientist* untuk mengolah dan menganalisis data di perusahaan tersebut sebagai dasar pengambilan kebijakan dan mengotomatisasi proses bisnis menjadi *data driven*. Bahasa pemrograman seperti Python dan R telah menjadi pilihan utama bagi para peneliti maupun praktisi di bidang *data science* untuk mengolah dan menganalisis data baik itu untuk kepentingan penelitian maupun bisnis. Oleh karena itu, bagi seorang pemula di bidang *data science*, R merupakan bahasa pemrograman yang sangat *recommended* untuk dikuasai.

## 4.2. Model ARIMA

### 4.2.1. Hasil Run Model Arima dan Eksplorasi data

Syarat utama dalam melakukan pemodelan ARIMA adalah kestasioneran data. Kestasioneran data dapat diketahui secara visual (plot datanya) dan juga dengan uji statistik. Dari plot data gambar 1 terlihat luas areal cengkeh tahun 1983-2019 memiliki trend sehingga terindikasi tidak stasioner.



Gambar 1. Plot Luas Areal Cengkeh tahun 1983-2019

Dengan uji statistik dapat dilakukan salah satunya yaitu dengan uji *Augmented Dickey-Fuller* yaitu jika test statistics lebih besar dari critical value maka data tersebut tidak stasioner. Untuk menstasionerkan data dapat dilakukan differencing. Differencing adalah menghitung perubahan atau selisih nilai observasi. Berdasarkan uji *Augmented Dickey-Fuller* dapat dilihat seperti hasil dibawah ini.

Hasil Uji Augmented Dickey-Fuller

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)

Residuals:
  Min   1Q Median   3Q   Max
-58574 -11620 -1096  7143  68131
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.023e+04	3.741e+04	0.541	0.5925
z.lag.1	-4.690e-02	5.669e-02	-0.827	0.4144
tt	1.851e+02	5.379e+02	0.344	0.7331
z.diff.lag	3.609e-01	1.673e-01	2.158	0.0388 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25060 on 31 degrees of freedom

Multiple R-squared: 0.1773, Adjusted R-squared: 0.09767

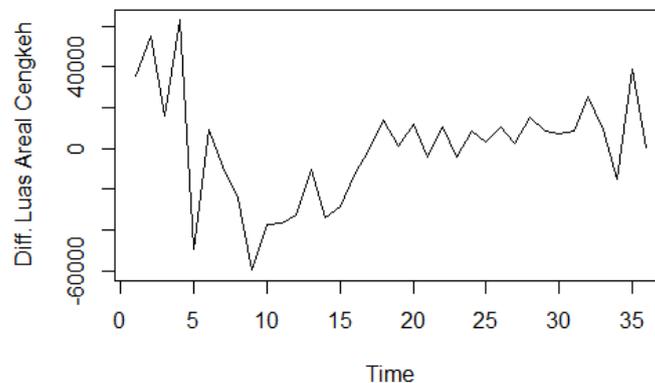
F-statistic: 2.227 on 3 and 31 DF, p-value: 0.1048

Value of test-statistic is: -0.8273 0.6381 0.9249

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-4.15	-3.50	-3.18
phi2	7.02	5.13	4.31
phi3	9.31	6.73	5.61

Dari hasil uji Augmented Dickey-Fuller diatas terlihat bahwa nilai tes statistic adalah -0,8273 dan nilai critical tau 3 untuk 1% sebesar -4,15, taraf uji 5% sebesar -3,50 dan taraf uji 10% sebesar -3,18. Dari hasil uji Augmented Dickey-Fuller ini dapat disimpulkan data tidak stasioner hingga pada taraf 1%, 5% dan 10% yang berarti tolak  $H_0$ . Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan differencing, maka akan dilakukan differencing 1 kali. Setelah differencing satu kali, secara visual sudah terlihat data sudah stasioner (Gambar 2) dan dari hasil uji Augmented Dickey-Fuller terlihat bahwa nilai tes statistic adalah -2.5044 dan nilai critical tau 1 untuk 1% sebesar -2,62, taraf uji 5% sebesar -1,95 dan taraf uji 10% sebesar -1,61% sehingga dapat disimpulkan data sudah stasioner pada taraf 5% dan 10%



Gambar 2. Plot Luas areal Cengkeh Tahun 1983-2019 dengan Differencing 1 kali

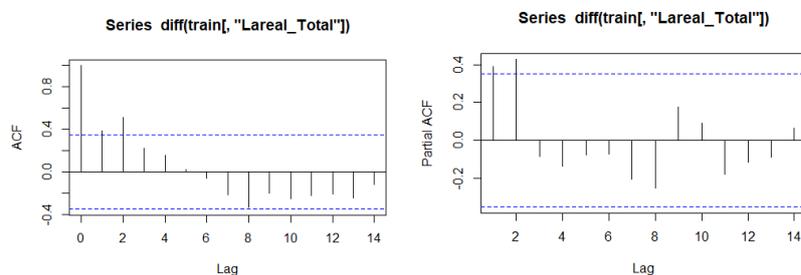
```

Hasil Augmented Dickey-Fuller diff 1
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
Residuals:
    Min     1Q Median     3Q     Max
-63961 -9887  2718  6506 35942
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1  -0.4147   0.1656  -2.504  0.01755 *
z.diff.lag -0.4752   0.1477  -3.217  0.00296 **
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 21040 on 32 degrees of freedom
Multiple R-squared:  0.5414,    Adjusted R-squared:  0.5127
F-statistic: 18.89 on 2 and 32 DF,  p-value: 3.829e-06
Value of test-statistic is: -2.5044
Critical values for test statistics:
    1pct 5pct 10pct
tau1 -2.62 -1.95 -1.61

```

#### 4.2.2. Identifikasi Model

Pada tahap ini akan ditentukan nilai p,d dan q dimana penentuan p dan q dapat dilakukan dengan beberapa cara yaitu dengan melihat korelogram autokorelasi (ACF) dan korelogram autokorelasi parsial (PACF), autoarima serta ar maselect. Sebelum menentukan nilai p,d,dan q terlebih dahulu data luas areal cengkeh dibagi menjadi data training yaitu data dari tahun 1983-2013 dan data testing mulai tahun 2014-2019. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk validasi dan pengepasan model. Dari gambar plot ACF pada Gambar 1 menunjukkan nyata pada lag=2 , sementara untuk lag=1 dan seterusnya berada dibawah garis selang kepercayaan dan tidak menunjukkan pola *cut off* maupun *tail off* sehingga agak kesulitan untuk mengidentifikasi model ARIMA nya.



Gambar 3. Plot ACF dan PACF Luas Areal Cengkeh

Apabila terjadi kesulitan mengidentifikasi model tentatif ARIMA menggunakan plot ACF dan PACFnya maka dapat digunakan *autoarima*. Hasil *autoarima* adalah ARIMA

(2,1,0) dengan Standard error model sebesar 0,1465 dan nilai MAPE data training sebesar 2,494327 seperti terlihat pada tabel dibawah ini.

```
Series: train[, "Lareal_Total"]
ARIMA(2,1,0)
Coefficients:
    ar1  ar2
    0.2262 0.5203
s.e. 0.1465 0.1561
sigma^2 estimated as 473834069: log likelihood=-353.03
AIC=712.06 AICc=712.95 BIC=716.36
Training set error measures:
      ME      RMSE    MAE    MPE      MAPE      MASE    ACF1
Training set -873.3783 20722.26 14146.12 -0.02679157 2.494327 0.700784
0.001571991
```

Selain cara diatas dapat juga dijalankan dengan menjalankan perintah armaselect, maka R-Studio akan menunjukkan sepuluh (10) model tentative yang diurutkan berdasarkan nilai SBC terkecil. Sepuluh model tentative yang ditampilkan berdasarkan armaselect yaitu:

```
p q  sbc
[1,] 2 4 526.8942
[2,] 3 4 529.3674
[3,] 2 5 529.9539
[4,] 4 4 532.2613
[5,] 3 5 533.2519
[6,] 5 4 533.8430
[7,] 5 5 534.5775
[8,] 4 5 536.4695
[9,] 3 1 544.0887
[10,] 0 2 544.3371
```

Dari hasil model autoarima dan 10 model tentative hasil armaselect dicobakan satu per satu dengan melihat mape terkecil dan kelayakan modelnya serta kelogisan hasil ramalan datanya. Setelah dicobakan satu persatu maka mape yang terkecil adalah ARIMA (2,1,0). Kemudian di panggil ARIMA (2,1,0)

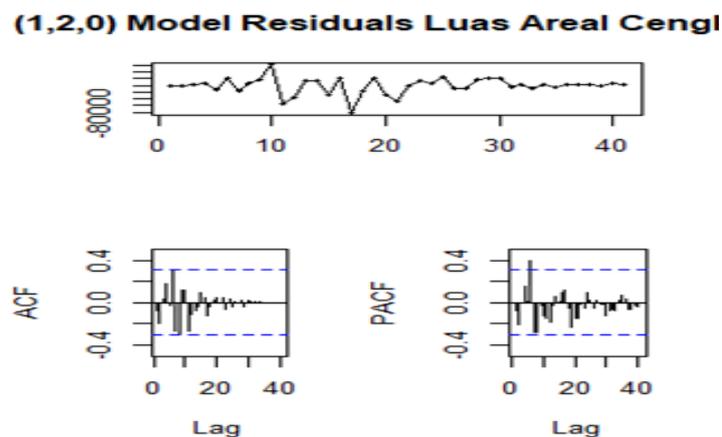
```
Call:
arima(x = train[, "Lareal_Total"], order = c(2, 1, 0))
Coefficients:
    ar1  ar2
    0.2262 0.5203
s.e. 0.1465 0.1561
sigma^2 estimated as 443253551: log likelihood = -353.03, aic = 712.06
z test of coefficients:
  Estimate Std. Error z value Pr(>|z|)
ar1 0.22619 0.14650 1.5440 0.1225952 *
ar2 0.52031 0.15610 3.3332 0.0008584 ***
```

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Dari Hasil penelusuran model ARIMA (2,1,0) diatas menunjukkan parameter ar 1 dan ar 2 adalah nyata

#### 4.2.3. Pemeriksaan sisaan

Suatu model layak dipakai apabila sisaan terdistribusi normal. Pemeriksaan sisaan bisa dilakukan dengan plot sisaan dan *Uji Ljung-Box*. Hasil plot plot sisaan adalah sebagai beriku



Gambar 4. Plot Sisaan Model Arima (1,2,0)

Hasil plot sisaan model ARIMA (2,1,0) menunjukkan bahwa plot ACF dan PACF sudah tidak menunjukkan pola *cut off* atau *tail off* yang merujuk ke model ARIMA tertentu sehingga model ARIMA (2,1,0) sudah cukup layak. Berdasarkan hasil *Uji Ljung-Box* kan mengindikasikan autokorelasi sisaan tidak signifikan hingga 30 lag karena nilai p-value diatas 5%. Oleh karenanya model ARIMA(2,1,0) sudah cukup baik mengepas data luas areal cengkeh Indonesia.

lags	statistic	df	p-value
5	3.362510	5	0.6442859
10	5.506986	10	0.8548459
15	6.722077	15	0.9649080
20	7.502395	20	0.9946814
25	7.737052	25	0.9996255
30	7.838538	30	0.9999842

#### 4.2.4. Peramalan set data testing dan training dengan model ARIMA (2,10)

Hasil peramalan data testing selama lima tahun kedepan (tahun 2020 – 2024) tersaji di bawah ini. Nilai MAPE yang yang diperoleh yaitu 2,49% untuk data training set dan 4,34% untuk data test set seperti dibawah ini

Time Series:

Start = 33

End = 38

Frequency = 1

[1] 516062.4 521971.3 526371.1 530440.8 533650.6 536494.1

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-873.3783	20722.26	14146.12	-0.02679157	2.494327	0.700784	0.001571991
Test set	24197.9622	27211.51	24197.96	4.34029566	4.340296	1.198742	NA

#### 4.2.5. Pengepasan model untuk seluruh data

Dengan model ARIMA (2,1,0) dilakukan pengepasan seluruh data dengan hasil sebagai berikut:

Time Series:

Start = 33

End = 38

Frequency = 1

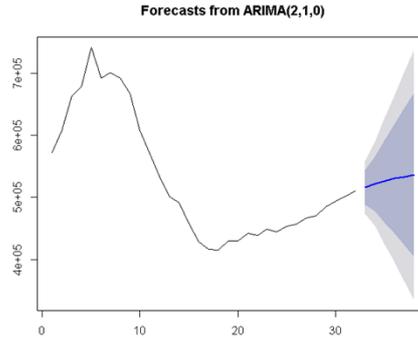
[1] 516062.4 521971.3 526371.1 530440.8 533650.6 536494.1

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-873.3783	20722.26	14146.12	-0.02679157	2.494327	0.700784	0.001571991
Test set	24197.9622	27211.51	24197.96	4.34029566	4.340296	1.198742	NA

Dengan menggunakan model ARIMA (2,1,0) diperoleh nilai mape yaitu sebesar 2,49% dan kemudian dilakukan peramalan untuk tahun 2020-2024. Hasil peramalan dengan ARIMA (2,1,0) adalah sebagai berikut:

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95	
38	589553.6	562173.2	616934.1	547678.8	631428.5
39	593779.4	550999.5	636559.2	528353.2	659205.5
40	605052.7	540646.2	669459.1	506551.5	703553.9
41	609500.6	525419.0	693582.3	480908.9	738092.4
42	616229.1	511077.1	721381.0	455413.1	777045.1

Hasil estimasi data luas areal cengkeh nasional untuk 5 tahun kedepan dari pengepasan model arima (2,1,0) adalah, pada tahun 2020 sebesar 589.553 ha, tahun 2021 sebesar 593.779 ha, tahun 2022 sebesar 605.052 ha, tahun 2023 sebesar 609.500 ha dan tahun 2024 sebesar 616.229 ha. Untuk plot estimasi luas areal cengkeh tahun 2020-2024 seperti gambar 6.



Gambar 6. Plot Hasil Estimasi Tahun 2020-2024

**4.2.6. Hasil Estimasi**

Hasil estimasi luas areal cengkeh tahun 2020-2024 dengan ARIMA (2,1,0) adalah sebagai berikut:

Hasil Estimasi Luas areal Cengkeh 2020-2024, Model ARIMA(1,1,2)

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	569.416	
2020	589.553	3,54
2021	593.779	0,72
2022	605.052	1,90
2023	609.500	0,74
2024	616.229	1,10
Rata-rata pertumbuhan (%)		1,60

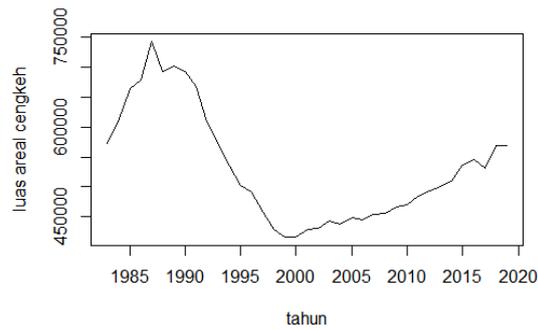
Tahun 2019 : Angka Sementara Ditjen Perkebunan

**4.3. Model Regresi**

Dalam kajian ini, regresi yang akan dilakukan adalah regresi linier dan regresi kuadrat

**4.3.1. Regresi Sederhana**

Series data yang digunakan untuk regresi sederhana dan kwadratik dimulai dari tahun 1983-2019. Plot data series luas areal cengkeh tahun 1983-2019



Gambar 7. Plot Luas Areal Cengkeh Tahun 1983-2019

b. Pembagian Data Testing dan Training

Data dibagi menjadi 2 bagian yaitu data training dimulai dari tahun 1983-2013 dan data testing dimulai dari tahun 2014-2019. Data training digunakan untuk membangun model, sedangkan data testing digunakan untuk validasi model.

c. Pendugaan Model

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dengan luas areal cengkeh. Data input yang digunakan untuk pendugaan model adalah data luas areal cengkeh dari tahun 1983-2019. Hasil pendugaan model adalah sebagai berikut

Call:

```
lm(formula = Lareal_Total ~ Tahun, data = train)
```

Residuals:

Min	1Q	Median	3Q	Max
-108903	-54929	-4736	58847	123228

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	16230085	2718519	5.970	1.52e-06 ***
Tahun	-7857	1360	-5.776	2.61e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 71050 on 30 degrees of freedom

Multiple R-squared: 0.5265, Adjusted R-squared: 0.5107

F-statistic: 33.36 on 1 and 30 DF, p-value: 2.613e-06

Dari hasil pendugaan model di atas dijelaskan bahwa tahun signifikan terhadap luas areal cengkeh. Nilai estimatannya adalah 7857 yang artinya ada kenaikan luas areal cengkeh rata-rata 7857 ha per tahun. Model ini kurang bagus dimana  $R^2$  Adjustednya sebesar 51,07%. Mape data training adalah 11,61% sementara mape data testing adalah 30,18 seperti dibawah ini:

Mape training

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
------	---------	--------	------	---------	------

0.1497 5.6410 11.5225 11.6192 16.5246 26.1875

Mape testing

Min. 1st Qu. Median Mean 3rd Qu. Max.  
 25.51 27.71 28.22 30.18 34.02 35.44

Evaluasi Asumsi

- Uji kenormalan dengan Jarque Bera (JB) yaitu salah satu uji normalitas jenis *gooness of fit test* yang mana mengukur apakah skewness dan kurtosis [sampel](#) sesuai dengan distribusi normal. Hasil uji JB= 2.126, p-value=0,155 yaitu data terdistribusi normal dimana p-value > 5% (Ho ditolak)
- Uji Durbin Watson (DW) yaitu untuk mengetahui ada tidaknya autokoreasi. DW =0,8673, p-value = 0,00 yaitu data ada autokorealsi antar residual (Ho ditolak)

Uji Non Constant Varians Score Test (NCVT) yaitu untuk melihat kehomogenan data.chisquare = 0,2650466, Df =1, p=0,6066 yaitu data homogen

Hasil Estimasi luas areal cengkeh

Tabel 3.2. Hasil Estimasi Luas areal Cengkeh Tahun 2020-2024

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	569.416	
2020	559.773	-1,69
2021	551.917	-1,40
2022	544.060	-1,42
2023	556.204	2,23
2024	528.347	-5,01
Rata-rata pertumbuhan (%)		-1,46

Tahun 2019 : Angka Sementara Ditjen Perkebunan

Regresi kwadratik

a. Pendugaan Model

Pendugaan model ini digunakan untuk mengetahui hubungan antara tahun dengan luas areal cengkeh. Data input yang digunakan untuk pendugaan model adalah data luas areal cengkeh dari tahun 1983-2019. Hasil pendugan model adalah sebagai berikut

Call:

lm(formula = Lareal\_Total ~ t + t2, data = train)

Residuals:

Min 1Q Median 3Q Max  
 -158514 -24577 -1017 21793 103979

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	731159.4	29461.0	24.818	< 2e-16 ***
t	-23994.9	4398.7	-5.455	7.17e-06 ***
t2	520.6	137.1	3.796	0.000693 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 59060 on 29 degrees of freedom  
Multiple R-squared: 0.6837, Adjusted R-squared: 0.6619  
F-statistic: 31.34 on 2 and 29 DF, p-value: 5.64e-08

Dari hasil pendugaan model diatas dijelaskan bahwa tahun signifikan terhadap luas areal cengkeh. Nilai estimatonya adalah 731159.4 yang artinya ada kenaikan luas areal rata-rata 731159.4 ha per tahun. Model ini kurang bagus dimana  $R^2$  Adjustednya sebesar 66,19%. Mape Data data training dan data testing adalah sebagai berikut:

#### Mape Data Training

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.4081	3.8529	5.0921	8.6129	8.8691	22.3774

#### Mape Data Testing

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.4824	8.810	7.028	15.753	7.114	7.334

#### Evaluasi Asumsi

- Uji kenormalan dengan Jarque Bera (JB) = 1,037 p-value=0,4755 yaitu data terdistribusi normal dimana p-value > 5% (Ho ditolak)
- Uji Durbin Watson (DW) = 0,7482354, p-value = 0,00 yaitu data ada autokorealsi antar residual (Ho ditolak)
- Uji Non Constant Varians Score Test (NCVT) = 2,56351, Df =1, p=0,000 yaitu data homogen

## Hasil Estimasi

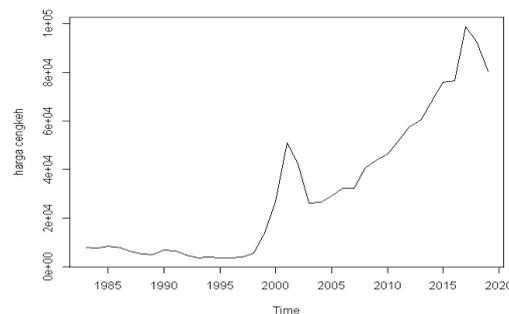
### Hasil Estimasi Luas areal Cengkeh Tahun 2020-2024

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	569.416	
2020	761.477	33,73
2021	791.103	3,89
2022	821.770	3,88
2023	853.478	3,86
2024	886.228	3,84
Rata-rata pertumbuhan (%)		9,84

Tahun 2019 : Angka Sementara Ditjen Perkebunan

#### 4.4. Model Fungsi Transfer

Seperti sudah dijelaskan diatas bahwa model fungsi transfer merupakan gabungan dari karakteristik analisis regresi berganda dengan karakteristik deret berkala ARIMA (*Autoregressive Integrated Moving Average*). Untuk model fungsi transfer ini menggunakan data harga untuk variabel inputnya. Plot data harga cengkeh.



Plot data harga Cengkeh tahun 1983-2019

##### 4.4.1. Pemeriksaan Kestasioneran

Dari plot data harga cengkeh terlihat bahwa adanya peningkatan dari tahun ke tahun sehingga terindikasi data tidak stasioner. Untuk pengujiannya dapat dilakukan dengan uji Augment Dickey-Fuller. Hasil Augmen Dickey Fuller adalah sebagai berikut:

Hasil Uji Augmented Dickey-Fuller

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression trend
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Residuals:
```

```

Min    1Q  Median    3Q    Max
-12092.3 -2864.2  296.1  3140.8 18435.3
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -3354.6614  2541.3747  -1.320  0.19879
z.lag.1      -0.3464   0.1228  -2.819  0.00928 **
tt           728.1875  251.1117   2.900  0.00767 **
z.diff.lag   0.4279   0.1792   2.388  0.02484 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 5859 on 25 degrees of freedom
Multiple R-squared:  0.3257,    Adjusted R-squared:  0.2448
F-statistic: 4.026 on 3 and 25 DF,  p-value: 0.01822
Value of test-statistic is: -2.8194 3.4211 4.4387
Critical values for test statistics:
      1pct  5pct 10pct
tau3 -4.15 -3.50 -3.18
phi2  7.02  5.13  4.31
phi3  9.31  6.73  5.61

```

Dari hasil uji Augmented Dickey-Fuller terlihat bahwa data tidak stasioner karena nilai *tes statistic* lebih besar dari nilai *critical* dimana nilai *tes statistic nya* adalah -2,8194 dan nilai *critical tau 3* untuk 1% sebesar -4,15, taraf uji 5% sebesar -3,50 dan taraf uji 10% sebesar -3,18. Dari hasil uji Augmented Dickey-Fuller ini dapat disimpulkan data tidak stasioner hingga pada taraf 1%, 5% dan 10% yang berarti tolak  $H_0$  sehingga dilakukan differencing 1. Hasil uji Augmented Dickey-Fuller Diff 1 adalah sebagai berikut:

#### Hasil Uji Augmented Dickey-Fuller Diff 1

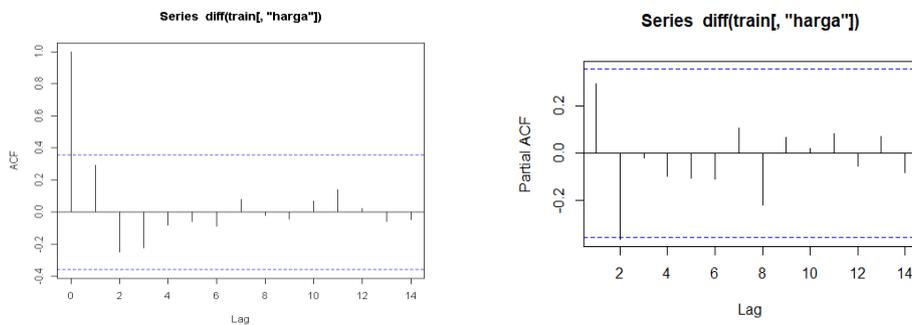
```

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
Test regression none
Call:
lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
Residuals:
      Min       1Q   Median       3Q      Max
-15671.8 -1012.8  -31.1  4121.8 21398.0
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
z.lag.1      -0.8789   0.2153  -4.082 0.000377 ***
z.diff.lag   0.3297   0.1876   1.757 0.090651 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 6448 on 26 degrees of freedom

```

Multiple R-squared: 0.399, Adjusted R-squared: 0.3528  
 F-statistic: 8.63 on 2 and 26 DF, p-value: 0.001335  
 Value of test-statistic is: -4.0821  
 Critical values for test statistics:  
 1pct 5pct 10pct  
 tau1 -2.62 -1.95 -1.61

Dari hasil uji Augmented Dickey-Fuller setelah di differencing 1 kali terlihat bahwa nilai *tes statistic* sudah lebih kecil dibandingkan nilai *critical*ya dimana nilai *tes statistic* adalah -4,0821 sementara nilai critical tau 1 untuk 1% sebesar -2,62, taraf uji 5% sebesar -1,95 dan taraf uji 10% sebesar -1,61%. Hal ini dapat disimpulkan data tidak sudah stasioner hingga pada taraf uji 1%, 5% dan 10%. Stasioneritas data dapat juga dilihat dari plot ACF dan PACF. Jika lag pada ACF maupun PACF banyak yang keluar dari garis *confidence interval* maka data tidak stasioner. Berikut adalah plot ACF dan PACF harga cengkeh setelah differencing 1.



Gambar. Plot ACF dan PACF Harga Cengkeh

Berdasarkan plot di atas, terlihat bahwa plot ACF cenderung cut off dimana hanya lag 0 yang keluar dari *confidence interval*. Plot PACF menunjukkan cut off pada lag 2, ditunjukkan berada di luar garis *confidence interval*. Hal tersebut memperkuat hasil uji ADF yaitu data harga cengkeh telah stasioner.

#### 4.4.2. Pencarian Model Tentatif Variabel Input

Untuk mencari model tentative variable harga, dapat dilakukan dengan *autoarima* maupun *arma selec*. Data yang digunakan untuk memilih model terbaik adalah data trainingnya. Hasil dari *autoarima* adalah sebagai berikut:

```
Series: train[, "harga"]
ARIMA(0,1,1)
Coefficients:
    ma1
    0.5113
s.e. 0.1733
sigma^2 estimated as 38183820: log likelihood=-304.08
AIC=612.16 AICc=612.6 BIC=614.96
Training set error measures:
      ME      RMSE     MAE      MPE      MAPE      MASE      ACF1
```

```
Training set 1135.225 5976.65 3397.916 3.092626 15.18112 0.867686 -
0.06233682
```

Hasil *autoarima* untuk harga cengkeh adalah ARIMA (0,1,1) dengan Standard error sebesar 0,1733 dan nilai MAPE data training sebesar 15.18112% seperti terlihat pada tabel diatas. Selanjutnya dilakukan pengujian *coefstest* pada model ARIMA (0,1,1) untuk mengetahui signifikansi dari koefisien MA pada model tersebut, Output *coefstest*nya adalah

```
z test of coefficients:
Estimate Std. Error z value Pr(>|z|)
ma1 0.51133 0.17328 2.9508 0.003169 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

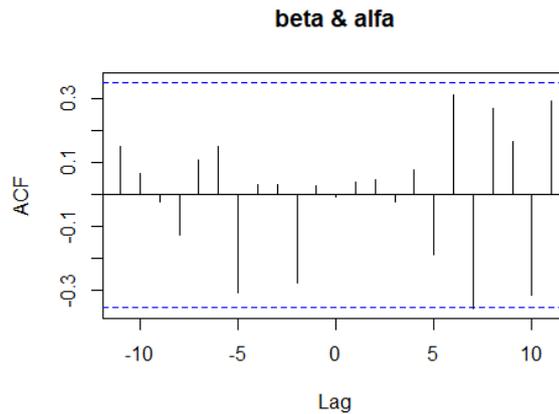
Berdasarkan output di atas terlihat bahwa Ma1 signifikan, pada taraf 1% Selain menggunakan *script auto.arima* dapat juga dijalankan dengan perintah *armaselect* dari ordo p=1 sd 5 dan q= 1sd 5 maka R-Studio akan menunjukkan sepuluh (10) model tentative yang diurutkan berdasarkan nilai AIC terkecil. Sepuluh model tentative yang ditampilkan adalah sebagai berikut:

```
  p q  sbc
[1,] 5 5 264.8169
[2,] 1 5 477.0809
[3,] 2 5 480.6898
[4,] 4 5 482.5022
[5,] 0 5 483.2223
[6,] 3 5 484.4578
[7,] 4 4 495.3675
[8,] 5 4 496.8568
[9,] 3 4 499.4510
[10,] 2 4 504.1149
```

Untuk mengetahui model terbaik dari 10 model tentative *auto.arima* dan model *armaselect* dapat diuji cobakan dengan membandingkan nilai AIC. Nilai AIC terkecil merupakan model terbaik. Setelah semua diuji cobakan maka nilai AIC terkecil adalah ARMA (0,1,1)

#### 4.4.3. Prewhitening dan korelasi silang

Korelasi silang menggambarkan struktur hubungan antara  $X_t$  dengan  $Y_t$ . Untuk mengidentifikasi pengaruh  $X_t$  terhadap  $Y_t$  maka deret  $X_t$  harus stasioner atau sudah distasionerkan. Dalam konteks pemodelan  $X_t$  terhadap  $Y_t$ , untuk membuat  $X_t$  stasioner tidak dengan pembedaan (*differencing*) namun dengan mengambil komponen *white noise* dari  $X_t$  (*prewhitening*). *Prewhitening* dilakukan terhadap deret input  $X_t$  yang didefinisikan sebagai alfa serta deret input  $Y_t$  yang didefinisikan sebagai beta. Hasil grafig ACF untuk beta dan alfa.



Gambar ACF untuk Beta dan alfa

Berdasarkan grafik ACF di atas, terlihat bahwa tidak ada yang keluar dari garis signifikansi,. Karena lag signifikan pertama kali pada lag 0 maka nilai  $b=0$ . Selain itu, tidak ada tambahan lagi lag yang signifikan maka nilai  $s=0$ . Mengingat data luas areal cengkeh dan harga cengkeh merupakan data tahunan yang tidak mengandung musiman maka diasumsikan nilai  $r=0$ . Nilai  $b=0$  menunjukkan tidak ada jeda pengaruh antara harga pada waktu  $t$  terhadap luas areal pada waktu  $t$ . Nilai  $s=0$  berarti ada korelasi antara luas areal dan harga cengkeh pada tahun yang sama. Dengan kata lain, dampak dari harga cengkeh terhadap luas areal dirasakan pada waktu yang sama ( $t$ ).

#### 4.4.4. Pengepasan Model

Pengepasan model dengan noise dilakukan dengan menggabungkan nilai  $r,s,b$  dan hasil identifikasi model noise yaitu ARIMA(2,1,0) Hasil pengepasan model sebagai berikut:

```
Series: train[, "Lareal_Total"]
Regression with ARIMA(2,1,0) errors
```

Coefficients:

```
      ma1  xreg
      0.2108 0.2173
s.e. 0.1267 0.7518
```

```
sigma^2 estimated as 732125967: log likelihood=-347.73
AIC=701.46 AICc=702.38 BIC=705.66
```

Training set error measures:

```
      ME  RMSE  MAE  MPE  MAPE  MASE  ACF1
Training set -2169.309 25715.27 18632.42 -0.4883843 3.275659 0.9059918 0.1006728
```

#### 4.4.5. Identifikasi Model Noise

Identifikasi model noise dilakukan untuk menentukan model ARIMA terbaik pada data residual hasil pengepasan model.. Arima yang direkomendasikan adalah (0,1,0):

Series: res  
ARIMA(0,1,0) with zero mean

sigma^2 estimated as 102692901: log likelihood=-329.92  
AIC=661.84 AICc=661.98 BIC=663.27

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	321.8076	10133.75	8560.288	100	100	0.7672982	-0.000929

3487

Berdasarkan data residual, model ARIMA yang direkomendasikan oleh fungsi *auto.arima* yaitu ARIMA(0,0,0), namun model ini menghasilkan MAPE yang cukup besar yaitu 100%. Maka akan dicari model alternatif dengan dengan *armaselect*. Ada 10 model tentative terbaik yang ditampilkan hasil dari *armaselect* yaitu sebagai berikut:

	p	q	sbc
[1,]	5	5	501.3197
[2,]	4	5	505.5649
[3,]	5	4	539.9377
[4,]	3	5	541.2054
[5,]	2	5	544.5020
[6,]	2	4	557.8965
[7,]	3	4	559.1056
[8,]	4	4	560.2821
[9,]	1	4	562.9512
[10,]	1	5	564.8626

Model alternative pertama yang disarankan untuk data residual adalah ARIMA(5,1,5). Kemudian di ujikan semua model yang direkomendasikan. Model yang dipakai adalah model ARIMA (1,1,4)

Output untuk model residual ARIMA (1,1,4)

```
modelres  
Series: res  
ARIMA(1,1,4)
```

Coefficients:

	ar1	ma1	ma2	ma3	ma4
	0.4836	-1.0105	0.0001	1.0102	-0.9997
s.e.	0.1793	0.2754	0.4331	0.3794	0.3454

sigma^2 estimated as 111911008: log likelihood=-322.14

AIC=656.28 AICc=659.93 BIC=664.68

```
> coeftest(modelres)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z )
ar1	0.4835611	0.1792812	2.6972	0.0069921 **

```

ma1 -1.0105164 0.2754057 -3.6692 0.0002433 ***
ma2 0.0001403 0.4331091 0.0003 0.9997415
ma3 1.0101849 0.3794309 2.6624 0.0077593 **
ma4 -0.9997162 0.3454042 -2.8943 0.0037996 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Berdasarkan output di atas, maka model residual untuk ARIMA(1,1,4) signifikan pada ar1, ma1, ma 3 dan ma 4 tapi tidak signifikan si ma 2.

#### 4.4.6. Pengepasan model dengan noise

Pengepasan model dengan noise dilakukan dengan menggabungkan nilai r,s,b dan hasil identifikasi model noise yaitu ARIMA(1,1,4). Setelah dihasilkan model noisenya maka dilakukan pengepasan dengan arima output yaitu ARIMA (2,1,0).  
Regression with ARIMA(2,1,0) errors

Coefficients:

```

      ma1  xreg
0.2108 0.2173
s.e. 0.1267 0.7518

```

sigma^2 estimated as 732125967: log likelihood=-347.73  
AIC=701.46 AICc=702.38 BIC=705.66

Training set error measures:

```

      ME  RMSE  MAE  MPE  MAPE  MASE  ACF1
Training set -2169.309 25715.27 18632.42 -0.4883843 3.275659 0.9059918 0.1006728

```

#### 4.4.7. Peramalan Fungsi Transfer data input nilai aktual

Berdasarkan model fungsi transfer dengan noise ARIMA (1,1,4), dilakukan peramalan berbasis nilai aktual dimana luas areal cengkeh diestimasi menggunakan data aktual harga cengkeh periode 2014-2020 dengan mape 2,483371 seperti berikut:  
Series: test[, "Lareal\_Total"]  
Regression with ARIMA(1,1,4) errors

Coefficients:

```

      ar1 ma1  ma2 ma3  ma4
-0.7757 0.7548 0.487 -0.1286 0.1667
s.e. 0.1793 0.2754 0.4331 0.3794 0.3454

```

sigma^2 estimated as 113226926: log likelihood=-56.54  
AIC=115.07 AICc=116.41 BIC=114.68

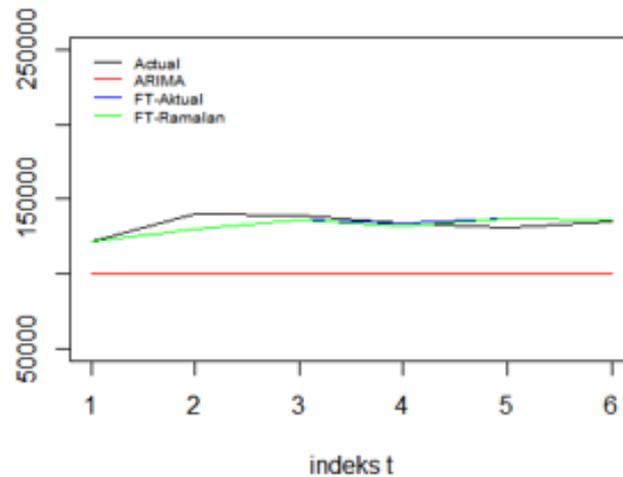
Training set error measures:

```

      ME  RMSE  MAE  MPE  MAPE  MASE  ACF1
Training set 8926.113 11736.07 8926.113 1.610059 1.610059 0.5031801 -0.5653505

```

Untuk membandingkan ketepatan model ramalan, dilakukan perbandingan hasil ramalan terhadap data aktual luas areal cengkeh pada tahun 2014-2019. Hasil ramalan yang dibandingkan adalah ramalan dengan model ARIMA (2,1,0), fungsi transfer ARIMA(1,1,4) dimana input harga cengkeh yang digunakan adalah data aktual maupun ramalan



Perbandingan Hasil Ramalan Luas areal Cengkeh Tahun 2014-2019

Dari grafik di atas terlihat jika dibandingkan dengan data aktual luas areal cengkeh 2014-2019 (warna hitam), maka hasil peramalan dengan model ARIMA tanpa fungsi transfer (warna merah) tidak mengikuti pola data aktualnya. Di sisi lain, peramalan dengan fungsi transfer khususnya jika input harga yang digunakan adalah data aktual maka hasil ramalan luas arealnya (warna biru) seperti pola data luas areal actual selama 6 tahun terakhir. Jika input harga cengkeh yang digunakan adalah hasil ramalan, maka estimasi luas arealnya (warna hijau) hampir menyerupai pola data asli.

#### Hasil Estimasi

Berikut adalah hasil ramalan untuk lima tahun kedepan yaitu tahun 2020-2024

Time Series:

Start = 32

End = 36

Frequency = 1

5[1] 566725.5 567855.0 578992.2 581603.0 589766.1

Tabel 3.2. Hasil Estimasi Luas areal Cengkeh Tahun 2020-2024

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	569,416	
2020	566,725	-0.47
2021	567,855	0.20
2022	578,992	1.96
2023	581,603	0.45
2024	589,766	1.40
<b>Rata-rata pertumbuhan (%)</b>		<b>0.71</b>

Tahun 2019 : Angka Sementara Ditjen Perkebunan

#### IV.5. Model VAR

Metode VAR memperlakukan seluruh variabel secara simetris tanpa memperlakukan variabel dependen dan independen atau dengan kata lain model ini memperlakukan seluruh variabel sebagai variabel endogen.

Dalam pemodelan ini menggunakan tiga variabel yaitu luas areal cengkeh, harga cengkeh dan volume ekspor cengkeh. Data yang digunakan adalah dari series tahun 1983-2019. Untuk mencari model var terbaik maka dilakukan uji coba dari VAR lag p=1 sampai dengan lag p=6. dengan type both dan trend. Setelah dilakukan uji coba maka model terpilih adalah VAR lag 1 dengan type both. Hasil dari VAR lag p=1 type both adalah sebagai berikut:

VAR Estimation Results:

=====

Endogenous variables: Lareal\_Total, harga, volek

Deterministic variables: both

Sample size: 30

Log Likelihood: -932.303

Roots of the characteristic polynomial:

0.8723 0.7587 0.05723

Call:

VAR(y = cengkeh\_luas[2:32, c(2, 5, 6)], p = 1, type = "both")

Estimation results for equation Lareal\_Total:

=====

Lareal\_Total = Lareal\_Total.l1 + harga.l1 + volek.l1 + const + trend

Estimate Std. Error t value Pr(>|t|)

Lareal\_Total.l1 0.8941 0.0714 12.522 2.87e-12 \*\*\*

harga.l1 1.6770 0.4761 3.522 0.00167 \*\*

volek.l1 0.8011 1.0359 0.773 0.44657

```

const      76689.5795 50943.3260  1.505  0.14475
trend      -3944.6620 1363.9729 -2.892  0.00781 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 23130 on 25 degrees of freedom  
Multiple R-Squared: 0.957, Adjusted R-squared: 0.9501  
F-statistic: 139.1 on 4 and 25 DF, p-value: < 2.2e-16

Estimation results for equation harga:

```

=====
harga = Lareal_Total.l1 + harga.l1 + volek.l1 + const + trend

```

```

              Estimate Std. Error t value Pr(>|t|)
Lareal_Total.l1 -2.641e-03  2.035e-02 -0.130  0.898
harga.l1         7.820e-01  1.357e-01  5.762  5.27e-06 ***
volek.l1        -9.033e-02  2.953e-01 -0.306  0.762
const           -1.455e+03  1.452e+04 -0.100  0.921
trend            6.152e+02  3.888e+02  1.583  0.126
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 6592 on 25 degrees of freedom  
Multiple R-Squared: 0.9141, Adjusted R-squared: 0.9003  
F-statistic: 66.47 on 4 and 25 DF, p-value: 5.916e-13

Estimation results for equation volek:

```

=====
volek = Lareal_Total.l1 + harga.l1 + volek.l1 + const + trend

```

```

              Estimate Std. Error t value Pr(>|t|)
Lareal_Total.l1 -2.700e-02  1.325e-02 -2.038  0.0523 .
harga.l1         1.053e-01  8.836e-02  1.192  0.2446
volek.l1         1.222e-02  1.922e-01  0.064  0.9498
const            1.958e+04  9.454e+03  2.071  0.0489 *
trend           -1.541e+02  2.531e+02 -0.609  0.5482
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 4292 on 25 degrees of freedom  
Multiple R-Squared: 0.3822, Adjusted R-squared: 0.2834  
F-statistic: 3.867 on 4 and 25 DF, p-value: 0.01403

Covariance matrix of residuals:

```

Lareal_Total  harga  volek

```

```
Lareal_Total 534919555 49794222 -9999358
harga      49794222 43457263 -9295251
volek     -9999358 -9295251 18421501
```

Correlation matrix of residuals:

```
      Lareal_Total  harga  volek
Lareal_Total  1.0000  0.3266 -0.1007
harga         0.3266  1.0000 -0.3285
volek        -0.1007 -0.3285  1.0000
```

#### 4.5.1. Uji asumsi untuk lag p=1 type both

##### a. Pemeriksaan autokorelasi residual

Portmanteau Test (asymptotic)

data: Residuals of VAR object varhsheet.b1

Chi-squared = 100.51, df = 135 p-value = 0.9884

Nilai p-value lebih besar dari 0,05 maka sisaan saling bebas atau asumsi non autokorelasi terpenuhi.

##### b.

emeriksaan kenormalan dengan normality.test” yang dengan uji Jarque-Bera tests

JB-Test (multivariate)

data: Residuals of VAR object varhsheet.b1

Chi-squared = 92.262, df = 6, p-value <2.2e-16

Skewness only (multivariate)

data: Residuals of VAR object varhsheet.b1

Chi-squared = 24.442, df = 3, p-value = 2.02e-05

\$Kurtosis

Kurtosis only (multivariate)

data: Residuals of VAR object varhsheet.b1

Chi-squared = 67.82, df = 3, p-value = 1.255e-14

Nilai pvalue lebih kecil dari 0,05, namun karena jumlah data yang digunakan cukup banyak, maka series tersebut dapat dianggap normal.

##### c. Pemeriksaan heteroskedastisitas dengan fungsi “arch.test”

ARCH (multivariate)

data: Residuals of VAR object varhsheet.b1

Chi-squared = 150, df = 180, p-value = 0,9498

Nilai p-value lebih besar dari 0,05 maka ragam sisaan homogen atau asumsi homoskedastisitas terpenuhi.

#### 4.5.2. Mape data testing dan data training adalah adalah sebagai berikut:

Mape data testing

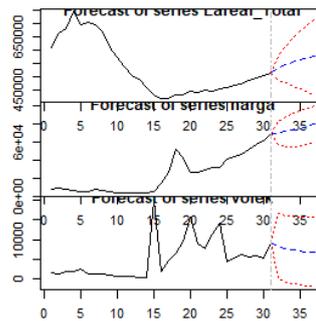
```
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
1.74 13.47 14.23 14.51 16.33 16.77 1
```

Mape data training adalah sebagai berikut

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
7.571 11.240 20.158 24.235 31.311 54.227
```

Data testing hasil ramalan luas areal cengkeh dengan nilai  $p=1$  type both MAPE =14,51%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data testing adalah 14,51%. Data training hasil ramalan luas areal cengkeh dengan nilai  $p=1$  MAPE =24,23%. Nilai ini dapat diartikan bahwa rata-rata seluruh persentase kesalahan antara data aktual dengan data hasil ramalan pada data training adalah 24,23%.

Plot hasil ramalan luas areal cengkeh, harga produsen cengkeh dan volume ekspor cengkeh dengan VAR  $p=1$  type both



Plot Hasil Ramalan Luas areal Cengkeh Tahun 2014-2019

Untuk model VAR (1) type both dapat dilihat dari gambar diatas bahwa luas areal cengkeh cenderung mengalami peningkatan. Nilai ramalan 5 tahun kedepan adalah sebagai berikut :

Hasil Ramalan VAR (1) type both

Tahun	Luas Areal (Ha)	Pertumbuhan (%)
2019	569.416	
2020	542.699	-4,69
2021	553.266	1,95
2022	561.523	1,49
2023	568.189	1,19
2024	573.788	0,99
Rata-rata pertumbuhan (%)		0,18

Tahun 2019 : Angka Sementara Ditjen Perkebunan

#### IV.6. Model Terbaik

- c. Kriteria Pemilihan Model Terbaik
  - Nilai MAPE Kecil
  - Nilai AIC dan SIC kecil
  - Uji asumsi terpenuhi

d. Model Terbaik

No	Model	Mape		Hasil Estimasi Luas Areal Cengkeh (ha)					Pertumb. (%)
		training	testing	2020	2021	2022	2023	2024	
1	Arima	2,49	4,34	589.553	593.779	605.052	609.500	616.229	1,11
2	Regresi								
	-Linier sederhana	11,61	30,18	659.773	651.917	544.060	556.204	528.347	-1,40
	- Kwadratik	8,61	15,75	761.477	791.103	821.770	853.478	886.228	3,87
3	Fungsi Transfer	2,61	3,27	566.725	567.855	578.992	581.603	589.766	1,00
4	VAR	14,51	24,23	542.699	553.266	561.523	568.189	573.788	1,40

Dari table diatas di dapatkan model terbaik adalah Fungsi transfer

vi. Kesimpulan dan Saran

Dari ke empat metode estimasi yang dikaji maka dapat disimpulkan bahwa model fungsi transfer adalah yang terbaik karena memiliki mape testing dan mape training terkecil dan angka estimasinya mendekati angka aktualnya. Mape training sebesar 2,61 persen dan mape testing sebesar 3,27 persen, sementara hasil estimasi tahun 2020 adalah 566.725 ha dan tahun 2019 (Angka Sementara) sebesar 569.416 ha.



# **Kajian Metode Estimasi Data Komoditas Perkebunan**



**Pusat Data dan Sistem Informasi Pertanian  
Sekretariat Jenderal - Kementerian Pertanian  
Tahun 2020**

**Jl. Harsono RM No. 3, Ragunan - Jakarta 12550  
Gedung D Lantai 4**