JURNAL MATEMATIKA, STATISTIKA & KOMPUTASI E-ISSN: 2614-8811
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 18, No. 2, 198 -223, January, 2022
DOI: 10.20956/j.v18i2.18382

Application of GARCH Forecasting Method in Predicting The Number of Rail Passengers (Thousands of People) in Jabodetabek Region

Penerapan Metode Peramalan GARCH dalam Memprediksi Jumlah Penumpang Kereta Api (Ribu Orang) di Wilayah Jabodetabek

Warosatul Anbiya[1], Farin Cyntiya Garini[2]

Abstract

PT. Kereta Api Indonesia and PT. KAI Commuter Jabodetabek records time series data in the form of the number of train passengers (thousand people) in Jabodetabek Region in 2011-2020. One of the time series methods that can be used to predict the number of train passengers (thousand people) in Jabodetabek area is ARIMA method. ARIMA or also known as Box-Jenkins time series analysis method is used for short-term forecasting and does not accommodate seasonal factors. If the assumption of residual homoscedasticity is violated, the ARCH / GARCH method can be used, which explicitly models changes in residual variety over time. This study aims to model and forecast the number of train passengers (thousand people) in Jabodetabek area in 2021. Based on data analysis and processing using ARIMA method, the best model is ARIMA (1,1,1) with an AIC value of 2,159.87 and with ARCH / GARCH method, the best model is GARCH (1,1) with an AIC value of 18.314. Forecasting results obtained based on the best model can be used as a reference for related parties in managing and providing public transportation facilities, especially trains.

Keywords: Trains Passengers, ARIMA, ARCH-GARCH, Forecasting

Abstrak

PT. Kereta Api Indonesia dan PT. KAI *Commuter* Jabodetabek mencatat data deret waktu berupa jumlah penumpang kereta api (ribu orang) di Wilayah Jabodetabek tahun 2011-2020. Salah satu metode deret waktu yang dapat digunakan untuk meramalkan jumlah penumpang kereta api (ribu orang) di Wilayah Jabodetabek adalah metode ARIMA. ARIMA atau disebut juga metode analisis runtun waktu *Box-Jenkins* digunakan untuk peramalan jangka pendek dan tidak mengakomodasi faktor musiman. Jika asumsi homoskedastisitas residual terlanggar, dapat digunakan metode ARCH/GARCH yang secara eksplisit memodelkan perubahan ragam residual dari waktu ke waktu. Penelitian ini bertujuan untuk melakukan pemodelan dan peramalan jumlah penumpang kereta api (ribu orang) di Wilayah Jabodetabek pada tahun 2021. Berdasarkan analisis dan pengolahan data dengan metode ARIMA, diperoleh model terbaik adalah model ARIMA (1,1,1) dengan nilai AIC sebesar 2.159,87 dan dengan metode ARCH/GARCH, diperoleh model terbaik adalah GARCH (1,1) dengan nilai AIC sebesar 18,314. Hasil peramalan yang diperoleh berdasarkan model terbaik dapat dijadikan referensi bagi pihak-pihak terkait dalam mengelola dan menyediakan fasilitas transportasi umum, terutama kereta api.

Kata Kunci: Penumpang Kereta Api, Peramalan, ARIMA, ARCH-GARCH

* *Program Studi Sarjana Statistika, FMIPA-UNPAD*
***Email address**: [1] warosatul19001@mail.unpad.ac.id, [2] farin19001@mail.unpad.ac.id*

1. PENDAHULUAN

Kegiatan peramalan dapat digunakan dalam berbagai bidang. Dalam penelitian ini, penulis melakukan suatu peramalan mengenai jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek pada masa yang akan datang dengan menggunakan analisis deret berkala (*time series*). Tujuan penelitian ini adalah untuk meramalkan (*forecasting*) jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek pada tahun 2021, mulai dari bulan Januari hingga bulan Desember dan diharapkan dapat bermanfaat dalam memberi gambaran jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek dan dapat menjadi bahan pertimbangan dan referensi pihak-pihak terkait dalam mengelola fasilitas transportasi umum, terutama kereta api.

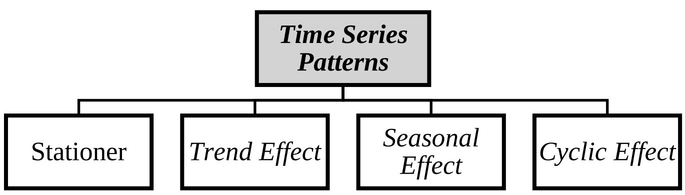
Ada beberapa metode peramalan pada model peramalan deret berkala, salah satu metode dari model deret berkala adalah metode *Box-Jenkins* atau ARIMA (*Autoregressive Integrated Moving Average*). Metode *Box-Jenkins* terdiri dari AR (*Autoregressive*), MA (*Moving Average*), ARMA (untuk data stasioner), ARIMA (untuk data yang tidak stasioner) dan ARIMA musiman atau SARIMA (*Seasonal Autoregressive Integrated Moving Average*, untuk data yang tidak stasioner dan musiman). Pada peramalan jumlah penumpang Kereta Api Indonesia pada masa yang akan datang akan digunakan metode peramalan ARIMA (*Autoregressive Integrated Moving Average*) ditambah dengan pemodelan ARCH/GARCH (*Autoregressive Conditional Heteroscedasticity/Autoregressive Conditional Heteroscedasticity*) karena terlanggarnya asumsi homoskedastisitas, dimana rata-rata dan ragam (residual) data yang dimodelkan secara simultan berubah-ubah. Untuk meramalkan jumlah penumpang Kereta Api Indonesia di Wilayah Jabodetabek 12 periode ke depan (Januari 2021-Desember 2021), dilakukan analisis berdasarkan data jumlah penumpang Kereta Api Indonesia di Wilayah Jabodetabek pada tahun 2011-2020.

1.1. Data Deret Waktu (*Time Series*)

Pada Hanke, J. E dan Wichern, D. W. [7], *Time series* atau runtun waktu adalah himpunan observasi data terurut dalam waktu. Metode *time series* adalah metode peramalan dengan menggunakan analisis pola hubungan antara variabel yang akan dipekirakan dengan variabel waktu. Peramalan suatu data *time series* perlu memperhatikan tipe atau pola data. Menurut Hanke, J. E dan Wichern, D. W. [7], secara umum terdapat empat macam pola data *time series*, yaitu horizontal, *trend*, musiman, dan siklis. Pola horizontal merupakan kejadian yang tidak terduga dan bersifat acak, tetapi kemunculannya dapat mempengaruhi fluktuasi data *time series*. Pola *trend* merupakan kecenderungan arah data dalam jangka panjang, dapat berupa kenaikan maupun penurunan. Pola musiman merupakan fluktuasi dari data yang terjadi secara periodik dalam kurun waktu satu tahun, seperti triwulan, kuartalan, bulanan, mingguan, atau harian, sedangkan pola siklis merupakan fluktuasi dari data untuk waktu yang lebih dari satu tahun.

Data *time series* merupakan serangkaian pengamatan yang terurut berdasarkan waktu dengan jarak yang sama. Jenis data ini sering ditemui dalam keseharian karena data tersebut dikumpulkan melalui waktu interval, yaitu harian, mingguan atau bulanan. Dari data yang terkumpul dapat dilihat ada suatu pola didalamnya. Pola gerakan data atau nilai-nilai variabel dari data *time series* yang diketahui dapat dijadikan sebagai dasar untuk:

a) Pembuatan keputusan saat ini;
b) Peramalan keadaan pada masa yang akan datang;
c) Perencanaan kegiatan untuk masa depan.



Dalam *time series*, pola tersebut secara umum dibagi menjadi empat, yaitu pola stasioner, *trend*, musiman (*seasonal*), dan siklis. Pola stasioner berarti bahwa tidak terdapat perubahan yang drastis pada data atau jika berdasarkan rata-rata atau variansnya bersifat tetap. Fluktuasi data berada di sekitar suatu nilai rata-rata yang konstan, serta tidak tergantung pada waktu dan varians dari fluktuasi tersebut.

1.2. Peramalan (*Forecasting*)

Peramalan adalah proses perkiraan (pengukuran) besarnya atau jumlah sesuatu pada waktu yang akan datang berdasarkan data pada masa lampau yang dianalisis secara ilmiah khususnya menggunakan metode statistika, seperti pada Sudjana [13]. Menurut Makridakis, dkk. [9], peramalan adalah dasar dari segala jenis perencanaan dimana hal ini sangat diperlukan untuk lingkungan yang tidak stabil, yaitu menjembatani antara sistem dengan lingkungan. Teknik peramalan diklasifikasikan menjadi tiga divisi utama: *time series*, *causal* dan *judgemental*. Model peramalan *time series* mencoba memprediksi masa depan dengan menggunakan data historis, model peramalan kausal (regresi) menggabungkan variabel atau faktor yang mungkin mempengaruhi kualitas menjadi peramalan menjadi model peramalan [4].

Perkiraan atau pengukuran dapat dilakukan secara kualitatif maupun kuantitatif. Perkiraan secara kualitatif biasanya menggunakan pendapat dari para ahli pada bidangnya, sedangkan perkiraan secara kuantitatif menggunakan metode statistik dan matematik yang selanjutnya metode ini banyak dipakai, salah satunya adalah metode deret berkala, seperti pada Awat [3].

Menurut Sudjana [13], terdapat dua jenis model peramalan yang utama, yaitu model deret berkala dan model regresi. Pada model deret berkala, pendugaan masa depan dilakukan berdasarkan nilai pada masa lalu dari suatu variabel dan kesalahan pada masa lalu. Tujuan model deret berkala adalah untuk menemukan pola dalam deret waktu historis dan mengekstrapolasikan pola tersebut ke masa depan, sedangkan model regresi mengasumsikan bahwa faktor yang diramalkan menunjukkan suatu hubungan sebab akibat dengan satu atau lebih variabel bebas.

1.3. Pemilihan Model Peramalan (*Forecasting*) Terbaik

Peramalan merupakan alat bantu yang penting dalam perencanaan yang efektif dan efisien, seperti pada Makridakis, dkk. [10]. Dalam membuat peramalan, diusahakan pengaruh ketidakpastian pada model peramalan dapat seminimum mungkin. Dengan kata lain, peramalan bertujuan agar perkiraan yang dibuat meminimumkan kesalahan prediksi (*forecast error*). Ukuran yang digunakan untuk menghitung kesalahan dalam peramalan atau prediksi adalah sebagai berikut.

a. ME (*Mean Error*)

$$\frac{1}{n}\sum_{i=1}^{n} e_i$$

b. MSE (*Mean Squared Error*)

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2$$

c. RMSE (*Root Mean Squared Error*)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$$

d. MAE (*Mean Absolute Error*)

$$\frac{1}{n}\sum_{i=1}^{n} |e_i|$$

e. MPE (*Mean Percentage Error*)

$$\frac{1}{n}\sum_{i=1}^{n}\frac{e_i}{y_i}\times 100$$

f. MAPE (*Mean Absolute Percentage Error*)

$$\frac{1}{n}\sum_{i=1}^{n}|e_i|\times 100$$

g. MASE (*Mean Absolute Scaled Error*)

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|e_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|}$$

Keterangan:

e_i = kesalahan (*error*)

y_i = data real ke-i;

y_{i-1} = data real ke-i-1;

n = banyaknya data hasil *forecasting*

Semakin kecil nilai yang dihasilkan oleh ukuran di atas, maka model peramalan yang digunakan akan semakin baik. Dari ketujuh alat ukur diatas, MAPE (*Mean Absolute Percentage Error*) merupakan yang paling sering digunakan dan merupakan ukuran ketetapan relatif berupa persentase penyimpangan hasil peramalan.

Banyak kandidat model yang bisa mewakili data sebagai hasil dari proses analisis data deret waktu. Untuk mengetahui model yang paling tepat atau terbaik, dapat dilihat dari perhitungan model residual berdasarkan *forecast error* atau kesalahan peramalan dengan ukurannya yang telah diuraikan sebelumnya. Selain dari ukuran-ukuran tersebut, untuk menentukan model yang lebih baik atau paling optimum, yaitu dengan melihat nilai AIC masing-masing model. Dalam membandingkan 2 atau lebih model ARIMA, maka model dengan AIC yang paling kecil merupakan model yang lebih baik. Rumusan AIC (*Akaikes Information Criterion*) sebagi berikut.

$$AIC = n\,log\left(\frac{RSS}{n}\right) + 2k$$

dengan:

$$RSS\ (Residual\ Sum\ of\ Square) = \sum_{i=1}^{n}\hat{e}_i^2$$

2. METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah data sekunder, yaitu data jumlah penumpang Kereta Api Indonesia (ribu orang) untuk wilayah Jabodetabek dari bulan Januari tahun 2011 sampai bulan Desember tahun 2020. Data ini bersumber dari PT. Kereta Api Indonesia dan PT. KAI *Commuter* Jabodetabek yang diunduh dari laman resmi Badan Pusat Statistika (www.bps.go.id) [4]. Pada analisis kasus ini, variabel yang digunakan adalah jumlah penumpang Kereta Api Indonesia di Wilayah Jabodetabek. Struktur data untuk analisis peramalan pada jumlah penumpang Kereta Api Indonesia di Wilayah Jabodetabek tahun 2011-2020 adalah sebagai berikut.

Tabel 2.1. *Struktur Data Jumlah Penumpang Kereta Api di Wilayah Jabodetabek Tahun 2011-2020 (dalam Ribuan Orang)*

Tahun	Bulan				
	Januari	Februari	Maret	…	Desember

2011	Z_{11}	Z_{12}	.	.	.
2012	Z_{21}	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
2020	.	.	.	.	$Z_{10;12}$

2.1. ARIMA (*Autoregressive Integrated Moving Average*)

ARIMA (*Autoregressive Integrated Moving Average*) merupakan salah satu metode peramalan yang disebut juga metode analisis runtun waktu *Box-Jenkins*. Ketepatan peramalan jangka pendek dengan ARIMA sangat baik, tetapi untuk peramalan jangka panjang, ketepatan peramalannya kurang baik karena biasanya untuk periode yang cukup panjang akan cenderung mendatar atau konstan. Untuk menghasilkan peramalan jangka pendek yang akurat, ARIMA menggunakan data masa lalu dan sekarang dari suatu variabel (*univariate*) deret waktu. ARIMA cocok digunakan jika observasi dari deret waktu (*time series*) secara statistik berhubungan satu sama lain (dependen).

Dalam peramalan menggunakan model ARIMA, variabel independen diabaikan secara penuh. Selain itu, diasumsikan juga bahwa data yang digunakan harus stasioner. Stasioneritas berarti tidak terdapat pertumbuhan atau penurunan pada data. 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.

Apabila data deret waktu yang digunakan tidak stasioner, perlu dilakukan penyesuaian untuk menghasilkan data yang stasioner, baik dalam varians, maupun dalam rata-ratanya. Pengubahannya menjadi data stasioner dapat dilakukan dengan *differencing*, yaitu 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.

Pembentukan 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.

Model *Box-Jenkins* (ARIMA) dibagi ke dalam 3 kelompok, yaitu: model *autoregressive* (AR), *moving average* (MA), dan model campuran ARIMA (*autoregresive moving average*) yang mempunyai karakteristik dari dua model pertama, seperti pada Ali [1].

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

Model *autoregressive* (AR) menyatakan suatu ramalan sebagai fungsi nilai-nilai sebelumnya dari time series tertentu (Makridakis, dkk. [11]). Model *autoregressive* dengan ordo p (AR(*p*)) atau model ARIMA (*p*,0,0) memiliki bentuk umum sebagai berikut.

$$X_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t$$

dengan:

μ' = suatu konstanta

ϕ_p = parameter *autoregressive* ke-*p*

e_t = nilai kesalahan pada saat *t*

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

Moving Average (MA) merupakan nilai *time series* pada waktu t yang dipengaruhi oleh unsur kesalahan pada saat ini dan unsur kesalahan terbobot pada masa lalu (Makridakis, dkk. [11]). Model *moving average* dengan ordo *q* (MA(*q*)) atau model ARIMA (0,0,*q*) memiliki bentuk umum sebagai berikut.

$$X_t = \mu' + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-k}$$

dengan:

μ' = suatu konstanta
θ_q = parameter *moving average* ke-q
e_{t-k} = nilai kesalahan pada saat $t - k$

3) Model Campuran

a. Proses ARMA

Model umum untuk campuran proses AR(1) murni dan MA(1) murni, misal ARIMA (1,0,1) dinyatakan sebagai berikut.

*Xt=μ′+ϕ1Xt−1+et−θ1et−1*$X_t = \mu' + \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1}$

atau

$$(1 - \phi_1 B)X_t = \mu' + (1 - \theta_1 B)e_t$$

AR(1) MA(1)

b. Proses ARIMA

Apabila nonstasioneritas ditambahkan pada campuran proses ARMA, maka model umum ARIMA (p,d,q) terpenuhi. Persamaan untuk kasus sederhana ARIMA (1,1,1) adalah sebagai berikut.

$$(1 - B)(1 - \phi_1 B)X_t = \mu' + (1 - \theta_1 B)e_t$$

AR(1) MA(1)

dengan telah dilakukannya *differencing* sebanyak satu kali.

2.2. ARCH (*Autoregressive Conditional Heteroscedasticity*)

ARCH (*Autoregressive Conditional Heteroscedasticity*) merupakan metode yang secara eksplisit memdelkan perubahan ragam (varians) dari waktu ke waktu dalam data deret waktu yang dikembangkan oleh *Engle*. Perubahan ragam residual ini merupakan akibat dari ragam residual yang tidak hanya fungsi dari peubah bebas, tetapi juga tergantung seberapa besar residual di masa lalu.

Pada model ARCH(m), diperoleh variansi dari r_t kondisional terhadap informasi masa lalu sebagai:

$$Var(r_t|\zeta_{t-1}) = E(\varepsilon_t^2|F_{t-1}) = \sigma_t^2$$

Dapat digambarkan dengan persamaan:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i \varepsilon_{t-i}^2$$

dengan $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1,2, \ldots, m$. Kondisi $\alpha_i \geq 0$ diperlukan agar persamaan vitalitas tidak negatif. Jika semua $\alpha_i = 0$, maka variansi bersyarat σ_t^2 akan menjadi konstanta α_0 dan ε_t akan merupakan proses IID (*Independent and Identically Distributed*) yang bersifat homoskedastis.

Pada model ARCH, proses ε_t menunjukkan komponen random dari model (sering disebut sebagai proses *White Noise*), dimana $(\varepsilon_t) = 0$ dan bersifat tidak berkorelasi dengan waktu lampau atau waktu yang akan datang, yakni:

$$E(\varepsilon_t, \varepsilon_s) = \begin{cases} 0, & t \neq s \\ \sigma_t^2, & t = s \end{cases}$$

Dari persamaan di atas, meskipun proses ε_t bersifat non autokorelasi, namun proses ini tidak bersifat independen. Dalam model ARCH(m), proses ε_t dibangkitkan oleh proses:

$$\varepsilon_t = \sigma_t \upsilon_t$$

dimana σ_t adalah akar positif dari σ_t^2 dan υ_t adalah suatu proses IID (*Independent and Identically Distributed*), υ_t merupakan *white noise* dengan mean nol dan variansi satu, dan υ_t sering diasumsikan berdistribusi normal standar N(0,1) atau berdistribusi *student t*. Selain itu, υ_t independen dengan ε_t, $s < t$. Asumsi ini mengakibatkan distribusi bersyarat dari ε_t akan berdistribusi normal dengan mean 0 dan variansi σ^2 (Wei [15]).

Model ARCH (q) dapat dinyatakan dalam bentuk persamaan berikut.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

2.3. GARCH (*Generalized Autoregressive Conditional Heteroscedasticity*)

GARCH (*Generalized Autoregressive Conditional Heteroscedasticity*) merupakan model yang dikembangkan oleh *Bollerslev*. Menurut Enders [6], model ini adalah pengembangan dari model ARCH yang dibangun untuk menghindari ordo yang terlalu tinggi dengan berdasar pada prinsip *parsimony* atau memilih model yang lebih sederhana sehingga akan menjamin variansinya selalu positif. Dengan prinsip *parsimony* ini juga menjadikan GARCH mencapai prediksi yang lebih baik dengan variabel yang lebih sedikit dan menghindari *overfitting* [5].

Model GARCH (p, q) dapat dinyatakan dalam bentuk persamaan berikut.

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \lambda_1 \sigma_{t-1}^2 + \cdots + \lambda_p \sigma_{t-p}^2$$

dengan:

σ_t^2 = varians kondisional
α_0, α_1 = konstanta
σ_{t-p}^2 = varians kondisional pada bulan $t-p$
ε_{t-q}^2 = residual pada bulan $t-q$
λ_p = parameter GARCH, untuk $p = 1, 2, \ldots, p$
α_q = parameter ARCH, untuk $q = 1, 2, \ldots, q$

GARCH adalah pengembangan dari model ARCH yang menggabungkan komponen *moving average* (MA) bersama dengan komponen *autoregressive* (AR). Model GARCH (p, q) merupakan model deret waktu yang menyatakan kondisi heteroskedastisitas dalam waktu t sebagai fungsi linier dari residual kuadrat ε^2 dan varians bersyarat waktu sebelumnya [2]. Secara khusus, model ini mencakup istilah varians *lag* (misalnya pengamatan jika memodelkan kesalahan residual *white noise* dari proses lain), bersama dengan *lag* kesalahan residual dari proses rata-rata.

Pengenalan *moving average* memungkinkan model untuk memodelkan perubahan bersyarat dalam varians atau ragam dari waktu ke waktu serta perubahan dalam varians yang bergantung pada waktu. Contohnya termasuk kenaikan dan penurunan bersyarat dalam varians. Dengan demikian, model GARCH memperkenalkan parameter baru "*p*" yang menjelaskan jumlah istilah varians *lag*:

1) *p* = Jumlah varians lag untuk disertakan dalam model GARCH
2) *q* = Jumlah kesalahan residual *lag* yang akan disertakan dalam model GARCH

Notasi yang diterima secara umum untuk model GARCH adalah dengan menentukan fungsi GARCH () dengan parameter *p* dan *q* GARCH (*p*, *q*); misalnya GARCH (1, 1) akan menjadi model GARCH orde pertama. Model GARCH memasukkan model ARCH, di mana GARCH (0, *q*) setara dengan model ARCH (*q*).

Untuk *p* = 0 proses, direduksi menjadi proses ARCH (*q*), dan untuk *p* = *q* = 0, E(*t*) hanyalah *white noise*. Dalam proses ARCH (*q*), varians kondisional ditentukan sebagai fungsi linier dari varians sampel masa lalu saja, sedangkan proses GARCH (*p*, *q*) memungkinkan varian kondisional tertinggal (*lag*) untuk masuk juga.

Seperti halnya ARCH, GARCH memprediksi varians masa depan dan mengharapkan bahwa rangkaian tersebut stasioner, selain perubahan dalam varians, yang berarti tidak memiliki tren atau komponen musiman. Selain itu, tidak semua data deret waktu mengandung efek heteroskedastisitas. Oleh karena itu, diperlukan adanya pengujian untuk mengetahui ada tidaknya efek heteroskedastisitas dalam data runtun waktu. Menurut Wei [15], pengujian dapat dilakukan dengan cara memeriksa fungsi autokorelasi dari kuadrat residual atau ε^2.

2.4. Tahapan Analisis

Data jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek akan dianalisis menggunakan ARIMA (*Autoregressive Integrated Moving Average*) ditambah dengan pemodelan ARCH/GARCH (*Autoregressive Conditional Heteroscedasticity/Autoregressive Conditional Heteroscedasticity*). Prosedur *Box-Jenkins* merupakan prosedur yang paling populer untuk pemodelan ARIMA.

Ringkasan langkah-langkah dalam pemodelan ARIMA dengan prosedur Box-Jenkins dan ARCH/GARCH (*Autoregressive Conditional Heteroscedasticity/Autoregressive Conditional Heteroscedasticity*) adalah sebagai berikut.

1. Tahap pertama adalah pemeriksaan pola data. *Plotting* data ditampilkan dalam bentuk grafik garis yang diperlukan untuk melihat tren dan pola dalam data sebagai gambaran awal.
2. Melakukan uji stasioneritas dalam varians dengan memperhatikan *lambda Box-Cox*. Jika data tidak stasioner dalam varians, maka dilakukan transformasi *Box-Cox* pada data.
3. Melakukan uji stationeritas dalam rata-rata dengan menggunakan uji *Augmented Dickey Fuller* (ADF) dari data yang sudah stasioner dalam varians. Namun, apabila data belum stasioner pada nilai rata-ratanya (*means*), maka dilakukan proses *differencing*. Jika pada hasil *differencing* satu kali data masih tidak stasioner, maka harus dilakukan kembali *differencing* hingga dihasilkan data yang stasioner.
4. Membuat *plot Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) untuk mengidentifikasikan atau menetapkan beberapa kemungkinan model sementara ARIMA yang sesuai.

Tabel 2.2. *Ketentuan Identifikasi Model ARIMA*

Model	ACF	PACF
MA(q)	*Cuts off after lag q*	*Dies down*
AR(q)	*Dies down*	*Cuts off after lag p*
ARMA(p,q)	*Dies down*	*Dies down*
AR(p) atau MA(q)	*Cuts off after lag q*	*Cuts off after lag p*
Bukan AR(p) atau MA(q)	*No spike*	*No spike*

5. Penaksiran parameter dari semua model sementara ARIMA apakah parameter yang didapat dari model ARIMA sementara signifikan atau tidak. Model signifikan jika nilai signifikansi kurang dari *alpha* (α) dengan nilai α adalah 0,05. Jika ada lebih dari satu model yang signifikan, maka dilihat dari nilai AIC, yaitu *Akaike's Information Criterion* yang bertujuan untuk mengidentifikasi model dari suatu kumpulan data. Model dengan nilai AIC terkecil yang akan dipilih untuk digunakan dalam langkah selanjutnya.

6. Melakukan *diagnostic checking* untuk membuktikan model cukup memadai atau sudah baik untuk digunakan dalam peramalan. Pemeriksaan diagnostik yang dilakukan dalam penelitian ini yaitu uji normalitas, uji *white noise*, dan uji homoskedastisitas.

   a. Pengujian Normalitas Residual

   H_0 : $F(a_t) = F_0(a_t) artinya\ residual\ berdistribusi\ normal$

   H_1: $F(a_t) \neq F_0(a_t)\ artinya\ residual\ tidak\ berdistribusi\ normal$

   $Statistik\ Uji$: *uji Kolmogorov-Smirnov*

   b. Pengujian Residual *White Noise* Non-Autokorelasi

   H_0 : $\rho_1 = \rho_2 = \ldots = \rho_k = 0$ artinya residual tidak berautokorelasi

   H_1 : minimal ada satu $\rho_k \neq 0$ artinya residual berautokorelasi

   Statistik Uji: uji *Ljung-Box*

   c. Pengujian Homoskedastisitas Residual

   H_0 : $\rho_1 = \rho_2 = \ldots = \rho_k = 0$ artinya residual homogen

   H_1 : minimal ada satu $\rho_k \neq 0$ artinya residual heterogen

   Statistik Uji: uji *Langrange Multiplier Engle*

7. Apabila *diagnostic checking* pada model terbaik ARIMA sementara yang telah ditentukan menghasilkan bahwa asumsi homoskedastisitas residual terlanggar, maka dilakukan pengujian untuk menentukan keberadaan efek ARCH (*Autoregressive Conditional Heteroscedasticity*) dari kuadrat residual model.

Pengujian Efek ARCH

H_0 : tidak terdapat efek ARCH

H_1 : terdapat efek ARCH

Statistik Uji : uji *Langrange Multiplier Engle* ARCH

8. Setelah diketahui bahwa data memiliki efek ARCH, dilakukan pembuatan *plot Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) dari residual kuadrat untuk mengidentifikasikan atau menetapkan beberapa kemungkinan model ARCH/GARCH yang sesuai.

9. Penaksiran parameter dari semua model ARCH/GARCH yang mungkin apakah parameter yang didapat dari model-model tersebut signifikan atau tidak dengan metode *maximum likelihood.* Model signifikan jika nilai signifikansi kurang dari *alpha* (α) dengan nilai α adalah 0,05. Jika ada lebih dari satu model yang signifikan, maka dilihat dari nilai AIC, yaitu *Akaike's Information Criterion* yang bertujuan untuk mengidentifikasi model dari suatu kumpulan data. Model dengan nilai AIC terkecil yang akan dipilih untuk digunakan dalam langkah selanjutnya.

10. Dalam pemilihan model terbaik, perlu diperhatikan beberapa hal sebagai berikut:

    a) Prinsip *parsimony*, yaitu model harus bisa sederhana mungkin. Dalam arti mengandung sesedikit mungkin parameternya, sehingga model lebih stabil;

    b) Model sebisa mungkin memenuhi asumsi-asumsi yang melandasinya;

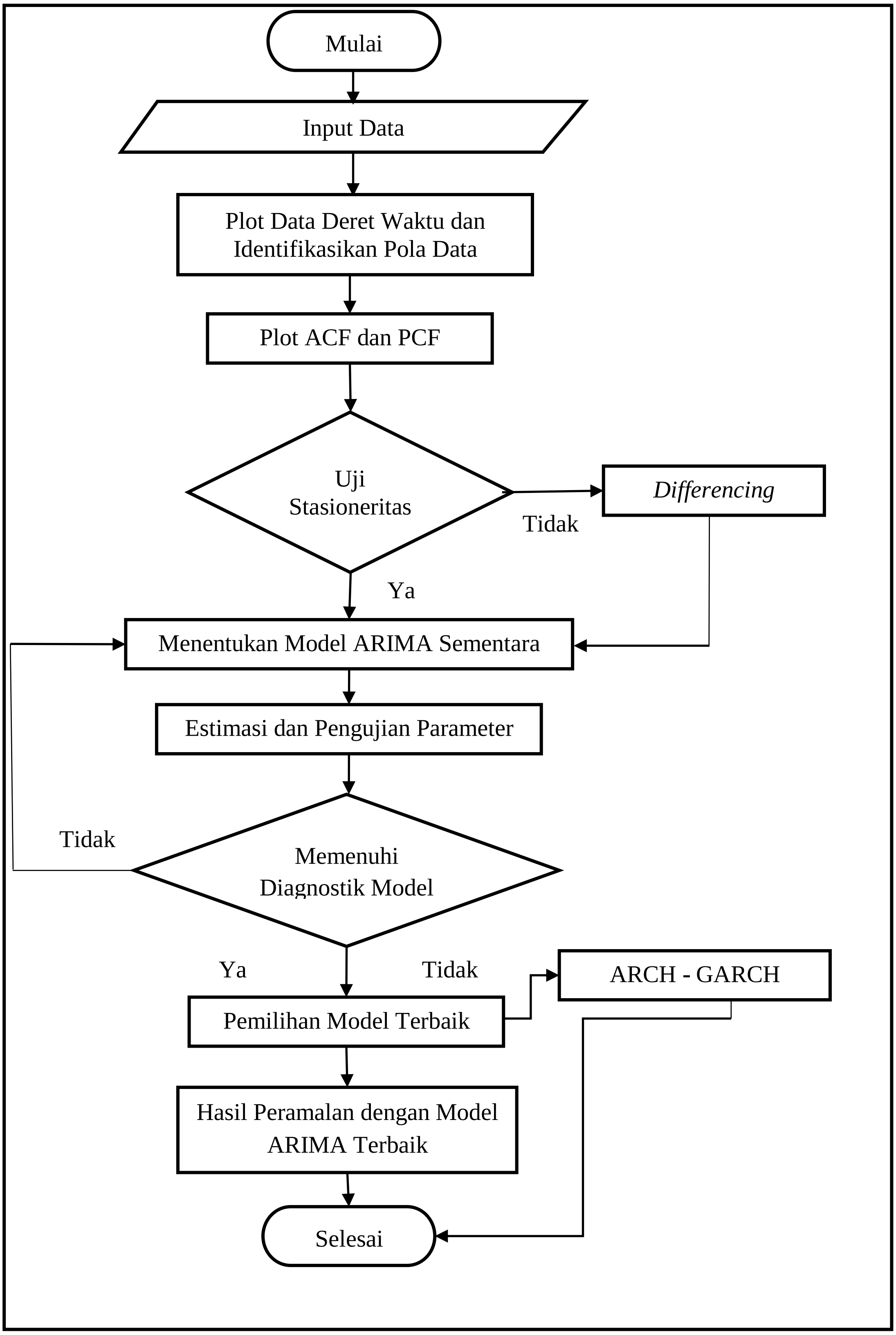
    c) Dalam perbandingan model, selalu pilih model yang paling tinggi akurasinya, yaitu yang memberikan galat (*error*) terkecil.

11. Melakukan peramalan untuk data berdasarkan model terbaik (model ARCH/GARCH dengan nilai AIC paling minimum).

12. Merumuskan kesimpulan dari hasil analisis yang diperoleh.

Dari langkah-langkah dalam pemodelan ARIMA dengan prosedur *Box-Jenkins* di atas disertai metode ARCH-GARCH, dapat dibuat *flowchart*-nya sebagai berikut.

Gambar 2.1. *Flowchart Pemodelan ARIMA dengan ARCH-GARCH*


Mulai
Input Data
Plot Data Deret Waktu dan Identifikasikan Pola Data
Plot ACF dan PCF
Uji Stasioneritas
Tidak
Differencing
Ya
Menentukan Model ARIMA Sementara
Estimasi dan Pengujian Parameter
Tidak
Memenuhi Diagnostik Model
Ya
Tidak
ARCH - GARCH
Pemilihan Model Terbaik
Hasil Peramalan dengan Model ARIMA Terbaik
Selesai

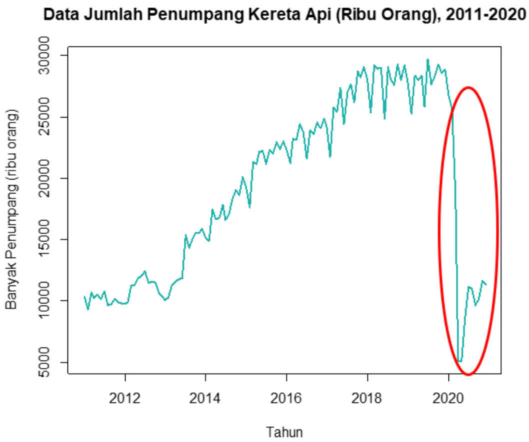
3. HASIL DAN PEMBAHASAN

Data jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek dianalisis menggunakan ARIMA ditambah dengan pemodelan ARCH/GARCH. Analisisnya dilakukan dengan menggunakan ***software* R *Studio***, langkah-langkahnya sebagai berikut.

3.1. *Plot* Data Awal

Untuk melakukan suatu peramalan atau *forecasting*, tahap pertama yang harus dilakukan adalah melakukan pemeriksaan pola data atau *plotting* data. Tujuan *plotting* data adalah untuk melihat *trend* dan pola dalam data.

Data yang digunakan pada penelitian ini adalah data jumlah penumpang Kereta Api Indonesia (ribu orang) untuk wilayah Jabodetabek dari periode bulan Januari tahun 2011 sampai bulan Desember tahun 2020 yang artinya terdapat sebanyak 120 data hasil pengamatan. Dengan menggunakan *software* R, dibentuk *plot* data seperti pada Gambar 3.1.

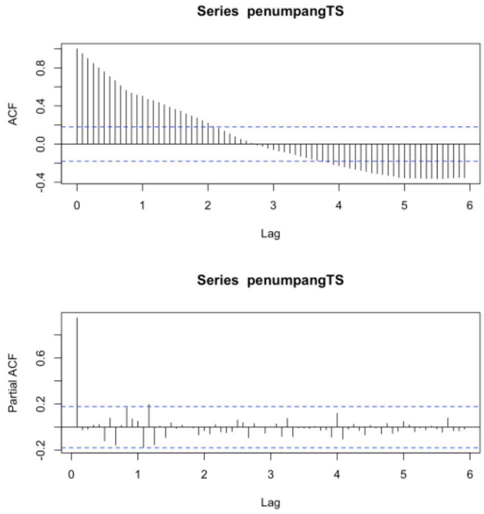


Gambar 3.1. *Plot Pola Sebaran Data Awal*

Berdasarkan grafik di atas, terlihat bahwa adanya indikasi data tidak stasioner. Hal ini dapat diketahui dari grafiknya yang memiliki pola *trend* naik.

Pada grafik yang dilingkari dengan warna merah terjadi perubahan pola data yang pada awalnya mengalami pola tren naik dari tahun ke tahun, berubah ke tingkat rata – rata baru pada tahun 2020. Hal ini diakibatkan oleh adanya kebijakan *social distancing* berkenaan dengan adanya pandemi *covid-19*. Sehingga, mengakibatkan pengaruh yang buruk pada sektor transportasi yang ditandai dengan penurunan drastis pada jumlah penumpang kereta api (ribu orang) Indonesia di wilayah Jabodetabek tahun 2020.

Gambar 3.2. *Plot ACF dan PACF Data Awal*



Gambar 3.2. *Plot ACF dan PACF Data Awal*

Berdasarkan grafik di atas, terlihat bahwa pada ACF data awal terjadi *dying down*, sedangkan pada PACF-nya terjadi *cut-off*.

3.2. Uji Stasioneritas

Stasioneritas berarti bahwa tidak adanya perubahan drastis pada data. Fluktuasi pada data tersebut biasanya berada disekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dari fluktuasi tersebut. Dalam peramalan atau *forecasting* dengan pemodelan ARIMA, data yang digunakan harus memenuhi kriteria stasioner atau konstan terhadap rata-rata dan varians. Berikut ini adalah pengujian stasioneritas dari data jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek tahun 2011-2020.

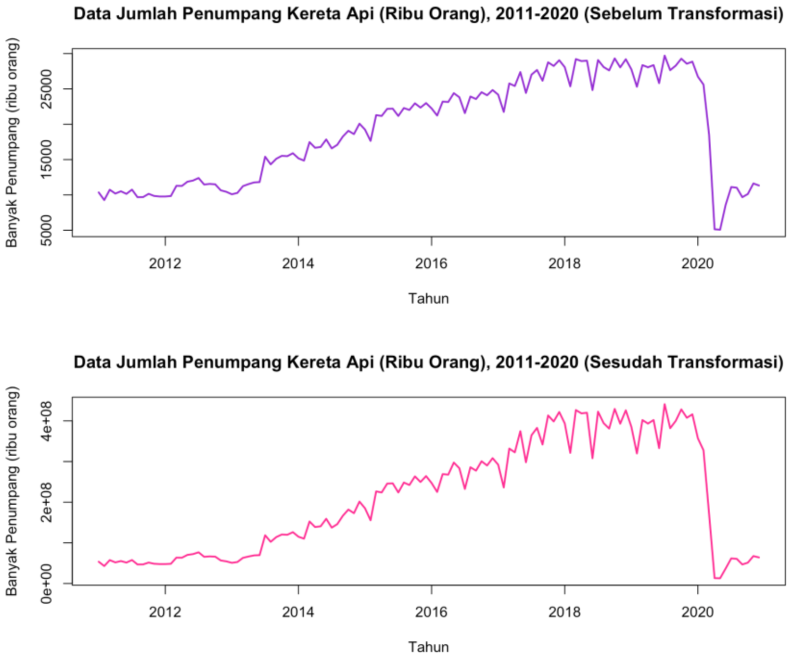
3.2.1. Uji Stasioneritas Varians

Suatu proses stasioner pada varians jika $Var(Z_t) = E(Z_t - \mu_t)^2 = \sigma^2$ adalah konstan untuk setiap t. Pengujian stasioneritas dalam varians dapat menggunakan uji *Box-Cox* dengan ketentuan apabila nilai *lambda* jauh dari satu, maka ada indikasi bahwa data tidak stasioner dalam varians. Jika data tidak stasioner dalam varians, maka dilakukan transformasi data.

Berdasarkan hasil perhitungan, diperoleh nilai *lambda* (λ) sebesar 1,999924 dimana nilai tersebut masih jauh dari 1 yang artinya data terbukti belum stasioner terhadap varians. Maka, data harus ditransformasi dengan menggunakan rumusan berikut.

$$Y_t^* = \frac{Y_t^{\lambda} - 1}{\lambda}$$

Kemudian, dihitung kembali nilai *lambda* (λ) setelah dilakukan transformasi data dan didapatkan nilai sebesar 1,528794. Nilai ini sudah dianggap cukup stasioner terhadap varians.



Gambar 3.3. *Plot Data Sebelum dan Sesudah Transformasi*

3.2.2. Uji Stasioneritas Rata-rata

Suatu proses stasioner dalam rata-rata jika $E(Z_t) = \mu_t = \mu$ adalah konstan untuk setiap t. Terlihat secara visual dari *plot* awal data terindikasi tidak stationer terhadap rata-rata. Namun, untuk memastikan sebuah data stationer atau tidak terhadap rata-ratanya, maka haruslah dilakukan pengujian mengenai kestasioneran data deret waktu dengan menggunakan uji *Augmented Dickey-Fuller* (ADF) pada data hasil transformasi yang telah diperoleh. Pengujian untuk stasioner rata-rata dilakukan sebagai berikut.

Hipotesis
$H_0 : \delta = 0$ (data tidak stasioner terhadap rata-rata)
$H_1 : \delta \neq 0$ (data stasioner terhadap rata-rata)

Taraf Signifikan
$\alpha = 5\% = 0{,}05$

Statistik Uji
Statistik uji yang digunakan untuk menguji stasioneritas terhadap rata-rata dari data *time-series* adalah *Augmented Dickey-Fuller* (ADF) *Test* dengan rumusan sebagai berikut.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + u_t$$

Kriteria Uji
Tolak H_0 jika $p\ value < \alpha$, terima dalam hal lainnya.

Hasil Pengujian
Dengan menggunakan *syntax adf.test()* pada *software* R dan dengan taraf signifikan sebesar 5%, diperoleh $p - value = 0{,}99 > \alpha = 0{,}05$. Maka, dapat disimpulkan bahwa H_0 diterima yang artinya data tidak stasioner terhadap rata-rata. Oleh karena itu, perlu dilakukan proses *differencing*.

3.3. *Differencing* Data

Proses *differencing* atau selisih dilakukan jika data tidak stasioner dalam rata-ratanya. Salah satu cara umum yang dipakai adalah metode pembedahan atau lebih dikenal dengan sebutan *differencing*, yaitu mengurangi nilai data pada suatu periode dengan nilai data pada periode sebelumnya, yang dapat ditulis sebagai berikut.

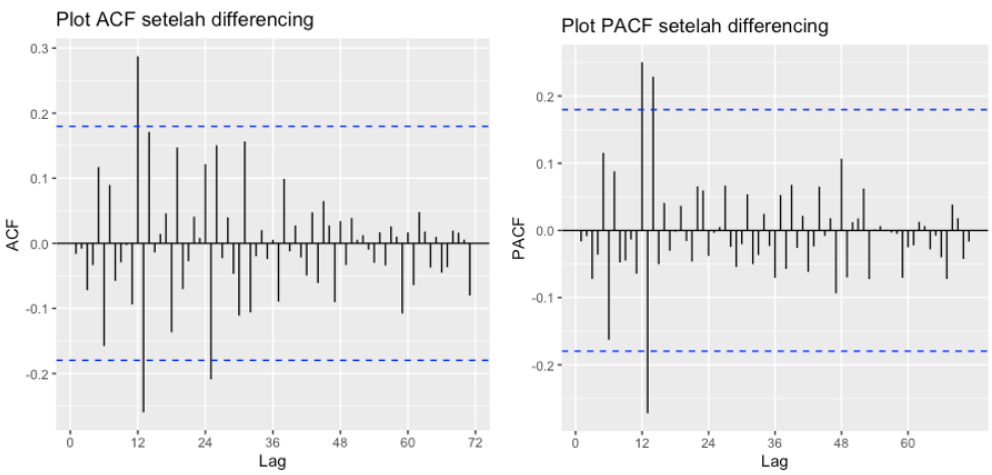
$$W_t = Z_t - Z_{t-1}$$

Setelah proses *differencing* sebanyak satu kali diterapkan pada data jumlah penumpang, selanjutnya dilakukan kembali pengujian stasioneritas terhadap rata-rata dengan *Augmented Dickey-Fuller* (ADF) *Test* dan telah diperoleh $p - value = 0{,}01 < \alpha = 0{,}05$. Maka, dapat disimpulkan bahwa H_0 ditolak yang artinya data sudah stasioner terhadap rata-rata. Oleh karena itu, proses *differencing* sudah cukup dilakukan sampai dengan satu kali proses saja.

Gambar 3.4. *Plot Data Sebelum dan Sesudah Differencing*

3.4. Identifikasi Model ARIMA

Setelah data stasioner dalam varians dan rata-ratanya, langkah selanjutnya adalah membuat *plot Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF) untuk mengidentifikasikan atau menetapkan model sementara ARIMA yang sesuai.



Gambar 3.5. *Plot ACF dan PACF dari Data yang Sudah Stasioner*

Dari *output* R di atas, dapat dilihat di sebelah kiri adalah *plot* ACF dan di sebelah kanan adalah *plot* PACF. Tampak bahwa pada kedua *plot*, yaitu ACF dan PACF terlihat *dying down* karena keduanya baru *cut off* di lag kedua belas. Maka, dapat diambil kesimpulan bahwa ACF dan PACF memiliki pola *dying down*.

Sesuai dengan dasar yang digunakan dalam penentuan model, jika ACF dan PACF menunjukkan *dying down*, maka dapat dikatakan bahwa modelnya adalah ARMA (p, q) yang merupakan gabungan AR (*autoregressive*) dan MA (*moving average*).

Berdasarkan identifikasi model di atas, dengan *differencing* sebanyak satu kali ($d = 1$), maka model ARIMA yang dapat dibentuk adalah ARIMA (0,1,1), ARIMA (0,1,2), ARIMA (1,1,0), ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (2,1,0), ARIMA (2,1,1).

Setelah melakukan proses identifikasi model dan berhasil menetapkan beberapa kemungkinan model ARIMA yang cocok serta mengestimasikan parameternya, selanjutnya tahap yang harus dilakukan adalah uji signifikansi pada koefisien parameter modelnya. Jika koefisien dari kemungkinan model ARIMA yang sudah ditetapkan sebelumnya tidak signifikan, maka model tersebut tidak tepat untuk digunakan dalam melakukan peramalan atau *forecasting* jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek.

Hipotesis
$H_0 : \theta = 0$ (parameter tidak signifikan)
$H_1 : \theta \neq 0$ (parameter signifikan)

Taraf Signifikan
$\alpha = 5\%$

Statistik Uji
Statistik uji yang digunakan untuk menguji signifikansi parameter adalah statistik *t* dengan rumusan sebagai berikut.

$$t_{Hitung} = \frac{\theta}{SE(\hat{\theta})}$$

Kriteria Uji
Tolak H_0 jika $\left(\frac{\alpha}{2}\right): df = n - p \leq |t| \leq -\left(\frac{\alpha}{2}\right): df = n - p$ atau $p - value < \alpha$, terima dalam hal lainnya.

Setelah melakukan pengujian parameter pada tujuh model tentatif di atas, diperoleh hasil dan dapat diambil kesimpulan dengan tabel sebagai berikut.

Tabel 3.1. *Hasil Pengujian Parameter Model ARIMA*

Model Tentatif	Signifikansi	AIC
ARIMA (0,1,1)	Tidak Signifikan	2157,89
ARIMA (0,1,2)	Tidak Signifikan	2159,88
ARIMA (1,1,0)	Tidak Signifikan	2157,89
ARIMA (1,1,1)	**Signifikan**	**2159,87**
ARIMA	Signifikan	2161,69

(1,1,2)		
ARIMA (2,1,0)	TIDAK SIGNIFIKAN	2159,89
ARIMA (2,1,1)	SIGNIFIKAN	2161,68

Berdasarkan tabel di atas, dapat dilihat bahwa model dengan parameter yang signifikan terdapat pada model ARIMA (1,1,1), ARIMA (1,1,2), dan ARIMA (2,1,1). Untuk menentukan model yang paling tepat untuk melakukan peramalan atau *forecasting* jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek dari ketiga model yang signifikan tersebut, dilakukan perbandingan nilai AIC, dimana model yang paling tepat ditandai dengan nilai AIC yang paling kecil dan model ARIMA dengan nilai AIC yang paling kecil adalah model ARIMA (1,1,1). Dengan demikian, model ARIMA (1,1,1) yang akan menjadi model yang lebih mungkin digunakan dalam peramalan atau *forecasting* jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021 dengan terlebih dahulu diuji asumsi-asumsi terhadap residualnya.

3.5. Uji Asumsi Residual

Setelah diperoleh model terbaik, yaitu model yang signifikan dengan nilai AIC (*Akaike Information Criterion*) paling kecil, langkah selanjutnya adalah melakukan uji asumsi residual yang meliputi uji normalitas, uji *white noise*, dan uji homoskedastisitas.

3.5.1. Uji Normalitas

Model ARIMA yang baik adalah model yang residualnya berdistribusi normal. Oleh karena itu, dilakukan uji normalitas residual untuk memeriksa apakah nilai residual yang dihasilkan terdistribusi secara normal atau tidak. Uji normalitas yang digunakan adalah Uji *Kolmogorov-Smirnov*, langkah-langkahnya sebagai berikut.

Hipotesis
$H_0 : F(a_t) = F_0(a_t)$ artinya residual berdistribusi normal
$H_1 : F(a_t) \neq F_0(a_t)$ artinya residual tidak berdistribusi normal

Taraf Signifikan
$\alpha = 5\% = 0{,}05$
Statistik Uji

$$D = \max|F_0(x) - S_n(x)|$$

Keterangan:
$F_0(x)$ = suatu fungsi distribusi kumulatif teoritis dibawah H_0
$S_n(x)$ = distribusi frekuensi kumulatif yang diobservasi dari suatu sampel random n observasi

Kriteria Uji
Tolak H_0 jika $p\ value < \alpha$, terima dalam hal lainnya.

Hasil Pengujian
Dengan menggunakan *syntax ks.test()* pada *software* R dan dengan taraf signifikan sebesar 5%, diperoleh $p-value = 0{,}07134 > \alpha = 0{,}05$. Maka, dapat disimpulkan bahwa H_0 diterima yang artinya residual berdistribusi normal. Dengan demikian, asumsi pertama dari residual model ARIMA yang dibentuk terpenuhi.

3.5.2. Uji *White Noise*

Suatu model ARIMA yang baik bersifat *white noise* yaitu jika nilai residual dari model tersebut bersifat acak dan menunjukkan tidak ada autokorelasi atau residual tidak berpola tertentu. Dengan kata lain, residualnya memenuhi asumsi identik, yaitu residual tidak berautokorelasi, independen dan varians residual homogen. Uji asumsi *white noise* ini dapat dilakukan dengan menggunakan Uji *Ljung-Box*, langkah-langkahnya sebagai berikut.

Hipotesis
$H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$ artinya residual tidak berautokorelasi
H_1 : minimal ada satu $\rho_k \neq 0$ artinya residual berautokorelasi

Taraf Signifikan
$\alpha = 5\% = 0{,}05$

Statistik Uji

$$Q = T(T+1)\sum_{k=1}^{m} \frac{\hat{\rho}_k{}^2}{(T-k)}$$

Keterangan:
T = banyaknya data residual
m = *lag* maksimum
$\hat{\rho}_k$ = autokorelasi pada *lag* ke-*k*

Kriteria Uji
Tolak H_0 jika $p\ value < \alpha$, terima dalam hal lainnya.

Hasil Pengujian
Dengan menggunakan *syntax Box.test()* pada *software* R dan dengan taraf signifikan sebesar 5%, diperoleh $p - value = 0{,}7469 > \alpha = 0{,}05$. Maka, dapat disimpulkan bahwa H_0 diterima yang artinya residual tidak berautokorelasi. Dengan demikian, asumsi kedua dari residual model ARIMA yang dibentuk terpenuhi.

3.5.3. Uji Homoskedastisitas

Model ARIMA dikatakan baik apabila residualnya mempunya varians yang konstan. Oleh karena itu, dilakukan uji homoskedastisitas untuk memeriksa kesamaan variansi residual. Uji asumsi homoskedastisitas ini dapat dilakukan dengan menggunakan Uji *Langrange Multiplier Engle*, langkah-langkahnya sebagai berikut.

Hipotesis
$H_0 : \rho_1 = \rho_2 = \ldots = \rho_k = 0$ artinya residual homogen
H_1 : minimal ada satu $\rho_k \neq 0$ artinya residual heterogen

Taraf Signifikan
$\alpha = 5\% = 0{,}05$

Statistik Uji

$$LM = TR^2$$

Keterangan:
T = banyaknya data residual
R^2 = koefisien determinasi

Kriteria Uji
Tolak H_0 jika $p\ value < \alpha$, terima dalam hal lainnya.

Hasil Pengujian

Dengan menggunakan *syntax Box.test()* dengan tipe *Ljung-Box* pada *software* R dan dengan taraf signifikan sebesar 5%, diperoleh $p - value = 0{,}01093 < \alpha = 0{,}05$. Maka, dapat disimpulkan bahwa H_0 ditolak yang artinya residual heterogen. Dengan demikian, asumsi ketiga dari residual model ARIMA yang dibentuk tidak terpenuhi.

3.6. Pengujian Efek ARCH

Dari uji asumsi residual pada model ARIMA terbaik yang diperoleh, yaitu model ARIMA (1,1,1) yang merupakan model campuran proses AR(1) dan MA(1) dengan satu kali proses *differencing*, didapatkan hasil bahwa model tersebut memenuhi asumsi normalitas residual dan residual bersifat *white noise*, tetapi tidak memenuhi asumsi homoskedastisitas yang artinya residual pada model bersifat heterogen. Dengan hal ini, perlu dilakukan identifikasi adanya efek ARCH pada model tersebut. Langkah-langkah pengujiannya sebagai berikut.

Hipotesis
H_0 : tidak terdapat efek ARCH
H_1 : terdapat efek ARCH

Taraf Signifikan
$\alpha = 5\% = 0{,}05$

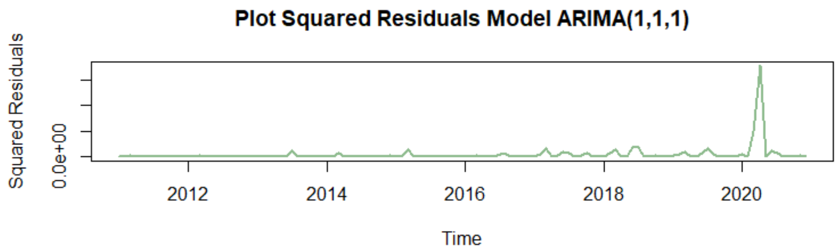
Statistik Uji
Langrange Multiplier Engle ARCH

Kriteria Uji
Tolak H_0 jika $p\ value < \alpha$, terima dalam hal lainnya.

Hasil Pengujian
Dengan menggunakan *syntax Box.test()* dengan tipe *Ljung-Box* pada *software* R dan dengan taraf signifikan sebesar 5%, diperoleh $p - value = 1{,}422 \times 10^{-15} < \alpha = 0{,}05$. Maka, dapat disimpulkan bahwa H_0 ditolak yang artinya terdapat efek ARCH pada data jumlah penumpang (ribu orang) Kereta Api Indonesia di Wilayah Jabodetabek tahun 2011-2020.

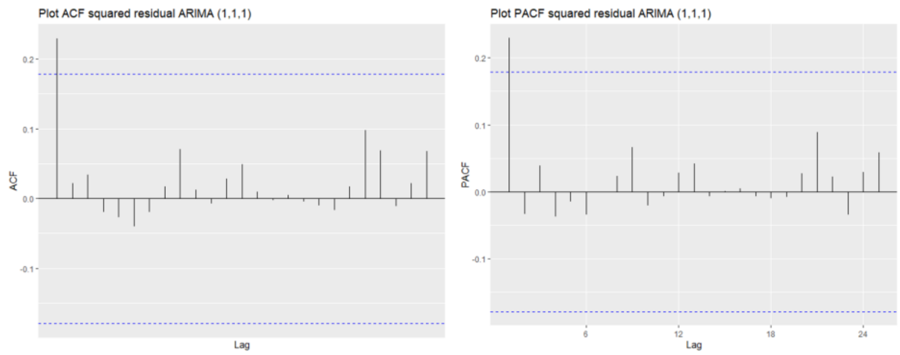
3.7. Identifikasi Model ARCH/GARCH

Setelah dilakukan pengujian efek ARCH dan dihasilkan bahwa adanya efek ARCH, selanjutnya dilakukan identifikasi model ARCH (*p*) dan GARCH (*p*,*q*). Identifikasi model ARCH (*p*) dan GARCH (*p*,*q*) dilakukan melalui pengamatan terhadap plot ACF dan PACF residual kuadrat model ARIMA (1,1,1).



Gambar 3.6. *Plot Residual Kuadrat Model ARIMA (1,1,1)*

Grafik *plot* kuadrat residual di atas (Gambar 3.6) menunjukkan adanya klaster volatilitas di beberapa titik, terutama di sebelah kanan setelah tahun 2020.



Gambar 3.7. *Plot ACF dan PACF Residual Kuadrat Model ARIMA (1,1,1)*

Dari *output* R di atas, dapat dilihat di Gambar 3.7 sebelah kiri adalah *plot* ACF dan di sebelah kanan adalah *plot* PACF. Tampak bahwa pada kedua *plot*, yaitu ACF dan PACF terlihat mengalami *cut-off* pada lag pertama. Maka, dapat diambil kesimpulan bahwa ACF dan PACF memiliki pola *cut-off* pada lag pertama.

Berdasarkan identifikasi model di atas, dapat dibentuk satu model ARCH (*p*) dan satu model GARCH (*p*,*q*). Dengan demikian, model dugaan ARCH dan GARCH yang dapat digunakan adalah ARCH (1) dan GARCH (1,1).

3.8. Penaksiran dan Pengujian Signifikansi Parameter ARCH/GARCH

Langkah yang dilakukan selanjutnya adalah penaksiran parameter untuk setiap model. Dari hasil penaksiran parameter tersebut dilakukan uji signifikansi parameter untuk dilihat apakah parameter tersebut layak masuk ke dalam model atau tidak. Langkah-langkah pengujiannya sebagai berikut.

Hipotesis
H_0 : θ = 0 (parameter tidak signifikan)
H_1 : θ ≠ 0 (parameter signifikan)

Taraf Signifikan
$\alpha = 5\% = 0{,}05$

Statistik Uji
Statistik uji yang digunakan untuk menguji signifikansi parameter adalah statistik *t* dengan rumusan sebagai berikut.

$$t_{hitung} = \frac{\hat{\theta}}{SE(\hat{\theta})}$$

Kriteria Uji
Tolak H_0 jika $p\ value < \alpha$, terima dalam hal lainnya.

Hasil Pengujian
Dengan menggunakan *software R*, didapatkan hasil penaksiran dan pengujian signifikansi parameter ARCH/GARCH yang disajikan dalam tabel berikut.

Tabel 3.2. *Hasil Penaksiran dan Pengujian Signifikansi Parameter ARCH/GARCH*

Model	Parameter	Estimasi	*Standard Error*	t-hitung	*p-value*	Keterangan
ARCH (1)	α_0	48298,11	1,5074e+05	0,32041	0,748657	Tidak Signifi

						kan
	α_1	0,960 30	7,5647e -02	12,69 451	0	Signifi kan
GAR CH (1,1)	α_0	48298 ,11	1,4930e +05	0,323 49	0,746 321	Tidak Signifi kan
	α_1	0,194 39	2,3673e -01	0,821 13	0,411 574	Tidak Signifi kan
	β_1	0,766 37	1,2244e -01	6,258 95	0	Signifi kan

Berdasarkan tabel di atas, dapat dilihat bahwa model ARCH (1) terdapat parameter α_1 yang signifikan/memiliki keberartian dalam model dan parameter α_0 yang tidak signifikan. Dan pada model GARCH (1,1) terdapat parameter α_0 dan α_1 yang tidak signifikan dalam model, tetapi parameter β_1 memiliki keberartian dalam model.

3.9. Pemilihan Model ARCH/GARCH Terbaik

Untuk menentukan model yang paling tepat untuk melakukan peramalan atau *forecasting* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek dari kedua model yang di atas, dilakukan perbandingan nilai AIC, dimana model yang paling tepat ditandai dengan nilai AIC yang paling kecil.

Tabel 3.3. *Nilai AIC Model ARCH/GARCH*

Model	AIC
ARCH (1)	18,496
GARCH (1,1)	18,314

Model dengan nilai AIC yang paling kecil adalah model GARCH (1,1) yaitu sebesar 18,314. Dengan demikian, model GARCH (1,1) yang akan menjadi model yang lebih mungkin digunakan dalam peramalan atau *forecasting* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021 dan diperoleh persamaan dari model GARCH (1,1) sebagai berikut:

$$\sigma_t^2 = 48298{,}11036 + 0{,}19439\varepsilon_{t-1}^2 + (0{,}76637)\sigma_{t-1}^2$$

Artinya, bahwa variansi jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek pada periode ke-t ditemukan oleh suatu konstanta (48298,11036) dan sisaan/residual kuadrat pada periode sebelumnya yaitu t – 1 dengan proporsi sebesar 19,439% serta varians residual pada periode sebelumnya yaitu t – 1 dengan proporsi sebesar 76,637%.

3.10. Peramalan (*Forecasting*)

Setelah didapatkan model terbaik, yaitu model ARIMA (1,1,1) dan varians residual GARCH (1,1), maka selanjutnya dilakukan peramalan atau *forecasting* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021.

Tabel 3.4. *Hasil Peramalan Jumlah Penumpang (Ribu Orang) Kereta Api Indonesia di Wilayah Jabodetabek 12 Periode ke Depan*

Periode	*Forecast*

Januari	2021	30173
Februari	2021	30268
Maret	2021	30362
April	2021	30455
Mei	2021	30547
Juni	2021	30638
Juli	2021	30727
Agustus	2021	30816
September	2021	30904
Oktober	2021	30991
November	2021	31077
Desember	2021	31162

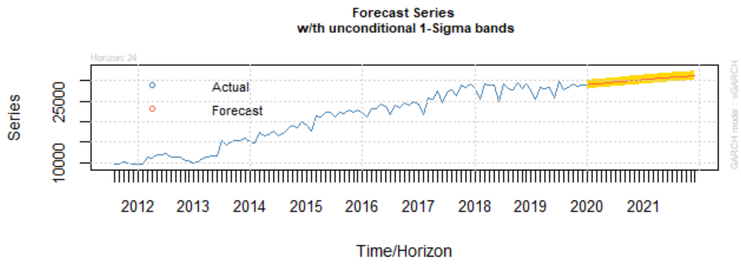
Tabel 3.4 di atas memuat hasil peramalan atau *forecasting* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021. Nilai hasil peramalan menunjukkan terus adanya kenaikan jumlah penumpang di setiap bulannya pada tahun 2021.

Untuk memeriksa kebenaran metode dan analisis yang dilakukan, data hasil peramalan dengan menggunakan model ARIMA (1,1,1) dan varians residual GARCH (1,1) dibandingkan dengan data asli terbaru, yaitu data jumlah penumpang Kereta Api Indonesia (ribu orang) untuk wilayah Jabodetabek dari bulan Januari tahun 2021 sampai bulan September tahun 2021 yang bersumber dari PT. Kereta Api Indonesia dan PT. KAI *Commuter* Jabodetabek dan diperoleh dari laman resmi Badan Pusat Statistika (www.bps.go.id) [4].

Tabel 3.5. *Perbandingan Hasil Peramalan Jumlah Penumpang (Ribu Orang) Kereta Api Indonesia di Wilayah Jabodetabek dengan Data Asli Terbaru*

Periode	*Forecast*	Data Asli	Selisih
Januari	30173	10149	20024
Februari	30268	9796	20472
Maret	30362	12041	18321
April	30455	12452	18003
Mei	30547	12230	18317
Juni	30638	11978	18660
Juli	30727	5102	25625
Agustus	30816	5947	24869
September	30904	8693	22211

Dari Tabel 3.5 di atas, sangat jelas terlihat bahwa adanya perbedaan (selisihnya) yang cukup besar antara nilai hasil peramalan dengan nilai data aslinya. Hal ini menunjukkan bahwa penggunaan model ARIMA (1,1,1) dan varians residual GARCH (1,1) masih kurang tepat digunakan pada peramalan jumlah penumpang Kereta Api Indonesia (ribu orang) untuk wilayah Jabodetabek dari bulan Januari tahun 2021 sampai bulan September tahun 2021.



Gambar 3.8. *Plot Peramalan Jumlah Penumpang Kereta Api (Ribu Orang) Indonesia di Wilayah Jabodetabek 12 Periode ke Depan*

Dapat dilihat pula plot peramalan 12 periode ke depan yang disajikan pada gambar 3.8. Dengan garis berwarna biru merupakan data actual jumlah penumpang kereta api (ribu orang) Indonesia di wilayah Jabodetabek yang digunakan untuk melatih model dari Januari 2012 hingga Desember 2020. Pada data *actual* atau data asli ini dapat dilihat bahwa pola pada Januari 2012 hingga Desember 2020 mengalami variasi musiman. Terdapat kenaikan plot yang disebabkan oleh adanya musim liburan dan liburan hari Raya. Dan penurunan plot setelah musim liburan berakhir.

Kemudian terdapat garis berwarna kuning adalah hasil peramalan jumlah penumpang kereta api (ribu orang) Indonesia di wilayah Jabodetabek 12 periode ke depan yaitu dari Januari 2021 hingga Desember 2021 menggunakan model GARCH (1,1).

3.11. Ketepatan Peramalan

Setelah didapatkan hasil peramalan atau *forecasting* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021, maka perlu dilakukan evaluasi mengenai ketepatan model peramalan antara persamaan model GARCH (1,1) yang telah terbentuk dengan data jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek. Dalam hal ini, MAPE digunakan sebagai ukurannya. Nilai MAPE ini sendiri mengindikasikan mengenai seberapa besar kesalahan dalam peramalan yang dibandingkan dengan nilai nyata pada deret.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{\left|Y_t - \hat{Y}_t\right|}{Y_t}$$

Nilai MAPE hasil peramalan atau *forecasting* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021 adalah sebesar 1,943536. Didapatkan nilai MAPE yang kecil, artinya nilai peramalan tidak jauh berbeda dengan data asli. Hal ini dapat dilihat dari nilai selisih data asli dengan data hasil peramalan yang terdapat pada tabel 3.5. Maka, model yang terbentuk sudah cukup baik untuk memprediksi jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021.

4. KESIMPULAN

Peramalan data jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek untuk 12 periode (1 tahun) ke depan, yaitu tahun 2021, dengan menggunakan model ARIMA (1,1,1) tidak dapat mengatasi adanya efek heteroskedastisitas. Maka, digunakan model ARCH/GARCH dengan model terbaik adalah GARCH (1,1).

Dengan nilai MAPE sebesar 1,943536, dapat dikatakan bahwa model GARCH (1,1) mampu memprediksi secara baik data jumlah penumpang (ribu orang) kereta api Indonesia di wilayah

Jabodetabek untuk 12 periode (satu tahun) ke depan, yaitu pada tahun 2021. Persamaan model peramalan GARCH (1,1) tersebut adalah sebagai berikut.

$$\sigma_t^2 = 48298{,}11036 + 0{,}19439\varepsilon_{t-1}^2 + (0{,}76637)\sigma_{t-1}^2$$

Artinya, bahwa variansi jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek pada periode ke-t ditemukan oleh suatu konstanta (48298,11036) yang merupakan nilai koefisien α_0 dan sisaan kuadrat pada periode sebelumnya dengan proporsi sebesar 19,439% yang merupakan nilai koefisien ARCH (1) (α_i) serta varians residual pada periode sebelumnya yaitu t – 1 dengan proporsi sebesar 76,637% yang merupakan nilai koefisien GARCH (1) (β_j).

Dari hasil peramalan menggunakan *software R*, didapatkan juga hasil peramalan untuk tahun 2020 selama 12 periode. Hasil ini dapat dibandingkan dengan nilai *actual* dari data jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek pada tahun 2020 yang disajikan sebagai berikut.

Tabel 4.1. *Perbandingan Data Actual dan Predicted/Forecast Tahun 2020*

Periode		*Actual*	*Forecast*
Januari	2020	26733	28950
Februari	2020	25616	29058
Maret	2020	18548	29165
April	2020	5138	29271
Mei	2020	5077	29376
Juni	2020	8591	29479
Juli	2020	11116	29582
Agustus	2020	11014	29683
September	2020	9678	29783
Oktober	2020	10128	29882
November	2020	11622	29980
Desember	2020	11330	30077

Pada data *actual* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek tahun 2020 dapat dilihat adanya penurunan angka yang drastis, hal ini disebabkan oleh adanya pandemi *covid-19* yang mengharuskan adanya *social distancing*. Masyarakat dihimbau untuk berada di rumah dan tidak bepergian jika bukan suatu hal yang penting. Hal ini berpengaruh buruk pada beberapa sektor di Indonesia salah satunya adalah sektor transportasi. Dapat dibuktikan dengan adanya perbedaan yang cukup besar pada hasil *forecast* jumlah penumpang (ribu orang) kereta api Indonesia di wilayah Jabodetabek tahun 2020 jika dibandingkan dengan data *actual* pada tahun 2020. Jika tidak terdapat pandemi *covid-19*, data *actual* pada beberapa bulan di tahun 2020 tidak akan berbeda jauh dengan hasil *forecast* yang didapat.

Untuk memeriksa kebenaran metode dan analisis yang dilakukan, data hasil peramalan dengan menggunakan model ARIMA (1,1,1) dan varians residual GARCH (1,1) dibandingkan dengan data asli terbaru, yaitu data jumlah penumpang Kereta Api Indonesia (ribu orang) untuk wilayah Jabodetabek dari bulan Januari tahun 2021 sampai bulan September tahun 2021 yang bersumber dari PT. Kereta Api Indonesia dan PT. KAI *Commuter* Jabodetabek dan diperoleh dari laman resmi Badan Pusat Statistika (www.bps.go.id) [4].

Tabel 4.2. *Perbandingan Hasil Peramalan Jumlah Penumpang (Ribu Orang) Kereta Api Indonesia di Wilayah Jabodetabek dengan Data Asli Terbaru*

Periode	*Forecast*	Data Asli	Selisih
Januari	30173	10149	20024
Februari	30268	9796	20472
Maret	30362	12041	18321
April	30455	12452	18003
Mei	30547	12230	18317
Juni	30638	11978	18660
Juli	30727	5102	25625
Agustus	30816	5947	24869
September	30904	8693	22211

Dari Tabel 4.2 di atas, sangat jelas terlihat bahwa adanya perbedaan (selisihnya) yang cukup besar antara nilai hasil peramalan dengan nilai data aslinya. Hal ini menunjukkan bahwa penggunaan model ARIMA (1,1,1) dan varians residual GARCH (1,1) masih kurang tepat digunakan pada peramalan jumlah penumpang Kereta Api Indonesia (ribu orang) untuk wilayah Jabodetabek dari bulan Januari tahun 2021 sampai bulan September tahun 2021. Dengan demikian, sebaiknya dilakukan analisis lebih lanjut mengenai peramalan ini, misalnya dengan menggunakan metode lain, agar bisa dihasilkan nilai peramalan yang lebih akurat.

DAFTAR PUSTAKA

[1]. Ali, S., & Andabdulaziztariq, H., 2018. *Forecasting Passenger Numbers in Saudi Arabian Airlines Flights*. *7*(4), 1–14.

[2]. Abdelhafez, M. E. M., 2018. *Using GARCH Models for Modelling and Forecasting Volatility an Empirical Study of the Egyptian Stock Market*. *57*(2), 167–178.

[3]. Awat, J. Napa, 1990. *Metode Peramalan Kuantitatif*. Yogyakarta: Liberty.

[4]. Badan Pusat Statistik, 2021. *Jumlah Penumpang Kereta Api (Ribu Orang), 2011-2020*. Badan Pusat Statistik. Diakses dari https://www.bps.go.id/linkTableDinamis/view/id/815 [Diakses 7 April 2021].

[5]. Brooks, C., 2014. Introductory econometrics for finance, volume 3. Cambridge University press.

[6]. Enders, W., 1995. *Applied Econometric Time Series*. New York: Jhon Wiley & Sons.

[7]. Hanke, J. E dan Wichern, D. W., 2005. *Business Forecasting, 8th Edition*. New Jersey: Prentice Hall.

[8]. Hendranata, Anton, 2003. *ARIMA (Autoregressive Moving Average)*. Manajemen Keuangan Sektor Publik FEUI. Diakses dari http://daps.bps.go.id/file_artikel/77/arima.pdf [Diakses 7 April 2021].

[9]. Makridakis, dkk., 1993. *Metode dan Aplikasi Peramalan Jilid 1, Edisi Pertama*. (Terjemahan Untung S., Andrianto). Jakarta: Erlangga.

[10]. Makridakis, dkk., 1995. *Metode dan Aplikasi Peramalan*. Jakarta: Erlangga.

[11]. Makridakis, S. dan Wheelwright, S.C., 1999. *Metode dan Aplikasi Peramalan, Edisi Kedua*. Jakarta: Binarupa Aksara.

[12]. Putri, A.R., 2019. *Pemodelan dan Peramalan Curah Hujan di Kabupaten Bogor menggunakan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA)*. Universitas Padjadjaran.

[13]. Sudjana, 1986. *Metode Statistika*. Bandung: Tarsito.

[14]. Syukrina, F., 2020. *Penerapan Metode Peramalan Model GARCH dalam Memprediksi Indeks DST*. Universitas Padjadjaran.

[15]. Wei, W., 2006. *Time Series Analysis: Univariate and Multivariate Methods, 2nd Edition*. Boston: Addison-Wesley Publishing Company.