

# Pemodelan dan Peramalan Volatilitas Harga Minyak Mentah Menggunakan Model ARIMA dan GARCH

## Laporan Kelompok

Muhammad Zaki Z

Dwiki Febrian

Ilmu Aktuaria

Matematika

muhammad.zaki22@ui.ac.id

dwiki.febrian@gmail.ui.ac.id

Stephanie Febriana E

Muhammad Fasya Syaifullah

Ilmu Aktuaria

Matematika

stephaniefbriana@sci.ui.ac.id

muhammad.fasya21@sci.ui.ac.id

Patrick Darian

Matematika

patrick.darian@sci.ui.ac.id

June 11, 2025

# 1 Pendahuluan

## 1.1 Latar Belakang

Minyak mentah (*crude oil*) merupakan salah satu komoditas strategis yang memegang peranan penting dalam perekonomian global. Menurut (Baumeister & Kilian, 2016), fluktuasi harga minyak mentah sangat dipengaruhi oleh berbagai faktor geopolitik seperti konflik di wilayah produsen utama di Timur Tengah, kebijakan organisasi produsen seperti OPEC (Organization of the Petroleum Exporter Country), serta ketegangan internasional yang memengaruhi pasokan dan permintaan global. Imbuhnya, krisis ekonomi, pandemi global, dan disrupsi rantai pasok juga turut memperburuk volatilitas harga minyak dunia.

Menghadapi kompleksitas tersebut, model berbasis deret waktu (*time series*) menjadi pendekatan yang umum digunakan dalam memprediksi harga komoditas. Metode ini memungkinkan peneliti untuk mempelajari pola historis dan dinamika temporal suatu variabel tanpa bergantung pada variabel eksternal (Hyndman & Athanasopoulos, 2018). Model *time series* seperti ARIMA (*Autoregressive Integrated Moving Average*) terbukti efektif dalam menangkap pola jangka pendek maupun jangka panjang dalam data ekonomi dan finansial (Box et al., 2015).

Model ARIMA bekerja dengan menggabungkan komponen autoregresif (AR), integrasi (I), dan moving average (MA), sehingga mampu membentuk model yang fleksibel dalam mengatasi data *non-stasioner*. Untuk menangani pola musiman, ARIMA dapat diperluas menjadi SARIMA (Seasonal ARIMA) yang mengakomodasi fluktuasi berkala dalam data (Makridakis et al., 2018). SARIMA bekerja dengan menambahkan komponen musiman pada model ARIMA standar, sehingga cocok diterapkan pada data dengan siklus atau pola musiman yang berulang, seperti harga komoditas pertanian yang sering dipengaruhi oleh musim tanam dan panen.

Namun demikian, model ARIMA dan SARIMA hanya berfokus pada peramalan nilai rata-rata (mean) dan mengasumsikan varians data konstan (homoskedastis). Dalam praktiknya, terutama pada data harga komoditas yang rentan terhadap fluktuasi tinggi, asumsi ini kerap tidak terpenuhi. Untuk itu, digunakan pendekatan tambahan seperti GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) yang dirancang khusus untuk memodelkan volatilitas atau perubahan varians secara dinamis. Model GARCH mampu menangkap fenomena heteroskedastisitas dan clustering volatilitas, yang sering ditemukan pada data keuangan dan komoditas (Engle, 1982).

Pendekatan awal dengan ARIMA dan perluasannya melalui SARIMA maupun integrasinya dengan model GARCH menjadi langkah penting untuk memahami struktur internal data secara menyeluruh sebelum melibatkan variabel eksternal seperti Google Trends atau indikator ekonomi lainnya. Model dasar ini juga berfungsi sebagai *benchmark* sebelum menerapkan pendekatan machine learning yang lebih kompleks.

## 1.2 Rumusan Masalah

1. Bagaimana membangun model prediktif deret waktu univariat yang sederhana untuk memprediksi harga komoditas Crude Oil?

2. Bagaimana karakteristik musiman dan tren dari data harga Crude Oil historis dapat dimodelkan secara efektif?

### 1.3 Tujuan

1. Menerapkan model *forecasting time series* untuk memodelkan dan memprediksi harga Crude Oil berdasarkan data historis univariat.
2. Menganalisis struktur musiman, tren, dan stasionaritas dalam data harga Crude Oil.

## 2 Landasan Teori

### 2.1 Model ARIMA

Pada tahun 1970, Box dan Jenkins mengenalkan model ***Integrated Autoregressive Moving Average*** atau ARIMA. Suatu runtun waktu  $\{Y_t\}$  dikatakan mengikuti sebuah model ARIMA jika diferensiasi ke- $d$  dari  $W_t = \nabla^d Y_t$  merupakan suatu proses stasioner ARMA. Jika  $\{W_t\}$  mengikuti model ARMA( $p, q$ ), kita katakan  $\{Y_t\}$  adalah sebuah proses ARIMA( $p, d, q$ ) di mana umumnya  $d$  mengambil nilai 1 atau 2. Sebagai contoh, model ARIMA( $p, 1, q$ ) memiliki bentuk umum:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + W_{t-p} + e_t + \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (1)$$

dengan  $W_t = Y_t - Y_{t-1}$

Sebelum membangun model ARIMA, data deret waktu harus bersifat stasioner, yaitu memiliki rata-rata, varians, dan kovarian yang konstan sepanjang waktu setelah melakukan *differencing*. *Differencing* adalah mengurangi setiap nilai dengan nilai sebelumnya. Proses ini diulang sampai data menjadi stasioner dengan jumlah pengulangan menentukan nilai  $d$  dalam model ARIMA. Untuk menguji stasioneritas digunakan uji statistik Augmented Dickey-Fuller (ADF). Jika hasil uji ADF menunjukkan nilai  $p$ -value lebih dari 0.05, maka data dianggap tidak stasioner pada taraf 5%.

### 2.2 Model GARCH

Meskipun runtun waktu konvensional dan model ekonometri beroperasi di bawah sebuah asumsi variansi konstan, proses ARCH yang dikenalkan oleh Engle (1982) memberikan varians bersyarat untuk berubah seiring waktu fungsi dari error masa lalu yang meninggalkan varians tanpa syarat yang konstan.

Lalu, Bollerslev (1986) mengajukan perumusan model ARCH yaitu model GARCH yang memodelkan variansi pada waktu ke- $t$  sebagai suatu fungsi dari residual kuadrat lag dan varians kondisional dari waktu sebelum  $t$ . Bentuk umum GARCH( $m, s$ ) yaitu

$$\sigma_t^2 = \omega + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2 \quad (2)$$

di mana

- $a_t = \sigma_t \epsilon_t$

- $\epsilon_t \sim \text{i.i.d. } (0,1)$  adalah proses white noise
- $\omega > 0, \alpha_i \geq 0$  (setidaknya satu  $\alpha_i > 0$ ),  $\beta_j \geq 0$
- $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$ , di mana  $\alpha_i = 0$  untuk  $i > m$  dan  $\beta_i = 0$  untuk  $j > s$ .

### 3 Dataset dan Metodologi

#### 3.1 Dataset: *Crude Oil WTI Futures Historical Data*

Data yang digunakan dalam penelitian ini adalah *Crude Oil WTI Futures Historical Data*, yang diperoleh dari situs *investing.com*. Dataset ini mencatat harga harian kontrak jangka minyak mentah jenis West Texas Intermediate (WTI), yang merupakan salah satu tolak ukur utama harga minyak dunia.

Dataset mencakup informasi historis harian dalam periode tertentu, terdiri atas beberapa variabel penting, yaitu:

Date	: Tanggal pencatatan data
Price	: Harga penutupan ( <i>closing price</i> ) WTI pada hari tersebut
Open	: Harga pembukaan
High	: Harga tertinggi dalam satu hari perdagangan
Low	: Harga terendah dalam satu hari perdagangan
Vol.	: Volume transaksi yang terjadi (dalam kontrak atau lot)
Change %	: Presentase perubahan harga dibandingkan hari sebelumnya.

Dalam penelitian ini, data difokuskan pada kolom *Price* karena merepresentasikan harga penutupan yang umum digunakan dalam analisis *time series* dan pemodelan finansial. Data kemudian dibersihkan, dibalik urutannya agar kronologis (dari lama ke baru), dan diubah ke format numerik agar dapat dianalisis lebih lanjut.

Sebagai langkah awal, dilakukan eksplorasi statistik dan visualisasi data untuk memahami distribusi, tren jangka panjang, serta potensi adanya komponen musiman dan volatilitas yang signifikan. Informasi ini menjadi dasar dalam pemilihan model time series seperti ARIMA dan GARCH yang mampu menangkap dinamika harga secara akurat.

Untuk melakukan *fitting* model, dilakukan pemisahan data *training* dan *testing* menggunakan metode *Fixed Split* yaitu memilih titik potong tunggal dalam deret waktu di mana data sebelum perpotongan adalah data latih dan data setelah perpotongan adalah data uji. Alasan penggunaan metode ini adalah karena metode ini sederhana dan sering digunakan.

#### 3.2 Analisis Eksplorasi Data

Melakukan Analisis eskplorasi data pada data ini, dilakukan dua hal yaitu melihat bagaimana tren dari runtun waktu dan melihat bagaimana stasioneritas dari data tersebut.

Dari data diketahui bahwa runtun waktu dimulai pada tanggal 1 Oktober 2022 lalu berakhir pada tanggal 30 September 2024. Berdasarkan pola sekilas gambar 1, dapat diasumsikan di awal bahwa runtun waktu tidak stasioner. Plot runtun waktu dibuat batas-batas seperti pada gambar 2 dengan tujuan untuk membuat tren dari runtun waktu tersebut.

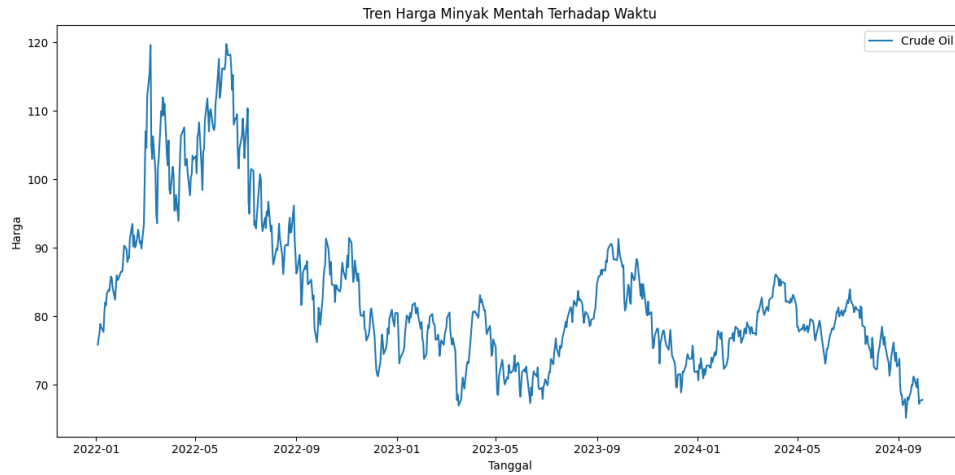


Figure 1: Plot Runtun Waktu *Crude Oil*

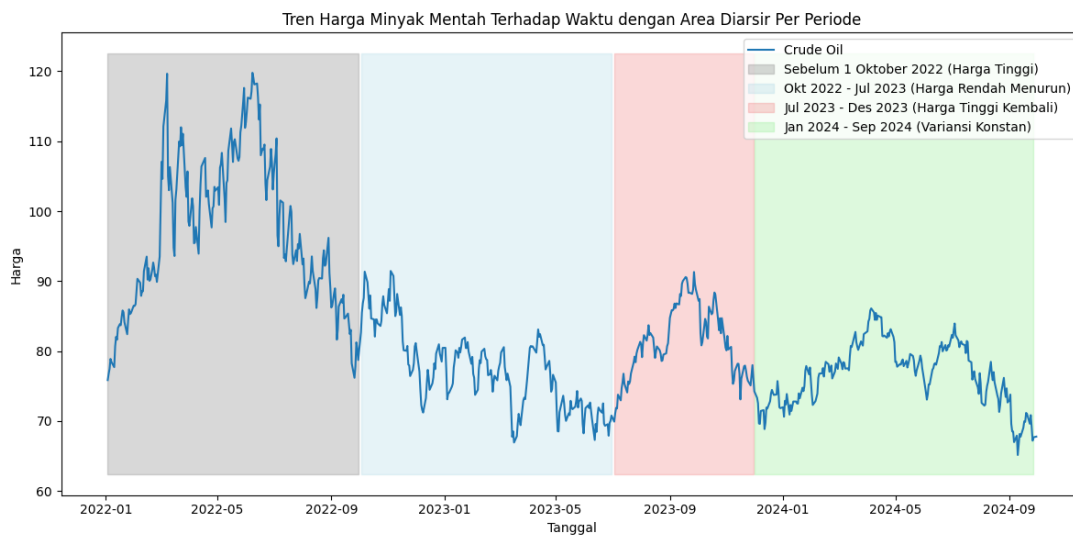


Figure 2: Plot Runtun Waktu dengan Batas

Batasan pada plot tersebut ada pada tanggal 1 Oktober 2022, 1 Juli 2023, 1 Desember 2023, dan 30 September 2024. Setelah membuat batasan tersebut diaplikasikan model regresi dengan partisi batas-batas waktu tersebut untuk mencari garis yang merepresentasikan kemiringan dari partisi waktu.

Dilihat pada gambar 3 masih belum terlihat pola tren yang jelas. Ini karena pada plot, garis pertama dan keempat masih kecenderungan horizontal. Oleh karena itu dibuat lagi batasan baru yaitu 9 Juni 2022, 1 Juli 2023, 1 Desember 2023, dan 30 September 2024.

Pada gambar 4 jelas terlihat tren untuk runtun waktu. Dari panjang garis, kemiringan, dan kecenderungan gelombang dapat dikatakan bahwa varians data tidak konstan. Pola tren yang berubah seiring berjalannya waktu mengindikasikan runtun waktu tersebut tidak stasioner.

Keberadaan tren yang berubah seiring waktu menjadi salah satu karakteristik utama dari runtun waktu non-stasioner. Namun, seringkali data dengan tren ini juga menunjukkan varians yang tidak konstan. Dapat dilihat pada plot harga minyak mentah bahwa periode dengan harga tinggi (periode sebelum Juni 2023 dan Juli-Desember 2023) cenderung

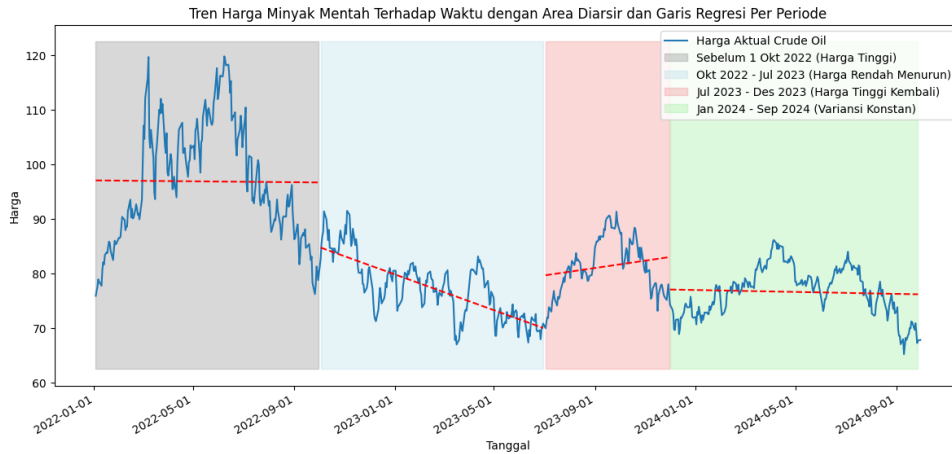


Figure 3: Tren partisi pertama

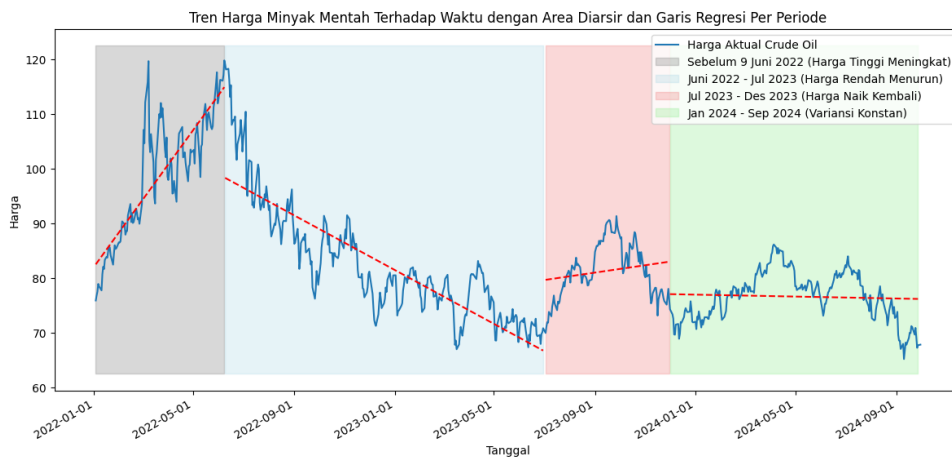


Figure 4: Tren partisi kedua

memiliki fluktuasi atau ”lebar” data yang lebih besar dibandingkan periode dengan harga yang lebih rendah (periode Juni 2022 - Juli 2023 dan Januari 2024 - September 2024). Artinya ada periode dengan variansi tinggi dan ada periode dengan variansi rendah.

Melihat gradien dari regresi untuk periode yang berbeda, setelah diamati terlihat bahwa kemiringan tren garis berubah secara signifikan. Tidak hanya menunjukkan perubahan tren, tetapi juga bisa mengindikasikan perubahan dalam ”perilaku” runtun waktu. Perubahan perilaku ini seringkali mencakup perubahan dalam volatilitas atau variansi. Periode dengan gradien yang lebih curam atau dengan fluktuasi yang lebih besar di sekitar garis regresi mungkin memiliki variansi yang lebih tinggi.

### 3.3 Uji Stasioneritas

Metode plot untuk menentukan stasioneritas dan variansi konstan masih belum cukup untuk digunakan sebagai acuan. Untuk itu dilakukan uji stasioneritas ADF untuk mengetahui apakah data stasioner atau tidak. Uji ADF yaitu

$H_0$  : Runtun waktu memiliki unit root dan non-stasioner.

$H_1$  : Runtun waktu tidak memiliki unit root dan stasioner.

Berikut adalah output dari uji stasioner pada data *Crude-Oil*

```

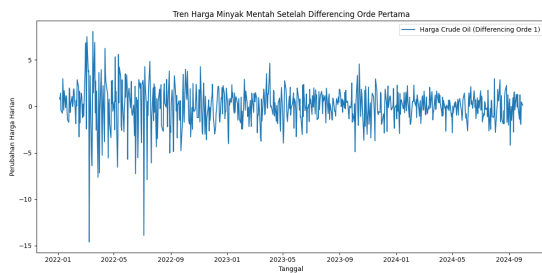
Hasil Uji ADF:
Statistik ADF: -1.4376154729359651
p-value: 0.564115927289219
Jumlah lag: 8
Jumlah observasi: 716
Critical Values:
1%: -3.439516060164992
5%: -2.8655850998755263
10%: -2.5689240826597173

Interpretasi:
p-value (0.5641) > alpha (0.05)
Gagal menolak hipotesis nol: Deret waktu kemungkinan NON-STASIONER.

```

Figure 5: Output uji ADF pertama

Karena nilai  $p$ -value dari uji ADF lebih besar dari taraf signifikansi 0.05 maka  $H_0$  tidak ditolak. Karena  $H_0$  tidak ditolak, masih kurang bukti untuk mengatakan bahwa asumsi stasioneritas data terpenuhi. Untuk menangani hal ini dilakukan diferensiasi pertama pada data lalu kembali dilakukan uji ADF. Diferensiasi yang dilakukan adalah pengurangan harga pada hari observasi  $t$  dengan harga sehari sebelum observasi  $t - 1$ .



(a) Plot nilai diferensiasi data

```

Hasil Uji ADF Setelah Differencing Orde Pertama:
Statistik ADF: -12.267831813688266
p-value: 8.77759270918779e-23
Jumlah lag: 7
Jumlah observasi: 716
Critical Values:
1%: -3.439516060164992
5%: -2.8655850998755263
10%: -2.5689240826597173

Interpretasi:
p-value (0.0000) <= alpha (0.05)
Tolak hipotesis nol: Deret waktu (setelah differencing) kemungkinan STASIONER.

```

(b) Output uji ADF kedua

Figure 6: Visualisasi hasil transformasi dan uji stasioneritas

Dari grafik 6a, dapat dilihat bahwa rata-rata data sudah konstan, dengan variansi yang tidak konstan. Ini memberikan ide untuk pemodelan dengan ARIMA dan GARCH untuk variansi. Selanjutnya, akan dilakukan uji ADF terhadap deret waktu harga minyak mentah setelah diferensiasi sekali. Hasil uji didapat sebagai berikut:

Statistik ADF : -12.2678  
 $p$ -value :  $8.78 \times 10^{-23}$   
 Nilai kritis : -3.4395 (1%)  
 High : -3.4395 (1%)  
 Low : -3.4395 (10%)

Karena nilai statistik ADF lebih kecil dari semua nilai kritis dan  $p$ -value sangat kecil (jauh di bawah 0.05), maka hipotesis nol ditolak. Artinya, deret waktu setelah diferensiasi pertama sudah bersifat stasioner.

### 3.4 Peramalan ARIMA

Setelah memperoleh deret waktu yang stasioner, orde  $p$  (untuk komponen AR) dan  $q$  (untuk komponen MA) dapat diperkirakan menggunakan plot ACF (Autocorrelation Func-

tion) dan PACF (Partial Autocorrelation Function):

Penentuan orde  $p$  pada model dapat dilakukan dengan memperhatikan plot PACF, yaitu melihat lag terakhir yang signifikan pada plot sebelum PACF *cut off* (tidak ada lagi nilai signifikan). Untuk menentukan orde  $q$  pada model dapat dilakukan dengan memperhatikan plot ACF, yaitu melihat lag terakhir yang signifikan sebelum terjadi *cut off*.

Beberapa pola khas yang umum dijumpai antara ACF dan PACF meliputi:

- ACF *cut off* dan PACF *tapering* menunjukkan model MA( $q$ ).
- PACF *cut off* dan ACF *tapering* menunjukkan model AR( $p$ ).
- Jika keduanya *cut off*, maka cocok menggunakan model ARMA( $p, q$ ).

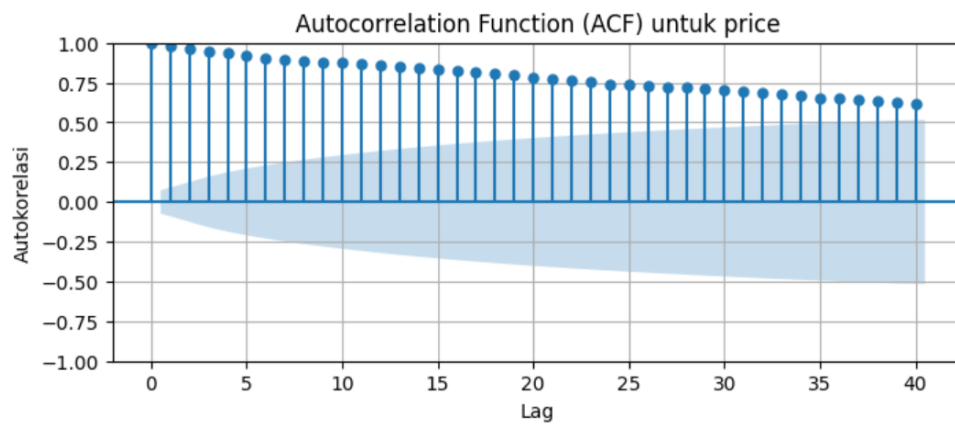


Figure 7: Plot ACF Price sebelum Dideferensiasi

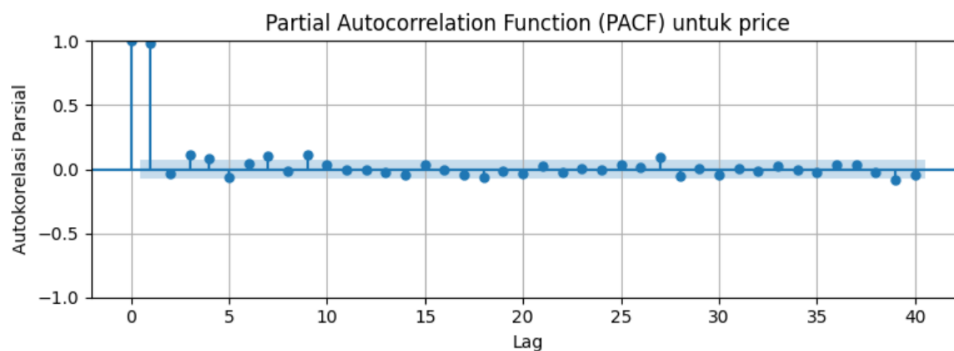


Figure 8: Plot PACF Price sebelum Dideferensiasi



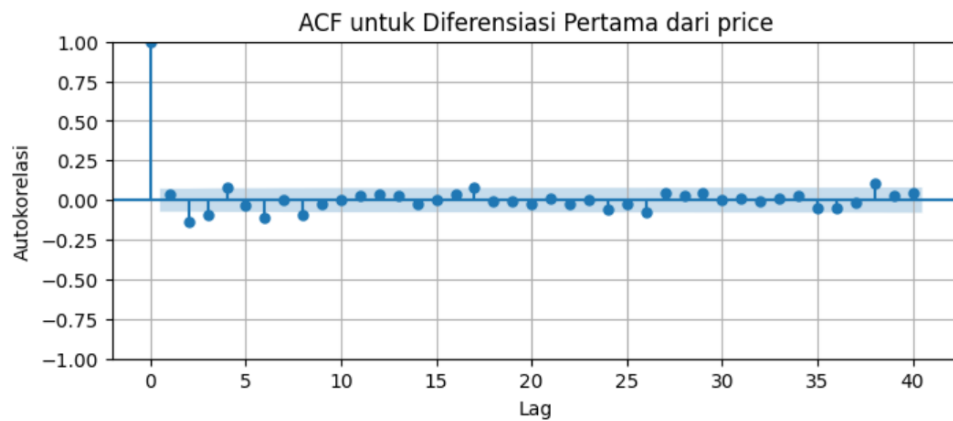


Figure 9: Plot ACF Price setelah Diferensiasi

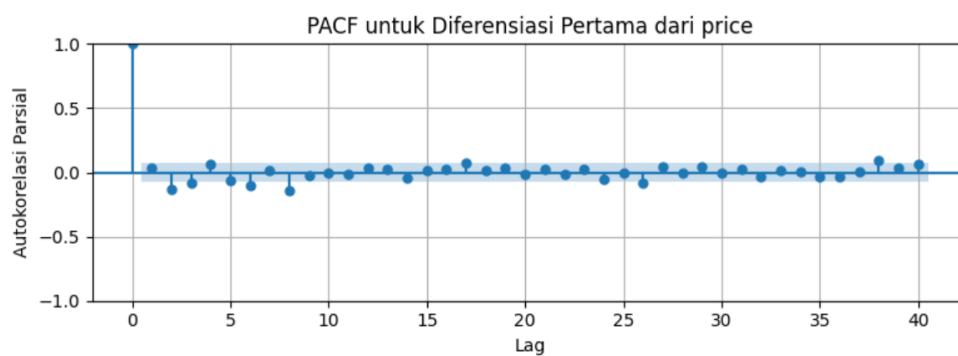


Figure 10: Plot PACF Price setelah didiferensiasi

Setelah meninjau hasil differensiasi dan plot ACF serta PACF, model kandidat yang dipilih adalah  $ARIMA(1,1,1)$ , dengan asumsi bahwa data memerlukan satu kali differencing ( $d = 1$ ), PACF mengindikasikan  $p = 1$ , dan ACF mengindikasikan  $q = 1$ .

SARIMAX Results						
=====						
Dep. Variable:	price		No. Observations:	728		
Model:	ARIMA(1, 1, 1)		Log Likelihood	-1537.283		
Date:	Wed, 11 Jun 2025		AIC	3080.567		
Time:	05:21:39		BIC	3094.333		
Sample:	01-03-2022		HQIC	3085.879		
	- 12-31-2023					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
ar.L1	-0.5484	0.412	-1.332	0.183	-1.355	0.258
ma.L1	0.5894	0.395	1.494	0.135	-0.184	1.363
sigma2	4.0197	0.101	39.840	0.000	3.822	4.217
=====						
Ljung-Box (L1) (Q):		0.32	Jarque-Bera (JB):	2615.86		
Prob(Q):		0.57	Prob(JB):	0.00		
Heteroskedasticity (H):		0.20	Skew:	-1.16		
Prob(H) (two-sided):		0.00	Kurtosis:	12.00		
=====						

Figure 11: Output model *fitting*  $ARIMA(1,1,1)$

Pada gambar 11 dapat dilihat bahwa koefisien-koefisien pada model tidak signifikan secara statistik ditunjukkan oleh nilai  $p$ -value masing-masing koefisien yang lebih dari 0.05.

Walaupun begitu varians dari error model signifikan, ini menunjukkan bahwa model memang menangkap beberapa variasi dalam data. Lalu untuk *Goodness of Fit* dapat dilihat pada informasi berikut.

Metode	Nilai	Informasi
Log Likelihood	-1537.283	Semakin tinggi (mendekati nol), semakin baik.
AIC	3080.567	Semakin kecil, semakin baik.
BIC	3094.333	Semakin kecil, semakin baik..

Lalu juga ditinjau pada residual. *Residual Autocorrelation* 0.57 ; 0.05 menyatakan bahwa model berhasil menangkap korelasi dalam data. *Residual Heteroskedasticity* 0.00;0.05 mengindikasikan ARIMA memiliki keterbatasan dalam menangani fluktuasi variabilitas yang menguatkan pertimbangan transformasi atau penggunaan ARCH/GARCH. *Residual Normality* 0.00 ; 0.05 menunjukkan asumsi normalitas residual tidak terpenuhi. Berdasarkan output model ARIMA(1,1,1), model masih belum cukup baik dan masih memiliki ruang untuk perbaikan. Untuk itu, data di-*fit* lagi ke dalam model ARIMA dengan parameter yang berbeda. Berikut adalah tabel ringkasannya.

Model	AR Coef	MA Coef	Signifikan
ARIMA(1,1,1)	AR(1): -0.5484	MA(1): 0.5894	Tidak
ARIMA(1,1,2)	AR(1): 0.8096	MA(2): -0.0751	Ya
ARIMA(2,1,1)	AR(1): -0.3458 AR(2): -0.0569	MA(1): 0.3657	Tidak
ARIMA(2,1,2)	AR(1): 0.4253 AR(2): 0.3354	MA(1): -0.4161 MA(2): -0.4279	Tidak
ARIMA(0,1,1)	-	MA(1): 0.0196	Tidak

Table 1: Koefisien ARIMA dan Signifikansinya

Model	LogLik	AIC	BIC	HQIC
ARIMA(1,1,1)	-1537.283	3080.567	3094.333	3085.879
ARIMA(1,1,2)	-1533.379	<b>3074.758</b>	<b>3093.113</b>	<b>3081.841</b>
ARIMA(2,1,1)	-1536.371	3080.741	3099.097	3087.824
ARIMA(2,1,2)	-1532.981	3075.963	3098.908	3084.817
ARIMA(0,1,1)	-1538.038	3080.075	3089.253	3083.617

Table 2: Kriteria Pemilihan Model (Goodness of Fit)

Dari ringkasan tabel ringkasan model, dapat dilihat bahwa satu-satunya model yang memiliki koefisien yang semuanya signifikan adalah model ARIMA(1,1,2). Meninjau residual, hanya bagian *Heteroskedasticity* yang menunjukkan masalah model yang terbatas dalam menangani fluktuasi variabilitas, hal ini menunjukkan saran transformasi atau

Model	Prob(Q)	Prob(JB)	Prob(H)	Skew	Kurtosis
ARIMA(1,1,1)	0.57	0.00	0.00	-1.16	12.00
ARIMA(1,1,2)	0.98	0.00	0.00	-1.06	11.19
ARIMA(2,1,1)	1.00	0.00	0.00	-1.16	11.95
ARIMA(2,1,2)	1.00	0.00	0.00	-1.07	11.31
ARIMA(0,1,1)	0.97	0.00	0.00	-1.13	11.76

Table 3: Diagnostik Residual (Normalitas, Heteroskedastisitas, Autokorelasi)

ARCH/GARCH masih berlaku dan *Normality* di mana residual tidak berdistribusi normal.

Jika model AR atau MA saja dikecualikan, model ARIMA(1,1,2) memiliki nilai AIC-BIC terkecil. Meskipun model AR atau MA saja memiliki AIC-BIC yang lebih kecil, melihat signifikansi koefisien model tetap dipertimbangkan untuk menggunakan model ARIMA sebagai model terbaik sejauh ini.

## 3.5 Peramalan ARIMA dengan GARCH

### 3.5.1 Transformasi Data

Pada bagian **Peramalan ARIMA** sebelumnya, dari model terbaik terdapat dua kendala yaitu residual yang tidak berdistribusi normal dan adanya heteroskedastisitas. Kedua kondisi ini dapat mempengaruhi validitas model dan akurasi peramalan. Untuk itu dilakukan transformasi data yang diharapkan dapat menstabilkan varians dan membuat distribusi data lebih mendekati normal, sehingga memenuhi asumsi model runtun waktu dengan lebih baik.

Transformasi data yang digunakan adalah transformasi logaritma. Transformasi ini dipilih berdasarkan karakteristik data harga *Crude-Oil* WTI yang menunjukkan volatilitas berkelompok. Transformasi digunakan untuk runtun waktu yang menunjukkan pertumbuhan eksponensial atau ketika varians meningkat seiring dengan nilai runtun waktu.

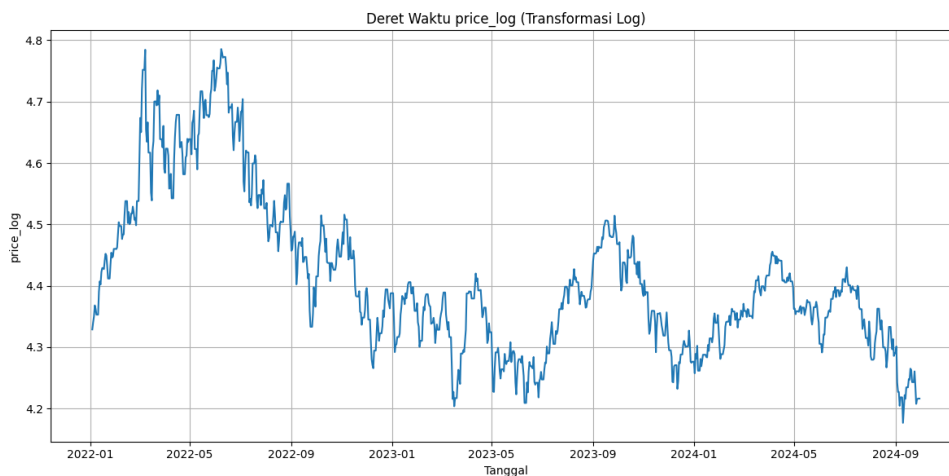


Figure 12: Plot Runtun Waktu Pasca Transformasi

Terlihat pada 12 setelah data ditransformasi, fluktuasi dalam varians tampak lebih stabil dibandingkan dengan data asli. Setelah transformasi dilakukan langkah berikutnya adalah melakukan uji stasioneritas, berikut adalah hasil pengujian ADF.

Statistik ADF    -1.3757  
 $p$ -value        0.5938768194

Hasil tersebut menunjukkan bahwa uji ADF gagal menolak hipotesis no. Maka, runtun waktu tidak dapat dikatakan sebagai runtun waktu yang stasioner. Dilakukan *differencing* pada data, lalu dilakukan uji ADF kembali didapatkan hasil.

Statistik ADF    -9.2771  
 $p$ -value        0.0000000000

Setelah dilakukan *differencing* data yang sudah ditransformasi sudah stasioner. Dengan ini dilakukan peramalan ARIMA dengan orde  $d = 1$ . Langkah berikutnya menentukan orde  $p$  dan  $q$  dari model, untuk itu dibuat plot ACF dan PACF.

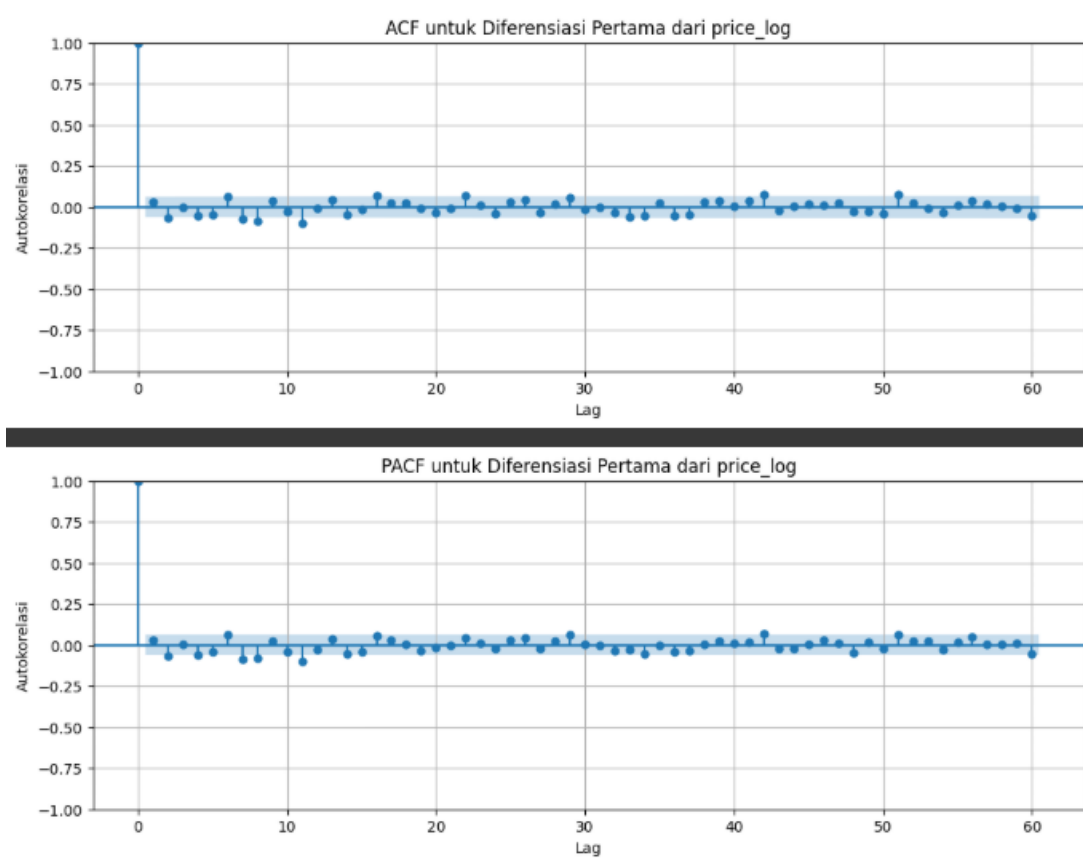


Figure 13: Plot ACF dan PACF

Sama seperti **Peramalan ARIMA** sebelumnya, estimasi nilai awal untuk  $p$  dan  $q$  melihat pada plot ACF dan PACF ketika terjadi *cut off*. Dari plot diambil orde  $p = 1$  dan  $q = 1$ . Perlu diingat bahwa tetap ada titik  $k$  yang signifikan pada plot ketika  $k = 6, 42, 51, \dots$  yang menandakan kemungkinan adanya komponen seasonal. Walaupun begitu, berdasarkan analisis plot, tidak ada pola berulang yang jelas pada kelipatan lag tertentu, maka dapat disimpulkan bahwa model meskipun dengan transformasi model ARIMA tanpa kompo-

nen musiman sudah cukup memadai.

### 3.5.2 Pemodelan Volatilitas dengan ARCH/GARCH

Secara statistik, residual model ARIMA(1,1,2) terkonfirmasi memiliki varians yang tidak konstan (heteroskedastisitas). Ini berarti bahwa periode tertentu memiliki variansi yang berbeda dengan periode lainnya. Periode dengan volatilitas tinggi cenderung diikuti oleh periode volatilitas tinggi, dan periode volatilitas rendah diikuti oleh periode volatilitas rendah. Untuk itu akan digunakan model ARCH/GARCH untuk memodelkan volatilitas ini. Pada pemodelan ini hanya diterapkan model GARCH pada residual dari model dengan harapan agar model lebih mudah untuk digunakan.

Untuk memodelkan volatilitas, hal pertama yang dilakukan adalah mengeskrak nilai-nilai residual dari model awal. Model yang digunakan untuk digunakan residualnya adalah model ARIMA(1,1,2). Sama seperti sebelumnya data yang digunakan adalah data uji hasil dari pemisahan data dengan Metode *Fixed Split*. Residual pada model ARIMA(1,1,2) dari data latih akan digunakan untuk memodelkan GARCH.

SARIMAX Results						
=====						
Dep. Variable:	price		No. Observations:	728		
Model:	ARIMA(1, 1, 2)		Log Likelihood	-1533.379		
Date:	Wed, 11 Jun 2025		AIC	3074.758		
Time:	07:06:22		BIC	3093.113		
Sample:	01-03-2022		HQIC	3081.841		
	- 12-31-2023					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
ar.L1	0.8096	0.070	11.560	0.000	0.672	0.947
ma.L1	-0.8008	0.072	-11.122	0.000	-0.942	-0.660
ma.L2	-0.0751	0.037	-2.010	0.044	-0.148	-0.002
sigma2	3.9765	0.105	37.697	0.000	3.770	4.183
=====						
Ljung-Box (L1) (Q):	0.00		Jarque-Bera (JB):	2168.01		
Prob(Q):	0.98		Prob(JB):	0.00		
Heteroskedasticity (H):	0.20		Skew:	-1.06		
Prob(H) (two-sided):	0.00		Kurtosis:	11.19		
=====						

Figure 14: Output Model ARIMA untuk Log Data

Model GARCH(1,1) digunakan dengan alasan bahwa model ini sudah umum dan seringkali memberikan hasil yang baik. Varians pada waktu ke- $t$  adalah suatu fungsi dari kuadrat residual dan varians ketika  $t-1$  dengan asumsi bahwa mean pada residual sudah nol.

Berdasarkan hasil model Zero Mean-GARCH dengan distribusi normal, dapat diambil informasi terkait model. Secara keseluruhan model signifikan ditandai oleh nilai omegge ; 0.3092 . Lalu, alpha[1] sebagai pengaruh guncangan/error periode sebelumnya juga sangat signifikan ( $P > |t| = 0.0002 < 0.05$ ) , koefisien 0.1685 menunjukkan guncangan masa lalu memengaruhi volatilitas terkini. Berikutnya beta[1] yaitu persistensi volatilitas juga sangat signifikan ( $P > |t| \approx 0.0000 < 0.05$ ) dengan nilai koefisien 0.7573 mengindikasikan volatilitas bersifat *persistent*. Total pengaruh volatilitas dilihat dari  $alpha[1] + beta[1] = 0.9258 \rightarrow 1$  volatilitas cenderung bertahan lama dan konsisten dengan karakteristik data.

Zero Mean - GARCH Model Results					
=====					
Dep. Variable:	None	R-squared:	0.000		
Mean Model:	Zero Mean	Adj. R-squared:	0.001		
Vol Model:	GARCH	Log-Likelihood:	-1504.12		
Distribution:	Normal	AIC:	3014.24		
Method:	Maximum Likelihood	BIC:	3028.01		
Date:	Wed, Jun 11 2025	No. Observations:	728		
Time:	07:06:22	Df Residuals:	728		
	Volatility Model	Df Model:	0		
=====					
	coef	std err	t	P> t	95.0% Conf. Int.
-----					
omega	0.3092	0.128	2.420	1.552e-02	[5.876e-02, 0.560]
alpha[1]	0.1685	4.540e-02	3.711	2.065e-04	[7.950e-02, 0.257]
beta[1]	0.7573	4.708e-02	16.085	3.260e-58	[ 0.665, 0.850]
=====					
Covariance estimator: robust					

Figure 15: Output Model GARCH(1,1)

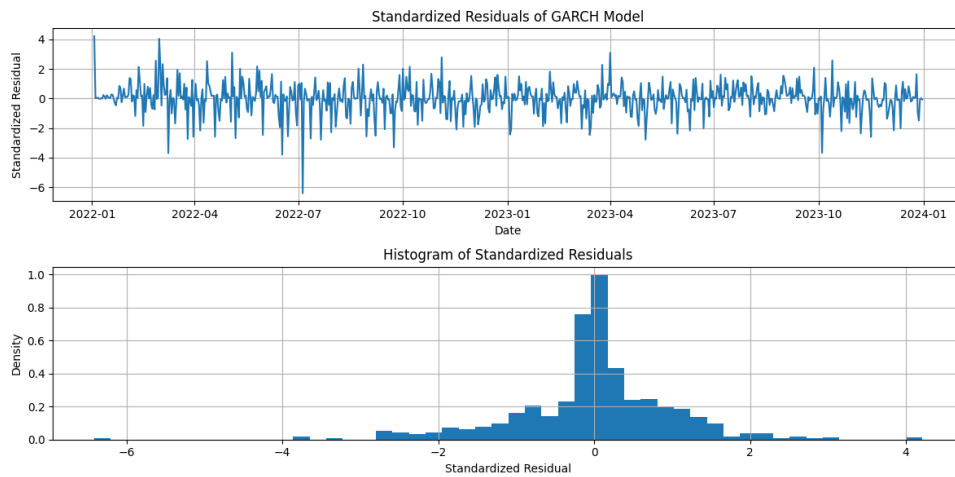


Figure 16: Runtun (atas) dan Distribusi (bawah) Residual

Pada lampiran notebook 3, dilakukan uji Ljung-Box untuk memeriksa sisa efek ARCH autokorelasi pada volatilitas. Dari uji tersebut didapatkan bahwa pada lag rendah  $k \in [1, 6]$  tidak terdeteksi autokorelasi signifikan yang artinya model berhasil menangkap heteroskedastisitas jangka pendek. Namun pada lag menengah hingga tinggi  $k \in [7, 20]$  terdapat autokorelasi tidak signifikan, mengindikasikan sisa efek ARCH yang tidak termodelkan terutama pada volatilitas jangka panjang.

Selain volatilitas, diuji juga distribusi residual. Berdasarkan uji Jarque-Bera (pada lampiran) residual tidak berdistribusi normal, menunjukkan karakteristik ekor gemuk yang umum pada data finansial.

Selain GARCH(1,1) dicoba juga model GARCH(1, 2), GARCH(2,1), dan GARCH(2,2). Informasi output masing-masing model dapat dilihat pada tabel 4 dan 5. Dari tabel tersebut model terbaik berdasarkan AIC-BIC tetap model GARCH(1,1). Untuk signifikansi parameter. Oleh karena itu dipilih GARCH(1,1) sebagai model terbaik untuk volatilitas.

Model	LogLik	AIC	BIC	omega	alpha	beta
GARCH(1,1)	-1424.95	<b>2857.91</b>	<b>2876.27</b>	20.05	1.0000	$\sim 0$
GARCH(2,1)	-1424.51	2859.02	2881.97	15.57	0.7784, 0.1941	0.0275
GARCH(1,2)	-1424.75	2859.49	2882.45	19.62	0.9910	$\sim 0$ , 0.0089
GARCH(2,2)	-1424.75	2861.49	2889.04	19.64	0.9911, $\sim 0$	$\sim 0$ , 0.0089

Table 4: Ringkasan Estimasi GARCH( $p, q$ ) dengan distribusi t (Student's t)

Model	$\nu$ (derajat bebas)	Ljung-Box ( $p > 0.05$ )	JB (p-value)
GARCH(1,1)	2.0910	Ya ( $p = 0.878$ )	0.0000
GARCH(2,1)	2.1111	Ya ( $p = 0.931$ )	0.0000
GARCH(1,2)	2.0917	Ya ( $p = 0.877$ )	0.0000
GARCH(2,2)	2.0917	Ya ( $p = 0.877$ )	0.0000

Table 5: Distribusi dan Diagnostik Residual GARCH( $p, q$ )

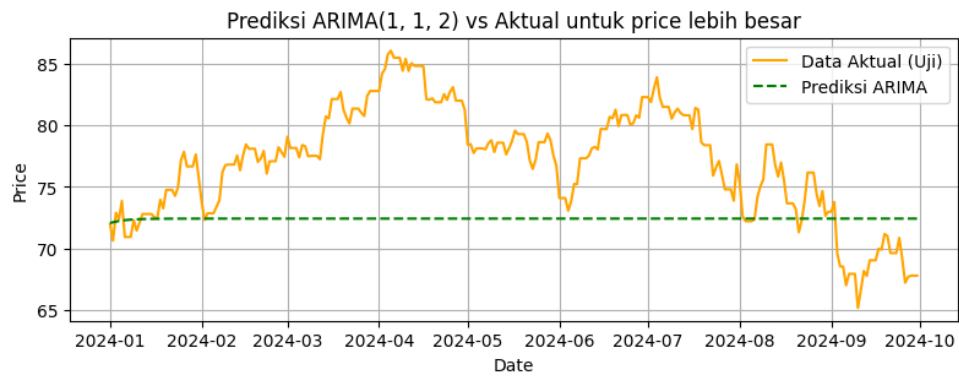
### 3.5.3 ARIMA(1,1,2)-GARCH(1,1)

Setelah mengidentifikasi model ARIMA dan GARCH yang cocok, langkah berikutnya adalah menggabungkan kedua model menjadi satu model yang utuh. Dalam konteks model, ARIMA(1,1,2) akan bertindak sebagai model mean dan GARCH(1,1) bertindak sebagai model volatilitas. Untuk melakukan peramalan dengan model gabungan ini, dapat diprediksi nilai mean melalui ARIMA dan setelahnya diprediksi varians (standar deviasi) menggunakan model GARCH. *Fit* pada model menggunakan data yang sama pada sebelumnya, untuk lebih lengkap pada langkah-langkah model dapat dilihat pada lampiran notebook 3. Untuk peramalan dibagi menjadi 3 cara yaitu pada harga langsung, dengan *log returns*, dan simulasi Monte Carlo.

## 4 Diskusi dan Hasil

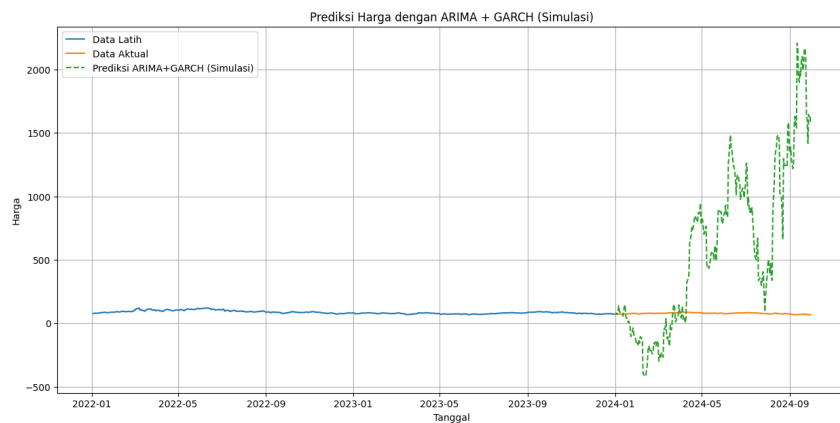
Berdasarkan bagian **Peramalan ARIMA** dan **Peramalan ARIMA dengan GARCH** ada dua model yang dapat kita bandingkan yaitu model ARIMA(1,1,2) tanpa volatilitas dan ARIMA(1,1,2) dengan volatilitas GARCH(1,1). Untuk mengevaluasi model digunakan data uji sebanyak 274 observasi dari data *Crude-Oil* dengan metrik evaluasi MAE, MSE, RMSE, dan, MAPE. Berikut adalah hasil evaluasi setiap model.

- (a) Model ARIMA(1,1,2) tanpa Volatilitas.



Metrik	Nilai
Mean Absolute Error	5.4927
Mean Square Error	41.7104
Root Mean Squared Error	6.4584
Mean Absolute Percentage Error	6.9405%

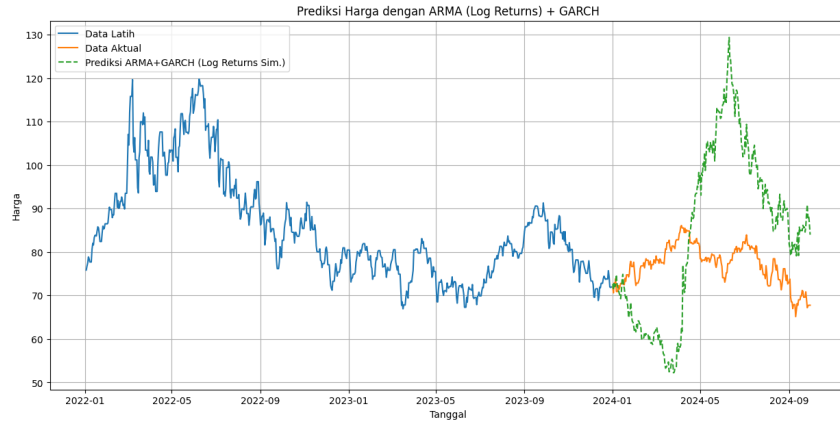
(b) Model ARIMA(1,1,2) Volatilitas Langsung



Metrik	Nilai
Mean Absolute Error	639.9058
Mean Square Error	693434.7657
Root Mean Squared Error	832.7273
Mean Absolute Percentage Error	850.3869%

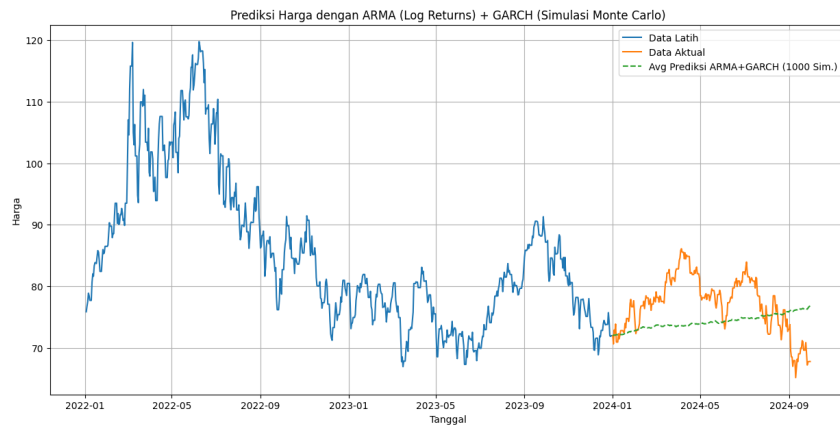
(c) Model ARIMA(1,1,2) Volatilitas Log-Return





Metrik	Nilai
Mean Absolute Error	18.7563
Mean Square Error	456.0797
Root Mean Squared Error	21.3560
Mean Absolute Percentage Error	24.2163%

(d) Model ARIMA(1,1,2) Volatilitas Log-Return Monte-Carlo



Metrik	Nilai
Mean Absolute Error	4.8213
Mean Square Error	32.7794
Root Mean Squared Error	5.7253
Mean Absolute Percentage Error	6.1824%

## 5 Kesimpulan

Berdasarkan analisis yang telah dilakukan terhadap data harga minyak mentah WTI, penelitian ini menyimpulkan bahwa pendekatan pemodelan ARIMA-GARCH memberikan hasil yang efektif dalam menangkap volatilitas harga minyak. Kombinasi model ARIMA(1,1,2) untuk komponen rata-rata dan GARCH(1,1) untuk komponen volatilitas berhasil mengidentifikasi pola heteroskedastisitas dan *clustering volatilitas* yang menjadi ciri khas data komoditas energi. Transformasi logaritmik terbukti meningkatkan stabilitas model dengan menstabilkan varians dan mendekatkan distribusi data ke kondisi normal, meskipun

residual akhir masih menunjukkan sifat *fat tails* yang umum ditemui pada data finansial.

Dari berbagai metode peramalan yang diuji, pendekatan simulasi Monte Carlo pada model ARIMA-GARCH dengan *log-return* menghasilkan prediksi terbaik, ditunjukkan oleh nilai MAPE sebesar 6,18% yang lebih unggul dibandingkan model ARIMA standalone (MAPE 6,94%) maupun prediksi langsung (MAPE 850,39%). Temuan ini mengindikasikan bahwa simulasi stokastik dapat meningkatkan akurasi peramalan dalam konteks data yang sangat fluktuatif. Selain itu, analisis parameter GARCH(1,1) mengungkap sifat persistensi volatilitas yang kuat ( $\alpha + \beta \approx 0,9258$ ), menunjukkan bahwa guncangan harga cenderung berdampak jangka panjang.

Untuk pengembangan penelitian selanjutnya, dapat dipertimbangkan integrasi variabel eksternal seperti indikator makroekonomi atau sentimen pasar guna memperkaya model. Eksplorasi model yang lebih kompleks seperti EGARCH atau pendekatan machine learning seperti LSTM juga layak diuji untuk menangkap asimetri volatilitas dan pola non-linier. Selain itu, penggunaan data dengan periode lebih panjang dapat memberikan pemahaman lebih komprehensif mengenai dinamika harga minyak dalam berbagai siklus ekonomi. Implementasi model ini dalam konteks manajemen risiko dan strategi lindung nilai komoditas juga dapat menjadi nilai tambah praktis dari penelitian lanjutan. Secara keseluruhan, penelitian ini telah memberikan kontribusi dalam demonstrasi aplikasi model time series klasik untuk peramalan komoditas volatil sekaligus menegaskan pentingnya transformasi data dan metode simulasi dalam meningkatkan akurasi prediksi.

## 6 Tabel Kontribusi

<b>Nama</b>	<b>No. NPM</b>	<b>Persentase</b>	<b>Kontribusi</b>
Muhammad Fasya Syaifullah	2206025496	100%	Mencari dan memilih dataset runtun waktu yang sesuai dengan topik. Selain itu, membersihkan data dan menyiapkan deskripsi awal dataset.
Dwiki Febrian	2206048493	100%	Mengolah data di Python, termasuk eksplorasi data awal dan visualisasi.
Muhammad Zaki Zulhamlizar	2206028610	100%	Membangun model runtun waktu dan mengevaluasi hasilnya. Serta menginterpretasi dan analisis statistik dari model.
Stephanie Febriana Effendie	2101234570	100%	Mengonversi grafik/tabel dari Python ke format LaTeX dan memastikan data siap untuk dimodelkan
Patrick Darian	2206028491	100%	Menulis bagian Hasil dan Pembahasan serta menyusun Daftar Pustaka dan Lampiran, serta memeriksa keseluruhan laporan untuk memastikan koherensi dan kerapian format.

## References

- Baumeister, C., & Kilian, L. (2016). Forty years of oil price fluctuations: Why the price of oil may still surprise us. *Journal of Economic Perspectives*, 30(1), 139–160.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th). John Wiley & Sons.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4), 987–1007.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd) [Available at <https://otexts.com/fpp2/>]. OTexts.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889.