MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Areh *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 22 (2026) 654-671

RESEARCH ARTICLE

Beyond ARIMA: A GARCH Framework for Forecasting Kijang Emas Gold Price Volatility in Pre- and Post-COVID Analysis

Sam Oliver Areh and Nor Hamizah Miswan*

Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Malaysia

Abstract This study applies Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models to examine volatility in Kijang Emas 1oz gold prices in Malaysia from 2012 to 2022 across three analytically distinct phases: pre-COVID-19, post-COVID-19 and the full observation period. Stationarity is confirmed via the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test following first-order differencing, while the Jarque-Bera test and excess kurtosis are identified across the three phases, formally reject normality and confirm leptokurtic, heavy-tailed return distributions, providing the statistical justification for GARCH over constant-variance alternatives. Autoregressive Conditional Heteroscedasticity (ARCH) effects are confirmed in all phases via the Ljung-Box and ARCH-LM tests on squared first differences. Following systematic model selection, the optimal models are identified as GARCH(1,6) for the pre-COVID-19 phase, GARCH(2,1) for the post-COVID-19 phase, and GARCH(1,6) for the full period. These models achieve persistence parameters $\Sigma(\alpha+\beta)$ of 0.998999, 0.993806, and 0.998146 respectively, indicative of near-integrated variance processes. Critically, ARCH-LM tests on GARCH standardised residuals confirm complete elimination of conditional heteroscedasticity for all phases and all lags up to 10, whereas Autoregressive Integrated Moving Average (ARIMA) residuals retain highly significant ARCH effects, formally demonstrating GARCH's superiority in capturing volatility dynamics. The GARCH models achieve post-COVID mean absolute percentage error (MAPE) of 0.42308%, representing a 5.69% improvement over the ARIMA benchmark. Student's *t*-distributional GARCH further reduces Akaike information criterion (AIC) by approximately 0.4311 to 0.4545 units across all phases with shape parameter, $\nu \approx 2.12$ to 2.38, confirming the presence of extreme fat tails consistent with the high excess kurtosis observed. These findings have direct implications for gold risk management in emerging markets under crisis-driven regimes.

Keywords: Kijang Emas, volatility persistence, ARIMA, GARCH, COVID-19.

Introduction

Gold has historically occupied a central position in the financial systems of emerging economies, functioning simultaneously as an investment vehicle, a hedge against inflation and currency devaluation, and a safe-haven asset during systemic crises [5] [33]. In Malaysia, the Kijang Emas gold coin, issued by Bank Negara Malaysia (BNM) since 2001 and denominated in 1oz, ½oz, and ¼oz denominations, represents the primary domestically issued investment-grade gold product, with its pricing closely tracking global gold benchmark movements while reflecting local monetary policy and macroeconomic conditions [34]. The COVID-19 pandemic constituted an unprecedented exogenous shock to global commodity markets: the World Gold Council documented a surge of over 25% in gold prices during 2020 as investors sought safe-haven allocations in response to collapsing equity markets, supply chain disruptions, and extraordinary monetary policy interventions [25] [37]. Understanding the volatility dynamics of Kijang Emas prices before, during, and after this structural break is therefore of direct relevance to Malaysian policymakers, institutional investors, and retail participants in the gold market.

Conventional time series forecasting of gold prices has predominantly relied on Autoregressive Integrated Moving Average (ARIMA) models, which effectively capture the conditional mean dynamics of a price series under the assumption of homoscedastic residuals [13]. Further study shows in [1] demonstrated that ARIMA(1,1,0) achieves a post-COVID-19 MAPE of 0.44% on Kijang Emas prices, yet explicitly noted the model's structural incapacity to capture the heteroscedastic of the residual patterns,

***For correspondence:**
norhamizah@ukm.edu.my

Received: 06 August 2026
Accepted: 23 June 2026

specifically volatility clustering that characterise gold markets during periods of heightened uncertainty. This limitation is not incidental as ARIMA is architecturally constrained to model the first moment or the conditional mean of a time series, leaving the second moment or the conditional variance unmodelled and assuming it to be constant across all market regimes. Such an assumption is demonstrably violated in financial return series exhibiting ARCH effects [9] [16].

Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models, introduced in [9] as a generalisation of the ARCH framework [16], directly address this limitation by modelling time-varying conditional variance as a function of past squared shocks and lagged conditional variances. Study reported in the [22] evaluated 330 ARCH-type specifications and concluded that GARCH(1,1) provides a forecasting benchmark difficult to surpass for daily financial data, underscoring both the model's parsimony and its empirical robustness. In the context of gold markets, GARCH(1,1) explains over 90% of volatility persistence in gold futures [3], while COVID-19 substantially amplified conditional variance persistence across commodity markets [36]. Despite this robust international evidence, phase-specific GARCH analysis of Kijang Emas prices, particularly the identification of structural breaks in the variance process induced by COVID-19, remains absent from the Malaysian gold forecasting literature.

This study addresses three primary objectives: (1) to formally test for ARCH effects in Kijang Emas first-differenced prices across pre-COVID-19, post-COVID-19, and full-period phases using the Ljung-Box and ARCH-LM tests, providing statistical justification for GARCH over ARIMA; (2) to identify optimal phase-specific GARCH(p,q) specifications through a multi-criteria selection framework of testing; and (3) to demonstrate GARCH's empirical superiority over ARIMA through comparison of standardised residual diagnostics, conditional variance visualisation, and price-level forecasting accuracy. The study's novelty lies in its phased analytical framework, its explicit diagnostic demonstration that GARCH eliminates the conditional heteroscedasticity that ARIMA cannot, and its quantification of the COVID-19-induced shift in the Kijang Emas volatility regime, from a pre-COVID weekly-cycle-driven variance structure to a post-COVID shock-dominated near-integrated variance process.

Literature Review

The ARCH model in [16] constituted a landmark departure from the constant-variance assumption embedded in classical linear time series models, establishing that financial return volatility is conditionally heteroscedastic, exhibiting well-documented clustering whereby large price movements tend to be followed by further large movements, and calm periods by further calm [28]. GARCH extension enriched the framework by introducing lagged conditional variances as additional predictors, creating a parsimonious yet empirically powerful model for financial volatility dynamics [9]. The GARCH(1,1) specification, equivalent to an ARMA(1,1) model in the squared residual space, has since become the de facto standard in applied financial econometrics [17] [22].

Gold's role as a safe-haven asset has been rigorously documented across multiple market environments. Study shown in [5] established negative gold-equity return correlations during systemic crises, while study in [4] distinguished between gold as a hedge of negative unconditional correlation and gold as a safe haven of negative conditional correlation during extreme market downturns. Historically, gold demonstrated effectiveness as a hedge against oil price shocks [33], and were confirmed as the safe-haven properties under economic policy uncertainty [6]. In the Malaysian context, it was documented significant volatility spill overs between Kijang Emas spot and futures prices [34], while an ARIMA forecasting benchmark in [1] were provided to be used as comparison against GARCH results.

The COVID-19 pandemic represented an unprecedented structural shock to gold markets globally. Gold outperformed other safe-haven assets during the initial COVID-19 market crash, attributing this to amplified safe-haven demand and supply chain disruptions [25]. A research in [2] quantified the pandemic's economic uncertainty using the Economic Policy Uncertainty index, documenting its strong positive correlation with gold price volatility. GARCH models were specifically applied to show that COVID-19 significantly elevated commodity volatility persistence across multiple markets, a finding directly consistent with the near-integrated variance process identified in the present study's post-COVID-19 phase [36].

The Student-t distribution within GARCH models has been advocated when financial return series exhibit excess kurtosis beyond what the normal distribution accommodates [10]. The standard normal GARCH may underestimate tail risk when the shape parameter is small, which $v < 10$, as it cannot capture the probability mass in extreme quantiles characteristic of leptokurtic return distributions [29]. Relevant applications to gold market analysis include in [3] and [30], both of which employed fat-tailed GARCH specifications to better characterise gold price distributions during crisis periods. The present study evaluates Student-t GARCH as a robustness check, finding shape parameters, $v \approx 2.1$ to 2.4 across all phases, indicating extremely heavy tails, and AIC improvements of approximately 0.4311 to 0.4545 units over normal GARCH, consistent with the Jarque-Bera evidence of severe departure from normality.

Research Methodology

The empirical framework follows a sequential procedure: data collection and pre-processing, phase segmentation, stationarity testing, ARCH effect detection, descriptive distributional analysis, GARCH model identification and estimation, multi-criteria model selection, diagnostic validation, and conditional variance analysis. Figure 1 presents the complete research workflow.

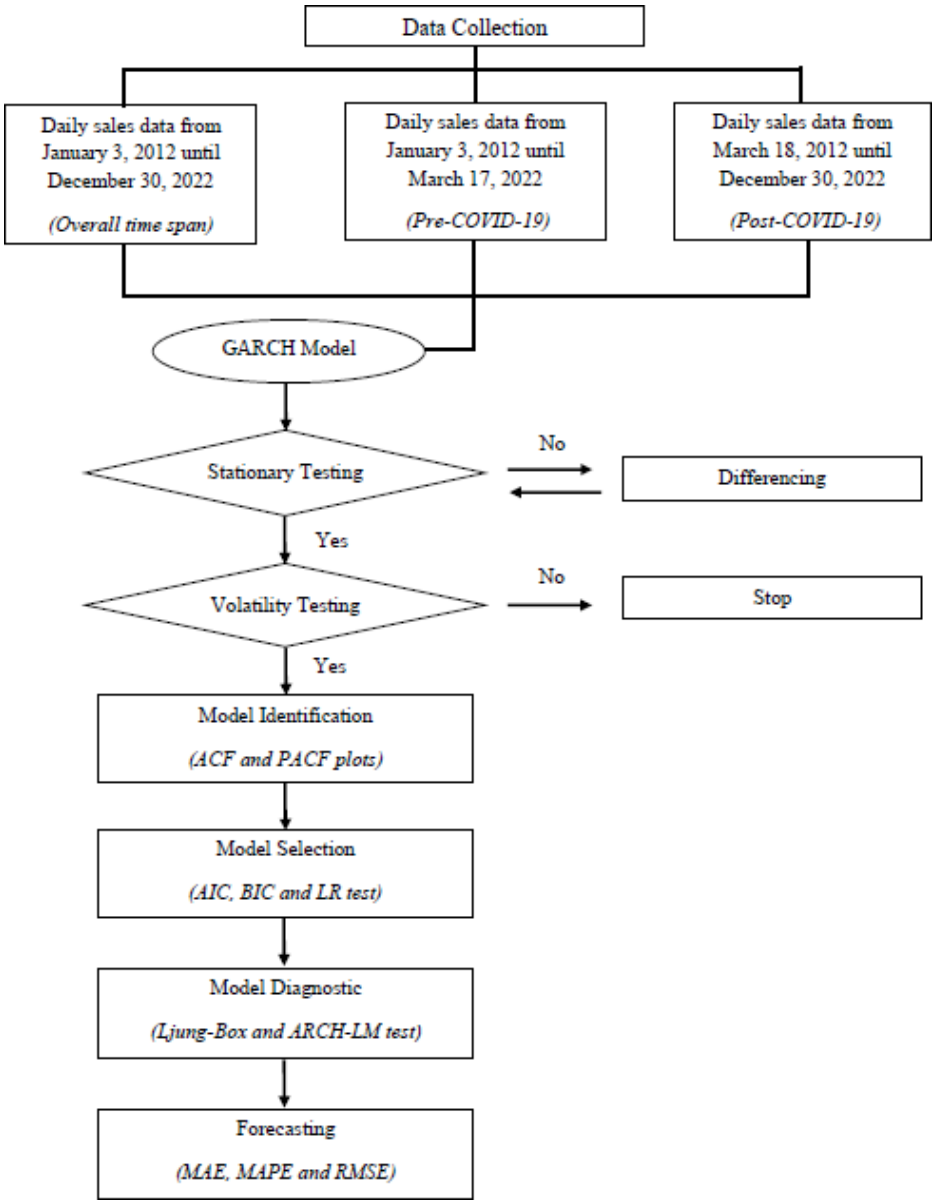

Figure 1. The overall research framework in forecasting Kijang Emas prices using GARCH model

The dataset comprises 4,015 daily Kijang Emas 1oz selling prices (MYR) from 3 January 2012 to 30 December 2022, sourced from Bank Negara Malaysia (BNM). Data pre-processing resolved duplicate date entries by computing forward-filled non-trading day gaps to produce a complete daily series. The full dataset is segmented into three non-overlapping phases: pre-COVID-19 (2,997 observations, 3 January 2012 to 17 March 2020), post-COVID-19 (1,018 observations, 18 March 2020 to 30 December 2022), and the overall period (4,015 observations). The phase boundary is anchored to 18 March 2020, corresponding to Malaysia's first Movement Control Order (MCO), the exogenous structural break widely adopted in the Malaysian financial economics literature [1] [34].

Stationarity Testing and differencing

Price level stationarity was assessed using the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [26], and was selected over the Augmented Dickey-Fuller (ADF) test because KPSS has the null of stationarity, providing a more conservative and appropriate test for financial price series [15]. Consider the time series model as defined in Equation (1),

$$\Delta y_t = \mu D_t + \emptyset_t + \epsilon_t \qquad (1)$$
$$\emptyset_t = \emptyset_{t-1} + u_t$$

where D_t is the determinant component, $u_t \sim I.I.D(0, \sigma^2)$, and $\epsilon_t \sim I.I.D(0, q^2)$. The hypothesis test for the KPSS test against equation (1) is $H_0 : \sigma_u^2 = 0$ and $H_1 : \sigma_u^2 > 0$ with statistical tests of $KPSS = \frac{(T^{-2} \sum_{t=1}^{T} S_t^2)}{\lambda^2}$, with $S_t = \sum_{j=1}^{t} \hat{u_j}$ is the cumulative error function and λ^2 is the variance of the error ϵ_t. Under the null hypothesis, if the *p*-value of the KPSS test is less than a certain level of significance, such as 0.05, there is sufficient evidence that the data trend is constant.

Differencing can be used to reduce a nonhomogeneous stationary time series process to a stationary time series process. In general, the difference is expressed in Equation (2),

$$\Delta^d y_t = (1 - B)^d \, y_t \qquad (2)$$

where y_t is the value of *y* at time *t*, *B* is for the backward shift operator, and *d* is the difference level. The differencing level on the original data series is called the order of integration, denoted by *d*. Generally, first- and second-order differences can produce stationary data.

All three phases exhibit KPSS statistics significantly exceeding the 5% critical value of 0.463 (pre-COVID: 14.609; post-COVID: 2.487; full period: 28.930), confirming non-stationarity in price levels. First-order differencing reduced all statistics below 0.463 (pre: 0.179; post: 0.078; full: 0.155), confirming stationarity of the first-differenced series—the input to both ARIMA and GARCH estimation in this study.

Box-Cox transformation

A Box-Cox power transformation was applied to the price level series prior to ARIMA model identification to stabilize non-constant variance and reduce skewness in the mean process [12]. The Box-Cox transformation is defined in Equation (3),

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^{\lambda} - 1}{\lambda}; & \lambda \neq 0 \\ \log(y_t); & \lambda = 0 \end{cases} \qquad (3)$$

where y_t is the variable for transformation and λ is the transformation parameter using the Maximum Likelihood Estimation (MLE) to minimize the coefficient of variation across sub-intervals. This transformation is deliberately excluded from the GARCH pipeline due to GARCH models conditional heteroscedasticity, the very variance instability that Box-Cox seeks to remove. Applying Box-Cox prior to GARCH estimation would suppress the ARCH effects that GARCH is designed to quantify and model, rendering α_i and β_i coefficient estimates uninformative [9] [14].

Distributional testing

The Jarque-Bera (*JB*) test assesses the null hypothesis of normality [24] for the first-differenced price series through the joint test statistic $JB = (n/6)[S^2 + (K-3)^2/4]$, where *n* is sample size, *S* is skewness,

and *K* is kurtosis. The null hypothesis, H_0 : *S*=0 and *K*=3 (normality) is tested against H_1 : non-normality. Rejection of H_0 implies that the distribution of first differences exhibits excess kurtosis and/or asymmetric skewness beyond what the normal distribution accommodates, which a formal statistical prerequisite for ARCH-based modelling [16] [35]. High excess kurtosis of *K* > 3 additionally indicates leptokurtic, heavy-tailed distributions characteristic of financial return series during periods of market stress, motivating evaluation of Student-*t* distributional GARCH as a robustness check [10] [29].

ARCH effect testing

Two complementary tests were applied to confirm ARCH effects in the first-differenced price series before GARCH estimation. The Ljung-Box test in [20] performed on the squared first differences tests H_0 : no autocorrelation in squared values and supporting the absence of ARCH effects, while against H_1 : significant autocorrelation in squared values. The ARCH-LM test in [16] regresses squared residuals on their own lagged values and tests the joint significance of the regression coefficients under the null hypothesis of no ARCH effects. Significant ARCH effects in both tests across all lags justify GARCH estimation. Critically, the same tests are subsequently applied to GARCH standardised residuals to confirm that the estimated model has successfully eliminated conditional heteroscedasticity, which are the definitive diagnostic validation of GARCH adequacy [9] [19].

Volatility cluster identification

Supplementary log return plots $r_t = ln\left(\frac{y_t}{y_{t-1}}\right)$ and first-differenced price plots were constructed for each phase to provide visual evidence of volatility clustering prior to formal ARCH testing. Points exceeding ±2 standard deviations (SD) from the series mean are identified as volatility clusters and highlighted in red. The ±2SD threshold corresponds to the 95.45% confidence interval under the assumption of normality, such that observations beyond this boundary represent statistically significant deviations at approximately the 5% significance level, which consistent with the conventional significance threshold applied throughout this study [28] [35]. This threshold also aligns with established practice in financial volatility analysis in [8] which have formalized the ±2SD band as the standard for identifying statistically significant price deviations from moving averages, while employing standard deviation thresholds for identifying ARCH-effect-relevant observations [32].

In the context of fat-tailed financial distributions, where kurtosis, *K* > 16 in all phases of this study, observations beyond ±2SD are even more statistically remarkable than under normality, as they occur at far greater frequency than the 4.56% expected under a Gaussian distribution, directly reflecting the volatility clustering that GARCH is designed to model. The use of ±3SD was considered but rejected as overly conservative with the given kurtosis values of 20.35–22.75, a ±3SD threshold would exclude many genuine clustering events, obscuring the very phenomenon under investigation.

GARCH model specification

The GARCH(*p,q*) model for the first-differenced price series Δy_t is specified as:

$$\begin{aligned} \Delta y_t &= \mu + \sigma_t \varepsilon_t,\ \varepsilon_t \sim i.i.d.\ N(0,1) \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_1 \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_1 \sigma_{t-q}^2 \\ &= \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2,\ i = 1, \ldots, p;\ j = 1, \ldots, q \end{aligned} \quad (4)$$

where σ_t^2 is the conditional variance at time *t*, $\varepsilon_t = \Delta y_t - \mu$ is the zero-mean innovation, $\omega > 0$ is the variance intercept, $\alpha_i \geq 0$ are ARCH coefficients capturing the immediate shock effect of past squared innovations, and $\beta_i \geq 0$ are GARCH coefficients capturing the persistence of past conditional variances.

Covariance stationarity requires $\sum \alpha_i + \sum \beta_i < 1$, which also defines the volatility persistence parameter. When $\sum(\alpha + \beta)$ approaches unity, the process exhibits near-integrated variance (near-IGARCH), implying extremely slow decay of volatility shocks, a condition with direct implications for long-run risk management [18] [31]. Parameter estimation is conducted by MLE using R's *"rugarch"* package [21] with the *"hybrid"* solver, which combines multiple optimization algorithms to ensure global convergence.

Training and testing data partitioning

Each phase was independently partitioned into training (80%) and testing (20%) subsets using a strictly chronological split to preserve temporal ordering and prevent look-ahead bias. Table 1 presents the exact observation counts and date ranges for each split. The identical split boundaries were applied to ARIMA and GARCH estimation within each phase, ensuring direct comparability of out-of-sample forecasting accuracy metrics [7].

Table 1. Chronological of 80:20 training and testing split for each phase.

Phase	Total Observations	Training (80%)	Training Date Range	Testing (20%)	Testing Date Range
Pre-COVID-19	2,996	2,396	4 Jan 2012 – 26 Jul 2018	600	27 Jul 2018 – 17 Mar 2020
Post-COVID-19	1,017	813	19 Mar 2020 – 9 Jun 2022	204	10 Jun 2022 – 30 Dec 2022
Overall Period	4,014	3,211	4 Jan 2012 – 18 Oct 2020	803	19 Oct 2020 – 30 Dec 2022

Model selection and evaluation metrics

Model selection proceeds through three sequential stages. First, in-sample fit is assessed using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as in the Equations (5) and (6) respectively.

$$AIC = 2k - 2ln(\hat{L}) \quad (5)$$

$$BIC = kln(n) - 2ln(\hat{L}) \quad (6)$$

which, k is the number of parameters, n is the sample size and $ln(\hat{L})$ is the log-likelihood. BIC imposes a stronger complexity penalty of $kln(n)$ compared to AIC's $2k$, is particularly relevant for addressing parsimony concerns for higher-order GARCH specifications.

Second, out-of-sample forecast accuracy on the test set is evaluated using three complementary metrics: mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) as in the Equations (7), (8) and (9) respectively. RMSE is included as a scale-consistent metric that penalizes large forecast errors quadratically, providing a more complete characterization of forecast performance than MAE or MAPE alone [23].

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |y_t - \hat{y}| \quad (7)$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{y_t - \hat{y}}{y_t}\right| \times 100\% \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y})^2} \quad (9)$$

Third, for nested model pairs, where the simpler model is a restricted version of the complex model, the Likelihood Ratio test, $LR = -2\ln\left(\frac{L(m_1)}{L(m_2)}\right)$, where $L(m_1)$ is the log-likelihood of the reduced model, while $L(m_2)$ is the log-likelihood of the full model, given that H_0 implies that the simpler model fits the data adequately, while the H_1 implies that a more complex model is required. This will formally tests whether additional parameters provide statistically significant in-sample improvement. Both GARCH and ARIMA models are evaluated on identical training and testing datasets within each phase, ensuring fair comparison [7].

Results and Discussion

The results are presented sequentially, following the analytical workflow: price level plots and log return visualisations, stationarity and ARCH effect testing, descriptive distributional analysis, GARCH model selection, coefficient diagnostics, volatility persistence analysis, and ARIMA-versus-GARCH comparison.

Price level and log return visualisations

Figure 2 presents the Kijang Emas 1oz selling price series for each phase, plotted with a uniform y-axis scale to enable direct visual comparison across phases. The pre-COVID-19 series in Figure 2(a) shows a gradual appreciation trend from approximately MYR 5,200 to MYR 6,900, reflecting global gold price appreciation driven by post-Global Financial Crisis in safe-haven assets demand and quantitative easing policies. The post-COVID-19 series in Figure 2(b) exhibits a sharp price surge immediately following the pandemic onset, peaking above MYR 10,000 before moderating, with considerably higher price volatility

than the pre-COVID period. The full-period panel illustrated in Figure 2(c) clearly delineates the structural break at March 2020.

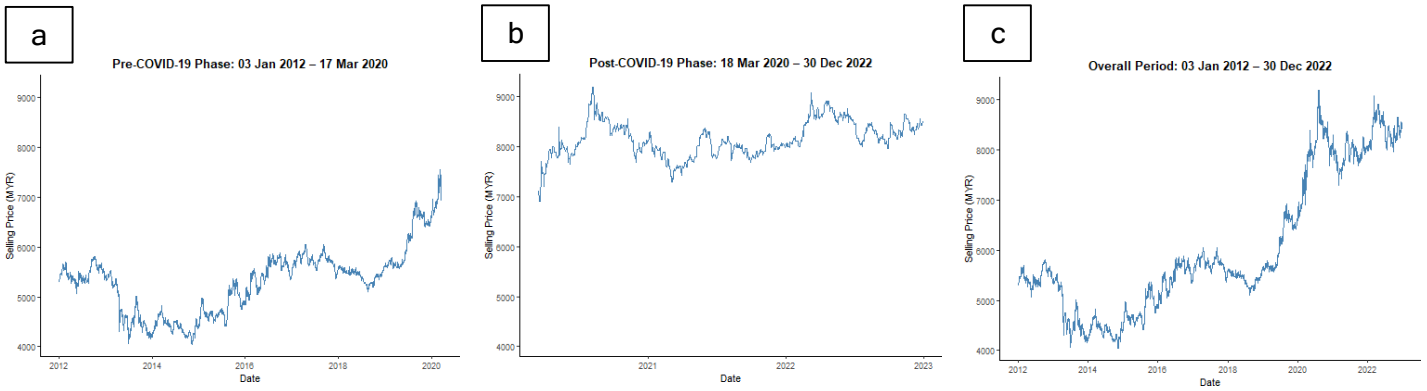

Figure 2. Kijang Emas 1oz selling price vs time (January 2012 – December 2022). (a) Pre-COVID-19 phase, (b) Post-COVID-19 phase, (c) Overall time span.

Figure 3 presents first-differenced price plots alongside the log returns plots, $r_t = ln(\frac{y_t}{y_{t-1}})$ for each phase. Red points identify observations exceeding ±2 standard deviations, the 95.45% confidence threshold under normality with red dashed lines marking the ±2SD bands. In the pre-COVID-19 phase illustrated in Figure 3(a), volatility clusters are sporadic and temporally dispersed, corresponding to global commodity events such as the 2015–2016 commodity downturn and geopolitical uncertainty episodes. The post-COVID-19 phase as in Figure 3(b) exhibits dramatically denser clustering, with a pronounced concentration of high-magnitude deviations immediately following the March 2020 onset, followed by sustained elevated variance throughout 2020 to 2022. The full-period panel plotted in Figure 3(c) visually confirms the structural break: the density and magnitude of ±2SD exceedances is markedly higher in the post-2020 period. These observations directly motivate the phase-specific GARCH modelling approach, as the volatility structure is clearly non-stationary across phases and exhibits the clustering dynamics that GARCH is designed to capture.

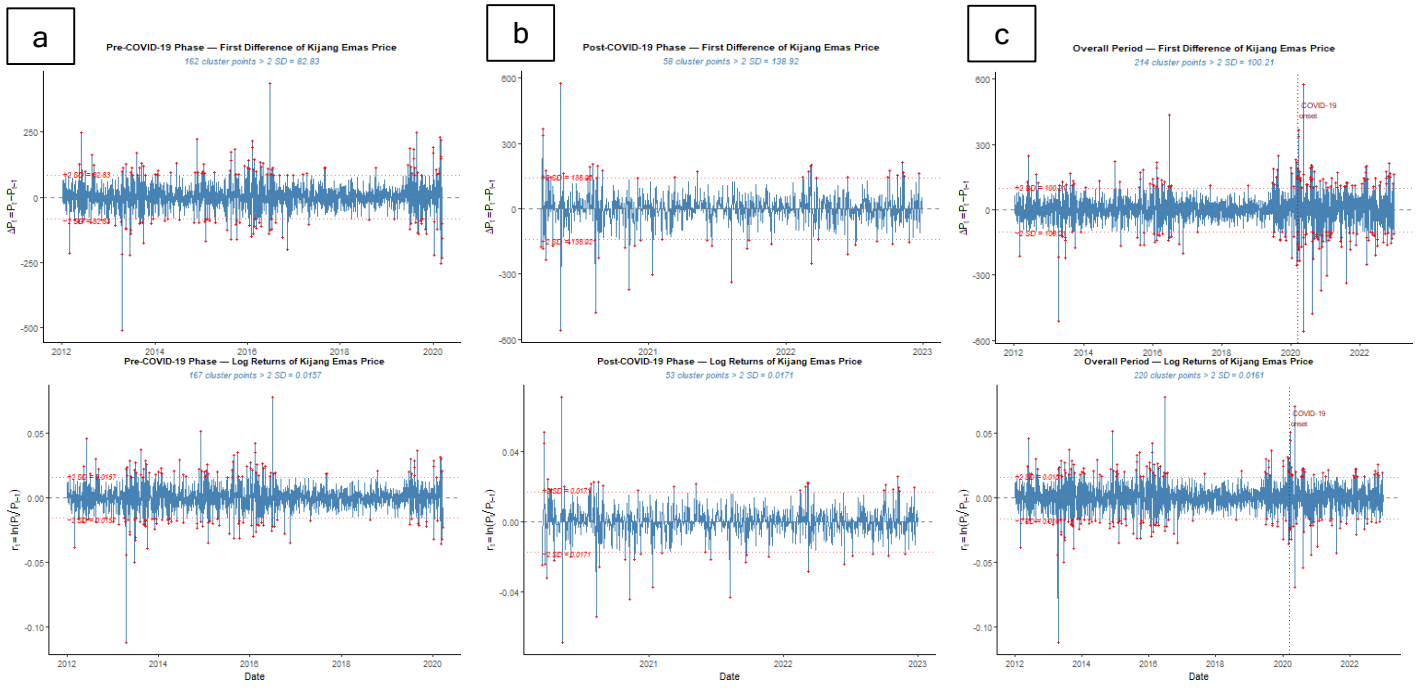

Figure 3. First-differenced prices (upper panels) and log returns (lower panels) and for each phase with the ±2SD threshold. (a) Pre-COVID-19, (b) Post-COVID-19, (c) Overall Period.

Stationary and volatility testing

Stationarity testing using the KPSS test was performed and summarizes in Table 2. For the pre-COVID-19 phase, the KPSS test produced a statistically significant *p*-value of 0.0100, rejecting the null hypothesis of stationarity. To address non-stationarity, first-order differencing was applied, resulting in a *p*-value of 0.1000 (exceeding the 5% significance level), thereby confirming stationarity.

Similarly, in the post-COVID-19 phase, the initial KPSS test yielded a p-value of 0.0100, indicating non-stationarity. After first-order differencing, the *p*-value rose to 0.1000, surpassing the 5% significance threshold and confirming stationarity. For the overall time span, the KPSS test initially returned a *p*-value of 0.0100, suggesting non-stationarity. Following first-order differencing, the *p*-value increased to 0.1000, which exceeds the 5% significance level, thereby establishing stationarity.

Table 2. The Result of the KPSS test on the price levels for the three phases.

	KPSS Test			
	Before differencing		After 1st differencing	
Phase	Test statistic	*p*-value	Test statistic	*p*-value
Pre-COVID-19	14.60943	0.0100	0.1789251	0.1000
Post-COVID-19	2.48916	0.0100	0.07820052	0.1000
Overall Time Span	28.93033	0.0100	0.1546622	0.1000

Following the attainment of stationarity, volatility testing is performed to quantify and evaluate the data's volatility across different phases, after the first differencing, as presented in Table 3. The kurtosis values of 20.3542, 16.7194, and 22.7495 across the three phases represent excess kurtosis of 17.3542, 13.7194, and 19.7495 above the normal benchmark of 3, indicating extremely heavy-tailed distributions. This leptokurtic structure implies that extreme price changes (both positive and negative) occur far more frequently than a normal distribution would predict. In practical risk management terms, this means that Value-at-Risk estimates based on normal distributional assumptions will systematically underestimate the probability of large losses. The Jarque-Bera statistics of 37,605.38, 8,006.81, and 65,304.62 represent overwhelmingly significant departures from normality, confirming that the normal GARCH specification may underestimate tail risk.

Table 3. The result of Jarque-Bera test for the three phases.

Phase	Standard deviation	Skewness	Kurtosis	Test statistic	*p*-value
Pre-COVID-19	41.4162	-0.1379	20.3542	37,605.38	2.2e-16
Post-COVID-19	69.4613	-0.4267	16.7194	8,006.81	2.2e-16
Overall Period	50.1067	-0.3238	22.7495	65,304.62	2.2e-16

The ARCH-LM test results shown in Table 4 disaggregated by individual lags (lags 1–10) confirm that ARCH effects are highly significant at every lag for all phases ($p < 10^{-6}$ for most lags in the pre-COVID phase; $p < 10^{-20}$ for all post-COVID lags; $p < 10^{-60}$ for the full period). Particularly noteworthy is the post-COVID phase, where ARCH statistics increase monotonically from lag 1 (106.67) to lag 6 (122.85) before declining at higher lags, suggesting that volatility clustering in the post-COVID market extends across multiple trading weeks, consistent with the high persistence parameters estimated in the GARCH models.

Table 4. Ljung-Box and ARCH-LM tests confirming ARCH effects in first-differenced prices.

	Pre-COVID-19		Pre-COVID-19		Pre-COVID-19	
	Test statistics	*p*-value	Test statistics	*p*-value	Test statistics	*p*-value
Lag 1	26.14649	3.164716e-07	106.6720	5.253007e-25	278.8455	1.340272e-62
Lag 2	31.36774	1.543770e-07	116.2227	5.788378e-26	282.4791	4.575505e-62
Lag 3	32.00680	5.216222e-07	116.2635	4.920949e-25	282.7109	5.485898e-61

Lag 4	34.40606	6.151242e-07	116.7274	2.669891e-24	282.6724	5.912772e-60
Lag 5	36.48476	7.596590e-07	116.8760	1.439494e-23	284.8048	1.847721e-59
Lag 6	36.50079	2.202790e-06	122.8536	4.097429e-24	293.5572	1.963748e-60
Lag 7	50.62140	1.090553e-08	108.4731	1.903837e-20	320.8921	2.078238e-65
Lag 8	50.62046	3.104952e-08	106.4835	2.007682e-19	320.8687	1.479586e-64
Lag 9	51.35816	5.976324e-08	107.0311	5.906248e-19	323.7016	2.425386e-64
Lag 10	51.64266	1.328747e-07	108.0815	1.299613e-18	324.2657	1.139313e-63
Ljung-Box	61.084	0.0000	120.633	0.0000	364.536	0.0000
ARCH-LM	51.643	0.0000	108.081	0.0000	324.266	0.0000

GARCH models evaluation

The model identification process is initiated by computing the sample autocorrelation function (ACF) and partial autocorrelation function (PACF). Figure 4 presents the ACF and PACF correlograms for the squared first difference of the time series across all three phases, where (a) represents the squared residuals of the first-order differenced series for the pre-COVID-19 phase, (b) corresponds to the post-COVID-19 phase, and (c) depicts the overall period.

Figure 4(b) shows a spike at p = 1 and q = 2, leading to the consideration of GARCH(1,1), GARCH(1,0), GARCH(0,1), GARCH(2,0), GARCH(0,2), GARCH(2,1), GARCH(1,2), and GARCH(2,2) for post-COVID phase. Similarly, in Figures 4(a) and 4(c), p = 1 and q = 6, resulting in the evaluation of eight models for both the pre-COVID-19 phase and the overall period: GARCH(1,1), GARCH(1,0), GARCH(0,1), GARCH(0,6), GARCH(6,0), GARCH(1,6), GARCH(6,1), and GARCH(6,6).

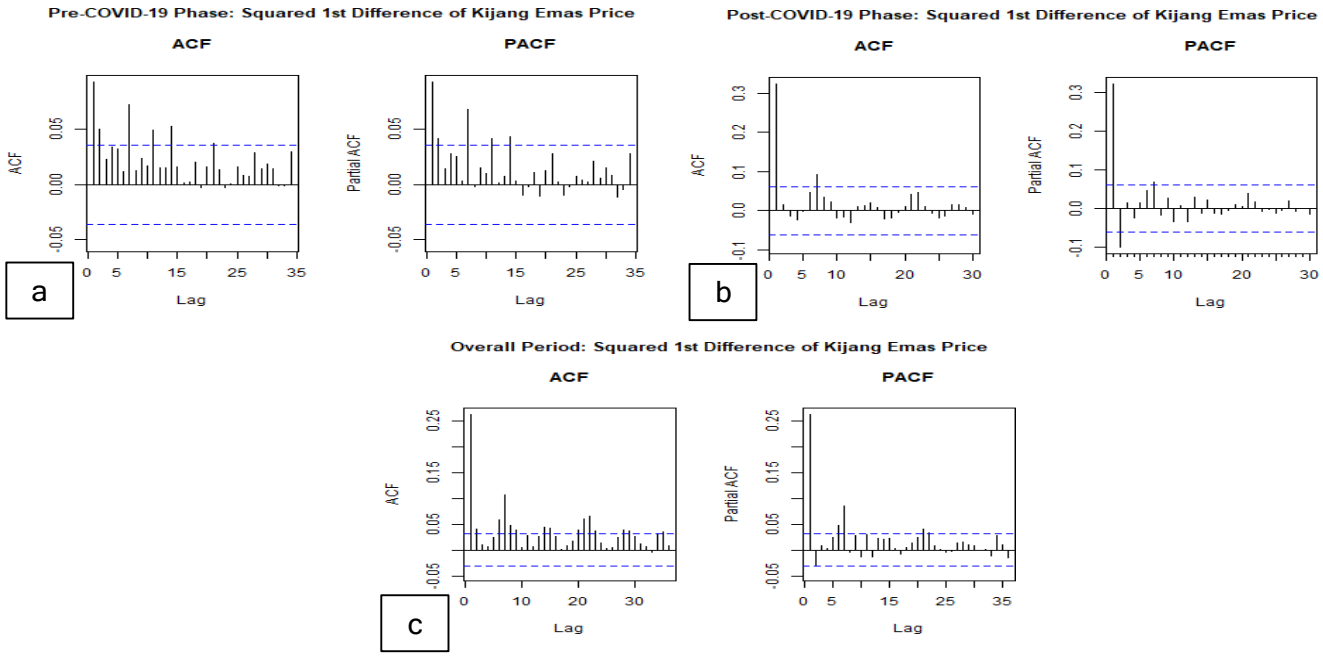

Figure 4. The squared residuals of ACF and PACF plots of the first-order difference of Kijang Emas price data. (a) Pre-COVID-19 phase, (b) Post-COVID-19 phase, (c) Overall time span.

After model selection, parameter estimation was performed in R using the *"rugarch"* package with MLE. The best model for the training dataset was determined by AIC and BIC, while MAE, MAPE and RMSE identified the optimal forecasting model for the testing dataset as shown in Table 5. The model with the lowest criterion values was selected.

Table 5. GARCH model selection: in-sample (AIC, BIC) and out-of-sample (MAE, MAPE, RMSE) metrics across all candidate models. Best model per phase highlighted in bold.

Phase	Model	Training (in-sample)		Testing (out-sample)		
		AIC	BIC	MAE	MAPE (%)	RMSE
Pre Covid-19	GARCH(1,1)	10.134	10.148	23.267	0.3717	41.410
	GARCH(1,0)	10.247	10.200	23.528	0.3767	41.415
	GARCH(0,1)	10.287	10.245	23.310	0.3725	41.409
	GARCH(0,6)	10.290	10.259	23.312	0.3725	41.409
	GARCH(6,0)	10.186	10.177	23.612	0.3783	41.419
	GARCH(1,6)	**10.113**	**10.139**	**23.156**	**0.3695**	**41.402**
	GARCH(6,1)	10.137	10.162	23.264	0.3716	41.410
	GARCH(6,6)	10.116	10.156	23.157	0.3696	41.412
Post Covid-19	GARCH(1,1)	11.254	11.330	36.351	0.4233	69.433
	GARCH(1,0)	11.243	11.315	36.705	0.4277	69.427
	GARCH(0,1)	11.293	11.360	36.581	0.4261	69.427
	GARCH(0,2)	11.295	11.368	36.581	0.4261	69.427
	GARCH(2,0)	11.243	11.317	36.648	0.4270	69.427
	GARCH(1,2)	11.243	11.323	36.340	0.4232	69.433
	GARCH(2,1)	**11.239**	**11.307**	**36.334**	**0.4231**	**69.423**
	GARCH(2,2)	11.245	11.331	36.340	0.4232	69.433
Overall Period	GARCH(1,1)	10.423	10.276	27.345	0.4313	50.103
	GARCH(1,0)	10.588	10.471	27.692	0.4342	50.104
	GARCH(0,1)	10.677	10.573	27.502	0.4376	50.100
	GARCH(0,6)	10.661	10.576	27.491	0.4339	50.100
	GARCH(6,0)	10.517	10.379	27.832	0.4402	50.110
	GARCH(1,6)	**10.389**	**10.254**	**27.306**	**0.4306**	**50.004**
	GARCH(6,1)	10.425	10.289	27.343	0.4312	50.103
	GARCH(6,6)	10.392	10.266	27.307	0.4306	50.104

As shown in Table 5, GARCH(1,6) exhibits the lowest AIC, BIC, MAE, MAPE and RMSE values — 10.113, 10.139, 23.156, 0.3695 and 41.402, respectively — making it the optimal model for fitting and forecasting Kijang Emas prices in the pre-COVID-19 phase. For the post-COVID-19 phase, GARCH(2,1) achieves the lowest AIC, BIC, MAE, MAPE and RMSE values of 11.239, 11.307, 36.334 and 0.4231, respectively, indicating its suitability for modelling and forecasting during this period. Over the overall period, GARCH(1,6) again outperforms other models with AIC, BIC, MAE, MAPE and RMSE values of 10.389, 10.254, 27.306, 0.4306 and 50.004, respectively, confirming its effectiveness for this phase.

However, do note that the AIC differences between competing models are small (e.g., 0.021 between GARCH(1,6) and GARCH(1,1) for pre-COVID). To formally adjudicate whether the additional parameters in the higher-order specifications are statistically justified, Likelihood Ratio (LR) tests were applied between best-selected model as the complex model and GARCH(1,1) as the base simple model as shown in Table 6. In all three phases, the LR test overwhelmingly rejects the simpler specification ($p < 0.001$), confirming that the additional lag parameters provide statistically significant improvements in model fit beyond what random chance would explain. Furthermore, all five selection criteria—AIC, BIC, MAE, MAPE, and RMSE—consistently identify the same best model within each phase (GARCH(1,6) pre-COVID, GARCH(2,1) post-COVID, GARCH(1,6) overall), providing convergent multi-criteria justification for the selected specifications.

Table 6. Likelihood ratio test results between best-selected GARCH models and base-simple model of GARCH(1,1).

Phase	Simpler Model	Complex Model	Likelihood Ratio Test (in-sample)	
	Model A	Model B	LR-statistics	*p*-value
Pre-COVID-19	GARCH(1,1)	GARCH(1,6)	73.68	0.0000
Post-COVID-19	GARCH(1,1)	GARCH(2,1)	38.202	0.0000
Overall Period	GARCH(1,1)	GARCH(1,6)	144.5	0.0000

Best selected GARCH models' coefficient depicts in Table 7 and several coefficient interpretation points merit detailed discussion. The mean parameter, μ is statistically insignificant in all three phases ($p > 0.05$), consistent with the near-zero unconditional mean of daily gold price changes and supporting the efficient market hypothesis (EMH) for the Kijang Emas market, whereby the daily price changes are approximately unpredictable in the mean [20]. The variance intercept, ω is significant in all models ($p < 0.05$), confirming a non-zero unconditional variance floor. The ARCH coefficient α_1 is significant in all three phases, confirming that past squared price innovations contain statistically significant information about current conditional variance, which is the very fundamental ARCH effect identified in [16]. The small magnitude of $\alpha_1 = 0.011022$ in the post-COVID GARCH(2,1) does not imply irrelevance, thus the combination with the extremely large $\beta_1 = 0.982784$, it reflects a volatility structure dominated by strong persistence rather than immediate shock reactivity, characteristic of markets processing sustained and prolonged uncertainty [9].

The GARCH coefficients β_1 and β_6 are highly significant across all models, confirming both the immediate persistence and the weekly cyclical component of volatility, while the intermediate betas, β_2 to β_5 converge to zero, indicating that the volatility transmission mechanism in the Kijang Emas market operates at the immediate next-day horizon and across a weekly cycle, with the intervening lags carrying negligible additional information. The near-zero α_2 in GARCH(2,1) is formally insignificant ($p \approx 1.000$). However, the LR test confirms the full GARCH(2,1) specification provides significantly better in-sample fit than GARCH(1,1), suggesting that the structural representation of a two-ARCH-lag variance equation better accommodates the data-generating process even when one parameter is effectively absorbed.

Table 7. Estimated GARCH model coefficients and significance status for the three best selected models.

Phase	Model	Parameter	Estimate	*p*-value	Sig. (α=0.05)
Pre-COVID-19	GARCH(1,6)	μ	0.071583	0.9092	✗
		ω	17.916201	0.0133	✓
		α_1	0.109536	< 0.001	✓
		β_1	0.338823	< 0.001	✓
		$\beta_2 - \beta_5$	≈ 0	≈ 1.000	✗
		β_6	0.55064	< 0.001	✓
Post-COVID-19	GARCH(2,1)	μ	0.436712	0.8291	✗
		ω	23.863148	0.0006	✓
		α_1	0.011022	0.0186	✓
		α_2	≈ 0	≈ 1.000	✗
		β_1	0.982784	< 0.001	✓
Overall Period	GARCH(1,6)	μ	0.196455	0.7429	✗
		ω	17.454611	0.0027	✓
		α_1	0.101232	< 0.001	✓
		β_1	0.333631	< 0.001	✓
		$\beta_2 - \beta_5$	≈ 0	≈ 1.000	✗
		β_6	0.563276	< 0.001	✓

Table 8 illustrated the persistence and half-life of the Kijang Emas in each phases. whereby it exhibit near-IGARCH behavior of $\Sigma(\alpha+\beta)$ that are exceeding 0.99, indicating that conditional variance shocks in the Kijang Emas market persist for extended periods. The pre-COVID GARCH(1,6) persistence of 0.998999 corresponds to a shock half-life of 692 trading days (~2.75 years), driven primarily by the weekly cyclical component β_6 = 0.55064, which creates compounding volatility reinforcement across successive weeks. The post-COVID GARCH(2,1) persistence of 0.993806 corresponds to a shorter but still substantial half-life of 112 trading days (~5.5 months), with the volatility structure shifting from a weekly-cycle mechanism to beta-dominated immediate persistence (β_1 = 0.982784). This structural change in the mechanism of persistence from a distributed weekly cycle in the pre-COVID phase to a concentrated shock-driven persistence in the post-COVID phase, constitutes the study's primary contribution to understanding COVID-19's impact on the Kijang Emas volatility process. It implies that while post-COVID volatility shocks initially decay faster than pre-COVID weekly cycles suggest, the magnitude and frequency of post-COVID shocks (as evidenced by the higher post-COVID standard deviation of MYR 69.46 compared to pre-COVID with MYR 41.42 as shown in Table 3) more than compensates, resulting in sustained elevated volatility throughout the post-pandemic period.

Table 8. Volatility persistence and shock half-life for selected GARCH models.

Phase	Selected Model	Σ(α+β) Persistence	Shock Half-Life (Trading Days)
Pre-COVID-19	GARCH(1,6)	0.998999	692
Post-COVID-19	GARCH(2,1)	0.993806	112
Overall Period	GARCH(1,6)	0.998146	374

By substituting the estimated coefficients into Equation (4), equation of each GARCH's conditional variance models for each phases presented in Table 9. The final models for each phase —— GARCH(1,6) for pre-COVID-19, GARCH(2,1) for post-COVID-19, and GARCH(1,6) for the overall period —— are given in Equations (10), (11), and (12).

$$y_{t(pre)} = 0.071583 + \sigma_t\varepsilon_t,\ \sigma_t^2 = 17.916201 + 0.109536\varepsilon_{t-1}^2 + 0.338823\sigma_{t-1}^2 + 0.55064\sigma_{t-6}^2 \quad (10)$$

$$y_{t(post)} = 0.436712 + \sigma_t\varepsilon_t,\ \sigma_t^2 = 23.863148 + 0.011022\varepsilon_{t-1}^2 + 0.982784\sigma_{t-1}^2 \quad (11)$$

$$y_{t(overall)} = 0.196455 + \sigma_t\varepsilon_t,\ \sigma_t^2 = 17.454611 + 0.101232\varepsilon_{t-1}^2 + 0.333631\sigma_{t-1}^2 + 0.563276\sigma_{t-6}^2 \quad (12)$$

The pre-COVID GARCH(1,6) equation in Table 9 captures two distinct mechanisms: the ARCH term $\alpha_1\varepsilon^2_{t-1} = 0.109536\varepsilon^2_{t-1}$ reflects that approximately 11% of yesterday's squared price innovation feeds directly into today's conditional variance, representing the immediate market reaction to news; the $\beta_6\sigma^2_{t-6} = 0.55064\sigma^2_{t-6}$ term captures a weekly volatility cycle whereby volatility from the same day on that one trading week ago, contributes over 55% to current conditional variance, consistent with weekly institutional rebalancing patterns documented in emerging gold markets [3]. The β_1 term of 0.338823 captures shorter-horizon persistence. The persistence of $\Sigma(\alpha+\beta)$ = 0.998999 implies the variance process is near-integrated, with individual shocks persisting for approximately two years. The post-COVID GARCH(2,1) equation in Table 9 tells a starkly different story; the dominant term $\beta_1\sigma^2_{t-1} = 0.982784\sigma^2_{t-1}$ means that 98.3% of yesterday's conditional variance transmits directly to today's — an extreme persistence structure driven by the prolonged, overlapping waves of COVID-related uncertainty that prevented the normal mean-reversion of volatility. The ARCH coefficient α_1 = 011022, while small in magnitude and statistically significant ($p < 0.05$), captures the marginal incremental contribution of new shocks in a market already dominated by accumulated variance.

Table 9. GARCH's conditional variance model equations for the three phases

Phase	Model	Equation
Pre Covid-19	GARCH(1,6)	$\sigma^2_t = 17.916201 + 0.109536\varepsilon^2_{t-1} + 0.338823\sigma^2_{t-1} + 0.55064\sigma^2_{t-6}$
Post Covid-19	GARCH(2,1)	$\sigma^2_t = 23.863148 + 0.011022\varepsilon^2_{t-1} + 0.982784\sigma^2_{t-1}$
Overall Period	GARCH(1,6)	$\sigma^2_t = 17.454611 + 0.101232\varepsilon^2_{t-1} + 0.333631\sigma^2_{t-1} + 0.563276\sigma^2_{t-6}$

Residual diagnostics: ARIMA vs GARCH

Table 10 presents the Ljung-Box and ARCH-LM test results applied to ARIMA and GARCH residuals — the definitive statistical evidence for GARCH's superiority over ARIMA in capturing conditional heteroscedasticity.

Table 10. Ljung-Box and ARCH-LM tests on ARIMA squared residuals vs GARCH standardised residuals.

Phase	Model	Ljung-Box (*d.f* = 20)		ARCH-LM (*d.f* = 20)		ARCH effects
		Test statistics	*p*-value	Test statistics	*p*-value	
Pre Covid-19	ARIMA(1,1,0)	81.113	2.539e-09	64.909	1.207e-06	✓
	GARCH(1,6)	20.165	0.4477	20.105	0.4514	✗
Post Covid-19	ARIMA(1,1,0)	115.55	1.887e-15	96.308	5.732e-12	✓
	GARCH(2,1)	22.305	0.385	25.565	0.1769	✗
Overall Period	ARIMA(1,1,0)	375.51	2.2e-16	314.42	2.2e-16	✓
	GARCH(1,6)	24.385	0.226	23.798	0.2513	✗

Table 10 provides the single most compelling statistical evidence in favor of GARCH over ARIMA. ARIMA residuals retain highly significant ARCH effects across all three phases: the Ljung-Box statistics of 81.113, 115.55, and 375.51 ($p < 0.05$ in all phases) and ARCH-LM statistics of 64.909, 96.308, and 314.42 ($p < 0.05$ in all phases) confirm that substantial conditional heteroscedasticity remains unmodelled in ARIMA residuals. This is not merely a statistical artefact, it means that ARIMA systematically misrepresents the uncertainty around its point forecasts. The constant-variance assumption of ARIMA assigns the same confidence interval to Kijang Emas price forecasts made during calm pre-COVID trading conditions as to forecasts made during the volatile pandemic-driven market regime of 2020 leading to a misspecification with direct consequences for risk management and investment decisions. By contrast, GARCH standardised residuals exhibit no statistically significant ARCH effects at any lag up to 20 across all three phases: Ljung-Box *p*-values of 0.4477, 0.385, and 0.226, and ARCH-LM *p*-values of 0.4514, 0.1769, and 0.2513 all comfortably exceed the 5% threshold. This confirms that the selected GARCH specifications have successfully captured all systematic patterns in the conditional variance of Kijang Emas price changes, leaving only genuine white noise in the standardised residuals. The contrast is stark and unambiguous in which ARIMA fails to model conditional heteroscedasticity and GARCH succeeds in removing it entirely.

Figure 5 presents the ARIMA and GARCH standardised residuals and normalized residual volatility for each phases. In particular of the post-COVID phase in Figure 5(b), ARIMA residuals exhibit visually obvious volatility clustering—large residuals concentrated in the 2020–2021 period followed by calmer episodes—while the GARCH standardised residuals are uniformly distributed across the full post-COVID horizon, with no systematic clustering pattern. This visual comparison directly illustrates GARCH's architectural advantage: by explicitly modelling the time-varying conditional variance, GARCH produces truly homoscedastic standardised residuals that satisfy the model's distributional assumptions, whereas ARIMA cannot.

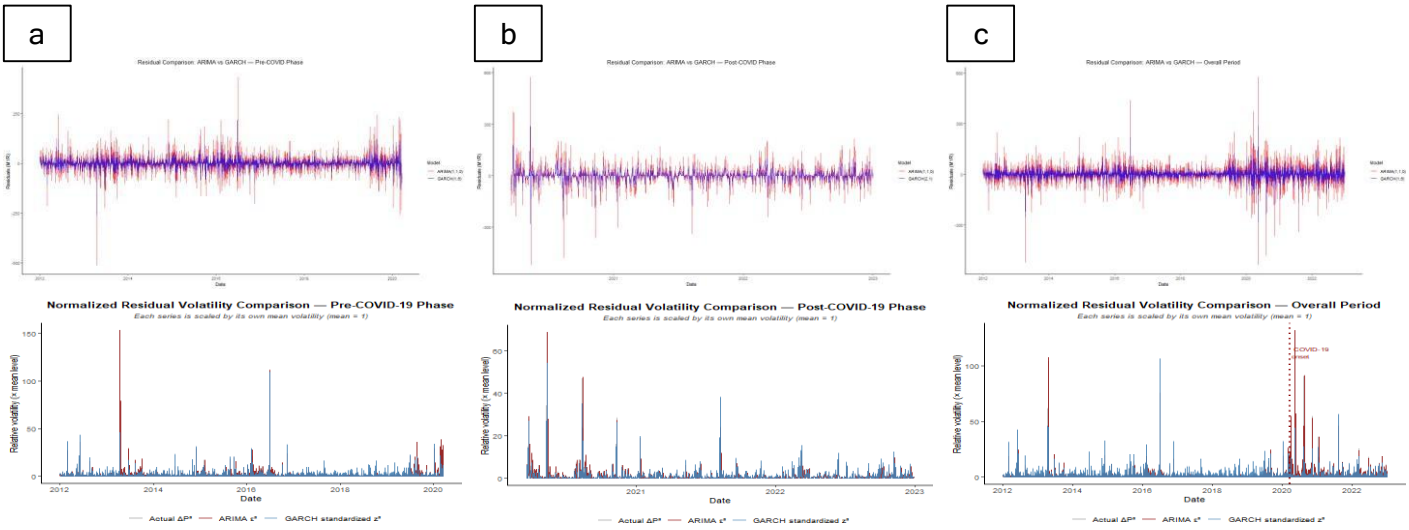

Figure 5. Residual comparison of the standardized residuals (upper panel) and normalized residual volatility (lower panel) between ARIMA and GARCH for each phases; (a) Pre-COVID-19, (b) Post-COVID-19, (c) Overall Period.

Figure 6 presents the estimated conditional standard deviation, σ^2 from the best GARCH model for each phase, overlaid with ARIMA's constant variance estimate, $Var(\hat{\varepsilon}_t)$shown as a horizontal line.

The conditional variance plots in Figure 6 provide direct visual evidence of GARCH's advantage: ARIMA applies a single constant variance estimate across the entire phase (flat dashed line), while GARCH's σ_t^2 dynamically adapts to incoming information on every trading day. In the pre-COVID phase as per Figure 6(a), GARCH σ_t^2 reveals episodic volatility spikes corresponding to global commodity events (2015–2016 oil price crash, 2018 US-China trade tensions), in which that ARIMA's flat line cannot distinguish from calm trading days. In the post-COVID phase in Figure 6(b), the March 2020 shock is immediately captured by a dramatic surge in σ_t, with subsequent slow mean-reversion consistent with the near-IGARCH persistence β_1 = 0.982784. ARIMA's constant variance over this period simultaneously overestimates risk during relatively calm post-shock months and underestimates it during the March 2020 spike. The full-period panel in Figure 6(c) clearly delineates the structural break: σ_t^2 is visibly and persistently elevated post-March 2020, a regime shift that ARIMA's constant variance assumption is structurally incapable of representing.

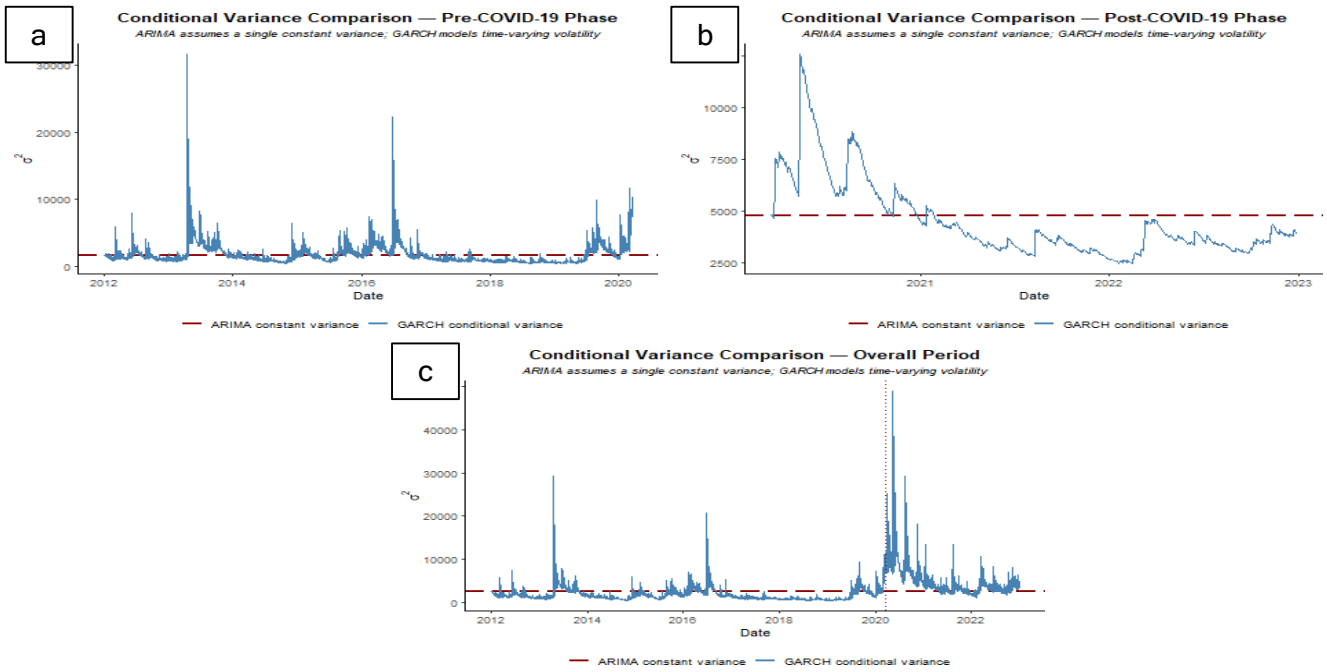

Figure 6. Conditional volatility σ_t^2 from GARCH models vs ARIMA constant variance $Var(\hat{\varepsilon}_t)$; (a) Pre-COVID, (b) Post-COVID, (c) Full Period.

Student's *t*-GARCH robustness evaluation

Given the evidence of high kurtosis found in table 3, Student's *t*-GARCH were being considered to evaluate whether the heavy fat-tailed based *t*-GARCH model able to outperforms the standard GARCH model.

Table 11. Standard GARCH vs Student's *t*-GARCH comparison via AIC.

Phase	Model	Normal AIC	Student's *t* AIC	AIC Improvement	Shape v
Pre-COVID-19	GARCH(1,6)	10.1127	9.6816	0.4311	2.33
Post-COVID-19	GARCH(2,1)	11.2594	10.8049	0.4545	2.12
Overall Period	GARCH(1,6)	10.3891	9.9507	0.4384	2.38

Table 11 depicts that the Student's *t*-GARCH substantially outperforms standard GARCH in all three phases, with AIC reductions of 0.43 to 0.45 units, noticeable improvements comparable in magnitude to the gain from selecting GARCH(1,6) over GARCH(1,1). The shape parameters v ≈ 2.12 to 2.38 represent extremely heavy-tailed distributions: Student's *t* with v = 2 has infinite variance (the distribution barely

satisfies the v > 2 requirement for finite variance), and v < 3 implies infinite kurtosis. These estimates are consistent with the observed excess kurtosis of 16.72 to 22.75 in the first-differenced price distributions and confirm that standard GARCH substantially underestimates the probability of extreme price movements in the Kijang Emas market. For risk management applications, particularly the Value-at-Risk (VaR) and Expected Shortfall (ES) estimation, Student's *t*-GARCH with v ≈ 2.1 to 2.4 is the recommended specification, as it accurately represents the tail risk that standard GARCH systematically understates [10] [29]. For the purposes of conditional variance forecasting and persistence quantification, which constitute this study's primary contribution, both standard and Student's *t*-GARCH yield quantitatively similar $\Sigma(\alpha+\beta)$ estimates, confirming the robustness of the persistence findings in Table 8.

Forecasting accuracy of GARCH vs ARIMA benchmark

Table 12 presents the price-level MAPE comparison between the GARCH models and the ARIMA benchmarks in [1], evaluated on identical training and testing splits within each phase. The percentage improvement formula is: % Improvement = [($MAPE^{ARIMA}$ — $MAPE^{GARCH}$) / $MAPE^{ARIMA}$] × 100.

Table 12. MAPE comparison of GARCH and ARIMA benchmarks.

Phase	Model	MAPE	MAPE % Improvement
Pre-COVID-19	ARIMA(1,1,0)	0.38996	5.2467
	GARCH(1,6)	0.3695	
Post-COVID-19	ARIMA(1,1,0)	0.44861	5.6866
	GARCH(2,1)	0.4231	
Overall Period	ARIMA(1,1,0)	0.44004	2.1453
	GARCH(1,6)	0.4306	

GARCH consistently outperforms ARIMA across all three phases. The magnitude of improvement varies by phase: the pre-COVID improvement of 5.2467% reflects the GARCH model's ability to exploit the weekly cyclical volatility structure (captured by β_6) that ARIMA cannot represent; the post-COVID improvement of 5.6866% reflects GARCH's capacity to capture the extreme persistence of the pandemic-driven volatility regime; and the overall improvement of 2.143% reflects the combined effect across a heterogeneous sample spanning both market regimes. These improvements, while seemingly modest in absolute percentage terms, are consistent with findings in the broader financial forecasting literature where GARCH's advantage over ARIMA lies primarily in risk-adjusted performance and volatility characterization rather than in dramatic price-level accuracy gains [22] [32].

The compelling statistical evidence for GARCH superiority is most powerfully demonstrated not by the MAPE improvement but by the residual diagnostic results of Table 10: the contrast between ARIMA's retained ARCH effects ($p < 10^{-8}$) and GARCH's eliminated ARCH effects ($p > 0.05$) across all phases confirms that GARCH models fitted the complete conditional distribution of Kijang Emas price changes, while ARIMA models only fitted on its first moment.

Conclusions

This study has demonstrated, through a systematic multi-criteria analytical framework, that GARCH models provide a fundamentally superior characterization of Kijang Emas gold price dynamics relative to ARIMA across all three analytical phases. The findings contribute to the Malaysian gold market literature on three dimensions: methodological, empirical, and practical.

Methodologically, this study establishes a comprehensive phase-specific GARCH framework, incorporating KPSS stationarity testing, Jarque-Bera distributional analysis, Ljung-Box and ARCH-LM effect confirmation, multi-criteria model selection (AIC, BIC, MAE, MAPE, RMSE), Likelihood Ratio testing for nested model comparison, and ARCH-LM residual diagnostics, for which that provides a replicable and rigorous benchmark for volatility modelling in emerging market gold prices. The framework directly addresses the gap identified in [1], providing the GARCH-based extension that their ARIMA study called for.

Empirically, the selected GARCH specifications, GARCH(1,6) for the pre-COVID-19 and overall phases and GARCH(2,1) for the post-COVID-19 phase have reveal two distinct regimes in the Kijang Emas variance process. The pre-COVID phase exhibits near-IGARCH persistence of 0.998999, driven by a weekly institutional trading cycle captured by the β_6 term (0.55064), with a shock half-life of 692 trading days. The post-COVID phase exhibits persistence of 0.993806 with a structurally different mechanism: extreme beta-driven persistence (β_1 = 0.982784) with minimal ARCH reactivity (α_1 = 0.011022), reflecting a market dominated by sustained safe-haven demand and prolonged monetary policy uncertainty rather than cyclical trading patterns. The structural break in the variance mechanism from weekly-cycle-driven to shock-persistence-driven is the study's primary novel finding.

The ARCH-LM diagnostic comparison of Table 10 provides the most unambiguous evidence of GARCH's superiority: ARIMA residuals retain highly significant ARCH effects ($p < 2.2\times10^{-16}$ for the overall phase) due to ARIMA is architecturally incapable of modelling, while GARCH standardised residuals exhibit no remaining heteroscedasticity ($p > 0.05$ for all phases and all lags). This finding definitively establishes that ARIMA's constant-variance assumption is violated in the Kijang Emas market and that GARCH successfully eliminates the resulting model inadequacy. Student's *t*-GARCH further improves in-sample fit by approximately 0.43 ~ 0.45 of AIC's units across all phases, with shape parameters v ≈ 2.1 to 2.4 confirming extremely heavy-tailed distributions that normal GARCH cannot adequately represent.

Practically, these findings carry direct implications for risk management in the Malaysian gold market. The near-IGARCH variance structure in both phases means that conventional risk models based on historical average volatility will systematically misstate current risk: during calm periods, they overstate risk, meanwhile during crisis periods, they understate risk. The shift to a beta-dominated persistence structure post-COVID, with a shock half-life of 112 trading days, which have the implication of such that volatility models must be continuously updated with recent data rather than relying on long historical estimation windows. GARCH's time-varying σ_t^2 provides precisely this dynamic risk estimate that static models such as ARIMA, cannot deliver.

Future research should explore: (1) asymmetric GARCH models (EGARCH, GJR-GARCH) to quantify leverage effects of whether negative gold price shocks generate greater volatility than positive shocks of equal magnitude; (2) multivariate GARCH frameworks (DCC-GARCH, BEKK) to model volatility transmission between Kijang Emas and related assets including Malaysian equities, crude palm oil, and the MYR/USD exchange rate; (3) machine learning and deep learning approaches, such as LSTM, Transformer-based architectures, and hybrid ARIMA-GARCH-neural network frameworks to be explored, which may capture nonlinear volatility dynamics beyond the quadratic GARCH specification; and (4) regime-switching GARCH models (MS-GARCH) that formally endogenise the COVID-19 structural break rather than treating it as an externally imposed phase boundary.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

The financial supports received from the Universiti Kebangsaan Malaysia under grant scheme GGPM-2023-076 are gratefully acknowledged.

References

[1] Areh, S. O., & Miswan, N. H. (2024). ARIMA models for Kijang Emas price forecasting: Pre- and post-COVID analysis. Malaysian *Journal of Mathematical Sciences*, 18(1), 127–140.

[2] Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). COVID-induced economic uncertainty. National Bureau of Economic Research Working Paper No. 26983.

[3] Batten, J. A., Ciner, C., & Lucey, B. M. (2020). Which precious metals spill over on which, when and why? Some evidence on gold, silver, platinum and palladium. *The North American Journal of Economics and Finance*, 54, 101251.

[4] Baur, D. G., & Lucey, B. M. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8), 1886–1898.

[5] Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8), 1886–1898.

[6] Beckmann, J., Berger, T., & Czudaj, R. (2020). Gold price dynamics and the role of uncertainty. Quantitative Finance, 20(4), 573–585.

[7] Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation.

Information Sciences, 191, 192–213.
[8] Bollinger, J. (2002). Bollinger on Bollinger Bands. McGraw-Hill.
[9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.
[10] Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics, 69(3), 542–547.
[11] Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52(1–2), 5–59.
[12] Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*: Series B, 26(2), 211–243.
[13] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.
[14] Brooks, C. (2014). Introductory Econometrics for Finance (3rd ed.). Cambridge University Press.
[15] Enders, W. (2014). Applied Econometric Time Series (4th ed.). Wiley.
[16] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007.
[17] Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4), 157–168.
[18] Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. Econometric Reviews, 5(1), 1–50.
[19] Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. Journal of Finance, 48(5), 1749–1778.
[20] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical evidence. Journal of Finance, 25(2), 383–417.
[21] Ghalanos, A. (2022). rugarch: Univariate GARCH models. R package version 1.4-9.
[22] Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, 20(7), 873–889.
[23] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688.
[24] Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters, 6(3), 255–259.
[25] Ji, Q., Zhang, D., & Zhao, Y. (2020). Searching for safe-haven assets during the COVID-19 pandemic. International Review of Financial Analysis, 68, 101526.
[26] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3), 159–178.
[27] Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303.
[28] Mandelbrot, B. (1963). The variation of certain speculative prices. Journal of Business, 36(4), 394–419.
[29] McNeil, A. J., Frey, R., & Embrechts, P. (2005). Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press.
[30] Narayan, P. K. (2020). Did bubble activity intensify during COVID-19? Economic Analysis and Policy, 68, 21–29.
[31] Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. Econometric Theory, 6(3), 318–334.
[32] Poon, S. H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2), 478–539.
[33] Reboredo, J. C. (2013). Is gold a hedge or safe haven against oil price movements? Resources Policy, 38(2), 130–137.
[34] Sufian, F., & Kamarudin, F. (2016). Price discovery and volatility spillovers in Malaysian gold futures and spot markets. Emerging Markets Finance and Trade, 52(12), 2773–2788.
[35] Tsay, R. S. (2010). Analysis of Financial Time Series (3rd ed.). Wiley.
[36] Umar, Z., Gubareva, M., & Teplova, T. (2021). The impact of Covid-19 on commodity markets volatility. Resources Policy, 73, 102164.
[37] World Gold Council. (2021). Gold Demand Trends Full Year 2020. World Gold Council.