Journal of Applied Science, Engineering, Technology, and Education *ISSN: 2685-0591 (online)*
Vol. 8 No. 2 (2026)
https://doi.org/10.35877/454RI.asci4894

Regime-Adaptive ARIMA–GARCH Framework under Structural Break: Evidence from the S&P 500 Index

Farah Hayati Mustapa[a,*] & Mohd Tahir Ismail[b]

[a]*Department of Computer & Mathematical Sciences, Universiti Teknologi MARA Cawangan Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang, Malaysia*
[b]*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Penang, Malaysia*

Abstract

Financial time series are vulnerable to structural instability caused by macroeconomic shocks and financial crises. Hybrid Autoregressive Integrated Moving Average-Generalized Autoregressive Conditional heteroskedasticity (ARIMA-GARCH) models effectively capture temporal dependencies and volatility in financial time series but typically assume constant model parameters throughout the sample period, potentially leading to model misspecification when structural breaks are present. This study develops a regime-adaptive ARIMA-GARCH modelling framework that incorporates structural break parameters into the conditional mean and conditional variance equations. Monthly closing price of S&P 500 index from January 2001 to December 2017 were analyzed. The Bai-Perron test identified the structural break in the time series, after which regime-specific ARIMA and GARCH models were estimated for the pre- and post-crisis periods. The empirical results identified a significant structural break in early 2009. Marking a transition from a moving-average process before the Global Financial Crisis to a random walk process during the post-crisis recovery period. The proposed regime-adaptive framework achieved substantially improved in-sample forecast accuracy, reducing the Mean Absolute Percentage Error (MAPE) from 59.8% for the conventional full-sample ARIMA-GARCH model to 13.0%. These findings demonstrate that incorporating structural breaks provides a more appropriate representation of changing market dynamics and improves the model's ability to capture historical behaviour during periods of heightened market volatility.

Keywords: ARIMA–GARCH; Financial Econometrics; Structural Breaks; S&P 500 Index; Volatility Persistence.

Received: 24 June 2026 Revised: 9 August 2026 Published: 31 August 2026

1. Introduction

Financial time series forecasting plays an important role in financial decision-making, from asset pricing to portfolio optimization, and macro-prudential policymaking. Investors use financial information to minimize portfolio risk, while policymakers use market indicators to monitor vulnerabilities in the financial system. Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models are among the most widely used in financial time-series analysis because they address different features of the data. ARIMA models represent conditional mean dynamics, while GARCH models represent time-varying conditional variance and volatility clustering (Bollerslev, 1986; Box & Jenkins, 1976).

1.1. Problem Statement

Standard ARIMA-GARCH models assume parameter stability, but economic crises, policy changes, or external shocks may alter the data-generating process. Previous studies have shown that unmodelled structural changes can exaggerate estimated GARCH persistence and affect volatility modelling (Lamoureux & Lastrapes, 1990; Hwang & Pereira, 2008). Evidence from S&P 500 research also suggests that volatility-model performance may differ across crisis periods (Chen, 2023). When such structural changes are ignored, the resulting full-sample model may provide a misleading

[*] Corresponding author.
E-mail address: farah.hayati@uitm.edu.my

representation of historical volatility dynamics and may affect the interpretation of the estimated ARCH and GARCH coefficients.

This issue is relevant to the S&P 500 because a sample that includes periods of financial disruption may combine substantially different volatility regimes. Mustapa and Ismail (2019) identified a structural break associated with the Global Financial Crisis of 2008 in their analysis of the S&P 500 using a hybrid ARIMA–GARCH model, but the identified break was not incorporated into the model specification. This illustrates a research gap between detecting evidence of instability and incorporating that instability into the model structure. The present study addresses this gap by examining whether a structural-break adjustment improves the in-sample representation of the monthly S&P 500 series compared with a conventional full-sample ARIMA–GARCH model.

1.2. Research Objectives

Motivated by evidence that unmodelled structural changes can exaggerate estimated GARCH persistence and affect the representation of historical volatility, this study proposes a break-adjusted ARIMA–GARCH modelling framework for the monthly S&P 500 index. The study first identifies structural breakpoint(s) in the monthly S&P 500 index data from January 2001 to December 2017 using a statistical breakpoint-detection procedure. Following the identification of these breakpoints, the study aims to develop an ARIMA-GARCH model that accounts for these structural breaks in both the conditional mean and conditional variance of the series. Finally, the study compares the in-sample fitted performance of the conventional and regime-adaptive specifications using MAE, RMSE, and MAPE.

1.3. Significance of Study

This study contributes to literature in three areas. Methodologically, it extends the conventional ARIMA-GARCH framework by incorporating structural break detection into the modelling process. Empirically, the study examines whether the S&P 500 exhibits different estimated conditional-mean and conditional-variance dynamics before and after the identified structural break. Practically, the proposed framework may assist investors, risk managers, and policymakers in understanding how structural breaks influence estimated volatility, persistence, and historical model fit. Since the study uses an in-sample evaluation, its practical contribution concerns the interpretation and representation of historical volatility dynamics rather than demonstrated predictive performance on future observations.

2. Literature Review

Robust forecasting of financial time series is challenging because financial returns commonly exhibit non-stationarity, heavy-tailed innovations, volatility clustering and time-varying dependence. ARIMA models are primarily designed to represent conditional-mean dynamics, whereas ARCH models and their generalized form, GARCH models represent conditional variance as a function of past shocks and past variance. Their combination has therefore been widely used to model the conditional mean and volatility of financial series. Applications to the S&P 500 include hybrid ARIMA–GARCH forecasting and ARIMA–SGARCH-based investment strategies (Mustapa & Ismail, 2019; Vo & Ślepaczuk, 2022). A further limitation is the assumption that model parameters remain stable over the full estimation period. This assumption is particularly questionable when the sample includes major market disruptions such as the Global Financial Crisis of 2008 or the COVID-19 crisis. Studies of financial volatility document structural changes in both the level and persistence of variance, while research on the S&P 500 and other equity markets shows that long-memory behaviour and apparent volatility persistence may partly reflect unmodelled breaks (Beltratti & Morana, 2004; Martens et al., 2004; Tsuji, 2018; de Oliveira et al., 2022).

Ignoring structural breaks can produce misspecification in both the conditional-mean and conditional-variance equations. Lamoureux and Lastrapes (1990) showed that failing to account for deterministic shifts can overstate GARCH volatility persistence in stock-return data. Similarly, Hwang and Pereira (2008) have observed that breaks in ARCH (α_1) and GARCH (β_1) coefficients can exaggerate estimated persistence, meaning that highly persistence in index volatility is not necessarily represent a stable characteristic throughout the whole sample. These findings offer an econometric explanation for how unmodelled variance shifts may produce the estimates persistence measure ($\alpha_1 + \beta_1$) approaches unity, creating spurious integrated-GARCH, (IGARCH) behaviour. The practical consequences include biased volatility forecasts, distorted risk measures, and models that adapt too slowly after a change in market conditions.

Research has also considered whether accounting for breaks improves volatility modelling and forecasting. Andreou et al. (2012) found that ignoring breaks in volatility parameters can produce biased and inefficient GARCH-type estimates, whereas the use of breaks improves the volatility forecasts especially in high-volatility periods. Recent study has shown that model performance depends on the crisis regimes: GJR-GARCH model had better performance during the 2008 financial crisis, whereas symmetric GARCH or EGARCH performed better during the COVID-19 crisis (Chen, 2023).

These findings motivate comparing latent-regime models and observable structural break adjustment procedures. Markov-switching model (Hamilton, 1989) and time-varying-parameter (TVP) models (Lubik & Matthes, 2015) can represent evolving market dynamics, but their latent states and additional parameters may complicate interpretation and estimation. Structural break adjustment methods instead explicitly identify changes in the underlying data-generating process by estimating the timing of parameter shifts and allow the sample to be divided into economically meaningful regimes. The present study therefore adopts the multiple-break procedure of Bai and Perron (1998, 2003) before estimating the ARIMA–GARCH model for the S&P 500. This design preserves the interpretability and parsimony of a break-adjusted ARIMA–GARCH framework while allowing the effects of major market shocks to be distinguished from persistent but otherwise stable volatility dynamics.

3. Methodology

3.1. Data Description and Differencing Transformation

This study utilized the historical dataset from Mustapa and Ismail (2019), which was originally obtained from Yahoo Finance (p. 2). The data set consisted of the monthly adjusted closing prices of the S&P 500 index spanning from January 2001 to December 2017 (T = 204 observations). Adjusted closing prices were selected because they accurately account for corporate actions such as stock splits and dividends.

To achieve stationarity in the mean and to preserve the structural shifts of the series, the raw price levels (P_t) were transformed via a first-difference operation (pp. 7-8):

$$\Delta P_t = P_t - P_{t-1} \tag{1}$$

EViews 13 software was used to estimate and diagnose structural breaks in the series.

3.2. Structural Breakpoint and Unit Root Testing

Standard Unit Root tests, namely the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests, may have low statistical power or produce misleading conclusions when structural breaks are present in the series (Perron & Vogelsang, 1992). Accordingly, stationarity was assessed in three-stages. First, the ADF and PP tests were applied to the original S&P 500 price series to assess its order of integration. Second, the Bai-Perron multiple breakpoint procedure (Bai & Perron, 1998, 2003) was applied using a single-break specification (m = 1) to identify the presence of structural change and its timing. Third, a Breakpoint Unit Root test was performed to the first-differenced price series (ΔP_t) to determine whether the differenced series was stationary after allowing for a structural shift.

3.3. The Benchmark Full-Sample ARIMA–GARCH Model

To provide a baseline for comparison with the regime-specific models, a full-sample model was estimated without any break adjustments. This model replicated the one from Mustapa and Ismail (2019) (p. 2). The model used an ARIMA (2,1,2) model for the price levels (ΔP_t):

$$\Delta P_t = c + \phi_1 \Delta P_{t-1} + \phi_2 \Delta P_{t-2} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} \tag{2}$$

where ϕ and θ denoted the autoregressive and moving average coefficients, respectively (p. 2).

To capture volatility clustering, the error term (ϵ_t) was modelled using a GARCH (1,1) conditional variance equation (Bollerslev, 1986) under a Generalized Error Distribution (GED) to account for residual non-normality (pp. 10, 11):

$$\epsilon_t = \sigma_t z_t, \tag{3}$$

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{4}$$

where $z_t \sim$ GED (0, 1, ν), $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$ (p. 2).

3.4. The Regime-Adaptive ARIMA–GARCH Framework

To account for the structural changes associated with the 2008 Global Financial Crisis and to reduce the potential variance distortion identified by Lamoureux and Lastrapes (1990), the identified breakpoint was incorporated into the ARIMA-GARCH framework using a regime indicator variable (D_t), where $D_t = 0$ before the estimated breakpoint and $D_t = 1$ from the breakpoint onward. The indicator variable represented the transition between market regimes and enabled the model parameters to adapt to changes in the underlying data-generating process.

The break-adjusted conditional mean equation was formalized as:

$$\Delta P_t = c_0 + c_1 D_t + \sum_{i=1}^{p} \phi_{i,j} \Delta P_{t-i} + \epsilon_t - \sum_{k=1}^{q} \theta_{k,j} \epsilon_{t-k} \tag{5}$$

The corresponding break-adjusted, regime-adaptive GARCH (1,1) equation allowed the baseline unconditional variance and volatility dynamics to vary across the pre-crisis and post-crisis regimes:

$$\sigma_t^2 = \omega_0 + \omega_1 D_t + \alpha_{1,j} \epsilon_{t-1}^2 + \beta_{1,j} \sigma_{t-1}^2 \tag{6}$$

where parameters subscripted with j indicated that the lag structures and coefficients were optimized independently within each distinct sub-period to maximize parameter stability.

3.5. In-Sample Model Evaluation Metrics

The in-sample fitted performance of the proposed regime-adaptive framework was compared with the conventional full-sample model. Specifically, the fitted values generated by each estimated model over the estimation sample were compared with the corresponding observed values. No out-of-sample forecasting procedure was employed. Three standard error measures were used, namely Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE):

$$MAE = \frac{1}{n} \sum_{t=1}^{n} \left| \Delta P_t - \Delta \hat{P}_t \right| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\Delta P_t - \Delta \hat{P}_t \right)^2} \tag{8}$$

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{\Delta P_t - \Delta \hat{P}_t}{\Delta P_t} \right| \tag{9}$$

4. Results and Forecast Evaluation

4.1. Descriptive Statistics

Figure 1 presents the descriptive statistics for the S&P 500 series. The skewness is 0.8748, indicating a positively skewed distribution with a longer right tail. The kurtosis is 2.7987, which is slightly below the normal-distribution benchmark of 3 and therefore, the statistic does not indicate excess kurtosis or particularly heavy tails in the series. The Jarque-Bera test has a p-value of 0.000, which is below the 0.05 significance level. Accordingly, the null hypothesis of normality is rejected, indicating that the series does not follow a normal distribution. These results provide evidence of non-normality, but they do not by themselves establish the presence of time-varying conditional variance or reject all homoscedastic models. Evidence of conditional heteroskedasticity should instead be assessed using formal residual diagnostics, such as the ARCH-LM test, which assesses whether the conditional variance depends on lagged squared disturbances (Lee, 1991).

4.2. Stationarity Test Results

As shown in Table 1, the ADF and PP tests indicate that the S&P 500 price series (P_t) is non-stationary at the level form. The p-values of the ADF and PP tests were 0.8984 and 0.8759, respectively. However, after applying a first-difference transformation (ΔP_t), the null hypothesis was rejected by both tests ($p < 0.01$). Thus, the S&P 500 price series is integrated of order one, I (1). The attainment of stationarity after applying the first difference transformation to the price series indicates that removal of the stochastic trend. This transformation reduces the likelihood of obtaining spurious regression results from modelling the price series. Therefore, the transformed price series can be appropriately modelled using ARIMA models and further analyzed for conditional volatility.

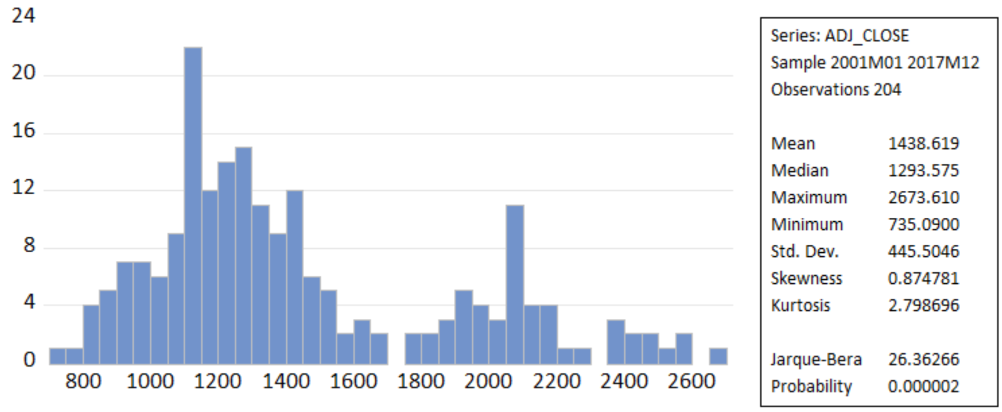

Figure 1. Histogram of the S&P index series.

Table 1. Stationarity Test Results

	Level		First Difference	
Test	**Probability**	**Stationarity**	**Probability**	**Stationarity**
ADF Statistic	0.8984	Non-stationary	0.0000	stationary
PP Statistic	0.8759	Non-stationary	0.0000	stationary

4.3. Structural Break Test Results

The Bai-Perron structural break analysis was conducted on the first-differenced S&P 500 series to determine whether its underlying mean dynamics exhibited a statistically significant change. The results are presented in Table 2. The supF test for "0 vs. 1" break produces an F-statistic of 11.8498, which exceeds the 5% critical value of 8.58. This result indicates a rejection of the null hypothesis of parameter stability and confirms the existence of at least one structural break in the series. However, the "1 vs. 2" break test yields an F-statistic of 2.5840, which is below the critical value of 10.13, suggesting that no additional structural break is statistically significant. Therefore, a single-break specification (m = 1) is sufficient to describe the S&P 500 price series. Both the sequential testing procedure and the repartition algorithm consistently identify March 2009 (2009M03) as the estimated breakpoint. This period coincides with the early recovery phase following the Global Financial Crisis, when the S&P 500 index began to recover from its prolonged decline. As the Bai-Perron procedure is specifically designed to estimate structural breakpoints for parameter stability and sample segmentation, the March 2009 breakpoint was adopted to divide the sample into the pre-crisis and post-crisis regimes used in the subsequent modelling.

Table 2. Bai-Perron Structural Break Test Results

Break Test	F-statistic	Critical Value (5%) *	Endogenous Break Date
0 vs. 1	11.8498	8.58	2009M03 (March 2009)
1 vs. 2	2.5840	10.13	—

*Note: Critical values are based on Bai-Perron (2003) using a trimming parameter of 0.15 and a maximum of 5 allowed breaks.

4.4. Breakpoint Unit Root Test Results

To assess the stationarity of the transformed series while allowing for a possible structural shift, a breakpoint unit-root test was applied to the first-differenced S&P 500 series.

Table 3. Breakpoint Unit Root Test Results (Innovational Outlier Model)

Variable	Endogenous Break Date	ADF t-Statistic	1% Critical Value	5% Critical Value	Probability Value
DSNP500	2009M02 (February 2009)	-13.67904	-5.347598	-4.859812	< 0.01

*Note: Trend specification includes trend and intercept; break specification isolates the intercept shift utilizing an automatic lag length of 0 based on t-statistic selection.

As shown in Table 3, the test rejects the null hypothesis of a unit root at the 1% significance level ($p < 0.01$), confirming that the transformed series is stationary or $I(0)$, under the test specification. The test identifies February 2009 (2009M02) as the breakpoint, which is closely aligns with the March 2009 (2009M03) breakpoint identified by the Bai-Perron test.

The one-month different is plausible because the two procedures serve different purposes. Bai–Perron estimates changes in regression parameters for structural-break detection and sample segmentation (Bai & Perron, 2003), whereas a breakpoint unit-root test estimates the break date within a test of the null hypothesis that the series contains a unit root (Zivot & Andrews, 1992). Break dates should therefore be interpreted as estimated locations rather than exact calendar boundaries, and modest differences between testing procedures may arise from their distinct objectives and statistical specifications (Hansen, 2001; Perron, 2006).

Therefore, the Bai-Perron breakpoint (March 2009) was retained as the primary breakpoint for regime classification, while the Breakpoint Unit Root test was used as a robustness check for the stationarity of the first-differenced series in the presence of structural change. Accordingly, the sample was divided into a pre-crisis period (January 2001-February 2009) and a post-crisis period (March 2009-December 2017) for the subsequent ARIMA-GARCH modelling.

4.5. ARIMA Model Identification and Regime Estimation

After the structural break has been identified, the stationarity and model-selection procedures were applied separately to the pre-crisis and post-crisis S&P 500 subsamples. The ADF test results in Table 4 indicate that the price series was non-stationary at levels but became stationary after first differencing, supporting an $I(1)$ classification for both periods.

Table 4. Augmented Dickey-Fuller Test. Results

Period	ADF at	Test Statistic	p-value	Diagnostic Status
Pre-crisis	Level (P_t)	-0.011208	0.9956	Non-Stationary
	First Difference (ΔP_t)	-7.540879	0.0000	Stationary — $I(0)$
Post-crisis	Level (P_t)	-2.817004	0.1946	Non-Stationary
	First Difference (ΔP_t)	-11.70702	0.0000	Stationary — $I(0)$

Candidate model orders were assessed using the autocorrelation function (ACF) and partial autocorrelation function (PACF), while the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) were used to compare competing specifications and select the preferred model for each regime. The parameters of the selected models were estimated by Maximum Likelihood (ML). For comparison, an unadjusted full-sample ARIMA (2,1,2) model, as identified by Mustapa and Ismail (2019), was used as the benchmark.

4.5.1. Pre-Crisis ARIMA Model Identification and Selection (January 2001 – February 2009)

The pre-crisis S&P 500 price series (P_t) shows clear deterministic cyclical patterns which prevent the data from returning to its average value, so it appears non-stationary. The observation is supported by the correlogram (Figure 2), where the ACF shows a decay pattern, accompanied by significant Ljung-Box Q-statistics across multiple lag periods.

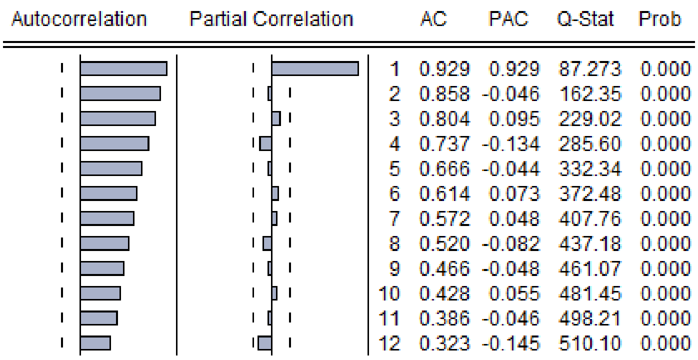

Figure 2. Correlogram of the level pre-crisis series

The autocorrelation structure of the transformed series (ΔP_t) was then examined to guide model specification. Based on Figure 3, both the ACF and PACF showed the strongest dependence at the first lag and motivated evaluation of low-order models.

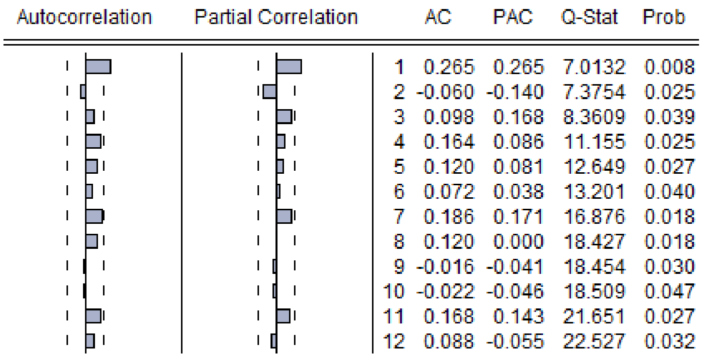

Figure 3. Correlogram of the first differenced pre-crisis series

As shown in Table 5, ARIMA(1,1,1) was excluded because its estimated parameters were not statistically significant. ARIMA (0,1,1) was preferred to ARIMA (1,1,0) because it produced the lowest reported information-criterion values (AIC = 10.6306; SIC = 10.6837). The initially included constant was removed because it was statistically insignificant ($p = 0.3511$). The final ARIMA(0,1,1) residuals did not show significant serial correlation according to the Ljung–Box test ($p > 0.05$), indicating that the conditional-mean specification adequately captured the remaining linear dependence for the subsequent GARCH analysis.

Table 5. Pre-Crisis ARIMA Model Selection Diagnostics

Model	AIC scores	SIC scores	Diagnostic Status
ARIMA(1,1,0)	10.6553	10.7084	Significant / White Noise
ARIMA(0,1,1)	10.6306	10.6837	Significant / White Noise
ARIMA(1,1,1)	—	—	Insignificant Coefficients

4.5.2. Post-Crisis ARIMA Model Identification and Selection (March 2009 – December 2017)

The same procedure was applied to the post-crisis S&P 500 series. The ACF, PACF and Ljung–Box Q-statistics indicate that the differenced series follows a white noise distribution with no significant autocorrelation, reflecting a weaker dependence structure post-crisis.

Following Maximum Likelihood estimation. ARIMA (0,1,0) was selected as the optimal specification (Table 6). It achieved the lowest information criteria (AIC = 10.7386; SIC = 10.7637) and satisfied the residual diagnostics, whereas alternative models including ARIMA (1,1,0) and (0,1,1) yielded insignificant parameters.

Table 6. Post-Crisis ARIMA Model Selection Diagnostics

Model	AIC scores	SIC scores	Diagnostic Status
ARIMA(0,1,0) [Random Walk]	10.7386	10.7637	Significant / White Noise
ARIMA(1,1,0)	—	—	Insignificant Coefficient
ARIMA(0,1,1)	—	—	Insignificant Coefficient

The difference between the selected models suggests that the estimated conditional-mean dynamics varied across the two periods. The shift from a full-sample ARIMA (2,1,2) and a shock-driven pre-crisis ARIMA (0,1,1) to a post-crisis ARIMA (0,1,0) random walk validates the structural change and emphasizes the necessity of regime-specific forecasting frameworks.

4.6. Conditional Heteroskedasticity, Parameter Estimation, and Regime Comparison

The ARCH-LM test is based on the Lagrange Multiplier framework for testing ARCH or GARCH-type conditional-variance dependence (Engle, 1982). The ARCH LM test was conducted and the results in Table 7 indicate different volatility behaviour across the two regimes. For the pre-crisis ARIMA(0,1,1) residuals, the ARCH-LM test is not statistically significant (Prob. F = 0.1001 > 0.05). The null hypothesis of no remaining ARCH effects is therefore not rejected. Accordingly, a separate GARCH component was not required for the pre-crisis model. In contrast, the post-crisis ARIMA(0,1,0) residuals exhibit significant ARCH effects (Prob. F = 0.0428 < 0.05). This result indicates that the conditional variance is related to previous squared disturbances and provides empirical motivation for incorporating a GARCH-type variance equation.

Table 7. ARCH-LM test of the residuals.

Heteroskedasticity Test: ARCH			
ARIMA(0,1,1) (Pre-crisis)	Before fitting GARCH(1,1)	Prob. F(3,90)	0.1001
ARIMA(0,1,0) (Post-crisis)	Before fitting GARCH(1,1)	Prob. F(3,99)	0.0428
	After fitting GARCH(1,1)	Prob. F(3,99)	0.4797

Following the estimation of the post-crisis ARIMA(0,1,0)-GARCH(1,1) model, the ARCH-LM test is no longer statistically significant (Prob. F = 0.4797 > 0.05). This result indicates that the fitted GARCH component removes the statistically detectable ARCH structure from the post-crisis residuals.

As summarized in Table 8, the full-sample series is represented by ARIMA(2,1,2), whereas the regime-specific analysis selects ARIMA(0,1,1) without a constant for the pre-crisis period and ARIMA(0,1,0) with a constant for the post-crisis period. These different specifications indicate that the estimated conditional-mean dynamics vary across the two sample periods.

Table 8. Comprehensive Model Parameters and Volatility Diagnostics Across Regimes

Estimation Horizon	Selected Model Framework	Mean/Variance Intercept (c / ω)	Lag Coefficients ($\phi, \theta / \alpha, \beta$)	AIC / SIC
Full Sample Baseline	ARIMA(2,1,2)–GARCH(1,1)	Significant	Significant	10.6374 / 10.7853
Pre-Crisis Regime	ARIMA(0,1,1)	—	θ_1: Significant	10.6306 / 10.6837
Post-Crisis Regime	ARIMA(0,1,0)–GARCH(1,1)	c = 20.68132*** ω = 757.1544	α_1=0.159420 β_1= 0.557367	10.7386 / 10.7637

Note: *** indicates mean significance at the 1% level.

The regime-adaptive framework specified in Equation (5) permits the lag structure to be selected separately for the two periods. The empirical conditional mean equations for each regime are expressed as follows:

Pre-crisis Regime ($D_t = 0$): $$\Delta P_t = \epsilon_t + 0.372470\epsilon_{t-1} \quad (10)$$

Post-crisis Regime ($D_t = 1$): $$\Delta P_t = 20.68132 + \epsilon_t \quad (11)$$

The post-crisis constant is statistically significant ($c = 20.68132$), indicating a positive estimated average change in the modelled price series during that period.

The corresponding conditional-variance specification in Equation (6) allows the variance dynamics to differ between regimes. In the pre-crisis period ($D_t = 0$), the absence of significant ARCH effects provided no empirical basis for adding a GARCH term. In the post-crisis period ($D_t = 1$), the significant ARCH-LM result supported the use of a GARCH(1,1) specification to model volatility clustering. The resulting variance equations are:

Pre-crisis Regime ($D_t = 0$): $$\sigma_t^2 = 2321.529 \quad (12)$$

Post-crisis Regime ($D_t = 1$): $$\sigma_t^2 = 757.1544 + 0.159420\epsilon_{t-1}^2 + 0.557367\sigma_{t-1}^2 \quad (13)$$

The estimated ARCH and GARCH coefficients are $\alpha_1 = 0.159420$ and $\beta_1 = 0.557367$, respectively. Although neither coefficient is statistically significant at the 5% level, the results should be interpreted together with the ARCH-LM diagnostics, persistence estimate, residual tests, and information criteria.

The estimated persistence measure is 0.7168 ($\alpha_1 + \beta_1 < 1$), which is consistent with a mean-reverting conditional-variance process under standard GARCH conditions. This suggests that the effect of a volatility shock declines over time rather than persisting indefinitely. The GARCH(1,1) specification is supported by the significant pre-estimation ARCH-LM test, the absence of significant remaining ARCH effects after estimation, the persistence estimate, the residual diagnostics, and the model's parsimony. Nelson (1990) emphasizes that GARCH stationarity and persistence depend on the complete variance process and not solely on the coefficient sum. Therefore, the results support the adequacy of the fitted GARCH(1,1) model but do not prove that it is the only valid specification or that it will improve forecasts of future observations. Since standard diagnostics may not detect structural breaks in GARCH dynamics, the ARCH-LM results should also be interpreted alongside the structural-break analysis (Smith, 2008).

The AIC and SIC values in Table 9 provide additional evidence for model selection within each regime, provided that the models are estimated using comparable likelihood samples. If the regime-specific and full-sample models use different observations, their information criteria should not be treated as definitive evidence that one model is superior to the other. Overall, the results indicate that the S&P 500 has different estimated mean and volatility dynamics across the two regimes. Since the analysis uses in-sample fitted values, the break-adjusted framework is interpreted as providing a more differentiated representation of historical market behaviour, not as demonstrating superior future forecasting performance.

4.7. In-Sample Fitted Performance Comparison and Empirical Discussion

This section compares the in-sample fitted performance of the conventional full-sample model and the proposed regime-adaptive model using MAE, RMSE and MAPE. Since the evaluation is based on fitted values generated from the same observations used for estimation, the results measure historical model fit and representation rather than out-of-sample forecasting performance.

The results are presented in Table 9. The conventional full-sample model produced an in-sample MAE of 805.5862, an RMSE of 881.4205, and a MAPE of 59.7674%. The proposed regime-adaptive model produced lower values, with an MAE of 233.1294, an RMSE of 277.4235, and a MAPE of 13.0377%. Relative to the conventional model, the regime-adaptive model reduced the MAE by approximately 71.1%, the RMSE by approximately 68.5%, and the MAPE by approximately 78.2%.

Table 9. In-Sample Model Fit Comparison

Model Specification	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)	Mean Absolute Percentage Error (MAPE %)	In-sample Fitted Outcome
Conventional Full-Sample Baseline	805.5862	881.4205	59.7674%	Higher fitted-value errors; weaker historical fit
Proposed Regime-Adaptive Model	233.1294	277.4235	13.0377%	Lower fitted-value errors; closer historical representation

These results indicate that the regime-adaptive model provides a closer in-sample representation of the observed S&P 500 series than the conventional full-sample model. The lower error values suggest that allowing the model structure to differ between the pre-crisis and post-crisis periods improves the representation of the historical series. This finding is consistent with the earlier results, which showed different conditional-mean specifications and volatility dynamics across the two regimes.

Figure 4 presents the fitted lines generated by the two models. The conventional full-sample model appears to deviate more substantially from the observed series during periods of heightened market uncertainty surrounding the Global Financial Crisis. In contrast, the regime-adaptive model follows the observed historical series more closely across the two regimes. The graphical evidence therefore supports the numerical comparison reported in Table 9. The findings suggest that incorporating structural-break information can improve in-sample model fit when the underlying mean and volatility dynamics vary across market regimes.

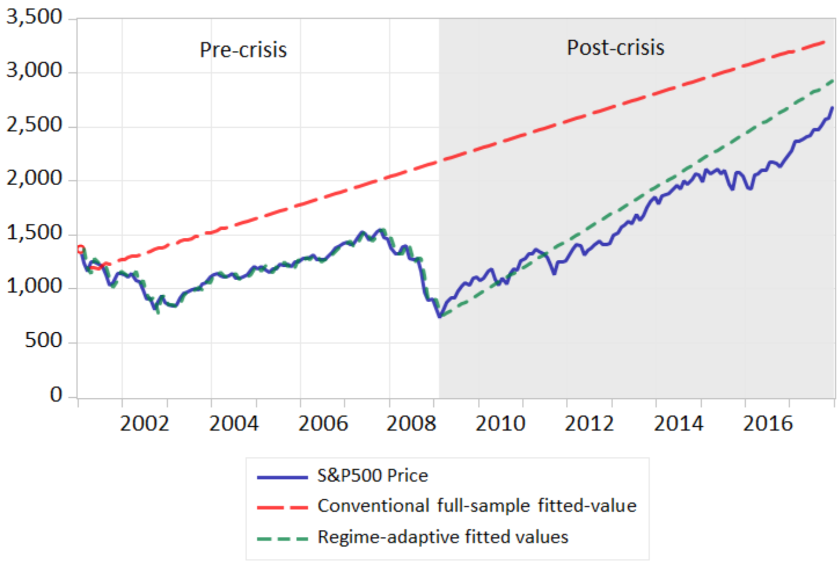

Figure 4. Comparison of Fitted and Observed Values for the Conventional and Regime-Adaptive Models

5. Conclusion

This study presents the development of a regime-adaptive ARIMA-GARCH framework for modelling structural changes in the S&P 500 index. The analysis was motivated by the limitation of conventional full-sample models, which assume that the underlying parameters remain stable throughout the estimation period. The Bai-Perron analysis identified a statistically significant breakpoint in March 2009, which was used to divide the sample into pre-crisis and post-crisis regimes and to investigate whether the conditional-mean and conditional-variance dynamics differed across the two periods. The analysis revealed that the two regimes exhibit distinct patterns of dependence and volatility. The post-crisis residual diagnostics provided evidence of conditional heteroskedasticity, supporting the inclusion of a GARCH-type variance specification. The estimated persistence measure was below one and was therefore consistent with a volatility process in which the effects of shocks gradually decline over time.

The regime-adaptive framework produced lower MAE, RMSE, and MAPE values than the conventional full-sample model. These results indicate that incorporating structural-break information into the ARIMA-GARCH framework improves the representation of changing market dynamics compared with the conventional full-sample approach. However, they should not be interpreted as evidence of higher forecasting accuracy. Since both models were evaluated using observations included in their estimation samples, the comparison may partly reflect differences in in-sample flexibility and parameterization. Demonstrating superior predictive performance would require an out-of-sample, rolling-window, or expanding-window evaluation. Future research should also examine multiple structural breaks and compare this framework with Markov-switching, time-varying-parameter, and other alternative models across different financial markets.

Acknowledgements: The authors would like to express sincere gratitude to Universiti Teknologi MARA Cawangan Pulau Pinang for providing financial support necessary to conduct this study. The suggestions of the reviewers for the improvement to the earlier version of this paper are gratefully acknowledged.

Funding Statement: The authors would like to express sincere gratitude to Universiti Teknologi MARA Cawangan Pulau Pinang for the opportunity to carry out this study by supporting financial aid.

References

Andreou, E., Ghysels, E., & Kourouyiannis, C. (2012). *Robust volatility forecasts in the presence of structural breaks* [Working paper]. University of Cyprus.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica, 66*(1), 47–78. https://doi.org/10.2307/2998540

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics, 18*(1), 1–22. https://doi.org/10.1002/jae.659

Beltratti, A., & Morana, C. (2004). Structural breaks in the volatility of macroeconomic and financial data: The rule, not the exception. In C. D. Aliprantis, K. J. Arrow, P. Hammond, F. Kubler, H. M. Wu, & N. C. Yannelis (Eds.), *Assets, beliefs, and equilibria in economic dynamics* (Studies in Economic Theory, Vol. 18). Springer. https://doi.org/10.1007/978-3-662-05858-9_27

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (2nd ed.). Holden-Day, San Francisco.

Chen, X. (2023). Comparing various GARCH-type models in the estimation and forecasts of volatility of S&P 500 returns during global finance crisis of 2008 and COVID-19 financial crisis. *SHS Web of Conferences, 169*, 01077. https://doi.org/10.1051/shsconf/202316901077

de Oliveira, A. M. B., Mandal, A., & Power, G. J. (2024). Impact of COVID-19 on stock indices volatility: Long-memory persistence, structural breaks, or both? *Annals of Data Science, 11*, 619–646. https://doi.org/10.1007/s40745-022-00446-0

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1007. https://doi.org/10.2307/1912773

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*(2), 357–384. https://doi.org/10.2307/1912559

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labour productivity. *Journal of Economic Perspectives, 15*(4), 117–128. https://doi.org/10.1257/jep.15.4.117

Hwang, S., & Valls Pereira, P. L. (2008). The Effects of Structural Breaks in ARCH and GARCH Parameters on Persistence of GARCH Models. *Communications in Statistics - Simulation and Computation*, *37*(3), 571–578. https://doi.org/10.1080/03610910701812378

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics, 8*(2), 225–234. https://doi.org/10.1080/07350015.1990.10509794

Lee, J. H. H. (1991). A Lagrange multiplier test for GARCH models. *Economics Letters, 37*(3), 265–271. https://doi.org/10.1016/0165-1765(91)90221-6

Lubik, T. A., & Matthes, C. (2015). Time-varying parameter vector autoregressions: Specification, estimation, and an application. *Economic Quarterly, 101*(4), 323–352. https://doi.org/10.21144/eq1010403

Martens, M., van Dijk, D., & de Pooter, M. (2004). *Modeling and forecasting S&P 500 volatility: Long memory, structural breaks and nonlinearity* (Tinbergen Institute Discussion Paper No. 04-067/4). Tinbergen Institute.

Mustapa, F. H., & Ismail, M. T. (2019). Modelling and forecasting S&P 500 stock prices using hybrid ARIMA-GARCH model. *Journal of Physics: Conference Series, 1366*(1), 012130. https://doi.org/10.1088/1742-6596/1366/1/012130

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory, 6*(3), 318–334. https://doi.org/10.1017/S0266466600005296

Perron, P. (2006). Dealing with structural breaks. In G. S. Maddala & D. E. L. McFadden (Eds.), *Palgrave handbook of econometrics: Volume 1: Econometric theory* (pp. 278–352). Palgrave Macmillan.

Perron, P., & Vogelsang, T. J. (1992). Nonstationarity and level shifts with an application to purchasing power parity. *Journal of Business & Economic Statistics, 10*(3), 301–320. https://doi.org/10.1080/07350015.1992.10509907

Smith, D. R. (2008). Testing for structural breaks in GARCH models. *Applied Financial Economics*, *18*(10), 845–862. https://doi.org/10.1080/09603100701262800

Tsuji, C. (2018). Structural breaks and volatility persistence of stock returns: Evidence from the US and UK equity markets. *Applied Economics and Finance, 5*(6), 76–83. https://doi.org/10.11114/aef.v5i6.3690

Vo, N., & Ślepaczuk, R. (2022). Applying Hybrid ARIMA-SGARCH in Algorithmic Investment Strategies on S&P500 Index. *Entropy*, *24*(2), 158. https://doi.org/10.3390/e24020158

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10(3), 251-270. https://doi.org/10.1080/07350015.1992.10509904