Management Studies and Business Journal (PRODUCTIVITY)
Vol 3 (9) 2026 : 318-332

ASYMMETRIC VOLATILITY AND EXOGENOUS FORECASTING OF CONTAINER FREIGHT RATES IN THE SHANGHAI-INDONESIA (CN.SHA-IDTPP) TRADE LANE

VOLATILITAS ASIMETRIS DAN PERAMALAN DENGAN VARIABEL EKSTERNAL PADA TARIF ANGKUT PETI KEMAS DI JALUR PERDAGANGAN SHANGHAI–INDONESIA (CN.SHA-IDTPP)

Ariya Primindra
UPN University
ariyaprimindra80@gmail.com

ABSTRACT
Container shipping rates are a critical cost signal for exporters in Shanghai and importers in Indonesia, but much of the forecasting literature treats South-East Asia as an aggregate market and relies on univariate models. This study consolidates a route-level econometric framework combining unit-root and seasonality diagnostics, ARCH-LM testing, symmetric and asymmetric GARCH models, and ARIMAX/SARIMAX-type forecasting with lagged exchange-rate, bunker-price, and port-dwell-time regressors. Because a verified weekly Shanghai Indonesia spot-rate series has not yet been assembled, the framework is evaluated using 300 weeks of synthetic data covering January 2021 to October 2026. The path represents a pandemic-era peak, a post-pandemic correction, and a later recovery, while the data-generating process embeds annual seasonality, exogenous effects, and EGARCH asymmetry. Results show mixed unit-root evidence, no statistically significant Fourier seasonality, weak-to-borderline ARCH evidence, and no detectable asymmetry despite a positive asymmetry parameter in the generating process. Exogenous regressors reduce one-step-ahead RMSE by 28.3% 31.9% relative to ARIMA, with Diebold Mariano p-values of 0.004 and 0.002; improvements at four weeks are smaller and statistically insignificant.
Keywords: container freight rates, Shanghai Indonesia trade lane, volatility, GARCH, EGARCH, ARIMAX, forecasting.

ABSTRAK
Tarif angkutan peti kemas merupakan indikator biaya yang penting bagi eksportir di Shanghai dan importir di Indonesia. Namun, sebagian besar literatur mengenai peramalan masih memperlakukan Asia Tenggara sebagai pasar agregat dan mengandalkan model univariat. Penelitian ini mengintegrasikan kerangka ekonometrika pada tingkat jalur perdagangan yang menggabungkan pengujian akar unit dan diagnostik musiman, pengujian ARCH-LM, model GARCH simetris dan asimetris, serta peramalan berbasis ARIMAX/SARIMAX dengan variabel eksternal tertunda, yaitu nilai tukar, harga bahan bakar kapal (bunker fuel), dan waktu tunggu di pelabuhan (port dwell time). Karena seri data tarif spot mingguan Shanghai–Indonesia yang terverifikasi belum tersedia, kerangka ini dievaluasi menggunakan 300 minggu data sintetis yang mencakup periode Januari 2021 hingga Oktober 2026. Pola data tersebut merepresentasikan puncak tarif pada era pandemi, koreksi pascapandemi, dan pemulihan pada periode berikutnya. Proses pembentukan data juga memasukkan unsur musiman tahunan, pengaruh variabel eksternal, dan asimetri EGARCH. Hasil penelitian menunjukkan adanya bukti akar unit yang beragam, tidak ditemukannya signifikansi statistik pada pola musiman Fourier, bukti ARCH yang lemah hingga mendekati batas signifikansi, serta tidak terdeteksinya asimetri meskipun parameter asimetri bernilai positif dalam proses pembentukan data. Variabel eksternal menurunkan nilai root mean squared error (RMSE) untuk peramalan satu langkah ke depan sebesar 28,3%–31,9% dibandingkan dengan model ARIMA, dengan nilai-p uji Diebold–Mariano masing-masing sebesar 0,004 dan 0,002. Sementara itu, peningkatan akurasi peramalan untuk horizon empat minggu lebih kecil dan tidak signifikan secara statistik.
Kata Kunci: tarif angkutan peti kemas, jalur perdagangan Shanghai–Indonesia, volatilitas, GARCH, EGARCH, ARIMAX.

Copyright © 2024 THE AUTHOR(S). This article is distributed under a a Creative Commons Attribution-NonCommercial 4.0 International license.

1. INTRODUCTION

Containerisation has made spot freight rates a central price signal in Asian supply chains. For exporters in Shanghai and importers in Indonesia, weekly rates affect landed costs, while their variability shapes decisions about contract coverage, budgeting, and risk management. Container flows have increasingly centred on Asia, and intra-Asian trade links are important to ASEAN economies. The 2024 2025 period also illustrates the potential for abrupt market changes: UNCTAD (2025) describes a strong rise in spot container rates during 2024, a peak around mid-year, and an easing by year-end while rates remained above pre-crisis levels; port congestion contributed to volatility.

The market narrative across 2021 2026 can be divided into three broad phases: the pandemic-era supply-chain crunch and elevated rates in 2021; post-pandemic normalization and downward correction during 2022 2023 as bottlenecks eased and capacity returned; and renewed disruption and regional volatility during 2024 2026, associated in the draft literature with Red Sea diversions around the Cape of Good Hope and seasonal demand surges. These phases provide motivation for a route-specific study, but they do not establish observed Shanghai Indonesia rate levels. The illustrative rate range used in the simulation is not an official Shanghai Shipping Exchange (SSE) series.

Prior work provides a methodological foundation. Munim and Schramm (2017) found that combining ARIMA with an ARCH component improved short-term forecasts for the Far East Northern Europe lane. Schramm and Munim (2021) argued that capacity, demand, and bunker prices are relevant information beyond historical freight-rate lags and reported gains from models incorporating practitioner information. Vemuri and Munim (2023) examined six South-East Asian Freight Index (SEAFI) routes using ARIMA and seasonal neural-network autoregression (SNNAR), finding that relative performance varied by forecast horizon and route. Liu et al. (2022) applied a GARCH-based approach to shipping-market volatility, while Totakura (2019) used HEGY methods to investigate freight-rate seasonality.

Three gaps motivate the present framework. First, regional aggregation can conceal route-specific features of the Shanghai Indonesia lane. Second, univariate models do not explicitly represent conditional variance or asymmetric responses to shocks. Third, they omit observable economic and operational variables that may affect freight rates. For Indonesia, plausible transmission channels include exchange-rate pass-through from dollar-denominated freight costs, bunker fuel costs incorporated into carrier surcharges, and port dwell time that ties up equipment and constrains effective capacity. The expected signs and lags of these relationships must be tested against observed data rather than assumed.

This study investigates the seasonal dynamics, conditional volatility, and forecasting performance of container freight rates on the Shanghai Indonesia trade lane in comparison with the broader South-East Asian Freight Index (SEAFI). Specifically, it examines whether seasonal fluctuations on the Shanghai Indonesia route differ in magnitude and pattern from those observed in the regional aggregate, the extent to which autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) effects, including potential asymmetric responses to positive and negative shocks, characterize route-specific freight rate volatility, and whether incorporating the Indonesian rupiah US dollar (IDR/USD) exchange rate, bunker fuel prices, and port dwell time improves forecasting accuracy relative to conventional univariate time-series benchmarks. Accordingly, the study formulates three hypotheses: H1, container freight spot rates on the Shanghai Indonesia trade lane exhibit statistically significant seasonal fluctuations whose intensity differs from the aggregate SEAFI pattern; H2, the residuals of freight rate models exhibit conditional

heteroskedasticity, indicating time-varying volatility; and H3, forecasting models incorporating exogenous regressors achieve lower out-of-sample root mean squared error (RMSE) and mean absolute percentage error (MAPE) than univariate benchmark models.

To address these research questions systematically, the study develops a pre-specified, route-level econometric framework integrating time-series diagnostics, conditional volatility estimation, exogenous-variable forecasting, and formal statistical comparisons of predictive accuracy. The framework is initially evaluated using a simulation-based dataset comprising approximately 300 weekly observations generated from a process with known parameters. This simulation design enables an assessment of the extent to which the proposed econometric procedures can recover underlying data-generating characteristics and identify volatility dynamics and forecasting improvements within a relatively limited sample. Importantly, the simulated results are used exclusively to evaluate the methodological performance of the proposed framework and are not interpreted as empirical estimates or evidence of actual freight market behaviour on the Shanghai Indonesia trade lane.

2. METHODS

2.1 Research design and data status

The analysis uses a controlled synthetic weekly series of T = 300 observations from 4 January 2021 to 26 October 2026. The real weekly route series has not yet been assembled. The synthetic regime path is log-linearly interpolated between stylised knots. It rises toward approximately USD 4,800 per 40-foot equivalent unit (FEU) in 2021, falls toward USD 950 in early 2024, recovers to around USD 2,300, occupies an intermediate band during 2025, and rises again to roughly USD 3,300 in 2026. These levels are illustrative assumptions, not SSE readings. The process also includes an annual seasonal cycle with log amplitude of about 5% and a semi-annual harmonic.

Three lagged drivers enter the generating process: log IDR/USD with coefficient −0.8, log bunker price with coefficient +0.35, and port dwell time in days with coefficient +0.08. The disturbance has AR(1) persistence of 0.6 and EGARCH(1,1) innovations with $\alpha = 0.22$, $\gamma = +0.14$, and $\beta = 0.93$. Positive shocks are therefore designed to raise variance more strongly. No deterministic variance regime is imposed; differences in yearly volatility arise from stochastic variation.

2.2 Diagnostics for integration, seasonality, and ARCH effects

Let $y_t = \ln(rate_t)$, and let Δy_t denote its first difference. Integration order is assessed using augmented Dickey Fuller (ADF) and Phillips Perron (PP) tests on both levels and first differences. Seasonality is examined with an F-test on two annual Fourier harmonics added to an AR(1) model for Δy_t. A HEGY seasonal unit-root test is specified for the real-data stage but is not run in this simulation. Conditional heteroscedasticity is assessed with Engle's ARCH-LM test on ARIMA(1,1,1) residuals at lags 1, 4, and 8.

2.3 Volatility models

The mean equation is $\Delta y_t = \mu + \varphi \Delta y_{t-1} + \beta' x_{t-1} + \varepsilon_t$, where $\varepsilon_t = \sigma_t z_t$, x_{t-1} contains lagged changes in log IDR/USD and log bunker price and lagged dwell time, and z_t follows a Student-t distribution. The variance specifications are:

$$\text{GARCH}(1,1): \sigma^2_t = \omega + \alpha \varepsilon^2_{t-1} + \beta \sigma^2_{t-1}.$$

$$\text{GJR-GARCH}(1,1,1): \sigma^2_t = \omega + \alpha \varepsilon^2_{t-1} + \gamma \varepsilon^2_{t-1} I(\varepsilon_{t-1} < 0) + \beta \sigma^2_{t-1}.$$

$$\text{EGARCH}(1,1,1): \ln \sigma^2_t = \omega + \alpha(|z_{t-1}| - E|z|) + \gamma z_{t-1} + \beta \ln \sigma^2_{t-1}.$$

Models are estimated by maximum likelihood on percent-scaled series. They are compared using log-likelihood, Akaike information criterion (AIC), Bayesian information criterion (BIC), and ARCH-LM tests on standardized residuals. The asymmetry coefficient is treated as an empirical quantity, not assigned a predetermined market interpretation.

2.4 Forecasting and evaluation

The first 240 weeks form the training sample and the final 60 weeks the test sample. The ARIMA order is selected by AIC over $p,q \in \{0,1,2\}$ with $d = 1$, yielding ARIMA(2,1,2). Five approaches are compared using rolling-origin re-estimation at each test week: a random-walk naive forecast; ARIMA; ARIMA with two Fourier harmonics; ARIMAX with the three lagged regressors; and ARIMAX with Fourier terms, referred to as SARIMAX-type. For one-step forecasts, lagged regressors are available at the forecast origin. For four-step forecasts, future regressor values are held at their last observed levels.

Forecast accuracy is evaluated in USD per FEU using root mean squared error (RMSE) and mean absolute percentage error (MAPE). Pairwise forecast differences against ARIMA are assessed using the Diebold Mariano test with squared-error loss and a small-sample adjustment for multi-step horizons. The proposed real-data protocol would use observed SSE spot rates, IDR/USD data from Bank Indonesia or an appropriate financial series, bunker prices, and measured Indonesian port dwell times; HEGY tests and polar plots would support comparison with SEAFI.

2.5 Market-structure context for future data collection

The second source draft identifies several carrier groups as potentially relevant to capacity and pricing: intra-Asia specialists (SITC, Wan Hai, KMTC, Interasia Lines, and Gold Star Line); global carriers (Evergreen, COSCO, CMA CGM/CNC, Maersk, and MSC); and operators whose networks connect regional ports through hub-and-spoke or feeder services. It proposes that capacity redeployment to longer-haul routes, blank sailings, equipment availability, carrier surcharges, terminal handling charges, and electronic delivery-order compliance may affect observed lane conditions. These carrier rankings and role descriptions have not been verified against schedules or capacity databases and are therefore treated as hypotheses for data collection, not as established findings. Carrier-capacity or blank-sailing indicators may be considered as additional regressors once reliable data are available.

3. RESULTS

3.1 Descriptive profile of the synthetic series

The synthetic series has a mean of approximately USD 2,183 per FEU and ranges from USD 913 to USD 4,807. Weekly log changes have a standard deviation of 4.9%, mild negative skewness (−0.45), and low excess kurtosis (0.65), indicating little fat-tail behaviour in this particular path.

Table 1. Annual volatility

Year	Stylised regime	Mean (USD)	Min	Max	SD weekly log change	Mean EGARCH volatility
2021	Pandemic peak	3,310	1,502	4,807	4.2%	4.0%
2022	Correction	2,754	1,366	4,477	6.0%	5.0%
2023	Trough	1,142	913	1,469	4.5%	4.3%
2024	Red Sea-era recovery	1,821	1,022	2,520	4.5%	3.6%
2025	Intermediate band	1,643	1,370	2,079	4.7%	4.1%
2026	Secondary rise	2,518	1,560	3,325	3.6%	3.3%

Source: Processed Data, 2026

Table 1 presents the annual descriptive statistics and estimated volatility of the synthetic container freight rate series for the 2021 2026 period. The results illustrate substantial variation in freight rate levels across the six years, reflecting the stylised market regimes incorporated into the simulation. Overall, the synthetic series has an average freight rate of approximately USD 2,183 per FEU, with observed values ranging from USD 913 to USD 4,807 per FEU. These figures indicate that the simulated freight market experiences both pronounced increases and substantial corrections over the observation period.

In 2021, freight rates reached their highest annual mean of USD 3,310 per FEU, with a maximum of USD 4,807 and a minimum of USD 1,502. This pattern is consistent with the simulation's pandemic peak regime, characterised by elevated freight rate levels. The standard deviation of weekly log changes was 4.2%, while mean EGARCH volatility was 4.0%, suggesting moderate week-to-week fluctuations relative to the high average rate level. In 2022, the annual mean declined to USD 2,754 per FEU, indicating a correction from the previous year's peak. Nevertheless, weekly volatility increased to 6.0%, the highest among all years, while mean EGARCH volatility rose to 5.0%. The divergence between declining average freight rates and increasing volatility suggests that the simulated correction phase was accompanied by greater short-term uncertainty and more pronounced weekly price movements.

The year 2023 recorded the lowest annual mean freight rate at USD 1,142 per FEU, with values ranging from USD 913 to USD 1,469. This corresponds to the simulated trough regime. Despite the substantial decline in freight rate levels, the standard deviation of weekly log changes remained at 4.5%, and mean EGARCH volatility was 4.3%. These findings suggest that a lower freight rate level does not necessarily imply proportionately lower relative volatility. In other words, freight rates may remain subject to meaningful percentage fluctuations even when their absolute values are comparatively low.

In 2024, the annual mean increased to USD 1,821 per FEU, representing a recovery from the previous year's trough. The maximum rate reached USD 2,520, while the minimum was USD 1,022. Both the standard deviation of weekly log changes and mean EGARCH volatility

declined to 4.5% and 3.6%, respectively. This pattern indicates that the simulated recovery occurred alongside a lower estimated level of conditional volatility than in 2022 and 2023. In 2025, the annual mean decreased slightly to USD 1,643 per FEU. Weekly volatility increased marginally to 4.7%, while mean EGARCH volatility rose to 4.1%. These results indicate that the simulated intermediate regime retained some short-term variability despite the relatively moderate range of freight rates, from USD 1,370 to USD 2,079 per FEU.

In 2026, the annual mean rose to USD 2,518 per FEU, representing a renewed increase in freight rate levels. The maximum reached USD 3,325, while the minimum was USD 1,560. Notably, this year exhibited the lowest standard deviation of weekly log changes, at 3.6%, and the lowest mean EGARCH volatility, at 3.3%. This combination suggests that the simulated secondary rise occurred with comparatively smaller week-to-week percentage fluctuations than those observed in the preceding years. However, because 2026 is marked with an asterisk, its interpretation should account for the possibility that the annual observations cover only part of the year, depending on the simulation's data construction.

Comparing the two volatility measures, the annual standard deviation of weekly log changes is higher than mean EGARCH volatility in every year except 2021, when the two measures are relatively close. This difference is consistent with the measures capturing distinct aspects of variability: the standard deviation summarises the unconditional dispersion of weekly changes within each year, whereas EGARCH volatility represents model-estimated conditional volatility that can vary over time in response to past shocks and information. The differences between the measures should therefore not automatically be interpreted as evidence of model inadequacy.

Overall, Table 1 demonstrates that the synthetic series exhibits changing freight rate levels and time-varying volatility across the stylised regimes. The most pronounced weekly fluctuations occur during the 2022 correction phase, whereas the lowest volatility is observed during the 2026 secondary rise. These descriptive patterns provide an initial motivation for applying conditional volatility models, such as EGARCH, to examine whether freight rate uncertainty varies over time. Nevertheless, because the series is synthetic, the results should be interpreted as characteristics of the simulated data-generating process rather than as empirical evidence of actual freight market conditions on the Shanghai Indonesia trade lane.

3.2 Integration, seasonality, and ARCH diagnostics

Table 2. Integration, seasonality, and ARCH diagnostics

Test	Statistic	p-value	Interpretation
ADF, log level	−2.95	0.040	Rejects unit root at 5% (marginal)
PP, log level	−1.34	0.609	Does not reject unit root
ADF, first difference	−3.77	0.003	Stationary
PP, first difference	−18.23	<0.001	Stationary
Fourier seasonality F-test (2 harmonics)	1.95	0.102	Not significant at 5%

Test	Statistic	p-value	Interpretation
ARCH-LM, lag 1	3.28	0.070	Borderline
ARCH-LM, lag 4	9.50	0.050	Borderline
ARCH-LM, lag 8	10.60	0.225	Not significant

Source: Processed Data, 2026

The results presented in the table provide diagnostic evidence regarding the stationarity, seasonal structure, and conditional heteroskedasticity of the synthetic container freight rate series. These tests are important for determining the appropriate econometric modelling strategy and assessing whether the series exhibits characteristics that justify the application of time-series and volatility models.

The Augmented Dickey Fuller (ADF) test applied to the log-level series produces a statistic of −2.95 with a p-value of 0.040, indicating rejection of the unit root null hypothesis at the 5% significance level. However, the result is relatively marginal. In contrast, the Phillips Perron (PP) test yields a statistic of −1.34 and a p-value of 0.609, providing insufficient evidence to reject the unit root null hypothesis. The disagreement between these two tests suggests that the stationarity of the log-level series is not conclusively established. Nevertheless, both tests indicate that the first-differenced series is stationary: the ADF test reports a statistic of −3.77 ($p = 0.003$), while the PP test produces a statistic of −18.23 ($p < 0.001$). Taken together, these results support the use of first differences, or log returns, in subsequent time-series modelling to reduce the risk of spurious regression associated with non-stationary data.

The Fourier seasonality F-test, based on two harmonics, yields an F-statistic of 1.95 and a p-value of 0.102. Since this p-value exceeds the 5% significance threshold, the null hypothesis of no statistically significant contribution from the tested Fourier seasonal terms cannot be rejected. Thus, the test does not provide sufficient evidence of a statistically significant seasonal pattern in the synthetic series under the specified Fourier specification. However, this finding does not establish that seasonality is entirely absent; rather, any seasonal component captured by the two-harmonic specification is not statistically significant at the conventional 5% level. Additional specifications or alternative seasonal diagnostics may be considered if seasonality remains theoretically relevant to the analysis.

The ARCH-LM tests assess whether the residuals exhibit autoregressive conditional heteroskedasticity, which would indicate that the conditional variance changes over time. At lag 1, the test statistic is 3.28 with a p-value of 0.070, while at lag 4, the statistic is 9.50 with a p-value of 0.050. These results indicate borderline evidence of ARCH effects. The lag-1 result is not statistically significant at the 5% level, whereas the lag-4 result lies at the conventional significance boundary and should be interpreted cautiously, particularly given the rounded p-value. At lag 8, the test yields a statistic of 10.60 and a p-value of 0.225, indicating that the null hypothesis of no ARCH effects cannot be rejected at the 5% level. Overall, the evidence for conditional heteroskedasticity is therefore mixed rather than conclusive across the tested lag structures.

From a modelling perspective, these diagnostic results have several implications. First, the evidence of stationarity in the first-differenced series supports modelling freight rate changes rather than relying exclusively on the log-level series. Second, the Fourier test does not establish statistically significant seasonality under the specified model, suggesting that seasonal components should be included only with appropriate empirical justification. Third, the borderline ARCH-LM results provide a reasonable basis for investigating conditional volatility models, such as GARCH or EGARCH, but do not independently establish that such models are necessary or superior. Their suitability should be evaluated through model estimation, residual diagnostics, and comparative forecasting performance.

Overall, the diagnostic tests suggest that the synthetic series is more consistently stationary after first differencing, while the evidence for seasonality and ARCH effects remains inconclusive at the 5% significance level. These findings support a cautious modelling strategy in which conditional volatility specifications are assessed empirically rather than assumed to be appropriate in advance. Because the underlying observations are synthetic, these conclusions describe the statistical properties of the simulated series and should not be interpreted as direct evidence of the behaviour of actual container freight rates on the Shanghai Indonesia trade lane.

3.3 GARCH-class estimates and asymmetry

Table 3. GARCH-class estimates and asymmetry

Model	Log-likelihood	AIC	BIC	α (p-value)	β (p-value)	γ (p-value)	Persistence
GARCH(1,1)	−833.1	1,684.1	1,717.4	0.122 (0.247)	0.806 (<0.001)	—	0.928
GJR-GARCH(1,1,1)	−833.1	1,686.1	1,723.0	0.116 (0.590)	0.810 (0.006)	0.006 (0.956)	0.929
EGARCH(1,1,1)	−833.2	1,686.3	1,723.3	0.230 (0.203)	0.920 (<0.001)	−0.018 (0.646)	0.920 (β)

Source: Processed Data, 2026

The results presented in the table compare three conditional volatility specifications: GARCH(1,1), GJR-GARCH(1,1,1), and EGARCH(1,1,1). The models are evaluated using their log-likelihood values, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), estimated coefficients, statistical significance, and volatility persistence. These measures provide insights into the dynamics of conditional variance and whether asymmetric responses to positive and negative shocks are supported by the synthetic freight rate series.

The standard GARCH(1,1) model produces a log-likelihood of −833.1, an AIC of 1,684.1, and a BIC of 1,717.4. The estimated ARCH coefficient, α, is 0.122, with a p-value of 0.247, indicating that it is not statistically significant at the 5% level. This suggests that the immediate impact of new shocks on conditional volatility is not statistically distinguishable from zero in this specification. In contrast, the GARCH coefficient, β, is 0.806 and statistically significant (p-value < 0.001), indicating that past conditional variance is an important predictor of current

volatility. The estimated persistence measure of 0.928 suggests that volatility shocks may dissipate gradually, implying considerable continuity in conditional variance over time.

The GJR-GARCH(1,1,1) model yields a log-likelihood of −833.1, an AIC of 1,686.1, and a BIC of 1,723.0. Its ARCH coefficient is 0.116 (p = 0.590), while the GARCH coefficient is 0.810 (p = 0.006). The ARCH coefficient is not statistically significant, whereas the GARCH coefficient remains significant, reinforcing the importance of past conditional variance in explaining volatility dynamics. The asymmetry coefficient, γ, is estimated at 0.006, with a p-value of 0.956. This result provides no statistical evidence of an asymmetric volatility response to positive and negative shocks within the fitted GJR-GARCH specification. Furthermore, the model's AIC and BIC are higher than those of the standard GARCH model, suggesting that the additional asymmetry parameter does not improve the balance between model fit and complexity according to these information criteria.

The EGARCH(1,1,1) model reports a log-likelihood of −833.2, an AIC of 1,686.3, and a BIC of 1,723.3. The estimated coefficient α is 0.230 (p = 0.203), indicating that it is not statistically significant at the 5% level. The coefficient β is 0.920 and highly statistically significant (p-value < 0.001), demonstrating strong persistence in the conditional volatility process. The asymmetry coefficient, γ, is −0.018 (p = 0.646), which is also statistically insignificant. Consequently, the results do not provide sufficient evidence that negative and positive shocks have systematically different effects on volatility in the EGARCH specification. Although EGARCH can accommodate asymmetric effects, the estimated asymmetry parameter does not indicate a statistically significant leverage effect in this synthetic series.

A comparison of the information criteria indicates that the standard GARCH(1,1) model performs best among the three specifications. It has the lowest AIC and BIC, while all three models produce very similar log-likelihood values. The marginal differences in log-likelihood suggest that the more complex GJR-GARCH and EGARCH models do not deliver a meaningful improvement in fit sufficient to offset their additional parameterisation. The persistence estimates are also relatively similar, ranging from 0.920 for EGARCH to 0.929 for GJR-GARCH, indicating that all three specifications capture a relatively persistent volatility process. However, persistence measures should be interpreted according to each model's specific parameterisation and should not automatically be treated as directly equivalent across model classes.

Overall, the estimation results indicate that volatility persistence is the most consistent feature across the fitted models, whereas the immediate ARCH effect and asymmetric responses are not statistically significant at the 5% level. Based on the reported AIC and BIC values, the standard GARCH(1,1) model is the preferred specification among the three alternatives. Nevertheless, this preference should be considered provisional and complemented by residual diagnostics, tests for remaining ARCH effects, and out-of-sample forecasting comparisons. Since the analysis uses synthetic observations, the findings describe the simulated volatility process rather than establish the presence of persistent or asymmetric volatility in actual freight rates on the Shanghai Indonesia trade lane.

3.4 Out-of-sample forecast performance

Table 4. Out-of-sample forecast performance

Horizon	Model	RMSE (USD/FEU)	MAPE (%)	DM statistic (p vs ARIMA)
1 week (n=60)	Naive	82.7	3.09	—
1 week (n=60)	ARIMA(2,1,2)	80.0	3.01	—
1 week (n=60)	ARIMA + Fourier	78.6	2.96	0.54 (0.589)
1 week (n=60)	ARIMAX	57.4	2.27	3.04 (0.004)
1 week (n=60)	SARIMAX-type	54.5	2.29	3.32 (0.002)
4 weeks (n=57)	Naive	192.0	7.10	—
4 weeks (n=57)	ARIMA(2,1,2)	163.7	6.45	—
4 weeks (n=57)	ARIMA + Fourier	158.7	5.90	0.32 (0.749)
4 weeks (n=57)	ARIMAX	155.1	6.25	0.71 (0.483)
4 weeks (n=57)	SARIMAX-type	141.5	5.79	1.65 (0.105)

Source: Processed Data, 2026

Table 4 presents the out-of-sample forecasting performance of five models at one-week and four-week horizons. Forecast accuracy is evaluated using the root mean squared error (RMSE) and mean absolute percentage error (MAPE), while the Diebold Mariano (DM) test is used to assess whether the predictive accuracy of alternative models differs significantly from that of the ARIMA(2,1,2) benchmark. Lower RMSE and MAPE values indicate better forecasting performance.

For the one-week forecasting horizon, the Naive model produces an RMSE of USD 82.7 per FEU and a MAPE of 3.09%, whereas the ARIMA(2,1,2) model improves these values to USD 80.0 and 3.01%, respectively. Incorporating Fourier terms into the ARIMA specification further reduces the RMSE to USD 78.6 and the MAPE to 2.96%. However, the DM test result for ARIMA + Fourier is not statistically significant (DM = 0.54, p = 0.589), indicating that its forecasting performance is not significantly different from that of the ARIMA benchmark at conventional significance levels.

More substantial improvements are observed when exogenous variables are incorporated into the forecasting models. The ARIMAX model achieves an RMSE of USD 57.4 per FEU and a MAPE of 2.27%, while the SARIMAX-type model records the lowest errors, with an RMSE of USD 54.5 and a MAPE of 2.29%. Relative to the ARIMA(2,1,2) benchmark, these results represent reductions in RMSE of approximately 28.3% for ARIMAX and 31.9% for SARIMAX-type. The corresponding DM statistics are 3.04 (p = 0.004) and 3.32 (p = 0.002), respectively, indicating statistically significant differences in predictive accuracy at the 5% level. Assuming the DM statistics are defined as benchmark loss minus alternative-model loss, the positive statistics support the interpretation that both models outperform the ARIMA benchmark. These findings suggest that incorporating relevant external information may improve short-term freight rate forecasts in the simulated setting.

At the four-week forecasting horizon, forecast errors increase across all models, reflecting the greater uncertainty associated with predictions further into the future. The Naive model records an RMSE of USD 192.0 per FEU and a MAPE of 7.10%, while the ARIMA(2,1,2) benchmark achieves an RMSE of USD 163.7 and a MAPE of 6.45%. The ARIMA + Fourier model improves on the benchmark, reducing RMSE to USD 158.7 and MAPE to 5.90%. Nevertheless, the DM statistic of 0.32 ($p = 0.749$) indicates that the improvement is not statistically significant. Similarly, ARIMAX achieves an RMSE of USD 155.1 and a MAPE of 6.25%, but its DM test result ($DM = 0.71$, $p = 0.483$) does not provide sufficient evidence of a statistically significant difference from the benchmark.

The SARIMAX-type model delivers the strongest overall performance at the four-week horizon, with an RMSE of USD 141.5 per FEU and a MAPE of 5.79%. Compared with ARIMA(2,1,2), this corresponds to an approximate 13.6% reduction in RMSE and a 10.2% reduction in MAPE. However, the DM statistic of 1.65 ($p = 0.105$) does not reach the conventional 5% significance threshold. Therefore, although SARIMAX-type produces the lowest forecast errors at this horizon, the available statistical evidence is insufficient to conclude that its predictive accuracy is significantly superior to that of the ARIMA benchmark.

Comparing the two forecast horizons reveals that the relative benefits of model complexity and exogenous information depend on the prediction horizon. At one week, both ARIMAX and SARIMAX-type demonstrate statistically significant improvements over the ARIMA benchmark according to the reported DM tests. At four weeks, the SARIMAX-type model maintains the lowest RMSE and MAPE, but its improvement is not statistically significant at the 5% level. This pattern suggests that incorporating external predictors may be particularly beneficial for short-term forecasts, while the gains become less conclusive as the forecasting horizon increases. The results also show that the model with the lowest RMSE does not necessarily have the lowest MAPE, as illustrated by the one-week ARIMAX and SARIMAX-type comparison. This distinction highlights the importance of evaluating forecasting performance using multiple complementary metrics.

Overall, Table 4 indicates that ARIMAX and SARIMAX-type are the most promising models for short-term forecasting, while SARIMAX-type achieves the lowest point-estimate errors at both forecasting horizons. However, statistical superiority is established only for the one-week ARIMAX and SARIMAX-type forecasts relative to the ARIMA benchmark, based on the reported DM tests. At the four-week horizon, the evidence of improved predictive accuracy remains inconclusive. These findings should be interpreted within the scope of the synthetic dataset and do not establish that the same forecasting advantages would necessarily hold for actual container freight rates on the Shanghai Indonesia trade lane.

3.5 Hypothesis assessment

Table 5. Hypothesis assessment

Hypothesis / issue	Evidence	Assessment
H1: seasonality differs from SEAFI	Fourier F-test p = 0.102; no SEAFI series in simulation	Not testable here; defer to observed data
H2: ARCH/GARCH effects	LM p = 0.050 at lag 4, 0.070 at lag 1, 0.225 at lag 8; β significant but α not	Weak and inconclusive

H3: exogenous variables improve accuracy	One-step RMSE falls 28% 32%, DM p ≤ 0.004; four-step p ≥ 0.105	Supported at h = 1 only
Asymmetric volatility	γ is not significant despite true γ = +0.14	Not detected; finite-sample power is limited

Source: Processed Data, 2026

Table 5 summarises the assessment of the study's hypotheses based on the diagnostic tests, volatility model estimates, and out-of-sample forecasting results obtained from the synthetic dataset. The findings provide varying levels of support for the proposed hypotheses, with some relationships remaining inconclusive because of statistical limitations and the absence of observed SEAFI data.

Hypothesis H1, which proposes that seasonal fluctuations in freight rates on the Shanghai Indonesia trade lane differ from those observed in the broader SEAFI, cannot be conclusively tested using the current simulation. The Fourier seasonality F-test yields a p-value of 0.102, indicating that the seasonal terms are not statistically significant at the 5% level. Furthermore, the simulation does not include an observed SEAFI series, preventing a direct comparison of seasonal patterns between the route-specific and regional indices. Therefore, H1 remains untested rather than rejected. Evaluating this hypothesis requires actual freight rate observations for both the Shanghai Indonesia route and the SEAFI over a comparable period.

Hypothesis H2 concerns the presence of ARCH/GARCH effects in freight rate volatility. The diagnostic results provide mixed evidence of conditional heteroskedasticity: the ARCH-LM test produces p-values of 0.070 at lag 1, 0.050 at lag 4, and 0.225 at lag 8. Although the lag-4 result lies at the conventional 5% significance boundary, the results are not consistently significant across the tested lag structures. In the estimated GARCH models, the volatility persistence coefficient, β, is statistically significant, whereas the ARCH coefficient, α, is not. These findings suggest that past conditional variance contributes to the estimated volatility dynamics, but the evidence for systematic ARCH effects remains weak and inconclusive. Accordingly, H2 receives only limited support and should not be regarded as definitively established.

Hypothesis H3 proposes that forecasting models incorporating exogenous variables achieve greater predictive accuracy than univariate benchmark models. The one-week forecasting results support this hypothesis: the ARIMAX and SARIMAX-type models reduce RMSE by approximately 28% 32% relative to the ARIMA(2,1,2) benchmark, with Diebold Mariano test p-values of 0.004 and 0.002, respectively. These statistically significant results indicate that the inclusion of external predictors improves short-term forecasting performance within the simulated setting. However, the evidence is less conclusive at the four-week horizon, where the reported DM test p-values are at least 0.105. Although the SARIMAX-type model achieves the lowest RMSE and MAPE at this horizon, its improvement over the benchmark is not statistically significant at the 5% level. Therefore, H3 is supported for one-week forecasts but not conclusively supported for four-week forecasts.

The assessment of asymmetric volatility also produces an inconclusive result. The asymmetry coefficient, γ, is statistically insignificant in both the GJR-GARCH and EGARCH specifications. Consequently, the estimated models do not detect a statistically significant difference in volatility responses to positive and negative shocks, despite the simulation's underlying asymmetry parameter being set to $\gamma = +0.14$. This discrepancy suggests that the estimation procedure may have insufficient statistical power to identify the imposed asymmetry in the available sample. However, the non-significant estimates do not prove that asymmetry is absent; they indicate only that it is not reliably detected under the current simulation design and model specifications.

Overall, the hypothesis assessment demonstrates that the proposed econometric framework produces encouraging results for short-term forecasting, while evidence concerning seasonality, conditional heteroskedasticity, and asymmetric volatility remains limited or inconclusive. The distinction between hypotheses that are unsupported and those that cannot be tested is particularly important: H1 cannot be assessed without observed SEAFI data, whereas H2 and the asymmetry analysis yield mixed or statistically insignificant results. Since the findings are derived from synthetic observations, they should be interpreted as an evaluation of the proposed methodological framework rather than as confirmation of empirical relationships in the actual Shanghai Indonesia container freight market.

4. DISCUSSION

The simulation results are directionally consistent with prior forecasting studies but do not replicate their empirical findings. Schramm and Munim (2021) reported that external information can improve forecasts over pure ARIMA, and the one-week gains observed here show that the estimation workflow can exploit regressors when they genuinely enter the generating process. Because those effects were deliberately built into the simulation, the result does not demonstrate that IDR/USD, bunker prices, or dwell time improve forecasts of actual Shanghai Indonesia rates. That claim requires observed data and out-of-sample validation.

The volatility results also counsel restraint. Munim and Schramm (2017) reported benefits from an ARCH component for a different trade lane. In this exercise, conditional heteroscedasticity is only weakly indicated, and the symmetric model is preferred by information criteria. More importantly, the failure to detect the known asymmetry parameter demonstrates that a non-significant estimate is not evidence that asymmetry is absent. Longer samples, higher-frequency data, repeated simulations, or more informative volatility covariates may be needed to assess statistical power.

The lack of significant Fourier seasonality does not establish that the actual route lacks seasonal effects. The seasonal amplitude is small relative to the regime path and weekly noise, and no SEAFI data are present for a direct comparison. In the real-data stage, polar plots and HEGY tests should be applied to observed route-level and SEAFI series. The proposed seasonal peaks around regional holidays, Golden Week, and pre-holiday stocking remain candidate mechanisms rather than results of the current analysis.

Market-structure variables may improve the explanatory framework if they can be measured reliably. The source draft proposes examining carrier capacity redeployment, blank

sailings, equipment availability, surcharges, terminal handling charges, and local operational requirements. However, its carrier rankings and strategic-role descriptions have not been independently verified. These should not be treated as established explanations of the synthetic results or as verified descriptions of the actual lane. For future estimation, carrier-capacity or blank-sailing indicators could be added to the exogenous set if consistent historical records are available.

4.1 Practical implications

For freight forwarders and Indonesian importers, the simulation supports testing lagged fuel, exchange-rate, and port-dwell indicators for short-horizon budgeting, while treating longer-horizon forecasts cautiously. For policymakers, port-efficiency metrics could be incorporated into freight-cost forecasting once measured consistently. Both implications are conditional on confirmation with real-data estimates; the present simulation cannot establish operational or policy effects in the market.

5. CONCLUSION

This article integrates two drafts into a single IMRAD study and specifies a route-level econometric protocol for the Shanghai Indonesia (CN.SHA IDTPP) container freight lane. The controlled experiment uses 300 synthetic weekly observations with a known regime path, seasonality, exogenous effects, and asymmetric EGARCH innovations. First, the comparison between the lane and SEAFI (RQ1) remains unanswered because the simulation contains no observed SEAFI counterpart. Second, for volatility (RQ2), ARCH evidence is borderline, persistence is recovered reasonably well, and asymmetry is not detected despite being present in the generating process. Third, for forecasting (RQ3), lagged exogenous regressors substantially improve one-week-ahead accuracy, but gains at four weeks are smaller and statistically insignificant.

The central limitation is that the rate series and exogenous effects are synthetic. The SNNAR benchmark and HEGY test were not implemented; future four-step regressor values were held at their last observations; volatility models were not evaluated using risk or interval-forecast scoring; and only one simulation path was analysed. The results must not be interpreted as actual Shanghai Indonesia market estimates.

The next stage should estimate the unchanged protocol using verified weekly SSE spot rates for the route, IDR/USD data, bunker prices, and port dwell-time series. It should add HEGY tests and polar plots for the SEAFI comparison, consider EGARCH-X or GJR-X specifications, verify carrier and capacity indicators against schedule databases, repeat the simulation to characterize estimator performance and asymmetry-test power, and evaluate volatility forecasts using appropriate scoring or loss functions.

6. REFERENCES

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

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427 431. https://doi.org/10.1080/01621459.1979.10482531

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253 263.

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

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance, 48, 1779 1801.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679 688.

Hylleberg, S., Engle, R. F., Granger, C. W. J., & Yoo, B. S. (1990). Seasonal integration and cointegration. Journal of Econometrics, 44(1 2), 215 238.

Liu, J., Li, Z., Sun, H., Yu, L., & Gao, W. (2022). Volatility forecasting for the shipping market indexes: An AR-SVR-GARCH approach. Maritime Policy & Management, 49(6), 864 881.

Munim, Z. H., & Schramm, H.-J. (2017). Forecasting container shipping freight rates for the Far East Northern Europe trade lane. Maritime Economics & Logistics, 19(1), 106 125. https://doi.org/10.1057/s41278-016-0051-7

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59(2), 347 370. https://www.jstor.org/stable/2938260

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2), 335 346. https://doi.org/10.1093/biomet/75.2.335

Schramm, H.-J., & Munim, Z. H. (2021). Container freight rate forecasting with improved accuracy by integrating soft facts from practitioners. Research in Transportation Business & Management, 41, Article 100662. https://doi.org/10.1016/j.rtbm.2021.100662

Totakura, B. R. (2019). Analysis of seasonality patterns in container freight rates using HEGY method. Journal of Maritime Research, 16(2), 87 95. https://www.jmr.unican.es/jmr/article/download/557/618/2028

United Nations Conference on Trade and Development. (2025). Review of maritime transport 2025: Staying the course in turbulent waters (Chapter III: Freight rates and maritime transport costs). https://unctad.org/system/files/official-document/rmt2025ch3_en.pdf

Vemuri, S., & Munim, Z. H. (2023). Seasonality and forecasting analysis of the South-East Asian container freight market. Maritime Business Review, 8(2), 121 138. https://doi.org/10.1108/MABR-06-2021-0047