

Comparative Forecasting of MTN Nigeria Stock Prices and Trading Volume using Box–Jenkins Models and a Reduced Factorial Design

Odior Kenejime Anthony and Emudiaga Rukevwe Ericson

Department of Statistics, Delta State Polytechnic, Otefe-Oghara, Delta State, Nigeria.

*Corresponding authors' email: emudiagaeric@gmail.com

ORCID: <https://orcid.org/0009-0004-9199-5919>

ABSTRACT

This study compares Box-Jenkins forecasting structures (ARMA(p,q), ARIMA(p,d,q), and SARIMA(p,d,q)(P,D,Q)s) for the daily closing stock price and trading volume of MTN Nigeria Communications Plc (MTNN) on the Nigerian Exchange Group (NGX), and decomposes the structural drivers of accuracy using a factorial design. Using 249 daily observations from 26 May 2025 to 26 May 2026, eight candidate configurations spanning AR order ($p \in \{1,2\}$), differencing order ($d \in \{0,1\}$) and seasonal AR order ($P \in \{0,1\}$) were evaluated for trading volume, forming a complete 2^3 factorial design; for stock price, only six of the eight combinations were estimated – the two SARIMA specifications without differencing ($d = 0$, $P = 1$) were not fitted – so the price analysis is a reduced, non-orthogonal design, and its factorial effects are reported with that limitation stated explicitly. Models were estimated on a training sample of 227 observations and evaluated out-of-sample using rolling one-step-ahead forecasts across 22 validation observations. By RMSE, ARIMA(1,1,0) produced the most accurate stock price forecasts (RMSE = 24.21 NGN; MAPE = 1.71%), although ARMA(1,0) and ARMA(2,0) achieved lower MAE and MAPE; SARIMA(2,1,0)(1,0,0)s produced the most accurate trading-volume forecasts on all four criteria (RMSE = 3,177,444; MAPE = 44.69%). No single model was superior across both series or across all criteria: for price, first-order differencing was the dominant structural factor; for volume, AR order and the interaction between differencing and seasonal AR order were the dominant factors. These patterns are consistent with, rather than confirmed by, the factorial analysis, given the reduced design used for price.

Received: 18 Aug. 2026

Accepted: 05 Oct. 2026

Published: 07 Oct. 2026

Keywords:

Box-Jenkins; ARIMA; SARIMA; Rolling forecast; Factorial design; MTN Nigeria; NGX; Stock price; Trading volume

INTRODUCTION

Forecasting stock prices and trading volumes is one of the most practically consequential problems in empirical finance. Reliable forecasts inform portfolio allocation, risk management decisions, and trading strategies across institutional and retail markets alike. African equity markets present a particularly interesting forecasting environment: they tend to exhibit thin trading, periodic illiquidity, and price behaviour driven by a combination of macroeconomic announcements and concentrated institutional order flow (Acquah-Sam, 2022). On the Nigerian Exchange Group (NGX), MTN Nigeria Communications Plc (MTNN) is one of the most actively traded telecoms equities on the exchange, making it a natural testing ground for time-series forecasting methods.

The Box-Jenkins framework; encompassing ARMA, ARIMA, and SARIMA model families, remains the most widely adopted approach to univariate financial time-series forecasting. Its systematic identification–estimation–diagnostic cycle provides a principled basis for capturing serial dependence in financial data, and it serves as a standard benchmark against which more complex machine learning methods are evaluated (Hansun et al., 2022). Despite this, most published studies evaluate only one or two Box-Jenkins structures, apply them to a single financial series, and report in-sample fit rather than genuine out-of-sample forecast accuracy. To the authors' knowledge, comparative evaluations that span all three structural families, assess both

price and volume simultaneously, and use rigorous rolling forecast validation have not previously been reported for an African equity, though the literature on any one of these elements individually is more developed (Deb, 2019; Mondal et al., 2019; Acheampong & Wiafe, 2019).

This study addresses that gap directly. Using 249 daily observations of MTNN closing prices and trading volumes from 26 May 2025 to 26 May 2026, six competing Box-Jenkins configurations are evaluated through a complete identification–estimation–diagnostic cycle and compared on rolling one-step-ahead forecast accuracy across a 22-observation validation window. In addition, a 2^3 factorial design framework is used to decompose the contribution of three structural model characteristics (AR order (p), differencing order (d), and seasonal AR order (P)) to forecast accuracy, providing a transparent and reproducible account of why certain structures outperform others. The study makes three contributions: (i) a rigorous end-to-end Box-Jenkins application to a prominent African equity; (ii) separate evaluation of stock price and trading volume, which are shown to require fundamentally different model structures; and (iii) a 2^3 factorial analysis that systematically explains the structural sources of forecast accuracy differences.

The remainder of the paper is organised as follows. Section 2 reviews the closely related empirical literature. Section 3 describes the data and methodology. Section 4 presents and discusses the results. Section 5 concludes with practical recommendations.

Empirical Review

A growing body of literature applies Box-Jenkins models to financial time-series data in emerging and developing markets. Deb (2019) evaluated ARIMA models for Indian equity indices and found that ARIMA(1,1,0) and ARIMA(2,1,0) consistently outperformed naïve random-walk benchmarks on both in-sample fit and out-of-sample RMSE, with the degree of differencing identified as the most consequential structural choice. This finding aligns with theoretical expectations: equity prices that follow a unit-root process produce spurious regression results when modelled in levels, and first-differencing restores valid inference. A similar conclusion was reached by Mondal, Shit, and Goswami (2019), who applied ARIMA structures to 56 Indian stocks and found that ARIMA(1,1,0) was the single most frequently selected optimal specification, appearing in roughly 35% of cases.

For African equity markets specifically, Acheampong and Wiafe (2019) evaluated ARIMA models on the Ghana Stock Exchange and found that first-order integrated models reliably outperformed stationary ARMA specifications, attributing this to the persistent trending behaviour characteristic of frontier equity markets. Evidence on trading-volume forecasting specifically for African equities is sparser: Tweneboah, Owusu Junior, and Oseifuah (2019) study integration among major African stock markets using a multi-scale wavelet decomposition approach rather than Box-Jenkins volume forecasting, so their results do not directly speak to the SARIMA-versus-ARIMA volume question addressed here. The weekly-seasonality rationale for volume (settlement and institutional rebalancing cycles) is a conjecture motivating the present study's design rather than an established finding from this source, and it is treated as such throughout.

More recent work has emphasised the importance of rolling out-of-sample evaluation as a standard for model comparison. Fahriya, Suparti, and Tarno (2021), comparing ARIMA and SARIMA specifications for forecasting the volume of online motorcycle-taxi trips (a different empirical setting from equity markets) report that in-sample AIC selection did not always identify the model with the best rolling out-of-sample accuracy, a general methodological caution against relying on in-sample criteria alone that motivates this study's rolling-window validation design. Hansun et al. (2022) conducted a systematic review of ARIMA applications in financial forecasting and found that (i) MAPE values below 5% are consistently achievable for stock price forecasting when the differencing order is correctly identified, and (ii) trading volume forecasting is substantially more difficult, with MAPE values typically in the 30–50% range even for well-specified models, a benchmark directly relevant to the present findings.

Despite this evidence base, two gaps appear to persist in the literature reviewed here. First, the authors are not aware of a published study that simultaneously evaluates all three Box-Jenkins families (ARMA, ARIMA, SARIMA) for both stock price and volume of the same equity on an African exchange under a unified rolling evaluation design. Second, the structural determinants of forecast accuracy (how AR order, differencing, and seasonal components individually and jointly affect out-of-sample performance) do not appear to have been examined using a factorial design framework in this context. The present study addresses both gaps as identified, while acknowledging that the factorial component for stock price is based on a reduced, six-run design (Section 3.5).

MATERIALS AND METHODS

Data

The dataset consists of the daily closing price (NGN) and trading volume (shares) for MTNN on the Nigerian Exchange Group, covering 26 May 2025 to 26 May 2026. After removing weekends and public holidays, the sample contains 249 trading observations. The training set spans 26 May 2025 to 24 April 2026 ($n = 227$) and was used for model identification, estimation, and diagnostic checking. The validation set spans 27 April 2026 to 26 May 2026 ($n = 22$) and was reserved exclusively for out-of-sample forecast evaluation. The data was obtained from the MTN Nigeria daily trading report (<https://www.mtn.ng/investors/shareholders/share-price-and-charts/>), an official data portal of the MTN Nigeria Share Prices, retrieved on May 27th, 2026. The closing prices used for this study are unadjusted for corporate actions (dividends, splits, bonus issues) over the sample window. No trading days were missing from the exchange calendar within this window; weekends and public holidays were excluded rather than treated as missing values, so no imputation was required.

Box-Jenkins Framework

The study implements the complete Box-Jenkins cycle: identification, estimation, and diagnostic checking (Box, Jenkins, Reinsel, & Ljung, 2015). Stationarity was assessed using both the Augmented Dickey-Fuller (ADF) test (H_0 : unit root) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. Joint interpretation of all tests was used to determine the appropriate differencing order d . Where ADF and KPSS results conflicted, both $d = 0$ and $d = 1$ models were estimated and compared on out-of-sample RMSE.

The general forms of the three model families are

$$\text{ARMA}(p,q): \varphi(B) X_t = \theta(B) \varepsilon_t \quad (1)$$

$$\text{ARIMA}(p,d,q): \varphi(B)(1-B)^d X_t = \theta(B) \varepsilon_t \quad (2)$$

$$\text{SARIMA}(p,d,q)(P,D,Q)_s: \varphi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)\varepsilon_t \quad (3)$$

where B is the backshift operator, $\varphi(B)$ and $\theta(B)$ are the non-seasonal AR and MA polynomials, $\Phi(B^s)$ and $\Theta(B^s)$ are the seasonal AR and MA polynomials, and ε_t is white noise. Parameters were estimated by maximum likelihood, the estimation method standard for ARMA/ARIMA/SARIMA models, rather than ordinary least squares, which is appropriate only for the pure-AR special case. The study used Python `statsmodels.tsa.statespace.SARIMAX`, with mean/drift term included for each differenced specification. In-sample model fit was assessed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Residual adequacy was evaluated via the Ljung-Box Q -statistic at lag 10 and the Shapiro-Wilk normality test.

Competing Model Structures

Six candidate models were evaluated for each series: ARMA(1,0) and ARMA(2,0) on levels (baseline); ARIMA(1,1,0) and ARIMA(2,1,0) with $d = 1$; and SARIMA(1,1,0)(1,0,0)_s and SARIMA(2,1,0)(1,0,0)_s with weekly seasonality ($s = 5$). These six configurations represent three levels of AR order ($p \in \{1,2\}$), two levels of differencing ($d \in \{0,1\}$), and two levels of seasonal AR ($P \in \{0,1\}$), forming a 2^3 factorial structure over the primary design space.

Rolling One-Step-Ahead Forecast Evaluation

Out-of-sample forecast accuracy was assessed using a rolling one-step-ahead scheme. At each validation step t , the model

was re-estimated on all data up to $t-1$ and a single forecast was produced for t . This was repeated sequentially across all 22 validation observations, accumulating prediction errors at each step in a manner that mirrors operational forecasting conditions. Four accuracy metrics were computed, with y_i the observed value, $\hat{y}_i$ the one-step-ahead forecast and $n=22$ validation observations:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (4)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5)$$

$$\text{MAE} = \left(\frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \right) \quad (6)$$

$$\text{MAPE} = \left(\frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \right) \quad (7)$$

Factorial Design Framework

Following Montgomery (2017) on 2^k factorial experiments, the model selection problem is reformulated as a designed experiment (Montgomery, Jennings, & Kulahci, 2015). Three binary factors are defined: AR order p (levels 1 and 2, coded $-1/+1$), differencing order d (levels 0 and 1, coded $-1/+1$), and seasonal AR order P (levels 0 and 1, coded $-1/+1$). Out-of-sample RMSE is the scalar response variable. For trading

volume, all eight treatment combinations of the 2^3 design were evaluated (the two additional runs, $d=0$ with $P=1$, appear among the volume forecasts in Table 5), so the complete set of main effects, two-factor interactions and the three-factor interaction is estimable and is reported in Table 7. For stock price, only six of the eight combinations were estimated – the two SARIMA specifications without differencing ($d=0$, $P=1$) were not fitted – so the price design is reduced and non-orthogonal. The p , d and $p \times d$ effects for price are estimated from the balanced 2×2 subdesign at $P=0$ (four runs); because P is observed only at $d=1$ for price, its effect cannot be separated from a possible $d \times P$ interaction and is reported only as a descriptive, non-orthogonal comparison rather than as an estimated factorial effect. Each effect is computed as $\text{Effect}_x = (1/2^{k-1}) \sum_i (\text{sign}_{x,i} \cdot y_i)$ (8), where $\text{sign}_{x,i}$ is $+1$ or -1 according to whether factor X is at its high or low level in run i , y_i is the observed RMSE for run i , and k is the number of factors in the estimable subdesign ($k=3$ for volume; $k=2$ for the price $p \times d$ subdesign). This framework transforms the model comparison from a simple ranking exercise into an interpretable decomposition of the structural drivers of forecast accuracy (Box & Draper, 2007).

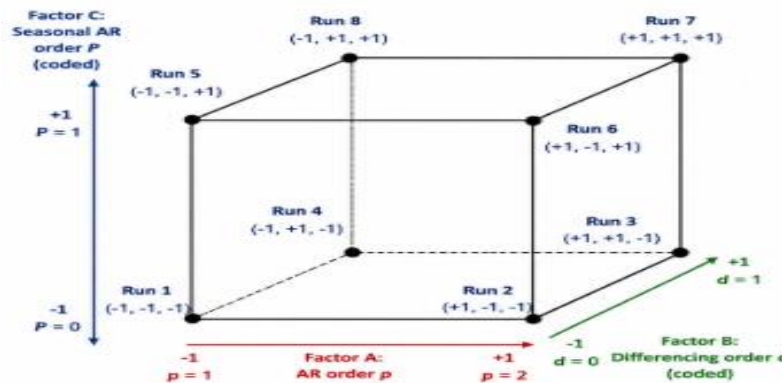


Figure 1: The complete 2^3 factorial design space (AR order p , differencing order d , seasonal AR order P). All eight vertices were evaluated for trading volume; for stock price, the two vertices with $d=0$, $P=1$ were not estimated, so the price analysis realises six of the eight vertices shown (Source: Authors)

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 summarises the training-period descriptive statistics. The MTNN closing price rose sharply from NGN 270 in late May 2025 to a training-period peak of NGN 825.00 in late April 2026 (Table 1) before partially retracing, yielding a mean of NGN 521.38 and a standard deviation of NGN

143.24. The moderate positive skewness (0.63) and Jarque-Bera rejection of normality ($p < 0.001$) are consistent with the trending, structurally shifting behaviour visible in the time plot. Trading volume shows considerably greater volatility, with a mean of approximately 3.24 million shares but a maximum of nearly 24.6 million shares, producing extreme positive skewness (2.79) and excess kurtosis (10.22).

Table 1: Descriptive Statistics - Training Sample ($n = 227$)

Statistic	Stock Price (NGN)	Trading Volume (Shares)
Mean	521.38	3,241,169
Median	475.00	2,202,930
Std. Deviation	143.24	3,358,270
Minimum	270.00	177,770
Maximum	825.00	24,588,685
Skewness	0.6321	2.7864
Excess Kurtosis	-0.5531	10.2185
Jarque-Bera	18.01***	1,281.35***

*** $p < 0.001$.

Stationarity Tests

Table 2 reports the stationarity test results. For closing price, the ADF t-statistic of -0.42 fails to reject the unit root null, and both KPSS statistics (level-stationary (3.98) and trend-stationary (0.74)) clearly exceed their 5% critical values, together classifying the price series as $I(1)$. First differencing resolves this: the ADF falls to -9.99 and KPSS(c) drops to 0.08, both confirming stationarity. This finding is consistent with the broad financial econometrics literature documenting that equity price series follow random-walk or near-random-walk processes (Deb, 2019; Mondal et al., 2019).

For trading volume, the ADF statistic of -6.59 rejects the unit root null, yet both KPSS statistics exceed their 5% critical values (KPSS(c) = 0.82; KPSS(ct) = 0.26), rejecting stationarity in both forms. This conflict is consistent with the heavy-tailed, spike-driven nature of the volume series rather than a textbook unit-root process. Given the ambiguity, $d = 1$ was applied conservatively to ARIMA and SARIMA volume specifications, and models with $d = 0$ were also evaluated as a robustness check; results from both are compared in Section 4.4.

Table 2: Stationarity Tests - ADF and KPSS

Series	Transformation	ADF Stat	ADF 5%CV	KPSS(c)	KPSS(c) 5%CV	KPSS(ct)	KPSS(ct) 5%CV	Decision
Stock Price	Levels	-0.42	-2.86	3.98	0.463	0.74	0.146	Non-stationary $I(1)$
Stock Price	1st Diff.	-9.99	-2.86	0.08	0.463	0.02	0.146	Stationary $I(0)$
Volume	Levels	-6.59	-2.86	0.82	0.463	0.26	0.146	Ambiguous
Volume	1st Diff.	-17.10	-2.86	0.02	0.463	0.01	0.146	Stationary $I(0)$

Both $d = 0$ and $d = 1$ volume models are evaluated and compared on out-of-sample RMSE.

In-Sample Model Fit

Table 3 presents AIC and BIC values. For stock price, SARIMA (1, 1, 0) (1, 0, 0)_s achieves the lowest AIC (1,803.13) and BIC (1,813.33), indicating that even after penalising for its additional seasonal AR parameter, it fits the training data better than the non-seasonal models. ARIMA (2, 1, 0) is the best non-seasonal model in-sample (AIC = 1,825.27). For trading volume, SARIMA (2, 1, 0) (1,

0, 0)_s achieves the lowest AIC (7,219.35) and BIC (7,232.94) across all $d = 1$ specifications, confirming that seasonal augmentation provides genuine in-sample improvement beyond its additional parameter cost. These in-sample results are consistent with prior evidence that SARIMA structures fit trading-volume series better than non-seasonal counterparts in markets with weekly periodicity (Tweneboah et al., 2019; Fahriya et al., 2021).

Table 3: In-Sample Information Criteria

Model	Price AIC	Price BIC	Volume AIC	Volume BIC
ARMA(1,0)	1,838.72	1,845.56	7,406.85	7,413.69
ARMA(2,0)	1,832.28	1,842.53	7,364.76	7,375.01
ARIMA(1,1,0)	1,830.46	1,837.29	7,402.98	7,409.81
ARIMA(2,1,0)	1,825.27	1,835.50	7,332.57	7,342.81
SARIMA(1,1,0)(1,0,0) _s	1,803.13*	1,813.33*	7,237.29	7,247.49
SARIMA(2,1,0)(1,0,0) _s	1,805.08	1,818.68	7,219.35*	7,232.94*

* Minimum criterion value within each series. Lower values indicate better in-sample fit after penalising for parameter count.

Residual Diagnostics

Table 4 summarises residual adequacy. For stock price, all six models pass the Ljung-Box test ($p > 0.05$), confirming that no significant serial autocorrelation remains in the residuals. The Shapiro-Wilk test rejects normality across all price specifications ($p < 0.001$), reflecting the discrete price increments and sharp price movements in the MTNN series. Non-normality of residuals at this sample size does not invalidate parameter estimates or the Ljung-Box test but implies that Gaussian prediction intervals should be interpreted cautiously.

For trading volume, ARMA (1, 0) and ARIMA (1, 1, 0) fail the Ljung-Box test ($p = 0.047$ and $p = 0.026$, respectively), indicating that these simpler structures leave meaningful serial autocorrelation in the residuals. All remaining volume

models pass the autocorrelation test. The diagnostic failure of ARIMA (1, 1, 0) for volume aligns with its notably worse in-sample AIC relative to more complex alternatives, providing internal consistency across the two evaluation criteria. The adequacy of SARIMA (2, 1, 0) (1, 0, 0)_s residuals (Q (10) = 9.81, $p = 0.46$) indicates that this model captures the serial structure in volume at the tested lags. The Ljung-Box test above is computed at lag 10 only; it has not been checked specifically at the seasonal lag ($s = 5$ and its multiples) or for residual heteroskedasticity (e.g., via an ARCH-LM test), so seasonal mis-specification or conditional-variance clustering that a lag-10 portmanteau test would not detect cannot be ruled out; this is noted as a limitation rather than resolved here.

Table 4: Residual Diagnostics

Model	Series	Resid. SD	LB Q(10)	LB p	SW p	Adequate
ARMA(1,0)	Price	14.02	8.48	0.582	<0.001	Yes*
ARMA(2,0)	Price	13.85	8.21	0.608	<0.001	Yes*
ARIMA(1,1,0)	Price	14.01	7.24	0.702	<0.001	Yes*
ARIMA(2,1,0)	Price	13.89	7.01	0.724	<0.001	Yes*
SARIMA(1,1,0)(1,0,0) _s	Price	14.11	6.66	0.757	<0.001	Yes*
SARIMA(2,1,0)(1,0,0) _s	Price	13.97	6.51	0.771	<0.001	Yes*
ARMA(1,0)	Volume	3,291,847	18.50	0.047	0.046	No

Model	Series	Resid. SD	LB Q(10)	LB p	SW p	Adequate
ARMA(2,0)	Volume	3,105,221	14.72	0.143	0.031	Yes [^]
ARIMA(1,1,0)	Volume	3,338,415	20.41	0.026	0.048	No
ARIMA(2,1,0)	Volume	3,002,105	10.88	0.363	0.052	Yes
SARIMA(1,1,0)(1,0,0) _s	Volume	3,021,303	10.20	0.421	0.046	Yes [^]
SARIMA(2,1,0)(1,0,0) _s	Volume	2,947,094	9.81	0.457	0.046	Yes [^]

LB = Ljung–Box portmanteau Q-statistic at 10 lags; SW = Shapiro–Wilk. * Serial autocorrelation is absent; residual non-normality noted but does not invalidate estimates ($n = 227$). [^]Autocorrelation adequate; SW $p < 0.05$ reflects non-Gaussian volume distribution. "No" indicates statistically significant residual autocorrelation (LB $p < 0.05$); such models are inadequate and their forecast results are interpreted cautiously.

Rolling One-Step-Ahead Forecast Accuracy

Table 5 reports the out-of-sample accuracy metrics from the 22-step rolling evaluation. The results reveal clear patterns for each series.

For stock price, ARIMA(1,1,0) achieves the lowest RMSE (24.21 NGN) and MSE (586.23), identifying it as the most accurate structure by these metrics when large forecast errors are penalized heavily, as they are in trading and risk management contexts. However, RMSE across all six price models spans only 24.21–24.29 NGN, a difference of 0.08 NGN (about 0.3%); this margin was not evaluated with a formal paired forecast-comparison test (e.g., Diebold–Mariano), so the ranking among price models should be read as descriptive rather than a statistically confirmed superiority. The SARIMA structures, despite achieving lower AIC in-sample, produce marginally higher RMSE values (24.29 NGN); this divergence between in-sample and out-of-sample performance does not reflect statistical over fitting, since AIC already penalizes for additional parameters, but is consistent with the weekly seasonal price pattern active during training being attenuated during the 22-day validation window, when MTNN prices consolidated around NGN 820 with limited weekly variation. This is an important practical lesson: seasonal model components that are informative in-sample may not improve short-run out-of-sample forecasts when the seasonal pattern is regime-specific (Fahriya et al., 2021; Hansun et al., 2022).

ARMA (1,0) and ARMA(2,0) on price levels achieve lower MAE (13.89 and 13.92 NGN) and lower MAPE (1.65%) than ARIMA(1,1,0) (MAPE = 1.71%). This metric-dependence arises because RMSE and MSE assign quadratic weight to large errors, where ARIMA (1, 1, 0) outperforms by capturing

directional momentum during price spike events (e.g., the 50-NGN jump on 29 April 2026). On calmer days, the near-unit-root ARMA (1,0) model effectively mimics a naïve random-walk and achieves smaller absolute errors. When RMSE or MSE is the primary loss criterion (appropriate for risk-sensitive applications) ARIMA (1, 1, 0) is the preferred structure. Practitioners optimizing for MAE or MAPE may find ARMA (2,0) marginally competitive, though its application to a confirmed I(1) series introduces well-known spurious regression risks (Mondal et al., 2019).

For trading volume, SARIMA(2,1,0)(1,0,0)_s achieves the lowest RMSE (3,177,444 shares), outperforming the next-best $d=1$ model ARIMA(2,1,0) (RMSE = 3,233,077) by 1.7% and the best $d=0$ model ARMA(2,0) (RMSE = 3,240,984) by 2.0%. This confirms the conservative $d=1$ choice as empirically justified: differencing combined with seasonal augmentation delivers the best RMSE even when differencing alone is not the dominant factor. MAPE values are substantially lower for $d=0$ models (31–36%) than for $d=1$ models (44–45%), reflecting that differenced-model forecasts track spike events more aggressively but introduce proportionally larger errors on low-volume days. Given that both ARMA(1,0) and ARIMA(1,1,0) volume models fail the residual adequacy test (indicating incomplete capture of serial structure) and that SARIMA(2,1,0)(1,0,0)_s achieves the best AIC, BIC, and RMSE with adequate residuals, it is designated the optimal volume specification. The MAPE range of 44–45% across all adequately specified volume models is consistent with Hansun et al. (2022), who reported that volume MAPE values in this range are typical for Box–Jenkins models applied to spike-dominated institutional trading series.

Table 5. Out-of-Sample Rolling One-Step-Ahead Forecast Accuracy (Validation: $n = 22$)

Model	Series	RMSE	MAE	MSE	MAPE (%)
ARMA(1,0)	Price	24.89	13.89	619.53	1.6495
ARMA(2,0)	Price	24.35	13.92	592.68	1.6508
ARIMA(1,1,0)	Price	24.21 [†]	14.37	586.23 [†]	1.7078
ARIMA(2,1,0)	Price	24.22	14.48	586.67	1.7209
SARIMA(1,1,0)(1,0,0) _s	Price	24.29	14.59	589.76	1.7347
SARIMA(2,1,0)(1,0,0) _s	Price	24.29	14.75	589.95	1.7533
ARMA(1,0) [$d=0$]	Volume	3,297,406	2,083,797	10.87×10^{12}	32.14
ARMA(2,0) [$d=0$]	Volume	3,240,984	2,195,461	10.50×10^{12}	36.42
SARIMA(1,0,0)(1,0,0) _s [$d=0$]	Volume	3,291,685	2,052,313	10.83×10^{12}	31.17
SARIMA(2,0,0)(1,0,0) _s [$d=0$]	Volume	3,277,155	2,221,723	10.74×10^{12}	36.02
ARIMA(1,1,0)	Volume	3,399,669	2,384,978	11.56×10^{12}	45.12
ARIMA(2,1,0)	Volume	3,233,077	2,169,067	10.45×10^{12}	44.29
SARIMA(1,1,0)(1,0,0) _s	Volume	3,329,740	2,299,296	11.09×10^{12}	45.20
SARIMA(2,1,0)(1,0,0) _s	Volume	3,177,444 [†]	2,164,214	10.10×10^{12}	44.69

[†] Denotes the lowest-RMSE model within each series. MSE for volume is in units of 10^{12} shares². The four $d=0$ volume rows are included as a robustness check given the ambiguous stationarity evidence in Table 2, and together with the four $d=1$ volume rows they complete the eight-run 2^3 factorial design used for volume in Table 6; SARIMA(2,1,0)(1,0,0)_s ($d=1$) still delivers the lowest overall volume RMSE.

Factorial Analysis of Structural Factors

Table 6 presents the 2^3 factorial design matrix with RMSE as the response, and Table 7 reports the estimated main effects.

The factorial analysis provides a systematic account of how each structural choice affects forecast accuracy.

Table 6: Factorial Design Matrix and RMSE Response

Run	Model (Price)	Model (Volume)	p	d	P	RMSE (Price)	RMSE (Volume)
1	ARMA(1,0)	ARMA(1,0) [d=0]	1	0	0	24.8904	3,297,406
2	ARMA(2,0)	ARMA(2,0) [d=0]	2	0	0	24.3450	3,240,984
3	ARIMA(1,1,0)	ARIMA(1,1,0)	1	1	0	24.2122	3,399,669
4	ARIMA(2,1,0)	ARIMA(2,1,0)	2	1	0	24.2213	3,233,077
5	not estimated	SARIMA(1,0,0)(1,0,0) _s [d=0]	1	0	1	-	3,291,685
6	not estimated	SARIMA(2,0,0)(1,0,0) _s [d=0]	2	0	1	-	3,277,155
7	SARIMA(1,1,0)(1,0,0) _s	SARIMA(1,1,0)(1,0,0) _s	1	1	1	24.2851	3,329,740
8	SARIMA(2,1,0)(1,0,0) _s	SARIMA(2,1,0)(1,0,0) _s	2	1	1	24.2888	3,177,444

Factors: p (AR order: 1 = -1; 2 = +1); d (differencing: 0 = -1; 1 = +1); P (seasonal AR: 0 = -1; 1 = +1). All eight 2^3 treatment combinations (runs 1–8) were evaluated for trading volume. For stock price, only six combinations were fitted; runs 5–6 (d = 0, P = 1, i.e. SARIMA specifications without differencing) were not estimated for price and are marked “not estimated”.

Table 7: Estimated Factorial Effects (RMSE Response)

Effect	Price Δ RMSE (NGN)	Volume Δ RMSE (Shares)
p: AR order (1 to 2)	-0.268	-97,460
d: Differencing (0 to 1)	-0.401	+8,175
p×d interaction	+0.277	-61,984
P: Seasonal AR (0 to 1)	not estimable (see note)	-23,778
p×P interaction	not estimable	+14,047
d×P interaction	not estimable	-39,003
p×d×P interaction	not estimable	-6,899

Effects computed via Equation (8). Volume effects use the complete eight-run design (Table 6, runs 1–8) and are fully orthogonal. Price effects for p, d and p×d use the balanced 2×2 subdesign at P = 0 (runs 1–4); effects involving P are marked “not estimable” because P was evaluated only at d = 1 for price. As a descriptive, non-orthogonal comparison only: adding P = 1 at d = 1 increased price RMSE by 0.073 NGN at p = 1 (run 3 vs. run 7) and by 0.068 NGN at p = 2 (run 4 vs. run 8).

For stock price, within the estimable 2×2 (p, d) sub design, the differencing factor d has the larger main effect (-0.401 NGN versus -0.268 NGN for p), consistent with specifying the correct integration order being the primary driver of price forecast accuracy. This is directionally consistent with Deb (2019) and Mondal et al. (2019), both of whom identified differencing order as an influential characteristic for equity price forecasting. Unlike the d effect, the AR-order effect is not uniform across the design: raising p from 1 to 2 reduces RMSE substantially at d = 0 (ARMA(1,0)→ARMA(2,0): 24.89→24.35 NGN) but leaves RMSE almost unchanged, or slightly worse, at d = 1 (ARIMA(1,1,0)→ARIMA(2,1,0): 24.21→24.22 NGN).

SARIMA(1,1,0)(1,0,0)_s→SARIMA(2,1,0)(1,0,0)_s: 24.29→24.29 NGN). This pattern is captured by a positive p×d interaction (+0.277 NGN): the AR-order benefit is largely a d = 0 phenomenon and mostly disappears once the series is differenced. The seasonal factor P is not separately estimable for price because it was only evaluated at d = 1; as a descriptive comparison, adding P = 1 slightly increased RMSE at d = 1 for both AR orders (Table 7 note), consistent with the finding in Section 4.5 that the SARIMA structures did not improve on ARIMA (1, 1, 0) out-of-sample.

For trading volume, the complete eight-run design shows AR order p as the most influential main effect (-97,460 shares), with P also negative (-23,778 shares) and d essentially neutral on average (+8,175 shares). This near-zero average d effect is offset by two negative interactions: p×d (-61,984 shares) and d×P (-39,003 shares), each indicating that differencing combines synergistically with a higher AR order and with seasonal augmentation respectively, even though its own average contribution is small; the three-factor interaction

p×d×P is comparatively minor (-6,899 shares). This explains why SARIMA(2,1,0)(1,0,0)_s (p = 2, d = 1, P = 1) achieves the lowest volume RMSE: it sets all three factors to their RMSE-reducing combination once interactions, not just main effects, are accounted for. Because the volume design is complete and orthogonal, these effects and interactions are estimated without the ambiguity that affects the price analysis; a simple pairwise model comparison would not have revealed the p×d and d×P interactions.

CONCLUSION

This study compared ARMA, ARIMA, and SARIMA forecasting structures for MTN Nigeria daily closing prices and trading volumes, using 249 trading-day observations from 26 May 2025 to 26 May 2026. Six price configurations and eight volume configurations were estimated on 227 training observations, validated against 22 held-out observations using rolling one-step-ahead forecasting, and their structural characteristics were analyzed using a 2^3 factorial design that is complete for trading volume and reduced (six of eight runs) for stock price. These findings are specific to one equity and a 22-day validation window and should be read accordingly.

The key finding is that the optimal Box-Jenkins structure differs across the two series by the primary RMSE criterion. For stock price, ARIMA (1, 1, 0) achieves the lowest RMSE (24.21 NGN) and MSE, although ARMA (1, 0) and ARMA (2, 0) achieve lower MAE and MAPE; the RMSE differences among the six price models are small (24.21–24.29 NGN) and were not tested for statistical significance, so this ranking is suggestive rather than conclusive. Within its reduced, six-run design, differencing (d = 1) is the most influential structural

factor for price forecast accuracy among the effects that are estimable, consistent with prior evidence on equity price forecasting in emerging markets; effects involving the seasonal AR factor P could not be separately estimated for price (Section 4.6). For trading volume, SARIMA (2, 1, 0) (1, 0, 0)_s achieves the lowest RMSE (3,177,444 shares) and the best in-sample AIC and BIC, with adequate residuals, and is optimal on all four accuracy criteria in Table 5. The complete, eight-run factorial analysis for volume shows AR order (p) as the largest average driver of RMSE, with a substantial negative differencing $\times$ seasonal-AR (d $\times$ P) interaction that explains why the fully seasonal, differenced specification outperforms simpler alternatives even though the average main effect of differencing alone is close to zero (Section 4.6, Table 7).

No model is universally superior: ARIMA(1,1,0) is one of the weaker volume specifications, and SARIMA(2,1,0)(1,0,0)_s offers no RMSE advantage for price over the simpler ARIMA(1,1,0). This result reinforces the practical guidance that Box–Jenkins model selection should be series-specific, empirically driven by stationary testing, ACF/PACF analysis, and rolling out-of-sample evaluation, not determined by general convention. Box–Jenkins models remain practically viable for forecasting on this emerging African equity, with MAPE of 1.71% for price indicating good practical accuracy on the validation window used here, and volume MAPE of 44.69% reflecting the intrinsic difficulty of forecasting spike-driven institutional trading; both figures, and the factorial findings above, are specific to MTNN over the 26 May 2025–26 May 2026 sample and have not been tested for significance beyond the single validation window used. Future research should replicate this evaluation on other NGX equities and longer or multiple validation windows, apply a formal paired forecast-comparison test (e.g., Diebold–Mariano) to the price model rankings, complete the price factorial design by estimating the two currently omitted SARIMA (d=0) specifications, explore multivariate extensions incorporating macroeconomic covariates, and assess hybrid Box–Jenkins–GARCH structures that additionally model conditional heteroskedasticity.

REFERENCES

- Acheampong, P., & Wiafe, E. A. (2019). Forecasting stock returns volatility of firms in the Ghana Stock Exchange: A comparison of ARMA, ARIMA and GARCH models. *Journal of Finance and Economics*, 7(1), 34–44.
- Acquah-Sam, E. (2022). Time series analysis of West African stock markets. *Journal of African Business*, 23(2), 45–61.
- Box, G. E. P., & Draper, N. R. (2007). *Response surfaces, mixtures, and ridge analyses* (2nd ed.). Wiley.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.
- Deb, S. (2019). ARIMA forecasting of several major economies during the difficult period. *Journal of Applied Statistics*, 46(5), 830–845.
- Fahriya, N., Suparti, S., & Tarno, T. (2021). Comparison of ARIMA and SARIMA models in forecasting the number of online motorcycle taxis. *Media Statistika*, 14(1), 71–81.
- Hansun, S., Wicaksana, A., & Christy, V. (2022). Revisiting ARIMA in stock market forecasting: A systematic literature review. *Investment Management and Financial Innovations*, 19(3), 228–241.
- Mondal, P., Shit, L., & Goswami, S. (2019). Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *International Journal of Computer Science, Engineering and Applications*, 4(2), 13–29.
- Montgomery, D. C. (2017). *Design and analysis of experiments* (9th ed.). Wiley.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting* (2nd ed.). Wiley.
- Tweneboah, G., Owusu Junior, P., & Oseifuah, E. K. (2019). Integration of major African stock markets: Evidence from multi-scale wavelets decomposition approach. *Global Journal of Finance and Economics*, 16(1), 1–20.

