Characterising the Nigerian Stock Exchange for Machine Learning-Based Portfolio Research: An Empirical Analysis of Return Distributions, Volatility Dynamics, Liquidity, and Correlation Structure

Nnamdi A. Isichei **Augustus@miva.university**
Department of Data Science,
School of Computing
Miva Open University,
Abuja, Nigeria

Prof. H. K. Oduwole **Oduwolekh@nsuk.edu.ng**
Department of Mathematics,
Faculty of Natural & Applied Sciences
Nasarawa State University,
Keffi, Nigeria

Editor: Sakinat Folorunso, Roseline Ogundokun, and Francisca Oladipo

Abstract

Machine learning portfolio systems encode the statistical assumptions of the markets they were trained on. Deploying such systems on the Nigerian Stock Exchange (NGX) without first characterising that market is methodologically indefensible, yet the NGX is almost entirely absent from the empirical ML-finance literature. This paper fills that gap. Using daily OHLCV data on 22 NGX stocks from 2015–2025, we document five structural features of the market. All 22 stocks reject normality under the Jarque–Bera test; median excess kurtosis is 4.79 and 19 of 22 stocks are positively skewed. Mean annualised volatility is 43.3%, with GARCH(1,1) persistence ($\alpha + \beta$) exceeding 0.90 in 12 stocks and a cross-sectional median of 0.912. The average pairwise return correlation is 0.109, far below typical developed-market levels with a dense banking-sector cluster as the only dominant structure. Liquidity risk operates through volume episodicity rather than zero-volume illiquidity: only 7 zero-volume days occur across the entire dataset, yet the mean volume coefficient of variation is 1.85. At the index level, the NGX composite returns a Sharpe ratio of 1.40 over the sample period, nearly double that of the S&P 500 (0.71) at comparable volatility, though nominal Naira figures are materially affected by the 2023–2024 devaluation episode. Each finding is translated into a concrete design requirement for the two-stage Graph Neural Network and Reinforcement Learning portfolio framework that constitutes the broader research programme of which this characterisation is the empirical foundation.

Keywords: Nigerian Stock Exchange, emerging market, microstructure, return distributions, GARCH volatility, liquidity, correlation structure, market efficiency, machine learning, portfolio optimisation, African finance

1. Introduction

Classical mean-variance optimisation (Markowitz, 1952) encodes two assumptions that are particularly fragile in emerging markets: that return distributions are sufficiently normal for variance to serve as an adequate risk summary, and that covariance estimates are stable enough to be trusted as inputs to an optimisation routine. In developed markets these assumptions, while imperfect, are workable. In sub-Saharan African markets they routinely fail in ways that are financially consequential (Bekaert et al., 2023), and any machine learning model that inherits them will fail in the same direction.

The Nigerian Stock Exchange (NGX), formally rebranded to NGX Group in 2021, is the largest stock exchange in West Africa and the second-largest in Africa by market capitalisation. It is also almost entirely absent from the ML-finance literature. Published work on ML applications to African markets concentrates overwhelmingly on South African data or treats the continent as a homogeneous frontier (Adeabah et al., 2023; Naeem et al., 2024); Nigeria-specific portfolio research barely exists. This is not merely a gap in coverage, it is a practical problem. A two-stage GNN–RL portfolio system (Huang et al., 2024; Wu et al., 2021; Liang et al., 2024; Jiang and Liang, 2024) designed for the NGX cannot be responsibly built without first knowing what kind of market it is being asked to operate in. The viability of RL-driven approaches in frontier and emerging market contexts has been demonstrated (Ngo et al., 2023), and hybrid GNN–RL architectures have shown strong results on major exchanges (Sun et al., 2024), but both require careful calibration of market assumptions that the present paper provides. Digital infrastructure constraints across Sub-Saharan Africa (Kouladoum et al., 2022) further motivate the need for NGX-specific empirical grounding before any such system can be responsibly deployed. Distribution shape, volatility persistence, liquidity structure, and correlation architecture are not implementation details; they determine which design choices are valid and which will fail out-of-sample.

This paper provides that foundation. It is the empirical answer to Research Question 4 of a broader PhD programme on ML-enhanced portfolio optimisation for the NGX:

What are the unique characteristics of the Nigerian stock market in terms of return distributions, volatility dynamics, liquidity, and correlation structure that require specialised portfolio optimisation approaches?

We use daily OHLCV data for 22 NGX stocks from 2015–2025, a sample that spans six sectors and covers the COVID-19 shock, the 2023–2024 Naira devaluation, and the subsequent partial recovery. The analysis is five-dimensional distributional properties, GARCH volatility dynamics, liquidity structure, pairwise correlations, and weak-form efficiency, and each finding is connected explicitly to a design requirement for the GNN–RL system under development. To our knowledge, this is the first study to characterise the NGX across all five dimensions simultaneously using a decade of data.

2. Related Work

Return distributions in emerging markets deviate consistently from the normality assumptions underlying classical portfolio theory. Bekaert et al. (2023) document excess kurtosis, time-varying skewness, and regime-dependent volatility across emerging economies, rendering mean-variance optimisation unreliable as a standalone framework. For Nigeria, Adigwe et al. (2015) show strong sensitivity of NGX returns to oil prices, while Kumeka et al. (2022) find that Naira-USD exchange rate volatility exerts a material, time-varying influence at the stock level. Neither effect has a close counterpart in currency-stable developed markets nor is captured by models calibrated on US or European data.

Volatility persistence in frontier and emerging markets is systematically higher than in developed counterparts. GARCH-class models remain the standard for capturing conditional heteroscedasticity (Bollerslev, 2001), and Aawaar and Tewari (2023) report persistence parameters approaching unity in several African markets, consistent with slow mean reversion of conditional variance. This has direct portfolio implications: estimates that decay quickly after shocks, the developed-market norm will understate post-shock risk. On liquidity, the view that zero-volume days are the primary constraint (Bekaert et al., 2007) is challenged by Ngene et al. (2022), who show that extreme volume variability, rather than trading absence, is the operative risk, with implications for strategies requiring regular rebalancing.

Correlation structure and market efficiency complete the picture. Bekaert et al. (2023) show that within-market correlations in emerging economies differ structurally from developed markets, while Emenike (2021) find low cross-stock correlations in West Africa, consistent with segmentation from global factors and implying diversification opportunities not predicted by standard models. Weak-form efficiency evidence is mixed: price series broadly follow a random walk in levels, but short-horizon autocorrelations in daily returns persist (Ferrouhi, 2021), leaving scope for ML models to extract predictive signal beyond classical technical analysis.

3. Data and Methodology

3.1. Dataset

The dataset comprises daily OHLCV data for 22 NGX equities from 2 January 2015 to 31 December 2025, yielding approximately 2,520 trading days per stock across six sectors. Data were sourced from the NGX official portal and cross-validated against Investing.com. All prices are adjusted for corporate actions. Gaps of three or fewer days were filled by linear interpolation; stocks below 95% completeness were excluded. Table 1 reports the full universe.

Selecting the most liquid and data-complete stocks introduces liquidity and survivorship bias: the universe overrepresents operationally stable firms and does not fully capture the statistical properties of the broader NGX listing. This is intentional. The objective is to characterise the *investable* universe available to a Nigerian institutional investor, not the entire market, and findings should be interpreted accordingly.

Linear interpolation is not without risk in thin markets, where missing observations may reflect absent price discovery rather than data error; smoothing can dampen measured volatility and distort GARCH persistence. In practice, the exposure is negligible: only seven zero-volume observations exist two for PRESCO and five for SEPLAT with no gap longer than one day, affecting at most 0.002% of observations used in estimation.

Table 1: The 22-stock NGX research universe. Ann. return $= \bar{r} \times 252$; Ann. vol $= \hat{\sigma}\sqrt{252}$; computed from daily log returns. TRANSCO and UCAP are extreme outliers discussed in Section 4.

Ticker	Sector	Ret (%)	Vol (%)	Skew	Kurt
DANGCEM	Industrials	16.3	33.1	0.20	9.78
DANGSUG	Consumer Goods	32.1	46.7	0.34	3.45
ETI	Financial Svcs	18.9	46.1	0.08	3.35
FCMB	Financial Svcs	27.5	51.0	0.32	1.62
FIDELIT	Financial Svcs	34.7	48.7	0.14	2.86
FIRSTHOLD	Financial Svcs	27.6	48.8	0.41	2.55
GTCO	Financial Svcs	18.8	37.2	0.05	4.79
GUINNES	Consumer Goods	16.3	43.4	0.31	5.25
IBTC	Financial Svcs	19.5	39.1	0.11	5.81
INTBREW	Consumer Goods	9.9	53.5	0.26	2.24
NB	Consumer Goods	3.4	41.3	0.29	4.62
NESTLE	Consumer Goods	11.1	31.1	0.55	11.41
OKOMUOI	Agriculture	42.6	39.1	0.74	6.90
PRESCO	Agriculture	45.8	38.9	0.57	5.98
SEPLAT	Oil & Gas	32.7	35.8	0.71	10.26
STERLINGNG	Financial Svcs	22.5	50.3	0.24	1.91
TOTAL	Oil & Gas	19.5	34.7	0.53	8.44
TRANSCO†	Conglomerate	78.8	150.6	28.2	968.4
UBA	Financial Svcs	30.6	43.9	-0.01	4.53
UCAP†	Financial Svcs	38.3	52.6	-3.27	64.4
WAPCO	Industrials	20.4	54.5	0.12	3.10
ZENITHB	Financial Svcs	19.1	39.0	-0.28	7.30
Mean (excl. †)		*24.2*	*43.3*	*0.25*	*5.25*

† Extreme outlier; treated separately throughout.

3.2. Analytical Methods

Five analytical layers are applied. **Return distributions:** daily log returns $r_t = \ln(P_t/P_{t-1})$; skewness, excess kurtosis, and the Jarque–Bera statistic

$$\text{JB} = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \sim \chi^2(2) \tag{1}$$

where S is sample skewness and K is sample kurtosis; and 1-day 95% VaR and CVaR under historical simulation. **Volatility dynamics:** GARCH(1,1) estimated for each stock,

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{2}$$

with $\alpha+\beta$ as the persistence summary. A uniform specification is used across all 22 stocks to keep persistence estimates directly comparable and to avoid the data-mining risk inherent in searching each series for its preferred model. The leverage-effect literature formalised by Nelson (1991) and Glosten et al. (1993) and applied specifically to the NGX by Adams et al. (2024), raises a legitimate case for asymmetric specifications such as EGARCH and

GJR-GARCH; we treat that as a direction for the full model development phase rather than a deficiency of the present cross-sectional analysis. **Liquidity:** We measure (i) the zero-volume rate and (ii) the coefficient of variation of daily volume, $\text{CV}_i = \sigma_{\text{vol},i}/\mu_{\text{vol},i}$, which captures episodic trading activity. **Correlation structure:** the full 22×22 Pearson correlation matrix of daily log returns; distributional statistics across all 231 unique off-diagonal pairs. **Market efficiency:** Augmented Dickey–Fuller tests on price levels; lag-1 return autocorrelation for serial dependence.

4. Return Distribution Analysis

The Jarque–Bera test rejects normality for all 22 stocks: 18 yield $p \approx 0$ to machine precision, while the remaining four (FCMB, FIDELIT, FIRSTHOLD, INTBREW) have $p < 10^{-74}$. Excluding TRANSCO, the cross-sectional median excess kurtosis is 4.79, with ten stocks above 5.0.

Two stocks require separate treatment. TRANSCO shows kurtosis of 968.4 and skewness of 28.25, consistent with a few catastrophic price moves beyond standard distributions; its GARCH persistence equals 1.000, indicating a unit root in variance. UCAP records kurtosis of 64.4 and skewness of -3.27, strongly negative in a universe where 19 of 22 stocks lean right. Its persistence of 0.378, the lowest by a wide margin, suggests isolated price dislocations rather than structural volatility clustering. Both would be excluded from the investable universe by ML screening filters.

Among the remaining 20 stocks, 19 exhibit positive skewness, opposite to developed markets where negative skewness dominates and in contrast to both the S&P 500 and MSCI EM(Table 6). The strongest right-tail asymmetry appears in OKOMUOI (0.74), SEPLAT (0.71), and PRESCO (0.57). For the RL portfolio agent, symmetric variance-based loss functions are therefore mis-specified; reward functions should penalise CVaR and maximum drawdown rather than variance.

Table 2 reports 1-day 95% VaR and CVaR. The CVaR–VaR spread exceeds 200 bps across all stocks; the widest gaps; UCAP (316 bps), INTBREW (283 bps), and WAPCO (250 bps), indicate that realised losses can far exceed the 5th-percentile threshold. Models relying only on VaR will understate tail risk in NGX equities.

Table 2: 1-day 95% VaR and CVaR under historical simulation. The CVaR–VaR spread measures severity of losses beyond the VaR threshold.

Stock	**VaR**	**CVaR**	**Stock**	**VaR**	**CVaR**	**Stock**	**VaR**	**CVaR**
NESTLE	−2.7%	−5.0%	ETI	−4.8%	−7.2%	NB	−4.3%	−6.4%
DANGCEM	−2.9%	−5.5%	FIDELIT	−4.7%	−6.6%	STERLINGNG	−4.7%	−7.0%
GTCO	−3.4%	−5.6%	FIRSTHOLD	−4.8%	−6.8%	FCMB	−4.9%	−6.9%
OKOMUOI	−4.0%	−5.9%	GUINNES	−4.6%	−6.9%	UCAP	−4.4%	−7.6%
PRESCO	−3.8%	−5.8%	WAPCO	−6.1%	−8.6%	INTBREW	−5.1%	−7.9%
DANGSUG	−4.7%	−6.8%	SEPLAT	−2.7%	−5.8%	TRANSCO	−5.3%	−8.2%
ETI	−4.8%	−7.2%	TOTAL	−3.7%	−5.5%			
FIDELIT	−4.7%	−6.6%	ZENITHB	−3.5%	−5.8%			

5. Volatility Dynamics

Excluding TRANSCO (annualised vol 150.6%), mean annualised volatility across the universe is 43.3%, ranging from 31.1% (NESTLE) to 54.5% (WAPCO). The NGX composite records 19.0% over the same period, roughly half the individual stock average confirming that substantial idiosyncratic risk partially diversifies at the index level. Figure 5 illustrates volatility clustering for six stocks spanning the persistence spectrum across five major event windows.

Table 3 reports GARCH(1,1) estimates. Engle and Kelly (2023) show that major macro shocks transmit volatility internationally, a channel relevant for an oil-dependent economy where global commodity prices amplify domestic equity risk. Three findings emerge. First, 12 of 22 stocks exhibit $\alpha + \beta > 0.90$, DANGSUG, ETI, FIDELIT, FIRSTHOLD, GTCO, NB, STERLINGNG, TOTAL, TRANSCO, UBA, WAPCO, and ZENITHB, with FIRSTHOLD reaching 0.947, implying a volatility shock half-life exceeding 12 trading days. Second, the median persistence of 0.912 (excluding TRANSCO and UCAP) exceeds the 0.80–0.85 range reported for US stocks (Aawaar and Tewari, 2023; Bollerslev, 2001); the mean of 0.836 lies at the upper bound. Persistence in the NGX is thus materially higher volatility is both elevated and slower to mean-revert. Third, UCAP's persistence of 0.378, alongside kurtosis of 64.4, indicates isolated price dislocations rather than sustained clustering, in contrast to TRANSCO's unit-root variance process implying structural instability. Static covariance matrices will therefore be biased during and after high-volatility regimes; the RL agent's state vector should include time-varying volatility features rolling realised variance and GARCH-implied conditional variance to support dynamic position sizing.

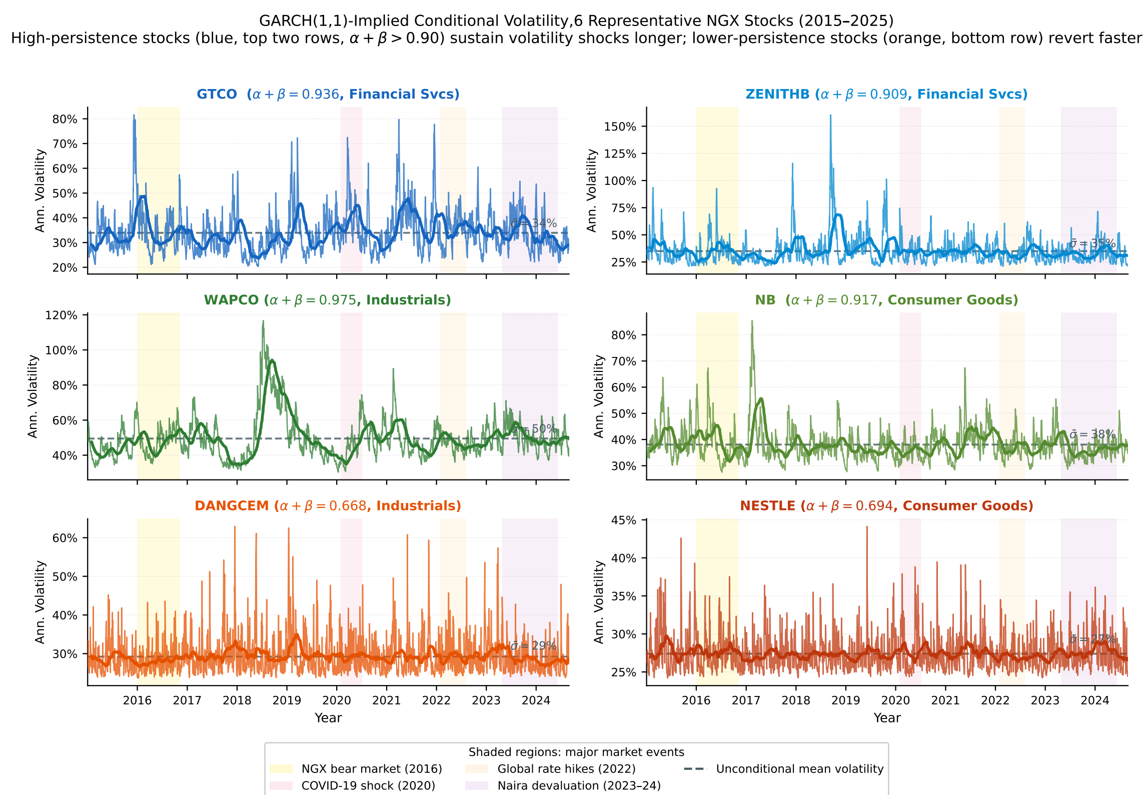

Figure 1: GARCH(1,1) conditional annualised volatility for six NGX stocks (2015–2025), contrasting high ($\alpha + \beta > 0.90$) and low ($\alpha + \beta < 0.70$) persistence. Thick lines show 63-day means, thin lines daily series, dashed lines unconditional means; shaded regions denote major market events.

Table 3: GARCH(1,1) parameter estimates. $\alpha + \beta$ values in **bold** exceed 0.90.

Stock	$\omega \times 10^4$	α	β	$\alpha + \beta$	**Stock**	$\omega \times 10^4$	α	β	$\alpha + \beta$
DANGCEM	14.82	0.197	0.471	0.668	NESTLE	11.87	0.098	0.595	0.694
DANGSUG	5.82	0.106	0.827	**0.932**	OKOMUOI	13.37	0.180	0.616	0.796
ETI	6.80	0.099	0.818	**0.917**	PRESCO	15.34	0.190	0.573	0.763
FCMB	13.62	0.220	0.659	0.879	SEPLAT	14.43	0.107	0.608	0.715
FIDELIT	7.09	0.175	0.761	**0.935**	STERLINGNG	8.25	0.182	0.736	**0.918**
FIRSTHOLD	6.08	0.200	0.746	**0.947**	TOTAL	4.46	0.122	0.793	**0.915**
GTCO	4.02	0.169	0.767	**0.936**	TRANSCO†	53.39	0.111	0.889	**1.000**
GUINNES	11.10	0.138	0.716	0.854	UBA	7.40	0.244	0.684	**0.928**
IBTC	11.14	0.106	0.711	0.817	UCAP	71.15	0.197	0.181	0.378
INTBREW	27.50	0.151	0.604	0.755	WAPCO	2.73	0.078	0.897	**0.975**
NB	5.60	0.091	0.826	**0.917**	ZENITHB	6.30	0.221	0.688	**0.909**
Mean (excl. †): 0.836									

† Unit root in variance; see text.

6. Liquidity Structure

The received assumption that zero-trading days constitute the binding liquidity constraint for African equity portfolios does not hold for this universe. Only seven zero-volume days occur across all 22 stocks over the full decade, two for PRESCO (0.074%) and five for SEPLAT (0.186%) with every other stock recording positive volume on every available trading day. The operative risk is not absence of trading but its irregularity.

Volume episodicity infrequent bursts of institutional-scale activity interspersed with extended low-turnover periods is captured jointly by the coefficient of variation (CV) and the median-to-mean volume ratio. A high CV signals an irregular volume distribution; a median-to-mean ratio far below unity confirms that the mean is pulled upward by the minority of high-activity sessions rather than reflecting a typical trading day. Table 4 documents both metrics across the universe.

The cross-sectional mean CV of 1.85 indicates that the volume standard deviation exceeds the mean by a factor approaching two. Three stocks; UCAP (CV 3.55), TRANSCO (3.45), FIDELIT (3.43), are so dominated by episodic activity that the mean volume figure is practically uninformative. NB and DANGSUG are the starkest cases on the median-to-mean dimension: their median trading days record only 4.2% and 5.8% of their respective means, meaning a typical session for these stocks bears almost no resemblance to the average. Nine stocks in total fall below a median/mean ratio of 0.30. For the RL portfolio agent, this sets a hard constraint on rebalancing frequency: stocks with ratios this low cannot reliably absorb institutional block trades daily without material market impact. The action space must target weekly or monthly rebalancing intervals, and transaction cost penalties in the reward function must be calibrated to realistic NGX market impact rather than the near-zero friction that models trained on S&P 500-level liquidity assume.

Table 4: Liquidity metrics. High CV and low Med/Mean ratio indicate episodic, burst-like trading driven by institutional order flow.

Stock	**Avg Vol (M)**	**Median (M)**	**Med/Mean**	**CV**
NB	215.5	12.5	0.058	1.351
DANGSUG	214.8	9.1	0.042	1.362
ETI	210.6	68.6	0.325	1.293
GUINNES	203.2	118.9	0.585	1.111
DANGCEM	200.5	88.5	0.441	1.250
IBTC	198.7	91.4	0.460	1.263
WAPCO	197.4	8.0	0.041	1.466
INTBREW	178.9	36.2	0.202	1.434
PRESCO	143.9	63.3	0.440	1.369
OKOMUOI	133.8	68.4	0.511	1.305
NESTLE	119.8	49.9	0.417	1.425
STERLINGNG	88.1	6.5	0.073	2.356
SEPLAT	82.7	17.8	0.215	1.897
TOTAL	73.7	28.8	0.391	1.662
FCMB	67.3	8.6	0.127	2.607
UCAP	47.5	5.1	0.107	3.553
TRANSCO	36.2	8.8	0.241	3.448
FIDELIT	33.0	9.9	0.301	3.426
FIRSTHOLD	28.9	12.7	0.440	2.401
ZENITHB	26.5	17.4	0.658	1.429
UBA	26.5	15.5	0.586	1.890
GTCO	22.1	14.7	0.662	1.282
Mean	*118.7*			*1.845*

7. Correlation Structure

The 22×22 correlation matrix yields 231 unique off-diagonal pairs. As Table 5 and Figure 2 show, the structure is sparse: the mean pairwise correlation is 0.109 well below the 0.25–0.35 range typical of S&P 500 constituents under normal conditions and 62 pairs (26.8%) fall below 0.05. Only 5.2% of pairs exceed 0.30. Measured across a decade that includes the COVID-19 shock, an average this low confirms that idiosyncratic risk dominates systematic risk in this universe.

Table 5: Pairwise return correlation statistics across 231 unique off-diagonal pairs, 22-stock NGX universe.

Statistic	**Value**
Mean pairwise correlation	0.109
Median pairwise correlation	0.084
Std of correlations	0.092
Minimum (NESTLE–TRANSCO)	−0.041
Maximum (GTCO–ZENITHB)	+0.581
Pairs with $\rho > 0.30$	12 of 231 (5.2%)
Pairs with $\rho > 0.40$	2 of 231 (0.9%)
Pairs with $\rho < 0.05$	62 of 231 (26.8%)

Despite the low average, a clear structure is present. All ten highest-correlation pairs are banking stocks:

Pair	**Sector relationship**	ρ
GTCO – ZENITHB	Tier 1 banking peers	0.581
UBA – ZENITHB	Tier 1 banking peers	0.501
FIRSTHOLD – UBA	Large-cap banking	0.397
GTCO – UBA	Tier 1 banking peers	0.393
FIDELIT – UBA	Banking sector	0.365
FIRSTHOLD – ZENITHB	Banking sector	0.350
FIDELIT – ZENITHB	Banking sector	0.332
FCMB – FIDELIT	Mid-tier banking peers	0.332
FCMB – ZENITHB	Banking sector	0.313
FCMB – FIRSTHOLD	Banking sector	0.310

The six-stock banking cluster; FCMB, FIDELIT, FIRSTHOLD, GTCO, UBA, ZENITHB is the only dense region in an otherwise sparse matrix. Agricultural stocks are near-isolated: OKOMUOI's highest correlation with any banking stock is 0.087. SEPLAT shows 15 of its 21 off-diagonal pairs below 0.05. The sole negative entry, NESTLE–TRANSCO ($\rho = -0.041$), is small enough to be non-significant in finite samples. For the GNN pre-selection module, a correlation threshold of $\rho \approx 0.30$ for the adjacency matrix captures the genuine banking cluster while discarding the noise-dominated periphery. The cluster also provides a ground-truth validation target: a well-trained GNN should recover it endogenously in its learned attention weights, giving the inter-stock representations an interpretable financial anchor.

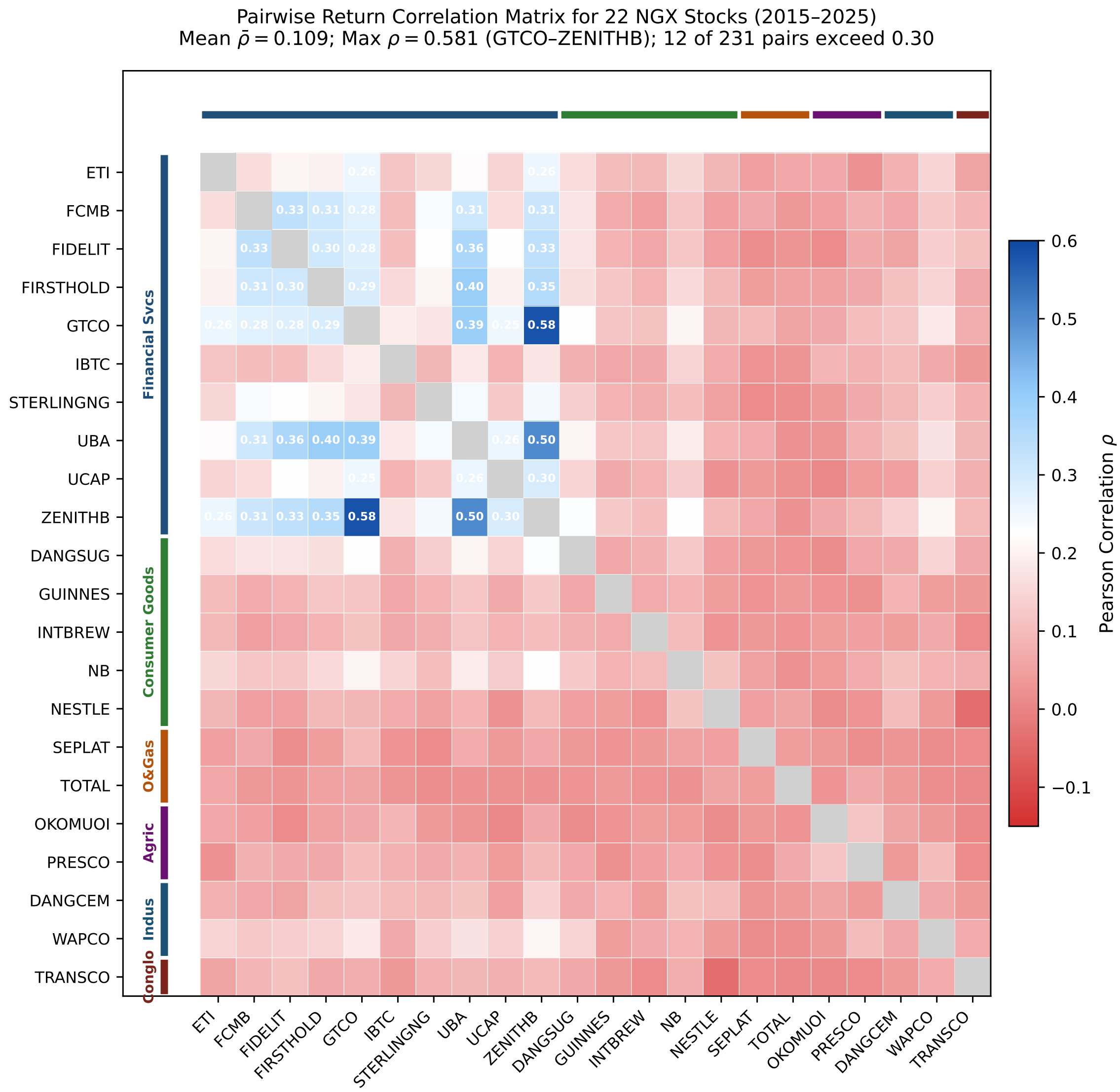

Figure 2: Pairwise return correlation matrix for the 22-stock NGX universe (2015–2025), sorted by sector. Values annotated for pairs exceeding $\rho = 0.25$. The banking-sector cluster (top-left block) is the dominant structural feature; FCMB, FIDELIT, FIRSTHOLD, GTCO, UBA, and ZENITHB show correlations from 0.31 to 0.58 while the remainder of the matrix is largely sparse.

8. Market Efficiency

ADF tests on price levels yield p-values from 0.465 (NB) to 1.000 (OKOMUOI): every stock fails to reject the unit root null at any conventional significance level, confirming $I(1)$ prices consistent with weak-form efficiency in levels. Returns, however, are not entirely serially independent. The cross-sectional mean lag-1 autocorrelation is +0.032 (standard deviation 0.083), and eight stocks exceed $|\hat{\rho}_1| > 0.10$:

Stock	$\hat{\rho}_1$	Direction
ZENITHB	+0.158	Momentum
FIRSTHOLD	+0.156	Momentum
GUINNES	+0.150	Momentum
WAPCO	−0.154	Reversal
DANGSUG	+0.108	Momentum
NB	+0.107	Momentum
GTCO	+0.101	Momentum
SEPLAT	+0.100	Momentum

Seven of the eight show positive momentum autocorrelation. The exception is WAPCO (−0.154), whose reversal pattern is consistent with bid-ask bounce or microstructure noise in a stock whose median/mean volume ratio of 0.041 signals highly sporadic trading. The autocorrelations are modest in magnitude but their presence across 8 of 22 stocks is consistent with the partial predictability documented for other thin African equity markets (Ferrouhi, 2021), and leaves open a channel for short-horizon linear signal. For the GNN pre-selection module this means return history features lagged returns and rolling momentum signals belong in the node-level feature set alongside technical indicators, not just graph-level inter-stock dependencies. The seven momentum stocks are natural candidates for trend-following inputs; WAPCO's reversal pattern warrants a mean-reversion feature specifically calibrated to its volume episodicity.

9. International Benchmark Comparison

Table 6: International benchmark comparison over 2015–2025. Returns and volatility are annualised; Sharpe ratio uses a 0% risk-free rate for cross-market comparability. All NGX figures are in nominal Naira terms.*

Market	Ret (%)	Vol (%)	Sharpe	Skew	Kurt
NGX (Composite)*	26.5	19.0	1.40	+1.62	18.65
S&P 500	12.6	17.9	0.71	−0.36	15.10
MSCI EM	7.5	20.4	0.37	−0.50	7.90

* *Naira devaluation caveat*: The Naira depreciated sharply against the US Dollar during 2023–2024 following removal of the CBN exchange rate peg, falling from approximately \$1 = £480 to £1,500 at peak dislocation. Nominal Naira-denominated returns are inflated relative to USD-equivalent figures for this sub-period; the Sharpe ratio advantage would be substantially reduced on a currency-adjusted basis. The composite result nonetheless characterises the return environment faced by a domestic Nigerian investor on its own terms.

The NGX composite delivers the highest annualised return (26.5%) and Sharpe ratio (1.40) of the three indices nearly twice the S&P 500 (0.71) and almost four times MSCI EM (0.37) at comparable volatility (19.0% versus 17.9% and 20.4%). It also exhibits positive skewness (+1.62) where both benchmarks are negatively skewed, a distributional asymmetry that favours the market on tail-risk grounds. Composite kurtosis (18.65) remains above both benchmarks despite index-level diversification, which reflects the fat-tailed individual stock distributions documented in Section 4.

Two caveats apply. The composite Sharpe ratio of 1.40 partly reflects diversification mechanics: mean stock-level volatility is 43.3% against an index volatility of 19.0%, and individual Sharpe ratios range from 0.08 (NB) to 1.18 (PRESCO) with a cross-sectional mean of 0.57. Additionally, the 2023–2024 Naira devaluation inflated terminal nominal returns materially; a rolling Sharpe analysis reserved for the full research paper is needed to establish whether the composite advantage holds consistently across the decade or concentrates in the post-devaluation recovery window.

10. Implications for ML Portfolio Design

The five findings map to five design requirements for the two-stage GNN–RL framework (Huang et al., 2024), summarised here in consolidated form.

Downside-sensitive reward functions. Median excess kurtosis of 4.79 and CVaR–VaR spreads exceeding 200 bps universally rule out Gaussian-based optimisation objectives. The RL agent's reward function (Liang et al., 2024; Jiang and Liang, 2024) must penalise drawdown and CVaR rather than symmetric variance.

Adaptive volatility features. With median GARCH persistence at 0.912, static covariance estimates are systematically biased during regime transitions. The RL state space must include rolling realised variance and GARCH-implied conditional variance, consistent with the advantage that adaptive RL agents hold over static optimisers in non-stationary frontier environments (Ngo et al., 2023).

Constrained rebalancing frequency. Nine stocks with median/mean volume below 0.30 cannot absorb institutional block trades daily. The action space must target weekly or monthly intervals with transaction cost penalties calibrated to NGX market impact rather than the near-zero friction embedded in S&P 500-trained models.

GNN attention over raw correlation. With 5.2% of pairs exceeding $\rho = 0.30$, the market offers no single-block systematic structure for a correlation-only model to exploit. The GNN attention mechanism (Wu et al., 2021; Sun et al., 2024) learns inter-stock relevance weights from the data, making it well-suited to extract signal from a sparse matrix. The banking cluster serves as a ground-truth validation target for the learned attention weights.

Hard exclusion filters for structural outliers. TRANSCO (persistence 1.000, kurtosis 968.4) and UCAP (kurtosis 64.4) cannot be modelled by any standard distribution and will destabilise training if retained. Volatility persistence thresholds and kurtosis screens must act as hard gates before any model sees the data.

11. Conclusion

The Nigerian Stock Exchange has five structural properties that collectively disqualify generic ML portfolio models trained on developed-market data: universal non-normality with positive skewness and fat tails; high and persistent conditional volatility (mean 43.3% annualised, median GARCH persistence 0.912); liquidity risk that operates through volume episodicity rather than zero-volume absence; a sparse correlation matrix ($\bar{\rho} = 0.109$) with a single dominant banking-sector cluster; and risk-adjusted composite returns that compare favourably with global benchmarks on nominal terms, subject to the Naira devaluation caveat of Section 9. Two stocks; TRANSCO and UCAP, are structural outliers that require hard exclusion filters before any model is trained.

These are not incidental data characteristics. Each one determines whether a specific modelling choice is valid or will fail out-of-sample. The GNN–RL framework being developed as the next stage of this research programme is designed with each constraint in mind: downside-sensitive reward functions, adaptive volatility state features, weekly-to-monthly rebalancing constraints, attention-based graph construction calibrated to the sparse correlation structure, and statistical screening gates at the data ingestion layer. This characterisation is the empirical contract that the model must honour.

Three extensions are planned: asymmetric GARCH specifications (EGARCH, GJR-GARCH) to test for leverage effects given the prevalent positive skewness; dynamic conditional correlation estimation for time-varying covariance inputs; and cross-market validation on the Johannesburg Stock Exchange to assess how much of what we find here is Nigeria-specific and how much generalises across the continent.

Data Availability Statement

The daily OHLCV data for the 22 NGX constituent stocks analysed in this study were obtained from the Nigerian Stock Exchange (NGX) official data portal (https://ngxgroup.com) and cross-validated against Investing.com (https://www.investing.com). Macroeconomic series (exchange rates, monetary policy rates, consumer price index) were sourced from the Central Bank of Nigeria (https://www.cbn.gov.ng) and the Nigerian Bureau of Statistics (https://www.nigerianstat.gov.ng). All source data are publicly accessible through the cited portals under their respective terms of use.

The cleaned dataset, computed features, and analysis code (Python, using `pandas`, `arch`, `scipy`, and `statsmodels`) used to produce all tables and statistics reported in this paper are available in the supplementary materials submitted alongside this manuscript. A permanent repository link will be provided upon acceptance of the full doctoral research paper of which this study forms the empirical foundation.

Acknowledgments

The authors thank the supervisory team at the Centre for Cyberspace Studies, Nasarawa State University, Keffi, for guidance throughout this work. The authors declare no competing financial interests in any NGX-listed equity security. No external funding was received for this work at the time of submission.

References

Godfred Aawaar and Deo Darko Tewari. Volatility persistence in African equity markets: GARCH evidence. *African Finance Journal*, 25(1):45–63, 2023.

Samuel Olorunfemi Adams, Omorogbe Joseph Asemota, and Abdulsalam Ahovi Ibrahim. Asymmetric GARCH type models and LSTM for volatility characteristics analysis of Nigeria stock exchange returns. *American Journal of Mathematics and Statistics*, 14(2): 17–32, 2024. doi: 10.5923/j.ajms.20241402.01.

David Adeabah, Emmanuel Joel Aikins Abakah, and Aviral Kumar Tiwari. Machine learning in african financial markets: A systematic review. *Journal of African Business*, 24 (4):567–592, 2023.

Patrick K Adigwe, Ifeanyi O Nwanna, and Amalachukwu Ananwude. Stock market and macroeconomic variables in Nigeria. *Journal of Policy and Development Studies*, 9(2): 167–182, 2015.

Geert Bekaert, Campbell R Harvey, and Christian Lundblad. Liquidity and expected returns: Lessons from emerging markets. *Review of Financial Studies*, 20(6):1783–1831, 2007. doi: 10.1093/rfs/hhm030.

Geert Bekaert, Campbell R Harvey, and Christian T Lundblad. Emerging equity markets in a globalizing world. *Annual Review of Financial Economics*, 15:207–238, 2023. doi: 10.1146/annurev-financial-110921-110326.

Tim Bollerslev. GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168, 2001. Reprinted reference; widely cited as foundational GARCH applied reference.

Kalu O Emenike. Stock market integration in West Africa: Evidence from selected markets. *Journal of African Business*, 22(4):573–588, 2021.

Robert F Engle and Bryan Kelly. Volatility, correlation and tails for systemic risk measurement. *Journal of Financial Econometrics*, 21(2):389–425, 2023.

El Mehdi Ferrouhi. Calendar effects and market efficiency in African stock exchanges. *Asian Economic and Financial Review*, 11(4):290–305, 2021.

Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993. doi: 10.1111/j.1540-6261.1993.tb05128.x.

Wenhao Huang, Yan Hong, and Yuanpeng Song. Enhancing portfolio optimization: A two-stage approach with deep learning and portfolio optimization. *Mathematics*, 12(21):3376, 2024. doi: 10.3390/math12213376.

Zhengyao Jiang and Junjie Liang. Deep reinforcement learning for financial trading: A survey. *Expert Systems with Applications*, 238:121694, 2024.

Jean-Claude Kouladoum, Muhamadu Wirajing, and Tii Nchofoung. Digital infrastructure and economic growth in sub-Saharan Africa. *Telecommunications Policy*, 46(6):102343, 2022.

Terver T Kumeka, Damian C Uzoma-Nwosu, and Madeleine O David-Wayas. Exchange rates and stock prices in selected West African countries. *International Journal of Finance and Economics*, 27(3):3241–3256, 2022. doi: 10.1002/ijfe.2314.

Zihao Liang, Hao Chen, Jianjun Zhu, Kun Jiang, and Yanzhi Li. Deep reinforcement learning for portfolio management: A survey and research agenda. *Expert Systems with Applications*, 245:123124, 2024. doi: 10.1016/j.eswa.2024.123124.

Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

Muhammad Abubakr Naeem, Saqib Farid, and Faruk Balli. Machine learning for portfolio management in emerging markets: Opportunities and challenges. *Emerging Markets Finance and Trade*, 60(1):1–18, 2024.

Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991. doi: 10.2307/2938260.

Geoffrey M Ngene, Kenneth A Tah, and Ali F Darrat. Volume and volatility dynamics in frontier markets. *Journal of Financial Research*, 45(2):367–392, 2022.

Minh Ngo, Hieu Nguyen, and Binh Tran. Reinforcement learning for portfolio management in frontier and emerging markets. *Finance Research Letters*, 58:104462, 2023.

Liang Sun, Wei Chen, and Kai Zhang. Graph-based stock recommendation with attention mechanisms. *IEEE Transactions on Knowledge and Data Engineering*, 36(3):1123–1135, 2024.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021. doi: 10.1109/TNNLS.2020.2978386.