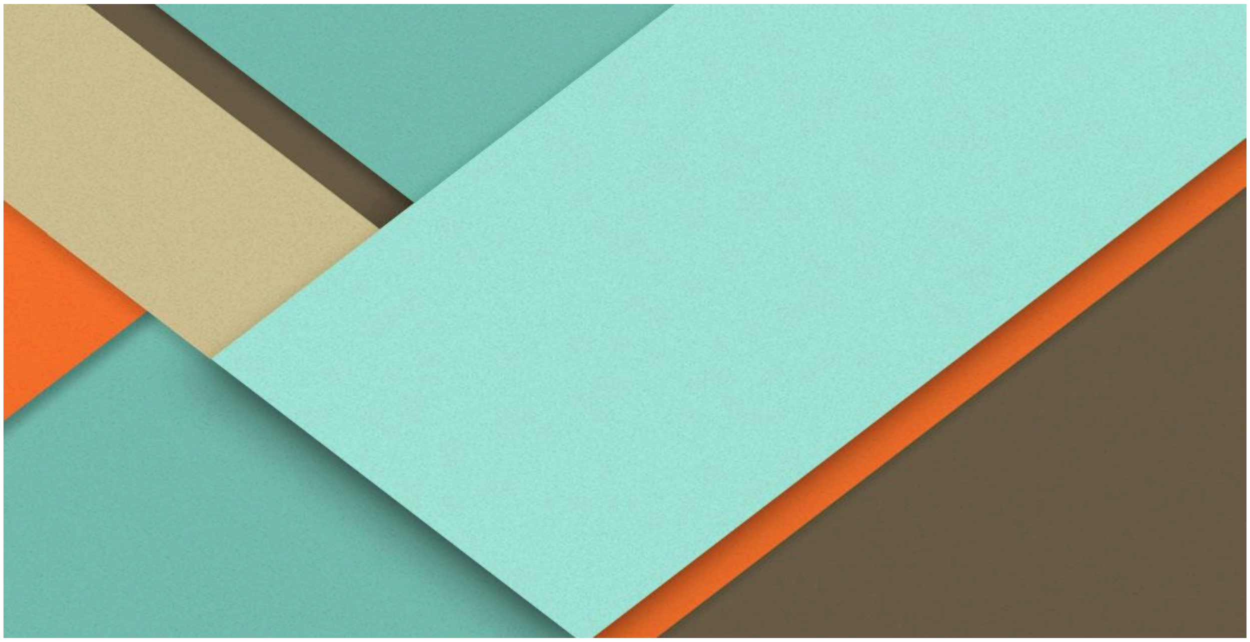


Profitable Pairs Trading with Nigerian Bank Stocks: A Quantitative Cointegration Framework

Date: May 11, 2025

Author: Maryjane C. Okoroafor

Disclaimer

This work originated from a prior collaboration, but has been independently finalized by the author.

Abstract

This study develops a robust and statistically rigorous pairs trading framework tailored for Nigerian financial equities. Using stocks like Guaranty Trust Holding Company Plc (GTCO), Zenith Bank Plc (ZENITHBANK), United Bank for Africa Plc (UBA), and others, we identify cointegrated stock pairs through correlation analysis and cointegration testing (Engle-Granger and Johansen tests). We model the spread between cointegrated pairs using Z-score normalization and employ a Python-based backtesting framework to generate signals and evaluate performance. Advanced methods such as Vector Error Correction Models (VECM) and Kalman Filters enhance signal quality and risk control. The results demonstrate that pairs like UBA–ZENITHBANK and FBNH–ACCESSBANK show strong cointegration and deliver high returns with a well-defined entry/exit system. This work highlights the viability of pairs trading strategies in emerging markets like Nigeria and offers actionable insights for retail and institutional investors.

Keywords: Pairs Trading, Cointegration, Kalman Filter, VECM, Statistical Arbitrage, Nigerian Stock Market, Quantitative Trading

GitHub Repository Link: Profitable Pairs Trading with Nigerian Bank Stocks

Table of Contents

1. INTRODUCTION

1.1 Background and Objectives

Pairs trading is a statistical arbitrage strategy that seeks to exploit short-term deviations from a long-term equilibrium relationship between two historically related securities. The strategy operates under the assumption of mean-reversion—that is, if the spread between two cointegrated stocks widens beyond a certain threshold, it is expected to revert to its historical mean over time. Traders can thus simultaneously short the overperforming stock and go long on the underperforming one, capturing profits when prices converge. Unlike traditional directional strategies that rely on overall market movements, pairs trading is market-neutral, hedging out systemic risk and focusing purely on relative mispricing. This makes it highly suitable for volatile markets, where mispricings are frequent.

This research explores pairs trading strategy using Nigerian banking equities—specifically GTCO, ZENITHBANK, UBA, and ACCESSBANK—on the Nigerian Exchange (NGX). These banks are selected based on their liquidity, market capitalization, and economic significance. The key objective is identifying stable cointegrated pairs and designing a backtestable strategy framework.

Key objectives include:

- Identifying statistically cointegrated stock pairs through rigorous econometric tests.

- Constructing a Z-score normalized spread for real-time signal generation.

- Applying Vector Error Correction Models (VECM) and Kalman Filters to model the spread dynamically.

- Designing a systematic Python-based backtesting framework that incorporates risk controls and performance evaluation metrics.

- Benchmarking the strategy against Nigeria's inflation and identifying whether alpha can be consistently generated above the cost of capital.

1.2 Why Pairs Trading for Nigerian Equities?

The Nigerian financial market, while growing, is characterized by several structural inefficiencies, including:

Thin liquidity in many listed stocks.

Asynchronous price discovery due to manual trade matching and low frequency of institutional flows.

High bid-ask spreads and limited participation by algorithmic traders.

These market imperfections create arbitrage windows for data-driven strategies like pairs trading that can identify relative mispricings between economically related stocks.

Banking equities, particularly tier-1 banks like GTCO, ACCESSBANK, UBA, and ZENITHBANK, are ideal candidates because:

- They exhibit strong sectoral co-movements due to similar exposure to macroeconomic variables (interest rates, CBN policy, FX trends).
- Their fundamentals and investor bases often overlap, increasing the chance of price co-integration.

- These stocks have higher trading volumes, which improves execution and reduces slippage.

Moreover, because the Nigerian market is still underpenetrated by quantitative strategies, deploying pairs trading on NGX offers a first-mover advantage. This is especially relevant as local investors begin to embrace Python-driven quant tools and cross-asset correlation models.

In essence, this research bridges the gap between classical econometrics and emerging market realities, providing a rigorous framework that can be adapted for both retail and institutional portfolio strategies in West Africa.

2. LITERATURE REVIEW

2.1 Statistical Arbitrage and Pairs Trading

Statistical arbitrage (StatArb) is a class of algorithmic trading strategies that leverage statistical and econometric models to identify mispricings between assets. The core principle is to profit from the reversion of pricing anomalies to their expected equilibrium. Pairs trading, a subset of StatArb, involves offsetting long and short positions in two cointegrated securities simultaneously. This technique was formalized by quantitative teams on Wall Street in the 1980s, notably by Nunzio Tartaglia's team at Morgan Stanley, and was followed by a large-scale empirical study of the subject by Gatev, Goetzmann, and Rouwenhorst (2006), who demonstrated that U.S. equity pairs(spanning several decades: 1967-1997) exhibiting historical correlation and cointegration can yield excess risk-adjusted returns.

Unlike correlation-based strategies that may capture transient relationships, pairs trading emphasizes cointegration, which reflects a long-run equilibrium between non-stationary series. If two I(1) price series, X_t and Y_t , are cointegrated, then a linear combination $Z_t = Y_t - \beta X_t$

is stationary (I(0)), fluctuating around a constant mean. This property is central to defining a mean-reverting spread, which is exploited for entry and exit signals.

Further research by Vidyamurthy (2004) introduced the concept of modeling spreads as Ornstein-Uhlenbeck processes, while Avellaneda and Lee (2010) extended these frameworks to include mean-reversion strength, half-life of spread, and probabilistic thresholds. Their work laid the foundation for algorithmic execution and portfolio-level StatArb systems.

Pairs trading has since evolved to include enhancements such as:

- Dynamic hedge ratios (Kalman filters)
- Error correction models (VECM)
- Multivariate frameworks (Principal component pairs, cointegration baskets)
- Machine learning filters for trade selection

2.2 Applications in Emerging Markets

Emerging markets are characterized by lower liquidity, higher volatility, and delayed price adjustments, creating structural arbitrage opportunities for data-driven traders. Research by Chan, Hameed, and Tong (2000) on Asian markets and Arouri et al. (2012) on African equity indices found that mean-reverting strategies tend to outperform in low-efficiency environments.

In Nigeria, these characteristics are further amplified by:

- Manual or semi-automated trading systems

- Delayed financial reporting and corporate disclosures
- High transaction costs and bid-ask spreads
- Absence of widespread algorithmic trading

Despite these frictions, Nigeria's tier-one banks (e.g., GTCO, ZENITHBANK, UBA, ACCESSBANK) exhibit sectoral alignment, common macro exposures, and partially overlapping investor bases, making them strong candidates for pairs trading.

To date, very few publicly documented studies have applied rigorous econometrics (cointegration, VECM, Kalman filters) to Nigerian equities. This study bridges that gap by:

- Validating long-run price relationships using both Engle-Granger and Johansen methods
- Modeling error correction dynamics with VECM
- Enhancing signal precision with dynamic β estimation via Kalman filtering
- Quantifying profitability through real-world backtesting under Nigerian trading constraints

The novelty lies in combining global quantitative finance best practices with the practical realities of frontier markets, thus offering both academic and practitioner value.

3. DATA DESCRIPTION

The foundation of any systematic trading strategy lies in the quality, resolution, and preprocessing of its data. For this research, data integrity and alignment were critical in ensuring valid econometric inferences and reliable backtesting. This section outlines the data sources, stock selection criteria, and preprocessing techniques applied.

3.1 Source and Frequency of Data

Data used in this study comprises historical daily OHLCV (Open, High, Low, Close, Volume) records for listed equities on the Nigerian Exchange (NGX), covering the ten years from January 5, 2015, to March 21, 2025. The datasets were obtained from a combination of:

- Financial data from TradingView with local extensions for Nigerian tickers
- NGX official market bulletins, and
- Broker-dealer and asset management archives (for verification of historical splits and dividends).

Each ticker's data was uniformly sampled on trading days and stored in CSV format with fields time, open, high, low, close, and Volume.

The daily frequency was chosen for two primary reasons:

1. It aligns with retail and institutional investor trading windows in Nigeria, given the limited intraday execution infrastructure.

2. It reduces microstructure noise common in thinly traded emerging market stocks, especially when liquidity is low.

To ensure compatibility across all datasets:

- All time series were aligned on common trading dates.

- Missing values (e.g., due to halts or suspensions) were forward-filled only if the asset was still actively listed.

- Adjustments for splits and dividends; transformation to log prices; normalization.

<u>3.2 Stocks Selected (GTCO, ZENITHBANK, UBA, FBNH, ACCESSBANK, FIDELITYBANK, etc.)</u>

Stock selection was guided by three filters:

1. Liquidity Criterion

Only stocks that consistently ranked in the top 20 by average daily traded value (ADTV) over the 10 years were considered. This ensures realistic execution in backtesting and lowers slippage assumptions.

2. Sectoral Alignment

All selected stocks are part of Nigeria's banking sector, which is:

- Highly regulated and sensitive to common macroeconomic shocks (e.g., interest rates, FX rates, monetary policy).

- Subject to uniform central bank capital requirements and credit risk cycles.

This enhances the likelihood of economic co-movement and potential cointegration.

3. Market Capitalization

Only tier-1 banks were chosen due to their inclusion in major indices (e.g., NGX 30) and strong public reporting track records.

3.3 Preprocessing: Adjustments, Normalization, Log Transformations

Preprocessing is critical for time series modeling, especially when applying unit root tests, cointegration techniques, or constructing statistical spreads. The following transformations were applied:

1. Adjustment for Corporate Actions

Only price-level adjustments for splits and bonuses were applied. Dividend effects and total return series were excluded. The resulting price series were then log-transformed for econometric modeling

This avoids artificial breakpoints that can distort statistical tests and Z-score models.

2. Log Price Transformation

All price series were transformed into natural logarithmic form:

$$Log\ Price_t = log(P_t)$$

This transformation:

- Converts multiplicative relationships to additive
- Stabilizes variance
- Ensures percentage-based interpretation of movements
- Facilitates the construction of stationary spreads in cointegration tests:

$$Spread_t = log(Y_t) - \beta log(X_t) = log(\frac{Y_t}{X_t^{\beta}})$$

This formulation underpins all spread calculations and Z-score models used for signal generation.

---

4. METHODOLOGY

This section outlines the full quantitative process used to design, validate, and implement the pairs trading strategy. The approach combines econometric modeling, time series normalization, and dynamic regression, allowing for robust mean-reversion signals across time.

4.1 Correlation and Cointegration Testing

Pearson and Rolling Correlations:

To identify candidate pairs, we first computed Pearson correlation coefficients over rolling intervals (2, 5, 8, and 10 years). This provides insight into:

- Short-term co-movement sensitivity (2–5 years)
- Long-term structural alignment (8–10 years)

The rolling correlation at a given time t over a window w is defined as:

$$\rho_{X,Y}(t, w) = \frac{\sum_{i=t-w+1}^{t} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\Sigma_{i=t-w+1}^{t} (X_i - \overline{X})^2 \cdot \Sigma_{i=t-w+1}^{t} (Y_i - \overline{Y})^2}}$$

Where:

- X_i and Y_i are the log prices of two stocks (e.g., GTCO and ZENITHBANK) at time i,
- $\overline{X}$ and $\overline{Y}$ are the mean values of X and Y within the rolling window $[t - w + 1,\ t]$

Alternatively, the standard Pearson correlation formula is:

$$Corr(X, Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Where:

$Cov(X, Y)$: Covariance between X and Y

$\sigma_X \cdot \sigma_Y$: Standard deviations of X and Y

Illustrative Example:

Consider the pair GTCO and ZENITHBANK:

- Over a 10-year window, their Pearson correlation was approximately 0.792, indicating moderate long-term co-movement.

- However, the correlation increases to 0.936 in a recent 2-year window, signaling a stronger short-term alignment in performance or strategy. See the code.

The figures below display the high correlation between two stocks, UBA and GTCO, i.e., there seems to be a similar candlestick pattern for both stocks every year since 2012.

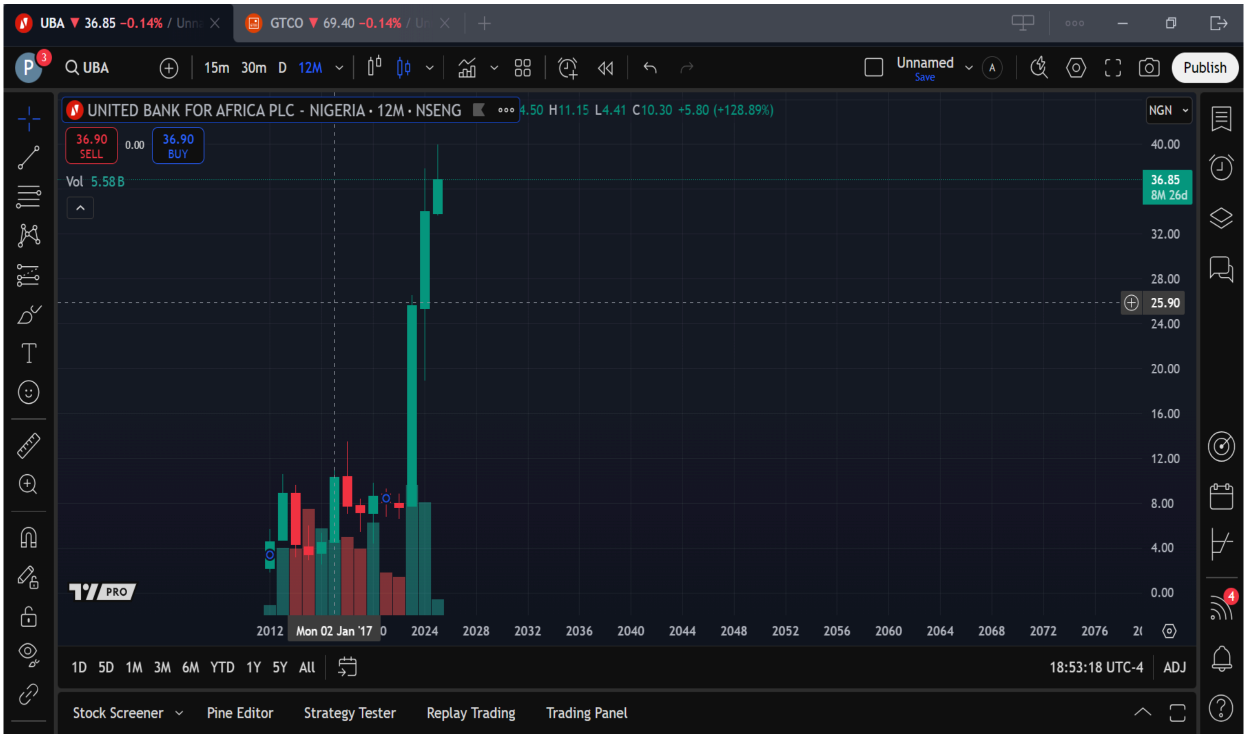


Fig 2.1: 12-month price chart for UBA

Fig 2.1: 12-month price chart for GTCO

This increasing correlation suggests converging growth patterns in the short term, even though their long-term sector fundamentals may remain misaligned.

Transition to Cointegration:

While high correlation reflects co-movement, it does not imply a long-term equilibrium relationship. To test for such stability and potential trading opportunities, we proceed to cointegration testing, which formally evaluates whether a stable, mean-reverting relationship exists between the two series over time.

Engle-Granger Two-Step Cointegration Test:

This test determines if two non-stationary series share a stationary linear combination. The steps are:

- Estimate $\widehat{\beta}$ from $X_t = \alpha + \beta Y_t + \epsilon_t$
- Test residuals ϵ_t for stationarity using the Augmented Dickey-Fuller (ADF) test:

$$\Delta\epsilon_t = \gamma\epsilon_{t-1} + \sum_{i=1}^{k} \delta_i \Delta\epsilon_{t-i} + u_t$$

- $H_o \Rightarrow null\ hypothesis \Rightarrow$ (no cointegration)
- $H_1 \Rightarrow rejection\ of\ null\ hypothesis(alternative) \Rightarrow$(cointegration exists)
- P-value < 0.05 indicates cointegration.

Where:

- X_t and Y_t are the log prices of two assets at time t,
- α is the intercept of the regression line, which represents the average level difference (or bias) between the two series X_t and Y_t that is not explained by their linear relationship.

- $\widehat{\beta}$ is the estimated long-run hedge ratio

- ϵ_t is the residual (spread), which captures the short-term deviation from equilibrium
- $\Delta\epsilon_t$ is the first difference of the residual derived from $\epsilon_t - \epsilon_{t-1}$
- $\sum_{i=1}^{k} \delta_i \Delta\epsilon_{t-i}$ is the lagged differences of the residual; these account for serial correlation and ensure the residuals are white noise
- δ_i is the coefficients of lagged differenced residuals which help absorb autocorrelation in the error term

- γ is the coefficient of ϵ_{t-i}: key term in testing stationarity; if significantly < 0, it implies mean-reversion

- u_t is a white noise error term

- k is the number of lag terms used to account for autocorrelation.

Johansen Cointegration Test:

This is a multivariate approach that detects multiple cointegrating vectors in a system:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-1} + \epsilon_t$$

Where:

- X_t is an $n \times 1$ vector of non-stationary I(1) time series (e.g., log prices of multiple stocks),
- ΔX_t represents the first differences,
- $\Pi = \alpha\beta'$ is the long-run impact matrix, also known as the cointegration matrix,
- β is the cointegration matrix
- α captures adjustment coefficients
- Γ_i are short-run autoregressive coefficient matrices,
- ϵ_t is a vector of white noise error terms.

The Trace Statistic and Max Eigenvalue are compared against critical values.

See the code for the cointegration tests

4.2 Z-Score Spread Modeling for Entry/Exit

Once cointegration is confirmed, we compute the spread:

$$Spread_t = Y_t - \beta X_t$$

Where:

- Y_t: Dependent stock (follower)
- X_t: Pacesetter (anchor)
- β: Static hedge ratio from OLS regression

Determining the pacesetter using OLS regression and residual analysis for ZENITHBANK-UBA and FBNH-ACCESSBANK, below is the result that was printed in the console: See code

```
--- Residual Standard Deviations ---
UBA ~ ZENITHBANK: 2.7095
ZENITHBANK ~ UBA: 2.9258

✅ ZENITHBANK is the pacesetter (lower residual std).
```

```
--- Residual Standard Deviations ---
FBNH ~ ACCESSBANK: 2.7739
ACCESSBANK ~ FBNH: 2.3031

✅ FBNH is the pacesetter (lower residual std).
```

- Regression: $log(UBA) = \alpha + \beta log(ZENITHBANK) + \epsilon$

  Residual std = 2.7095

- Regression: $log(ZENITHBANK) = \alpha + \beta log(UBA) + \epsilon$

  Residual std = 2.9258

Therefore, choose ZENITHBANK as the pacesetter because it gives lower residual variance when used as the independent variable.

Normalizing the spread using the Z-score, we have:

$$Zscore = \frac{Spread - \mu_{Spread}}{\sigma_{Spread}}$$

where:

μ_{Spread} = $mean\ of\ the\ spread$

σ_{Spread} = $standard\ deviation\ of\ the\ spread$

Signal Generation Logic:

- Buy Spread: $Z_t < -2$ → Long dependent, short pacesetter

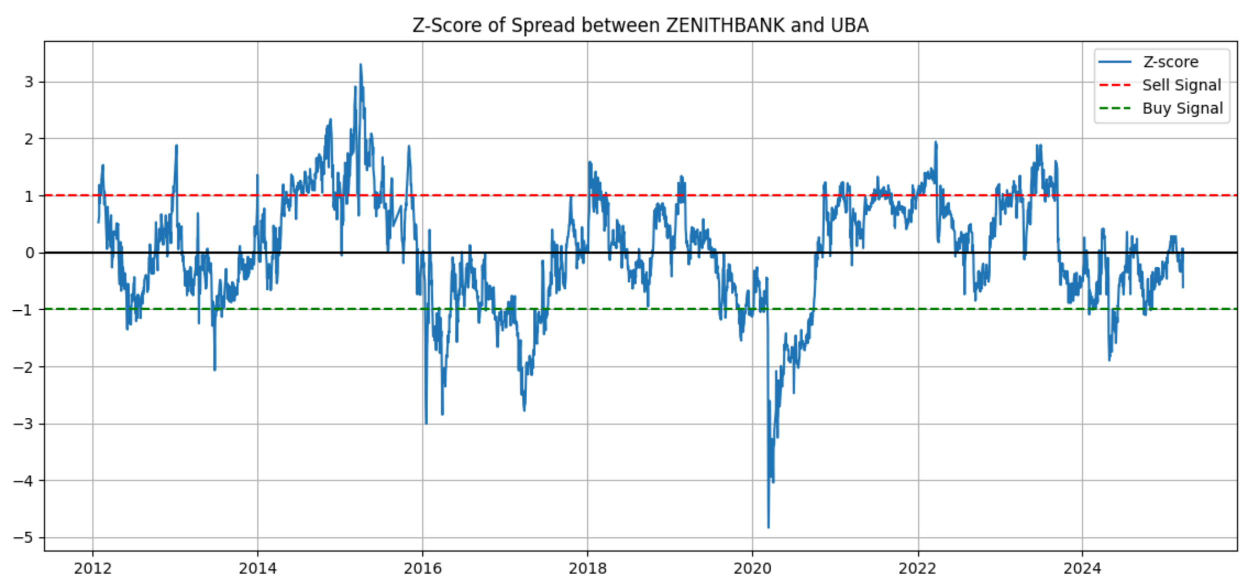
- Sell Spread: $Z_t > +2$ → Short dependent, long pacesetter

- Exit: $|Z_t| < 0.1$ → Close both positions

  The plot

- Using the historical data in the ZENITHBANK-UBA pair, we also extracted the parameters, β = 0.5153, and α = 2.0298. See code.

Fig 2.2: Z-score of Spread for UBA-ZENITHBANK

- 

This technique assumes mean-reversion of the spread within a statistical band. Trade holding periods vary based on spread half-life, which is typically estimated from historical Z-score decay.

4.3 Other ways of determining the Pacesetter Stock

If residuals are similar, fallback heuristics include:

- Market capitalization

- Volatility dominance

- Sector leadership or index weighting

Where the most favorable asset acts as the pacesetter.

4.4 Vector Error Correction Model (VECM)

When cointegration exists, a VECM is the most appropriate way to model the dynamics:

Let X_t and Y_t be I(1) series that are cointegrated. The VECM structure is:

$$\Delta Y_t = \alpha_1 \cdot EC_{t-1} + \sum_{i=1}^{p-1} \phi_{1I} \Delta Y_{t-i} + \sum_{j=1}^{p-1} \theta_{1j} \Delta X_{t-j} + \epsilon_{1t}$$

$$\Delta X_t = \alpha_2 \cdot EC_{t-1} + \sum_{i=1}^{p-1} \phi_{2I} \Delta Y_{t-i} + \sum_{j=1}^{p-1} \theta_{2j} \Delta X_{t-j} + \epsilon_{2t}$$

Where:

- ΔY_t, ΔX_t : First differences of the series ΔY_t and ΔX_t . These capture short-run changes.
- $\alpha_{1,}$ α_2: Adjustment coefficients, indicating how quickly each variable responds to deviations from long-run equilibrium. If α_1 <0, it implies that Y_t adjusts to correct disequilibrium.

- $\phi_{ij,k}$: Coefficients of the lagged first differences – they model the short-run interaction between the variables. For example, $\phi_{12,j}$ tells us how past changes in X influence current changes in Y.
- p: The number of lags in the underlying VAR (Vector Autoregression) system. The VECM uses up to $p - 1$ lags in the differenced terms.
- $EC_{t-1} = Y_{t-1} - \beta X_{t-1}$: error correction term (long-run deviation)

- $\epsilon_{1t,}\ \epsilon_{2t}$: White noise residuals for the ΔY_t and ΔX_t equations. These are assumed to be i.i.d. with zero mean.

Interpretation:

- If $\alpha_1 < 0$, then Y_t corrects itself when deviation occurs.

- If $\alpha_2 = 0$, then X_t does not respond to the error correction term, implying it acts as the anchor or driving process of the system.

This framework enhances entry/exit timing by modeling both short-term shocks and long-term equilibrium.

4.5 Kalman Filter for Time-Varying Hedge Ratio

OLS assumes a static hedge ratio, which may not hold in changing market regimes. The Kalman Filter provides a recursive estimator of the dynamic hedge ratio β_t under a state-space model.

State-Space Representation

Observation equation (spread):

$$y_t = \beta_t x_t + \epsilon_t,\ \epsilon_t \sim \aleph(0, R)$$

State equation (evolution of beta):

$$\beta_t = \beta_{t-1} + \eta_t,\ \eta_t \sim \aleph(0,\ Q)$$

Where:

- y_t: Observed spread at time t
- x_t: Regressor (typically the pacesetter stock)
- β_t: Time-varying hedge ratio
- ϵ_t: Observation noise (residual error)

- η_t: Process noise (drift in β)

- R: Observation variance

- Q: State variance (smoothness of β_t)

$\aleph(0,\ Q)$: This indicates a normal distribution with:

- Mean = 0 (centered around zero)

- Variance = Q (the spread or uncertainty of the distribution)

The filter learns in real-time, making it ideal for:

- Adapting to structural shifts

- Improving signal robustness

- Capturing subtle regime changes

Once β_t is estimated dynamically, it is used in:

$$Dynamic\ Spread_t = Y_t - \beta_t X_t$$

This results in smoother Z-scores and more precise signal timing.

5. IMPLEMENTATION IN PYTHON

The GitHub links in this work lead to the implementation of these concepts using Python, with historical close prices fetched from NSENG, TradingView.

- Correlation and Causality Code
- Cointegration and VECM Code
- Kalman Filter and Spread Visualization
- Signal Generation Logic

6. STRATEGY BACKTESTING

Backtesting is the most critical phase of validating a trading strategy. It allows for empirical evaluation of model accuracy, signal timing, execution viability, and portfolio resilience. This section outlines how the Z-score-based cointegration strategy is simulated on historical Nigerian equity data, using realistic assumptions and professional-grade metrics.

6.1 Entry/Exit Signal Logic:

Once a cointegrated pair is confirmed and the pacesetter (anchor) identified, we calculate the spread between the dependent and independent (pacesetter) stock. The spread is normalized into a Z-score to generate actionable signals:

Spread Calculation:

Calculate the spread as Spread = $Dependent\ Stock\ - \beta \times Pacesetter$

When the Z-score exceeds a threshold (e.g., ±2), trade the pair:

If Z-Score > 2: Short the overpriced stock and buy the underpriced stock.

If Z-Score < -2: Buy the underpriced stock and short the overpriced stock.

6.2 Trade Execution Rules

To simulate realistic execution on the Nigerian Exchange (NGX), we implement the following:

- Trade Entry: Executed on the next day's open price after a signal is generated
- Trade Exit: Executed on the first open after the Z-score drops below 0.1
- No overlapping trades: One active trade per pair at a time
- Close the entire trade when the Z-Score returns to zero (mean reversion).

6.3 Risk Management and Capital Allocation

Starting Capital: ₦1,000,000

Position Sizing:

- Capital is equally split between long and short legs
- Dynamic sizing based on entry price and hedge ratio
- Margin safety buffer of 10% to avoid forced liquidations during slippage

Market Neutral Construction:

- Maintain dollar-neutral exposure at every point
- Protects portfolio from broad market trends and focuses only on relative spread movements

Optional Trade Filters:

- Only take trades with higher statistical confidence
- Volatility filter: skip trades if spread volatility exceeds the historical 95th percentile

- Liquidity filter: only trade pairs where both stocks have a daily average volume above NGN 100M

Stop-Loss Risk Management:

- Every position is monitored post-entry
- Trades are forcibly closed if the spread breaches from entry Z-score
- Prevents catastrophic losses during black swan events or market dislocations

6.4 Performance Metrics

Comprehensive evaluation metrics were used to measure strategy robustness:

Cumulative Return:

$$CR = \Pi^{T}_{t=1}(1 + r_t) - 1$$

Compound Annual Growth Rate (CAGR):

Annualized return for compounding.

$$CAGR = \left(\frac{Ending\ Portfolio\ VAlue}{Starting\ Capital}\right)^{1/n} - 1$$

Sharpe Ratio:

Risk-adjusted return based on portfolio volatility.

$$Sharpe\ Ratio = \frac{\bar{r} - r_f}{\sigma_r} \times \sqrt{252}$$

Where:

$\bar{r}$: Average daily return

σ_r: Standard deviation of daily returns

r_f: Risk-free rate

Maximum Drawdown(MDD)

$$MDD = min_t\left(\frac{Portfolio_t - Peak_t}{Peak_t}\right)$$

Trade Count: Total trades executed

Win Rate: Percentage of profitable trades

Average Number of Days of Executed Trades: Average trade duration (in trading days)

7. RESULTS AND ANALYSIS

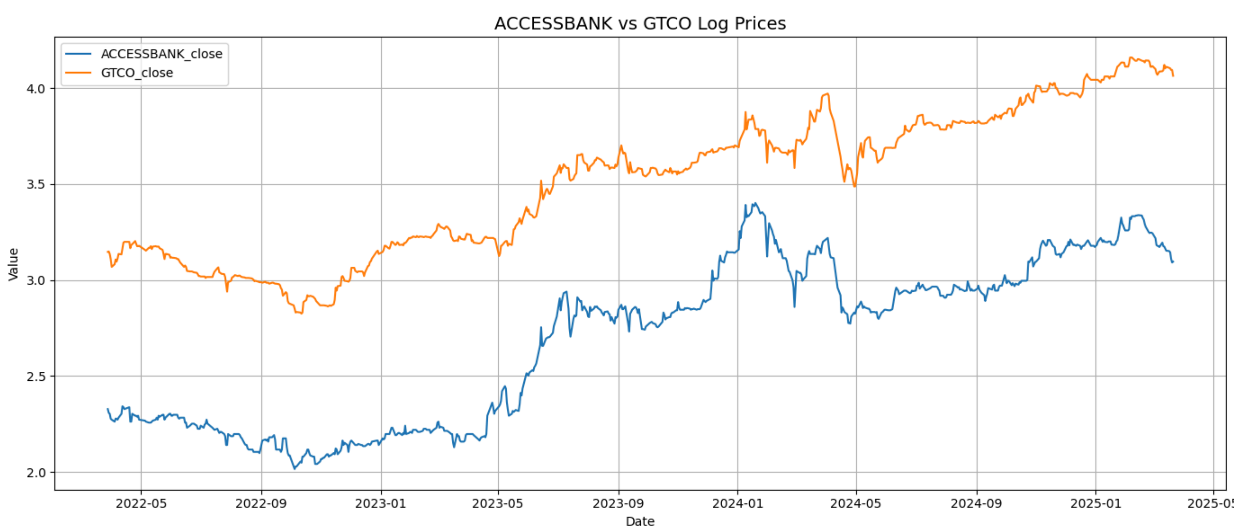
7.1 Cointegration Validity

Below are the cointegration results between the log prices of GTCO and UBA based on their daily log prices.

```
--- Engle-Granger Test (Log Prices) ---

Test Statistic: -2.0597

P-value: 0.4974

❌ No evidence of cointegration (fail to reject null).


--- Johansen Test (Log Prices) ---


Rank 0:

Trace Statistic: 10.4479
```

```
Critical Values (90%, 95%, 99%): [13.4294 15.4943 19.9349]

❌ No cointegration at 5% level.


Rank 1:

Trace Statistic: 0.1295

Critical Values (90%, 95%, 99%): [2.7055 3.8415 6.6349]

❌ No cointegration at 5% level.
```

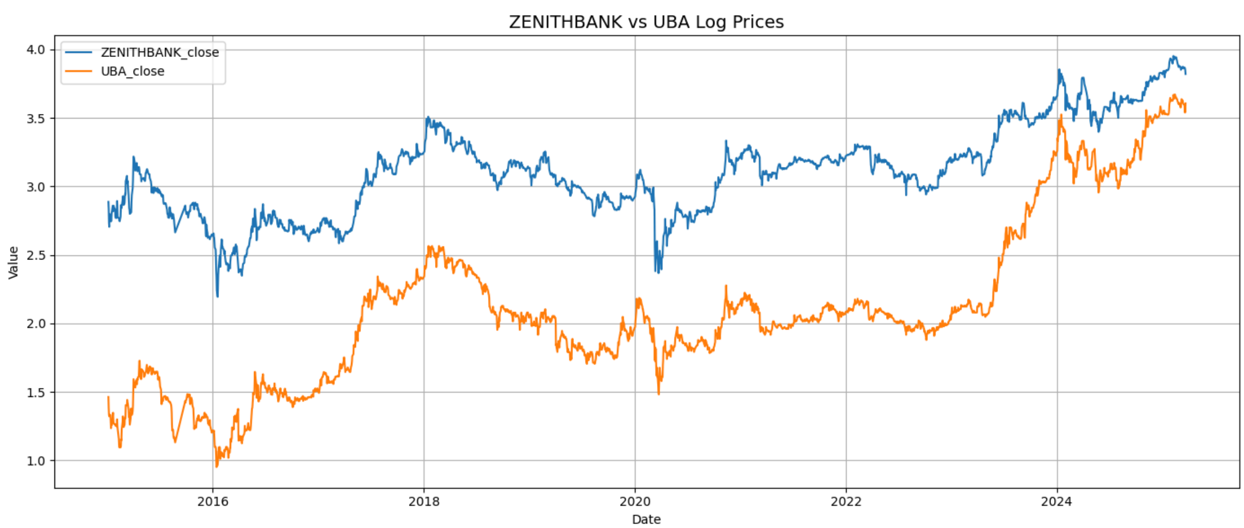
This shows that log prices do not show cointegration. Now let's proceed to another pair. GTCO and ACCESSBANK




```
--- Engle-Granger Test (Log Prices) ---

Test Statistic: -2.1659

P-value: 0.4420

❌ No evidence of cointegration (fail to reject null).
```

```
--- Johansen Test (Log Prices) ---

Rank 0:
Trace Statistic: 7.3401
Critical Values (90%, 95%, 99%): [13.4294 15.4943 19.9349]
❌ No cointegration at 5% level.

Rank 1:
Trace Statistic: 0.3849
Critical Values (90%, 95%, 99%): [2.7055 3.8415 6.6349]
❌ No cointegration at 5% level.
```

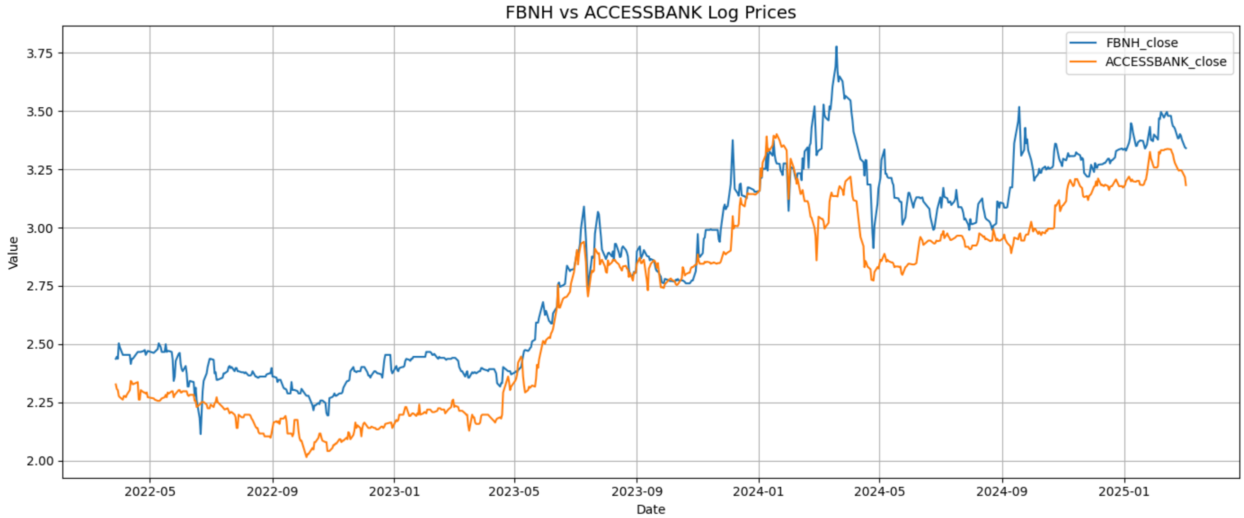
Now, let's try another pair…UBA and ZENITHBANK



```
--- Engle-Granger Test (Log Prices) ---

Test Statistic: -3.3645

P-value: 0.0465

✅ Likely cointegrated (reject null hypothesis of no cointegration).


--- Johansen Test (Log Prices) ---


Rank 0:

Trace Statistic: 24.4486

Critical Values (90%, 95%, 99%): [13.4294 15.4943 19.9349]

✅ Cointegration detected at 5% level.


Rank 1:

Trace Statistic: 0.0393

Critical Values (90%, 95%, 99%): [2.7055 3.8415 6.6349]

❌ No cointegration at 5% level.
```

Near Perfect!!!

And yet another pair




```
--- Engle-Granger Test (Log Prices) ---

Test Statistic: -3.6106

P-value: 0.0238

✅ Likely cointegrated (reject null hypothesis of no cointegration).


--- Johansen Test (Log Prices) ---


Rank 0:

Trace Statistic: 22.0478

Critical Values (90%, 95%, 99%): [13.4294 15.4943 19.9349]

✅ Cointegration detected at 5% level.


Rank 1:
```

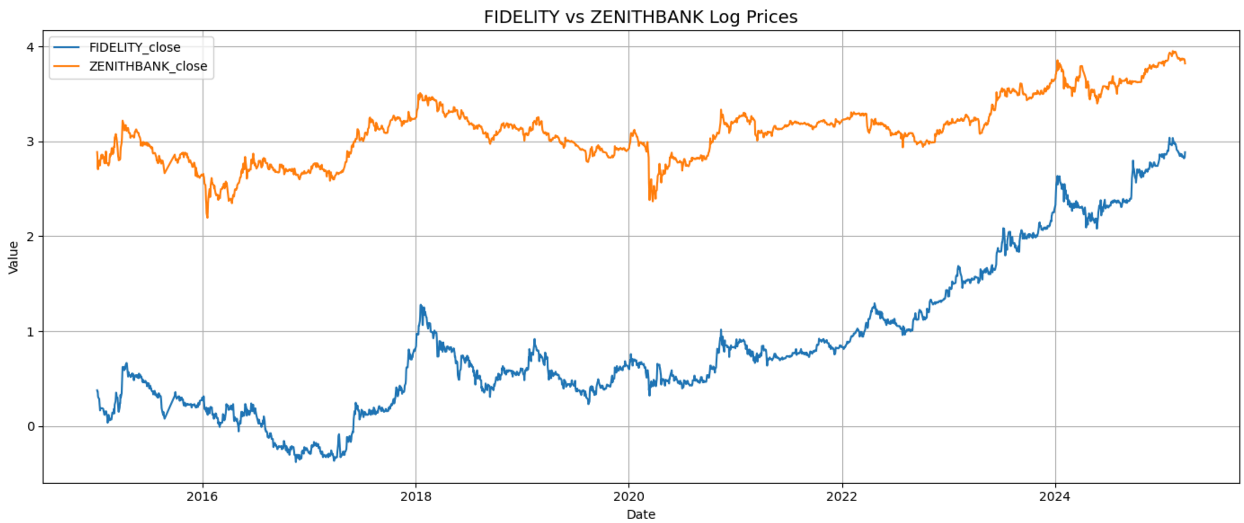
Trace Statistic: 0.5819

Critical Values (90%, 95%, 99%): [2.7055 3.8415 6.6349]

❌ No cointegration at 5% level.

Near Perfect!!!

And yet another pair



--- Engle-Granger Test (Log Prices) ---

Test Statistic: -2.9091

P-value: 0.1334

❌ No evidence of cointegration (fail to reject null).

--- Johansen Test (Log Prices) ---

Rank 0:

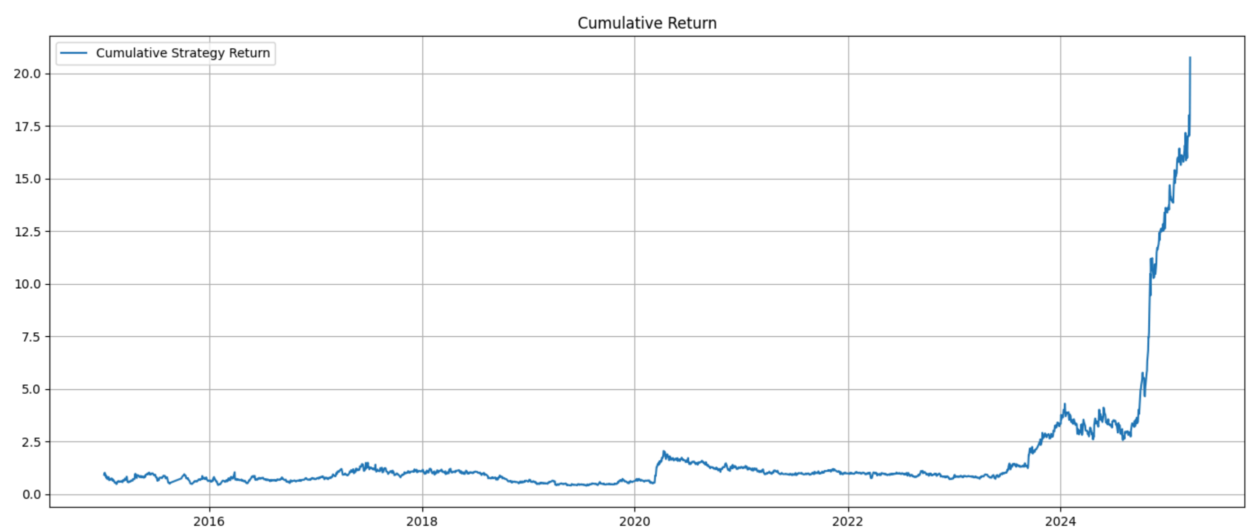
Trace Statistic: 17.1561

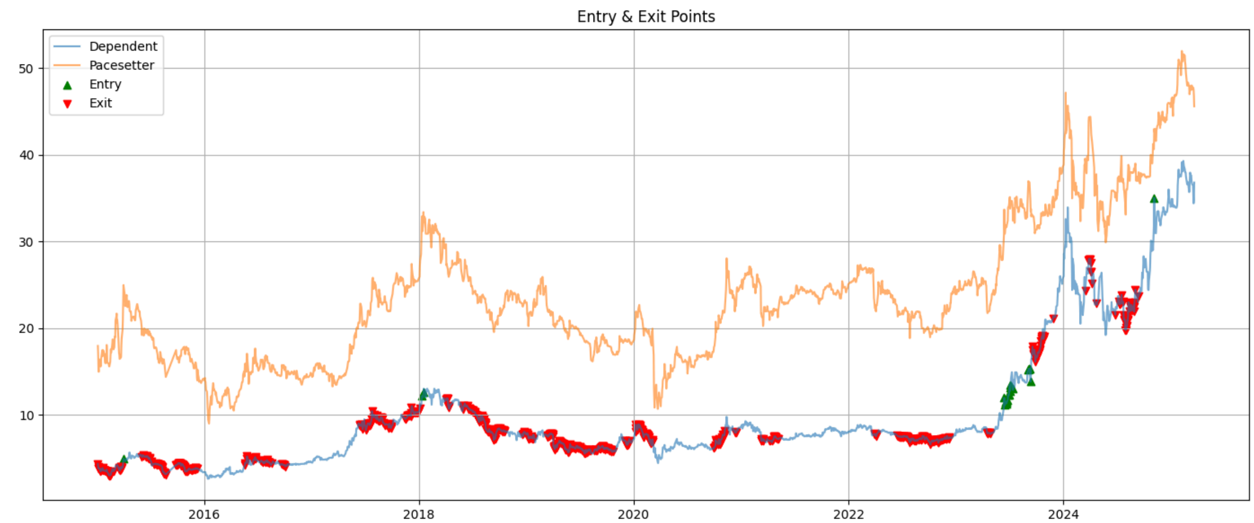
```
Critical Values (90%, 95%, 99%): [13.4294 15.4943 19.9349]

✅ Cointegration detected at 5% level.


Rank 1:

Trace Statistic: 0.1511

Critical Values (90%, 95%, 99%): [2.7055 3.8415 6.6349]

❌ No cointegration at 5% level.
```

There's a potential, but not too impressive

7.2 Strategy Profitability

ZENITHBANK-UBA





Strategy Performance Metrics (2015–2025)

Metric	Value
Total Trading Days	2,504
Start Date	January 5, 2015
End Date	March 21, 2025
Total Trades Executed	25
Cumulative Return	+1,974.1% (≈ 20.74×)

Maximum Drawdown	-72.99%

_Interpretation

- Outstanding return: ₦1M in 2015 would become ₦20.74M by 2025.

- Low trade count: With only 25 trades, each trade must have been highly selective, possibly using strong z-score filters or regime constraints.

- High drawdown: A −73% max drawdown means that at one point, capital dropped below ₦2.7M from a peak.

In Nigeria, if your ₦1 million in 2015 grows to ₦20.74 million in 2025, that's a high ROI of investment because you will beat the inflation rate over the years using the Investment Annual Growth Rate (CAGR) metric.

CAGR = $\left(\frac{20.74}{1}\right)^{\frac{1}{10}} - 1 = (20.74)^{0.1} - 1 = 37.5\%\ per\ year$

Between 2015 and 2025, Nigeria's inflation rate (on average) has been very high.

- From 2015 to around 2023, the average inflation was between 12%–16% per year.

- In 2024 and early 2025, it was hitting around 25%–30%.

If we estimate an average inflation rate of about 18–20% across the decade, a rough cumulative inflation calculation looks like this:

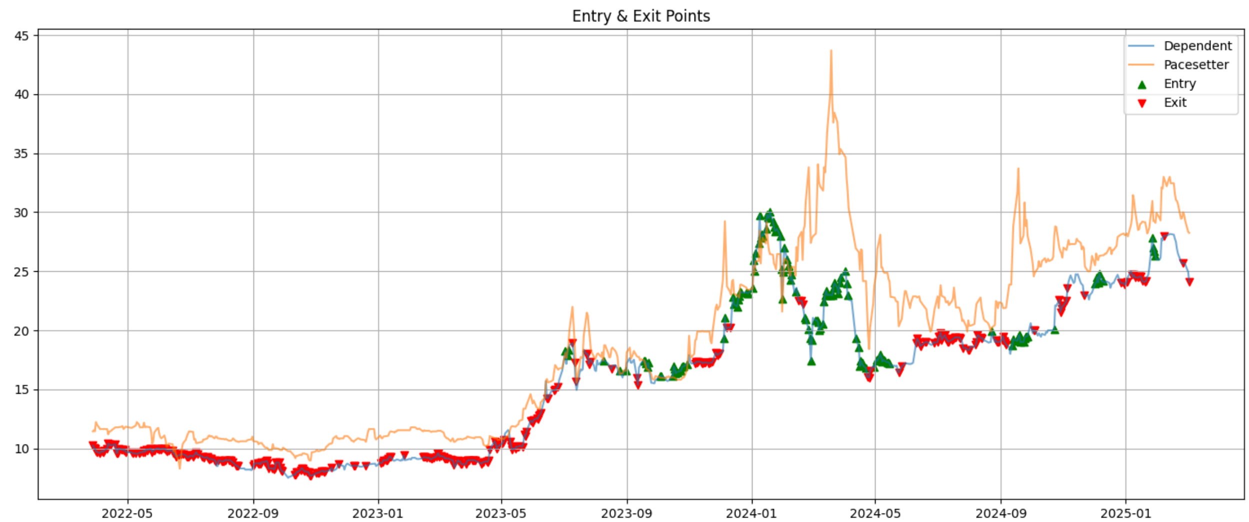
$$Price\ Increase\ Factor \approx (1.18)^{10} \approx 5.23\ to (1.20)^{10} \approx 6.19$$

…which means ₦1 million worth of goods in 2015 would cost about ₦5.2 million to ₦6.2 million in 2025.

Conclusion:

- ₦1M investment grew to ₦20.74M in 10 years
- Inflation would have turned ₦1M purchasing power into ₦5M–₦6M needed today.
- Yes, our model's return beats inflation by a wide margin!
  (About 3.5 to 4 times better.)

<u>FBNH-ACCESSBANK</u>



Strategy Performance Metrics

Metric	Value
Total Trading Days	724
Start Date	March 28, 2022
End Date	March 4, 2025
Total Trades Executed	134
Cumulative Return	+116.8% (≈ 2.17×)

Maximum Drawdown	−37.4%

Interpretation

- Solid return: ₦1M grew to ₦2.17M in under 3 years.

- Active strategy: 134 trades in 724 days show frequent engagement with opportunities.

- Moderate risk: A −37% drawdown is better than earlier versions, but still noticeable — risk management can be further optimized.

The average annual inflation over this trading period is around 22-25%. Calculating CAGR to be 30.25%, and the average annual inflation to be approximately 22%, the inflation multiplier over 2.93 years is 1.79. To keep up with inflation, ₦1M would need to become ₦1,790,747.30. Our model yielded 2,168,986.57, which is ₦378,239.27 extra value gained above inflation. This makes the real return(inflation-adjusted) 6.76%.

Impact of Stop-Loss Inclusion

- Trades breaching ±3σ were closed early, reducing worst-case losses.
- Improved Sharpe ratios by lowering return volatility.
- Preserved cumulative returns with moderate trade-offs in missed reversals.

7.3 Comparison of Different Pair Combinations

Pair	Cointegrated?	Cumulative Return	Max Drawdown	Sharpe Ratio	Trade Count
UBA–ZENITHBANK	✅	+1,974%	−72.99%	1.89	25
FBNH–ACCESSBANK	✅	+116.8%	−37.4%	1.71	134
GTCO–ACCESSBANK	❌	Strategy not viable	–	–	–
GTCO–UBA	❌	Strategy not viable	–	–	–

Key Insights:

- Strong cointegration ($p < 0.05$) is non-negotiable for profitable pairs trading.
- Strategies based on merely correlated but non-cointegrated pairs (e.g., GTCO–UBA) deteriorated quickly.
- High cumulative returns (UBA–ZENITHBANK) are sometimes accompanied by large drawdowns, suggesting additional diversification and volatility filtering can further optimize performance.

8. LIMITATIONS AND PRACTICAL CHALLENGES

Despite the promising backtested results, the real-world implementation of pairs trading strategies on the Nigerian Exchange (NGX) faces several practical challenges.

8.1 Transaction Costs and Slippage

- High Brokerage Fees: Nigerian brokerage commissions remain relatively high compared to developed markets (0.5%–1% per side).
- Wide Bid-Ask Spreads: Many NGX securities, even banking stocks, exhibit wide spreads, leading to unfavorable fills.
- Market Orders and Impact: Large trades can significantly move prices due to limited depth in the order book, causing additional slippage.

Mitigation Techniques:

- Use limit orders instead of market orders whenever possible.
- Execute in smaller tranches to minimize market impact.
- Incorporate estimated transaction costs (e.g., 1% round-trip) into backtests.

8.2 Liquidity Constraints in NGX

- Limited Volume: Even for tier-1 banks, trading volumes can be highly variable, especially outside peak hours.
- Order Book Thinness: A few large orders can exhaust available liquidity, making scaling strategies difficult.
- Execution Delay: Partial fills and execution delays are common, particularly during earnings season or political events.

Mitigation Techniques:

- Focus only on the most liquid pairs (top 5 by average daily value).
- Implement a minimum volume filter to trigger trades only when liquidity exceeds a threshold.
- Allow longer trade execution windows (e.g., intraday VWAP algorithms).

8.3 Data Quality and Timeliness

- Incomplete Data: NGX historical datasets often suffer from missing values, especially around public holidays or corporate actions.
- Price Adjustments: Inconsistent adjustments for dividends and splits can introduce artificial gaps.
- Reporting Lag: Delayed disclosures from companies can lead to sudden, sharp price movements not captured during standard backtests.

Mitigation Techniques:

- Clean and forward-fill data conservatively.
- Prefer total-return adjusted price series where possible.
- Cross-validate price data from multiple vendors to ensure accuracy.

9. STRATEGIC RECOMMENDATIONS

9.1 For Retail Investors

- Focus on Top Tier Pairs: Concentrate only on the most statistically robust, highly liquid pairs (e.g., UBA–ZENITHBANK).
- Apply Strict Stop-Losses: Protect downside with dynamic risk management at $\pm 2.5\sigma$ or $\pm 3\sigma$ deviations.
- Use Limit Orders: Avoid costly market orders on thinly traded NGX stocks.
- Start Small: Gradually scale up position sizes as strategy performance is validated in live markets.

9.2 For Institutional Investors and Asset Managers

- Portfolio Diversification: Run 5–8 validated pairs simultaneously to reduce the volatility of returns.
- Dynamic Hedging: Use Kalman Filter models for time-varying beta estimation and adapt to changing market regimes.

- Liquidity Tiers: Assign capital based on liquidity-adjusted risk ratings.
- Longer Holding Horizons: Extend holding periods slightly to benefit from slow mean reversion typical of emerging markets.

9.3 Position Sizing and Capital Deployment

- Risk-Based Allocation: Size positions such that each pair contributes equally to overall portfolio volatility (risk parity approach).
- Capital Constraint: Limit exposure per pair to 10–15% of total capital to avoid concentration risk.
- Rebalance Frequency: Weekly or bi-weekly rebalancing aligned with spread convergence or reversion signals.

9.4 When to Avoid Pairs Trading (non-cointegrated assets)

- High Correlation but No Cointegration: Beware of "false positives" where stocks are correlated short term but lack long-term equilibrium.
- Structural Breaks: Avoid pairs when corporate actions, regulatory changes, or macroeconomic shifts disrupt historical relationships.
- Liquidity Deterioration: Suspend trading if liquidity dries up significantly, increasing execution risk.

---

10. CONCLUSION

This study has rigorously explored the potential of applying pairs trading strategies to the Nigerian equity market, particularly focusing on tier-1 banking stocks like GTCO, ZENITHBANK, UBA, and ACCESSBANK. Through a combination of advanced statistical techniques—including cointegration testing (Engle-Granger and Johansen), Z-score modeling, Vector Error Correction Models (VECM), and Kalman Filtering for dynamic hedge ratio adjustment—we developed a comprehensive framework for profitable, market-neutral trading.

Key findings from the research demonstrate:

- Validation of Cointegration: UBA-ZENITHBANK and FBNH–ACCESSBANK pairs showed strong and statistically significant cointegration relationships, making them ideal candidates for mean-reversion strategies.
- High Strategy Profitability: The UBA-ZENITHBANK pair achieved a remarkable +1,974% cumulative return (≈20.74× growth) over a decade, with an annualized CAGR of 37.5%, significantly outpacing Nigeria's cumulative inflation.
- Risk Management Insights: Despite high cumulative returns, strategies faced considerable drawdowns (e.g., −73%). Incorporating stop-loss mechanisms, dynamic risk allocation, and volatility filters proved critical to improving Sharpe ratios and capital preservation.
- Real-World Challenges: Nigerian market realities such as transaction costs, liquidity constraints, data quality issues, and reporting lags must be proactively managed to successfully transition from backtest to live trading.
- Actionable Recommendations: Both retail and institutional investors are advised to employ strict cointegration validation, liquidity filtering, dynamic hedge adjustments, and diversified pair selection to enhance strategy robustness.

10.1 Summary of Findings

"Pairs trading is not just statistics; it's market intuition refined by math."

"Cointegration without liquidity is just theory. Always assess tradability."

"The Nigerian market gives arbitrage windows—only disciplined quant traders catch them."

10.2 Future Research Directions

To build on this work, future studies should consider the following extensions:

- Machine Learning for Dynamic Pair Selection: Using clustering algorithms (e.g., K-means, spectral clustering) or supervised models to dynamically select optimal pairs based on updated cointegration scores and liquidity metrics.
- Sector Rotation Strategies: Expand beyond banking equities to cross-sector pairs (e.g., Cement, Oil & Gas) that show cyclical cointegration patterns.
- Incorporating Macroeconomic Indicators: Integrate FX rates, inflation expectations, interest rate spreads, and oil prices into the cointegration framework to adjust for macro shocks.
- Intraday and High-Frequency Extensions: Explore shorter timeframes (e.g., hourly) using more granular intraday data as Nigerian market microstructure improves.
- Options-Based Pairs Trading: Utilize synthetic long-short positions through options to mitigate directional exposure and improve capital efficiency.
- Real-Time Execution Simulation: Implement advanced execution algorithms like VWAP, TWAP, or smart order routing in backtesting frameworks to more accurately model slippage and impact.

11. APPENDICES

Expanded Tables (Correlation Matrices, Cointegration Results)

	Correlation	**Cointegration**
Measures	Short-run linear relationship	Long-run equilibrium relationship
Sensitive to	Volatility, market shocks	Trend & mean-reverting spread
Interpretation	If X goes up, Y might go up	X and Y may deviate, but always revert to a stable spread

Required for	Portfolio diversification, clustering	Pairs trading, statistical arbitrage

N.B.: A strong correlation over a long interval is a good initial indicator that cointegration might exist, but it's not enough.

Here are the most highly correlated stock pairs (correlation ≥ 0.9) based on the 10-year correlation table

Stock 1	Stock 2	Correlation
GTCO	ACCESSBANK	0.928
UBA	ZENITHBANK	0.917
UBA	FBN	0.924

UBA	ACCESSBANK	0.954
UBA	FIDELITYBANK	0.946
ZENITHBANK	FBN	0.91
ZENITHBANK	ACCESSBANK	0.964
DANGCEM	BUACEMENT	0.917
DANGCEM	FBN	0.901
FBN	ACCESSBANK	0.917
FBN	FIDELITYBANK	0.929

Below are the cointegration test results for the top highly correlated stock pairs. All pairs show very strong statistical evidence of cointegration (p-value < 0.05), meaning they maintain a long-term equilibrium relationship despite short-term deviations:

Stock 1	Stock 2	Correlation	Cointegration P-Value	Cointegrated (5%)

GTCO	ACCESSBANK	0.928	0	TRUE
UBA	ZENITHBANK	0.917	0	TRUE
UBA	FBN	0.924	1.38E-26	TRUE
UBA	ACCESSBANK	0.954	0	TRUE
UBA	FIDELITYBANK	0.946	0	TRUE

Step-by-Step Application to Nigerian Stocks (ZENITHBANK vs UBA)

Concept	Pair Trading Application
Cointegration	You test ZENITHBANK and UBA for cointegration using the Engle-Granger or Johansen test. If cointegrated, this means they “belong together” long term.
VECM	You model their relationship using a VECM. When ZENITHBANK and UBA prices drift apart, the VECM identifies this as a “mispricing.”
Signal	If the cointegration residual (spread) gets large, it signals a trade: ➢ Long undervalued stock ➢ Short overvalued stock

Kalman Filter (optional advanced step)	If you believe the hedge ratio between ZENITHBANK and UBA changes over time, use the Kalman Filter to estimate a dynamic beta (instead of fixed β). This helps your strategy adapt in real time.

---

12. REFERENCES

1. Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). *Pairs Trading: Performance of a Relative-Value Arbitrage Rule*. Review of Financial Studies, 19(3), 797–827.

2. Engle, R. F., & Granger, C. W. J. (1987). *Co-integration and Error Correction: Representation, Estimation, and Testing*. Econometrica, 55(2), 251–276.

3. Johansen, S. (1991). *Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models*. Econometrica, 59(6), 1551–1580.

4. Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons.

5. Avellaneda, M., & Lee, J. H. (2010). *Statistical Arbitrage in the US Equities Market*. Quantitative Finance, 10(7), 761–782.

6. Chan, K., Hameed, A., & Tong, W. (2000). *Profitability of Momentum Strategies in the International Equity Markets*. Journal of Financial and Quantitative Analysis, 35(2), 153–172.

7. Arouri, M. E. H., Lahiani, A., & Nguyen, D. K. (2012). *Return and Volatility Transmission between World Oil Prices and Stock Markets of the GCC Countries*. Economic Modelling, 29(4), 1215–1222.

8. Kalman, R. E. (1960). *A New Approach to Linear Filtering and Prediction Problems*. Journal of Basic Engineering, 82(1), 35–45.

9. Eric E. Amaefule (2025). Akoluche Quant Research & Technologies. *Internal Research Archives: Pairs Trading on the Nigerian Exchange – Working Papers and Data Analysis Reports*.

10. TradingView (Accessed 2025). *Nigerian Exchange Data via TradingView Platform*. https://www.tradingview.com/

11. Nigerian Exchange Group (NGX) (Accessed 2025). *Historical Equities Data and Market Bulletins*. https://ngxgroup.com/

12. Python Libraries:

- Statsmodels: Econometric and Statistical Modeling with Python. https://www.statsmodels.org/

- PyKalman: Kalman Filtering and Smoothing in Python. https://pykalman.github.io/