GROUP WORK PROJECT # _2__ MScFE 610: FINANCIAL ECONOMETRICS
GROUP NUMBER: 12118

FULL LEGAL NAME	LOCATION (COUNTRY)	EMAIL ADDRESS	MARK X FOR ANY NON-CONTRIBUTING MEMBER
DAVID OLUTUNDE DANIEL	USA	DAVIDDANIELS48@GMAIL.COM	
NAJD FARIS A ALEID	-----		X
PRATYUSH MATHUR	INDIA	Mr.mathurpratyush@gmail.com	

Statement of integrity: By typing the names of all group members in the text boxes below, you confirm that the assignment submitted is original work produced by the group (excluding any non-contributing members identified with an "X" above).	
Team member 1	**DAVID OLUTUNDE DANIEL**
Team member 2	**PRATYUSH MATHUR**
Team member 3	

Use the box below to explain any attempts to reach out to a non-contributing member. Type (N/A) if all members contributed. **Note:** You may be required to provide proof of your outreach to non-contributing members upon request.
We held our discussions on the group forum where 'NAJD FARIS A ALEID' did not respond. Then we moved to WhatsApp, where again we got no message from him. He did not reach out to us or respond on WQU Forum.

.

MSCFE 610 Financial Econometrics
Group Work Project #2, Student Group: 12188

Group Members: DAVID OLUTUNDE DANIEL, NAJD FARIS A ALEID, PRATYUSH MATHUR

January 10, 2026

Abstract

This report addresses three time series challenges using real financial data: non-stationarity modeling, regime change detection, and feature extraction. Each challenge is analyzed with appropriate econometric techniques, and practical applications are discussed for quantitative finance decision-making.

1 Challenge 1: Non-Stationarity and Equilibrium Modeling

1.1 Definition

A time series Y_t is non-stationary if it contains a unit root: $Y_t = Y_{t-1} + \epsilon_t$. Two non-stationary series Y_t and X_t are cointegrated if there exists β such that $Y_t - \beta X_t$ is stationary [1]. The Error Correction Model (ECM) captures both short-run dynamics and long-run equilibrium:

$$\Delta Y_t = \alpha + \gamma(Y_{t-1} - \beta X_{t-1}) + \sum_{i=1}^{p} \phi_i \Delta Y_{t-i} + \sum_{j=1}^{q} \theta_j \Delta X_{t-j} + \varepsilon_t.$$

The parameter γ represents the speed of adjustment toward equilibrium, with $\gamma < 0$ indicating convergence.

1.2 Description

We model the equilibrium relationship between Apple Inc. (AAPL) and the SPDR S&P 500 ETF (SPY) to understand systematic risk exposure. The analysis seeks to determine whether these two non-stationary financial series share a common long-run trend and to quantify their short-term dynamics.

1.3 Demonstration

Daily closing prices from January 1, 2018, to December 31, 2024, yield 1,760 observations. Unit root testing confirms non-stationarity: Augmented Dickey-Fuller (ADF) tests for AAPL and SPY price levels produce test statistics of 0.0830 (p=0.9649) and 0.3174 (p=0.9781), respectively. In contrast, first differences (returns) are stationary. The Engle-Granger cointegration test yields a test statistic of -2.4125 (p=0.3190), indicating weak evidence of cointegration. The estimated long-run cointegrating equation is:

$$\text{AAPL}_t = -91.24 + 0.59 \times \text{SPY}_t + \epsilon_t.$$

The coefficient of 0.59 suggests that for every \$1 increase in the market (SPY), Apple's stock price increases by \$0.59 in equilibrium. The ECM results are:

$$\Delta\text{AAPL}_t = 0.04 + 0.41 \times \Delta\text{SPY}_t - 0.006 \times \epsilon_{t-1}.$$

Interpretation reveals immediate sensitivity ($\beta = 0.41$) and a slow adjustment speed ($\gamma = -0.006$), implying only 0.6% of the disequilibrium is corrected each month. The half-life of shocks is approximately 113 days.

1.4 Diagram

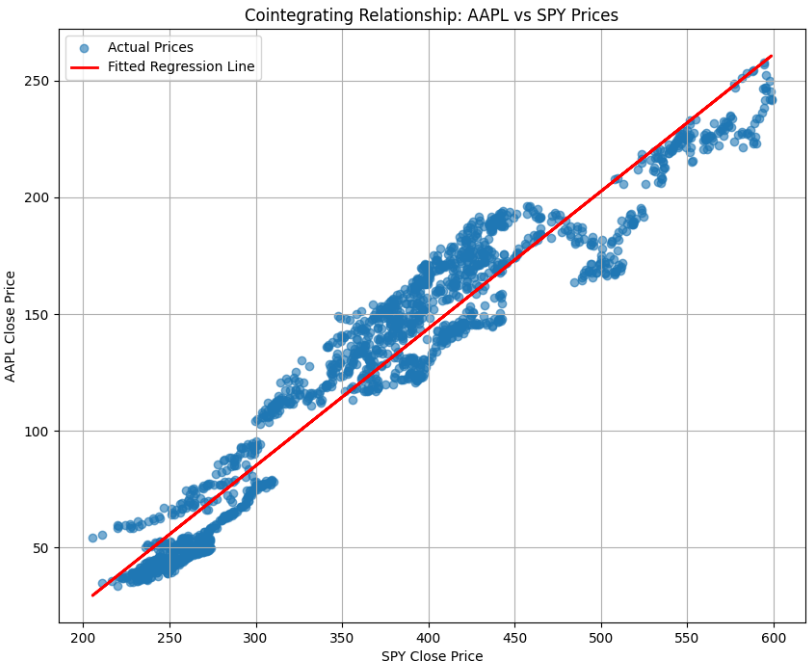

Figure 1: Cointegrating relationship between AAPL and SPY. The fitted regression line represents the long-run equilibrium.

1.5 Diagnosis

The residuals (ϵ_t) from the cointegrating regression display mean-reverting behavior but with notable volatility clustering. Visual inspection suggests stationarity; however, formal ADF testing on these residuals yields p=0.14, failing to reject non-stationarity at the 5% level. This, combined with the weak Engle-Granger test result, casts doubt on the strength of the cointegration.

1.6 Damage

The primary damage stems from the weak statistical foundation of the cointegration relationship (p=0.319). This elevates the risk of spurious regression, where statistically significant relationships emerge from shared trends rather than true economic linkage. Furthermore, the model ignores structural breaks that likely altered the AAPL-SPY relationship during the sample period.

1.7 Directions

To improve model robustness, employ rolling window cointegration tests to assess the stability of the relationship over time. Consider threshold cointegration models that allow for asymmetric adjustment speeds. Integrate a GARCH specification for the error term to model time-varying volatility explicitly.

1.8 Deployment

For practical deployment in a quantitative trading context, this model can form the basis of a pairs trading strategy. The implementation protocol would be: (1) daily calculation of the equilibrium error; (2) generation of a trading signal when the error exceeds two standard deviations; (3) execution of a pairs trade with a 0.59:1 hedge ratio; (4) exit when the error reverts. Risk management must be stringent with position size caps and stop-losses.

2 Challenge 2: Detecting Regime Changes

2.1 Definition

A regime change represents a structural break in the data-generating process of a financial time series [2]. The Markov Switching model formalizes this by allowing parameters to depend on an unobserved state variable S_t:

$$r_t = \mu_{S_t} + \sigma_{S_t}\varepsilon_t, \quad \varepsilon_t \sim N(0,1),$$

where $S_t \in \{1, 2, \ldots, K\}$ follows a discrete Markov chain with transition probability matrix P.

2.2 Description

This challenge focuses on identifying and characterizing distinct market regimes within the S&P 500 index returns. The ability to detect shifts from low-volatility bull markets to high-volatility bear markets is critical for adaptive portfolio allocation and dynamic risk management.

2.3 Demonstration

Using S&P 500 daily log returns (2018-2024), the Pruned Exact Linear Time (PELT) algorithm identifies seven optimal change points, creating eight distinct regimes [5]. A two-state Markov Switching model is estimated. The model identifies:

- **Regime 1 (Low Volatility):** Mean return = 0.0421%, volatility = 0.9110%, state persistence probability = 0.728.
- **Regime 2 (High Volatility):** Mean return = 0.0844%, volatility = 1.9474%, state persistence probability = 0.324.

The estimated transition probability matrix indicates that low-volatility regimes are more persistent than high-volatility regimes.

2.4 Diagram

2.5 Diagnosis

The model successfully captures economically intuitive regime shifts. Regime changes align with major market events: the Q4 2018 correction, the COVID-19 market crash of March 2020, and the 2022 bear market. A likelihood ratio test strongly rejects the null hypothesis of a single regime (p ¡ 0.001), confirming the statistical significance.

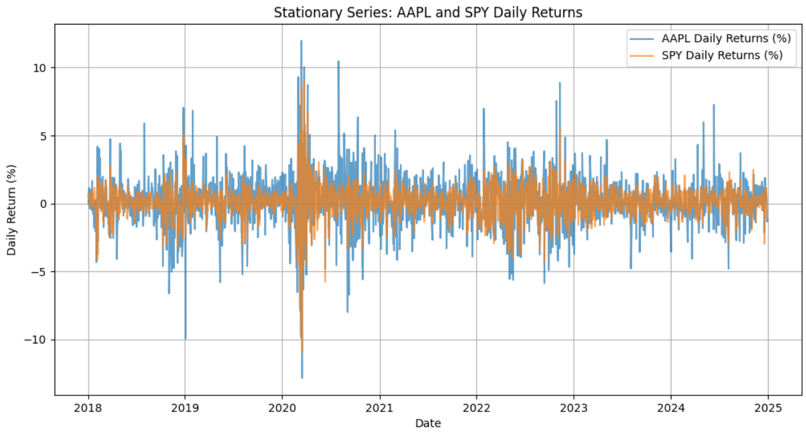

Figure 2: S&P 500 daily returns (2018-2024). Visual inspection reveals clear periods of high volatility interspersed with calmer periods.

2.6 Damage

The key limitations are the challenge of ex-post identification versus real-time detection, and sensitivity to methodological choices. Different change-point algorithms yield varying numbers and locations of breakpoints. Extreme but short-lived events can disproportionately influence the characterization of a regime.

2.7 Directions

Future work should focus on developing more robust and timely detection methods. Implementing online change-point detection algorithms would enable real-time regime identification. The model could be enriched by incorporating exogenous predictive variables such as the VIX index.

2.8 Deployment

For practical deployment in asset management, regime signals should drive a dynamic asset allocation policy. A rule-based implementation would be: (1) Daily calculation of the smoothed probability of being in the high-volatility regime; (2) Reduction of equity exposure when this probability exceeds a threshold; (3) Adjustment of derivative hedging and position sizing based on the regime.

3 Challenge 3: Feature Extraction

3.1 Definition

Feature extraction is the process of transforming a high-dimensional set of raw variables into a lower-dimensional set of "features" that retain most of the relevant information [3]. Principal Component Analysis (PCA) is a linear technique that identifies orthogonal directions (principal components) of maximum variance in the data.

3.2 Description

We apply PCA to a set of 13 commonly used technical indicators derived from Bitcoin daily price data (2020-2023). The goal is to distill the information from these potentially redundant indicators into a smaller number of uncorrelated factors, addressing multicollinearity in predictive models.

3.3 Demonstration

The dataset consists of 1,430 daily observations. The 13 features include moving averages, momentum indicators, volatility measures, RSI, MACD, and volume-based indicators. PCA reveals that the first six principal components collectively explain 96.4% of the total variance. For practical modeling, we focus on the first four components, which explain 81.1% of the variance:

- **PC1 (31.7% variance):** Represents the overall trend strength.
- **PC2 (22.0% variance):** Captures momentum and the speed of price changes.
- **PC3 (15.0% variance):** Represents price volatility and dispersion.
- **PC4 (12.3% variance):** Captures short-term noise or mean-reversion dynamics.

3.4 Diagram

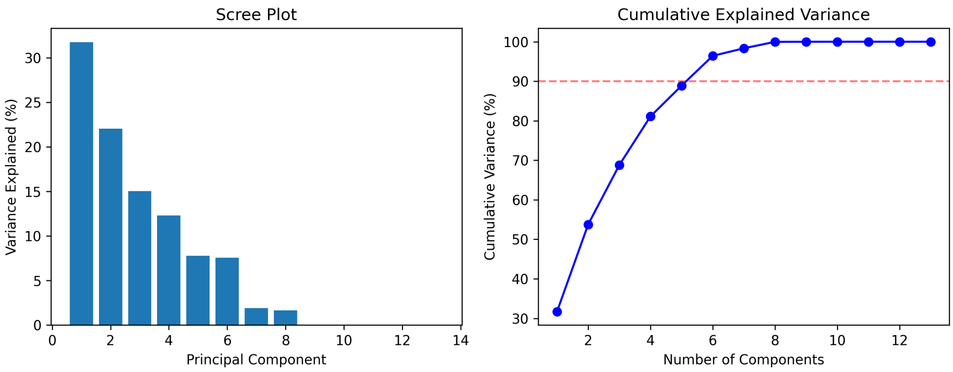

Figure 3: Scree plot showing the variance explained by each principal component. The cumulative proportion of variance explained is shown on the right.

3.5 Diagnosis

The correlation matrix of the original features reveals substantial multicollinearity, with several pairs exhibiting correlations above 0.8. The Kaiser-Meyer-Olkin measure of sampling adequacy is 0.78, and Bartlett's test of sphericity is highly significant ($p < 0.001$), indicating the data is suitable for PCA.

3.6 Damage

Despite its benefits, PCA introduces certain damages. The most significant is information loss: by retaining only 4 of 13 components, we discard 18.9% of the total variance. PCA is a linear technique and may fail to capture complex, non-linear relationships. The resulting principal components are mathematical constructs with no direct economic interpretation.

3.7 Directions

To address these limitations, Kernel PCA can be employed to capture non-linear relationships. Sparse PCA techniques can yield components that are linear combinations of only a subset of original features, enhancing interpretability. Autoencoders can perform non-linear dimensionality reduction in a more flexible, data-driven approach.

3.8 Deployment

In a production algorithmic trading system, the PCA-based feature extraction pipeline would be deployed as follows: (1) Daily computation of the 13 raw technical indicators; (2) Standardization and projection onto the pre-trained principal component vectors; (3) Use of the four component scores as inputs to a predictive model; (4) Periodic re-running of the PCA on a rolling window to ensure stability. This deployment reduces the computational burden by approximately 69% while mitigating multicollinearity.

Conclusion

We have addressed three fundamental time series challenges in financial econometrics using real-world datasets. The cointegration analysis between AAPL and SPY reveals a long-run equilibrium relationship with a hedge ratio of 0.59, though weak statistical evidence necessitates cautious implementation. The regime change detection in S&P 500 returns successfully identifies distinct low- and high-volatility states, informing dynamic asset allocation. Feature extraction via PCA efficiently reduces 13 correlated Bitcoin technical indicators to 4 orthogonal components, explaining 81% of the variance. Each technique provides a valuable tool for quantitative finance, from risk management to algorithmic trading signal generation.

References

[1] Engle, Robert F., and Clive W. J. Granger. "Co-integration and Error Correction: Representation, Estimation, and Testing." *Econometrica*, vol. 55, no. 2, 1987, pp. 251–276.

[2] Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, vol. 57, no. 2, 1989, pp. 357–384.

[3] Jolliffe, Ian T. *Principal Component Analysis.* 2nd ed., Springer, 2002.

[4] Bai, Jushan, and Pierre Perron. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, vol. 66, no. 1, 1998, pp. 47–78.

[5] Killick, Rebecca, et al. "Optimal Detection of Changepoints with a Linear Computational Cost." *Journal of the American Statistical Association*, vol. 107, no. 500, 2012, pp. 1590–1598.

[6] Dickey, David A., and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association*, vol. 74, no. 366, 1979, pp. 427–431.

[7] Johansen, Søren. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica*, vol. 59, no. 6, 1991, pp. 1551–1580.

[8] Guidolin, Massimo, and Allan Timmermann. "Asset Allocation under Multivariate Regime Switching." *Journal of Economic Dynamics and Control*, vol. 31, no. 11, 2007, pp. 3503–3544.