Revised Project Proposal: A Time Series Approach to Pair Trading on the Indian Stock Market

Course: DA341 — Applied Time Series Analysis

Prepared by:
Aryan Gupta (Roll Number: 230150003)
Billa Cherish (Roll Number: 230150007)
Vibha Gupta (Roll Number: 230150029)

Abstract

We propose a market-neutral pair trading strategy for highly liquid stocks on the National Stock Exchange (NSE) of India. The project identifies cointegrated pairs, and we compare the performance of single-series forecasting (applying ARIMA independently to each stock) against several pairs trading strategies utilizing **Z-score signals**: Spread using Ordinary Least Squares (OLS), Spread using **Kalman Filter** Hedge Ratios, and a **Vector Error Correction Model** (VECM) based approach. The core novelty lies in comparing the robustness and profitability—measured through Net Profit and Loss and Compound Annual Growth Rate—of the **single-series approach versus pairs trading strategies**, incorporating a rigorous transaction cost model.

1 Introduction

Pair trading exploits temporary divergence between two historically co-moving securities by trading their spread (price difference) rather than betting on market direction. By simultaneously maintaining a long position in one stock and a short position in the other (at the hedge ratio, η), the strategy remains market-neutral and aims to capture mean reversion. The spread is defined as:

$$\text{Spread}_t = \text{Price}_Y - (\eta \times \text{Price}_X)$$

2 Research Questions

1. Which NSE stock pairs exhibit a strong, stable, and statistically significant long-term relationship (**cointegration**) suitable for pair trading?
2. Does a dynamic hedge ratio (**Kalman Filter**) yield superior out-of-sample performance compared to a static hedge (**Ordinary Least Squares**)?
3. Does defining the spread using the cointegrating vector from VECM lead to different performance compared to Ordinary Least Squares or Kalman Filter approaches when using Z-score signals?

3 Data and Scope

Data Source Historical daily closing prices from the National Stock Exchange (NSE), accessed using Python libraries such as `yfinance`.

Period 5 to 10 years of daily data to capture multiple market regimes.

Universe Top ∼100 most liquid and frequently traded stocks to minimize slippage and ensure execution feasibility.

Candidate Pairs
Sector-based combinations such as TCS–INFY (IT) and HDFCBANK–ICICIBANK (Banking), as they share similar business models and exposure to common industry risks.

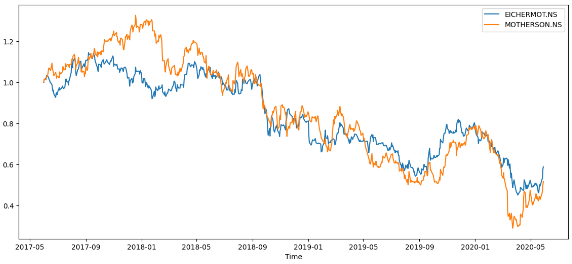

Figure 1: *

Relative closing prices of two highly correlated Indian automobile stocks.

4 Choice of Technology: Python

Python is selected for its robust, integrated, and industry-standard ecosystem, allowing an automated and reproducible workflow from data acquisition to backtesting. Libraries include:

- `pandas` and `numpy` for data handling,
- `statsmodels` for econometric modeling (cointegration, VECM),
- `pykalman` for dynamic parameter estimation (Kalman Filter),
- `matplotlib` and `seaborn` for visualization.

This ensures reproducibility, scalability, and alignment with industry practices in quantitative research.

5 Methodology and Approach

A. Identifying Potential Pairs

- Apply liquidity filters and rolling correlation screens to shortlist co-moving stocks.
- Perform the **Engle–Granger Cointegration Test** (Ordinary Least Squares regression followed by Augmented Dickey–Fuller test on residuals) to confirm a stable long-run relationship, forming the statistical foundation for mean reversion.

B. Modelling and Trading Signal Generation

We employ a phased, comparative approach using progressively advanced time series models:

- **Phase 1: Univariate Benchmark (ARIMA/SARIMA)** — Apply to individual stock series to establish a baseline for forecasting accuracy.
- **Phase 2: Static Benchmark (Ordinary Least Squares)** — Estimate a constant hedge ratio (η) using Ordinary Least Squares (OLS). Normalize the spread into a Z-score ($Z_t = (\text{Spread}_t - \mu)/\sigma$) and generate entry/exit signals based on Z-score thresholds.
- **Phase 3: Dynamic Model (Kalman Filter)** — Use the Kalman Filter to recursively estimate the time-varying hedge ratio ($\hat{\eta}_t$), maintaining maximal stationarity in the spread and feeding into the Z-score-based signal generation.
- **Phase 4: Multivariate System Analysis (VECM)** — Model the joint dynamics of both stock prices to validate the cointegrating vector and analyze long-term equilibrium.

6 Performance Evaluation

All backtests will use a rolling (walk-forward) methodology to ensure robustness and prevent lookahead bias. Results will be reported as both **Gross Returns** (before costs) and **Net Returns** (after transaction costs).

A. Core Financial Metrics

We will report cumulative Profit and Loss (P&L), Compound Annual Growth Rate (CAGR), annualized Sharpe Ratio, and Maximum Drawdown (MDD):

- **Net Profit and Loss (Net P&L):** Net P&L = Gross P&L − $\sum$(Transaction Costs), represents total profit after deducting all trading costs.
- **Compound Annual Growth Rate (CAGR):** CAGR = (End Value/Start Value)$^{1/\text{Years}}$ − 1, represents the average annual growth rate of the portfolio.
- **Annualized Sharpe Ratio (SR):** $SR = \frac{E[R_p - R_f]}{\sigma_p}\sqrt{T}$, measures risk-adjusted excess return ($E[R_p - R_f]$) per unit of risk (σ_p, portfolio return std dev), annualized using $T \approx 252$.
- **Maximum Drawdown (MDD):** $MDD = \max_{0 \le t < \tau \le T}((PV_t - PV_\tau)/PV_t)$, measures the largest peak (PV_t) to trough (PV_τ) percentage decline in portfolio value.

B. Transaction Cost Modeling

We adopt a **Combined Fixed–Variable Transaction Cost Model**:

- **Fixed Costs:** Brokerage + statutory fees (STT, etc.) modeled as a percentage $C_{fixed} \approx 0.05\%$ of turnover.
- **Variable Costs:** Slippage/impact estimated as $\alpha \times$ Transaction Value, where α is estimated empirically.
- **Total Cost:** Applied per trade side (buy/sell), Total Cost = $(C_{fixed} + \alpha) \times$ Transaction Value.

Key Reference

- Palomar, D. P. *Pairs Trading.* Lecture Slides, MAFS5310, Hong Kong University of Science and Technology. Available: `https://palomar.home.ece.ust.hk/MAFS5310_lectures/slides_pairs_trading.pdf`