Thesis Proposal: Forecasting JSE Equity Volatility (ML vs. GARCH/HAR Models)

Liam Andrew Beattie[a]

[a] *Stellenbosch University, South Africa*

Abstract

Compare the performance of Machine Learning (ML) models (e.g., LSTM, XGBoost) and traditional econometric models (GARCH, HAR) in forecasting equity volatility on the Johannesburg Stock Exchange (JSE), contextualized to South Africa's unique market dynamics (e.g., commodity dependence, political risk).

1. Introduction

Here are the papers I intend to read and give a brief summary of them:

Gunnarsson, Isern, Kaloudis, Risstad, Vigdel & Westgaard (2024)

Notes that ML and AI methods rank high and on par with econometric predicitons of realised volatility. Using explainable AI or creating explainable AI might be interesting in the context of my paper.

This paper looks at realised volatility which is inferreed from the sum of square intradaily hihg-frequency returns.

> We find that, going forward, the use of XAI and implied volatility forecasting forecasting proves an interesting direction for further research, as well as combination of the promsing performing models and Bayesian approaches to quanitify the uncertainty introduced by models to explain trustworthiness of predictions.

The top three testing messures found in Gunnarsson *et al.* (2024) to compare models and forecasting was - Diebold-Mariano - Clark-West - Model Confidence Set

On a slight tangent, this paper references another paper by Bouri et al. (2020) who used google search volume intensity associated with the US-China trade war to capture uncertainty aspects. In terms of

Email address: `22562435@sun.ac.za` (Liam Andrew Beattie)

data input, if we could get access to this, would be extremely valuable to add some google data to my research.

Horvath, Muguruza & Tomas (2021) Petrozziello, Troiano, Serra, Jordanov, Storti, Tagliaferri & Rocca (2022) Wu, Cheng, Jankovic & Kolanovic (2024) Wu, Jankovic, Kaplan & Lee (2025) Zeng & Klabjan (2019) Zhang, Zhang, Cucuringu & Qian (2023)

Zhang *et al.* (2023)

Compare the performance of **Machine Learning (ML)** models (e.g., LSTM, XGBoost) and **traditional econometric models** (GARCH, HAR) in forecasting equity volatility on the Johannesburg Stock Exchange (JSE), contextualized to South Africa's unique market dynamics (e.g., commodity dependence, political risk).

2. Key Steps

2.1. Data Collection & Preparation

- **Data Sources**:
  - **JSE Equity Data**: Daily/index returns (e.g., JSE Top 40) from Bloomberg, Yahoo Finance, or `Quantmod` in R.
  - **Volatility Proxy**: Realized volatility (using intraday data) or daily squared returns.
  - **Exogenous Variables**:
    * Commodity prices (platinum, gold) via `Quandl`.
    * SA-specific risks (Eskom load-shedding schedules, political risk indices).
    * Global factors (USD/ZAR exchange rate, Fed rates).
- **Preprocessing**:
  - Handle missing data, outliers, and structural breaks (e.g., COVID-19, 2021 riots).
  - Normalize/standardize features for ML models.

2.2. Model Implementation

- **Traditional Models**:
  - **GARCH**: Use `rugarch` in R to model volatility clustering (e.g., EGARCH for asymmetry).
  - **HAR-RV**: Implement Heterogeneous Autoregressive model with `highfrequency` package.
- **ML Models**:
  - **LSTM/GRU**: Sequence-based models using `keras/tensorflow`.
  - **Tree-Based**: XGBoost (`xgboost`), Random Forest (`randomForest`).
  - **Hybrids**: Combine HAR residuals with ML predictions.

2.3. Feature Engineering

- **Lag Features**: Historical volatility, returns, trading volume.
- **SA-Specific Features**:
  - Load-shedding stages (scraped via `rvest`).
  - SA sovereign credit rating changes.
  - Commodity price shocks (platinum/gold).

2.4. Training & Validation

- **Train-Test Split**: Time-series cross-validation (e.g., rolling window).
- **Hyperparameter Tuning**: Grid search for ML models (e.g., `mlr3` in R).

2.5. Evaluation

- **Metrics**:
  - Statistical: MSE, MAE, QLIKE.
  - Economic: Value-at-Risk (VaR) backtesting, Sharpe ratio of volatility-based strategies.
- **Statistical Tests**:
  - Diebold-Mariano test for forecast superiority.
  - Model Confidence Set (MCS) to rank models.

2.6. Interpretation & Context

- **XAI**: Use SHAP values (`DALEX`) to interpret ML drivers (e.g., "How much do platinum prices impact mining stock forecasts?").
- **Comparative Analysis**: Contrast results with global markets (e.g., "Why does XGBoost underperform in SA vs. the S&P 500?").

3. Deliverables

1. **Visualizations**:
   - Time-series plots of volatility with crisis annotations.
   - SHAP summary plots for feature importance.
   - Heatmaps of cross-correlations between variables.
2. **Code**: Reproducible R scripts for data, models, and visuals.
3. **Policy Insights**: Practical recommendations for SA investors/policymakers (e.g., hedging strategies during load-shedding).

4. Challenges & Solutions

- **Data Scarcity**: Use alternative data (news sentiment, Eskom reports) to augment limited JSE history.

- **Overfitting**: Regularization (L1/L2 for ML) and parsimony checks for GARCH.

- **Computational Cost**: Optimize code with `data.table` or cloud computing (AWS/GCP).

5. Why This Matters

- **Emerging Markets Gap**: Most volatility studies focus on developed markets; SA's unique risks (commodities, politics) offer fresh insights.

- **Practical Value**: Helps SA investors manage risk in a turbulent market.

Gunnarsson, E.S., Isern, H.R., Kaloudis, A., Risstad, M., Vigdel, B. & Westgaard, S. 2024. Prediction of realized volatility and implied volatility indices using AI and machine learning: A review. *International Review of Financial Analysis*. 93:103221.

Horvath, B., Muguruza, A. & Tomas, M. 2021. Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models. *Quantitative Finance*. 21(1):11–27.

Petrozziello, A., Troiano, L., Serra, A., Jordanov, I., Storti, G., Tagliaferri, R. & Rocca, M.L. 2022. Deep learning for volatility forecasting in asset management. *Soft Computing*. 26(17):8553–8574.

Wu, E., Cheng, P., Jankovic, L. & Kolanovic, M. 2024. *Investable AI for volatility trading deep learning model for cross asset volatility strategies*. (Research Report). J.P. Morgan Global Quantitative & Derivatives Strategy. [Online], Available: httpswww.jpmorganmarkets.com.

Wu, E., Jankovic, L., Kaplan, B. & Lee, T.S. 2025. *Cross asset volatility: Machine learning based trade recommendations*. (Research Report). J.P. Morgan Global Markets Strategy. [Online], Available: https://www.jpmorganmarkets.com.

Zeng, Y. & Klabjan, D. 2019. Online adaptive machine learning based algorithm for implied volatility surface modeling. *Knowledge-Based Systems*. 163:376–391.

Zhang, C., Zhang, Y., Cucuringu, M. & Qian, Z. 2023. Volatility forecasting with machine learning and intraday commonality. *Journal of Financial Econometrics*. 22(2):492–530.