Entanglement, Hedging, and Predictability of the ASX-AUD/USD Relationship: An Integrated Econometric and Machine Learning Analysis

Student Name: [YU-WEI LIN] Student ID: [1537312]

1 Introduction

Driven by a background in finance and interest in market dynamics, This study explores the crucial relationship between the Australian Securities Exchange (ASX200) and the AUD/USD exchange rate, which is essential for Australia's export-driven economy. Using an integrated statistical and machine learning (ML) approach, we examine their entanglement, hedging ratios, and predictability of next-day returns. This study investigates whether econometric insights can improve ML forecasting and whether ML models, when combined with statistical measures such as Granger causality and MVHR, can predict returns. For this particular market pair, this fills a gap in the literature by integrating ML forecasting with hedging evaluation. Over 5000 daily ASX200 and AUD/USD observations from Yahoo Finance `https://finance.yahoo.com/` covering 20 years are used in the analysis.

2 Literature Review

2.1 Theoretical Background and Research Motivation: Market Interlinkages & Hedging Importance

The interplay between exchange rates and stock markets is essential for global investors and policymakers to understand. In Australia, the commodity export dependence and sensitivity to international capital flows make the ASX and the AUD/USD exchange rate economic variables in sync. Little is known about their time-varying comovements and hedging implications when forecast modelling techniques that are viewed as advanced are taken into account.

2.2 Traditional Models for Market Interlinkage and Hedging: GARCH & MVHR

Narayan et al. (2016) showed significant transmission effects, especially through times of economic crises, in the context of volatility spillovers between the Australian equity market and the currency market, using asymmetric GARCH models. Nguyen and Bellalah (2008) also calculated hedge ratios for international markets such as the AUD/USD; the methods used were multivariate GARCH and the Minimum Variance Hedge Ratio (MVHR). Their results further show changes over time in the effectiveness of hedges.

Limitation: The linear or parametric assumptions that underlie these econometric models may hinder their ability to capture complex non-linear relationships and sudden structural shifts in financial markets.

2.3 Applications of Machine Learning in Financial Forecasting

Machine learning models such as Ridge Regression, Support Vector Regression (SVR), and Gradient Boosting Machines (GBMs) have been increasingly applied in asset return prediction studies (e.g., Lahmiri (2019); Gu et al. (2020)). In particular, Gu et al. (2020) demonstrated that ML-based models consistently outperform conventional statistical models in forecasting cross-sectional returns across various asset classes.

Gap: While these studies work to better forecast returns, however, they completely ignore the evaluation of hedging strategies; further, they do not consider the Australian market or the relation between the ASX and AUD/USD exchange rate.

This study combines in a unique way:

Predictive Modelling and Entanglement Analysis: Granger Causality (Granger) and **Vector Autoregression** (VAR) models are used to analyse lag relationships and dynamic interdependencies, while machine learning models (Ridge, SVR, and GBMs) are used to forecast returns. The feature importance of ML models is used to further detail market entanglement.

Integrated Hedging Strategy Evaluation: Estimating the Minimum Variance Hedge Ratio (MVHR) and investigating how information from time-series econometrics (such as VAR) and predictive models can contribute to a more thorough comprehension of the ASX's ability to hedge against changes in the AUD/USD exchange rate.

3 Methodology

This study is a combination of statistical analysis and machine learning techniques to empirically investigate the log returns of the stock market in Australia (ASX All Ordinaries Index) and of the AUD/USD exchange rate. Specifically, the study tries to mainly investigate the hedge relationships and future return entanglement as well as predict their joint log returns for the next day.

3.1 Data Source and Preprocessing

For this purpose, the study uses one dataset (`merged_data_with_target.csv`) created using historical daily price data (Open, High, Low, Close, and Volume) for the two assets-wild-type variables-as AUD/USD and ASX200, from Yahoo Finance. Initial preprocessing included date parsing, setting up a common date index, and numeric consistency.

Utilizing `pandas_ta` library, generating a comprehensive set of over 100 technical indicators for each series were generated-the moving and exponential moving averages, RSI, MACD, stochastics, Bollinger Bands, ATR, and so forth. These distinct feature-rich datasets were then merged on the date attribute.

To keep the data clean, the columns that contained NaN values above a (threshold: 40%) originated from the indicator calculation were discarded. Also, if there were any NaNs remaining, they were deleted row-wise.

3.2 Analytical Approaches

3.2.1 Statistical and Econometric Analysis

- **Granger Causality tests**: (Up to 5 lags) in investigating a daily log return series bidirectional lead-lag predictability.
- **The Minimum Variance Hedge Ratio (MVHR)**: The ratio is calculated via OLS regression of the ASX returns with the daily returns of the AUD/USD, thus giving a static hedge ratio β and a hedge effectiveness R^2.
- Data is prepared for potential **Vector Autoregression (VAR)** analysis to explore joint dynamics.

3.2.2 Machine Learning Pipeline (for Next-Day Log Return Prediction)

- **Models**: Ridge Regression, Random Forest Regressor, Gradient Boosting Regressor, and Support Vector Regression (SVR with linear and RBF kernels) are considered.
- **Process**: Each model uses a pipeline constructed with `SelectKBest` (f_regression, k as a hyperparameter) and `StandardScaler` for (feature scaling).
- **Validation & Tuning**: `GridSearchCV` searches for hyperparameters minimising negative RMSE, using nested `TimeSeriesSplit` (5 outer folds, 3 inner folds).
- **Metrics**: R^2, MAE, and RMSE are used to assess the model's performance.
- **Entanglement Insight**: To comprehend inter-market influences, cross-market feature importances (from RF/GBM) and coefficients (from Ridge/linear SVR) are examined.

4 Results and Discussion

This study aims to forecast the next-day log returns of the Australian stock market (ASX All Ordinaries Index) and the AUD/USD exchange rate by first examining the predictive interdependencies and hedging relationship between the two. This is accomplished by using an integrated methodology that combines machine learning models with statistical techniques. Furthermore, by contrasting modern machine learning models with conventional econometric techniques, the study seeks to clarify how well these disparate methodologies can capture non-linear factors.

For detailed empirical data and references, please visit `https://github.com/I-ueRya-n/ML-project/tree/main/Machine%20learning%20analysis%20data`

4.1 Market Entanglement Analysis

Numerous analytical techniques consistently show that the ASX and AUD/USD markets interact significantly and dynamically.

- **Granger Causality tests** revealed showed compelling evidence of a Granger-causal relationship in both directions between the daily log returns of the AUD/USD and ASX. The p-values were highly significant ($p \leq 0.001$) (lesser than or equal to 0.001) for all the considered lags (1 to 5 days). This, from a linear, time-lagged point of view, supports the notions of inter-market predictability and inter-twinning, by implying that price movements in the past in one market statistically possess information for forecasting future short-term movements in the other.
- The **Vector Autoregression (VAR) model** (using 10 lags as selected by AIC) provided a more detailed view of these dynamic interdependencies.
  - In explaining the current ASX returns, lagged AUD/USD returns at several horizons were statistically significant. For instance, L1.AUDUSD_Daily_Log_Return (coefficient: 0.2597, $p = 0.000$), L2.AUDUSD_Daily_Log_Return (coefficient: 0.0498, $p = 0.011$), and L5.AUDUSD_Daily_Log_Return (coefficient: -0.0527, $p = 0.007$) had significant effects. ASX's own past returns, such as L1.ASX_Daily_Log_Return ($p = 0.000$) and L5.ASX_Daily_Log_Return ($p = 0.010$), were also significant.
  - Lagged ASX returns were also significant in the equation for current AUD/USD returns, particularly L1.ASX_Daily_Log_Return (coefficient: 0.0394, $p = 0.000$) and L10.ASX_Daily_Log_Return (coefficient: 0.0265, $p = 0.015$). Historical returns of the AUD/USD at different lags (L1, L5, L6, L7) proved to be significant, as well.

- Required residual contemporaneous correlation between the two VAR equations was estimated at around 0.267, indicating a simultaneous impact of unobserved common shocks on both markets beyond the lagged indicators.
- These VAR results provide a more comprehensive, system-wide view of the dynamic, multi-lagged character of the entanglement and support the Granger causality findings.

- **Machine Learning (ML) models' feature importance and coefficient analysis** (from `all_model_performance_summary_integrated_plus_gbm.csv` and `all_model_fold_details_integrated_plus_gbm.txt`) further clarified this entanglement, especially when viewed from a non-linear angle:
  - **Predicting ASX Returns:** Ridge, RandomForest, and Gradient Boosting models consistently ranked AUD/USD features as important when predicting ASX. For instance, features such as `audusd_HLC_CCI_20`, `audusd_Open`, and `audusd_Close_ROC_5` were highlighted by the aggregated Ridge model coefficients (`top_10_coeffs_Ridge_agg`). AUD/USD features such as `audusd_Close_ROC_60`, `audusd_VTXP_14`, and `audusd_ICS_26` were found to be highly influential by RandomForest (`top_10_features_RandomForestRegressor_agg`) and GBM (`top_10_features_GradientBoostingRegressor_agg`).
  - **Predicting AUD/USD Returns:** Similarly, ASX-related features like `asx_Close_BBB_50_2.5` (Ridge), `asx_Close_ROC_5` (RandomForest, GBM), and `asx_HLC_ATR_7` (RandomForest) were considered significant by the corresponding models when predicting AUD/USD.
  - **Significance:** Even when overall return predictability was low (as discussed below), these ML models consistently selected cross-market features, which strongly suggests that the models value information from one market when attempting to comprehend the dynamics of the other. In addition to supporting the "entanglement" hypothesis, this machine learning-based evidence may be able to capture more intricate, non-linear interactions than the linear Granger or VAR models.

4.2 Hedging Relationship and Effectiveness

- A static hedge ratio (β) of roughly 0.3295 was obtained from the **Minimum Variance Hedge Ratio (MVHR)** analysis using OLS regression (`mvhr_results.txt`). With a p-value of 0.000, this coefficient was statistically significant.
- About 0.061 (or 0.0606) of the variance in ASX daily returns can be explained linearly by AUD/USD daily returns, according to the regression's **R-Squared** of 0.061 (or 0.0606). This implies that a basic static linear hedge has a **limited but statistically significant hedge effectiveness**.
- The **MVHR's positive sign** is significant. It suggests that, on average, during the sample period, the daily returns of the ASX and AUD/USD moved in the same direction. According to this model, a short position in AUD/USD would be recommended to achieve the least amount of variance when hedging a long ASX position. This calls into question the idea that the AUD/USD exchange rate serves as a straightforward negative correlator hedge.
- OLS residual diagnostics revealed fat tails and non-normality, which are typical in financial data and should be taken into account when evaluating the accuracy of OLS estimates.

4.3 Challenges in Future Daily Return Prediction with Machine Learning

`all_model_performance_summary_integrated_plus_gbm.csv` describes the difficult outcomes of the attempt to forecast next-day log returns using a suite of ML models (Ridge, SVRs, RandomForest, and GBM).

- For predicting ASX returns, all models yielded negative average R^2 values, suggesting that their performance outside of the sample was inferior to that of a naive mean forecast.
- For predicting AUD/USD returns, the Ridge model was the only one to achieve a slightly positive average R^2 (0.0829); the other models all had negative R^2 values.

Learning curve analysis (from the `learning_curves` files) provided further insights:

- **SVR models** generally substandard learning habits for both targets.
- **RandomForest and Gradient Boosting models** exhibited overfitting (high training scores but slightly lower and more erratic cross-validation scores) for both the ASX and the AUD/USD predictions.
- **Ridge Regression** presented the most consistent learning trends, especially for the prediction of the AUD/USD- the cross-validation score maintained an upward trajectory and came near to the training score.

Conclusion: This implies that one of the most demanding tasks is the relative accuracy of high-frequency (daily) return prediction, even with a complete set of technical indicators and solid machine learning methods, including hyperparameter tuning and nested cross-validation. Both the inherent noise and market efficiency probably stood tall.

4.4 Discussion: Methodological Evaluation, Market Insights, and Research Value

This multifaceted research encompasses the main perspectives of the assumptions:

- **Market Entanglement:** The hypothesis of deep entanglement between the ASX and AUD/USD markets is robustly supported by the combined evidence of significant bidirectional Granger causality, dynamic interactions shown by VAR model coefficients (significant cross-lags), and the consistent selection of cross-market features by multiple ML models (Ridge, RF, GBM). The limitations of conventional linear models mentioned in the literature review are addressed by the ML approaches, which provide insights into possibly non-linear aspects of this relationship.
- **Hedging Relationship:** The MVHR analysis provided the limited efficacy of a quantitative static hedge ratio. It would be difficult to use these particular return forecasts directly to inform dynamic hedging strategies (a novel aspect identified in the literature review) given the generally poor performance of ML models in return prediction. Nonetheless, qualitative aspects of risk management may be indirectly informed by the feature importance analysis, which determines which AUD/USD features have an impact on ASX.
- **Reflection on Machine Learning Concepts:** These demonstrate the training of different ML algorithms and reflections upon fundamental machine learning concepts through illustration of nested time-series cross-validation; feature selection (SelectKBest); investigations of feature importances/coefficient values, seemingly systematic parameter tuning (GridSearchCV); and learning curve interpretation. Secondly, this reflection looks at some of the challenges encountered in prediction.
- **Addressing Literature Gaps & Innovation:** The research contributes by concentrating on feature-driven entanglement in the analysis of the ASX/AUD-USD pair using a variety of machine learning techniques. Although the low forecast accuracy made it challenging to directly integrate ML return forecasts into quantitative hedging, the study establishes the foundation and identifies obstacles for this innovative strategy. For this market pair, non-linear models go beyond conventional linear analyses by identifying particular cross-market technical indicators that they consider significant.
- **Rationale for Machine Learning Algorithm Selection: Ridge Regression** was preferred over the more straightforward **Ordinary Least Squares (OLS)** owing to its capability of producing coefficient stability and reducing the possibility of overfitting, common in financial datasets with the plurality of potentially multicollinear technical indicators. By a similar logic, single Decision Trees are interpretable and capture nonlinearities but possess limited generalization power because of that same propensity for overfitting. Being contrary to that, **Random-ForestRegressor** and **GradientBoostingRegressor** ensemble methods were utilized. These model actual nonlinearities and feature interactions occurring in financial data, combining many trees to reduce variance and increase stability and predictive power.
- **Reflection on Hyperparameter Tuning: GridSearchCV** was used to automate hyperparameter tuning (HPT) for all machine learning models (Ridge, SVRs, RandomForest, and GBM) within the **nested TimeSeriesSplit** framework (5 outer, 3 inner folds), optimising against negative RMSE. The regularisation strength **(Ridge's alpha), SVR's C and gamma, tree complexity for RF/GBM (e.g., `n_estimators`, `max_depth`)**, and the number of selected features (k for **SelectKBest**) were all adjusted during this process. As per standard ML practice, HPT sought to maximise each model's potential based on pre-defined grids (e.g., Ridge alpha from 0.1 to 100; k from 50 to 150). However, a thorough examination of whether optimal parameters frequently hit grid boundaries would be required to confirm the sufficiency of search ranges for future work (Honorio, 30). The inherent challenge of daily financial return forecasting severely constrained overall predictive gains in spite of this methodical tuning, demonstrating that HPT, although essential, might not be able to overcome basic data or task complexities.

5 Conclusion and Future Outlook

Both statistical tests and machine learning feature analysis demonstrate the strong bidirectional dynamic entanglement between the ASX and AUD/USD markets. The positive hedge ratio indicates a co-movement tendency, even though static linear hedging with AUD/USD for ASX offers limited effectiveness. These demonstrate the training of different ML algorithms and reflections upon fundamental machine learning concepts through illustration of nested time-series cross-validation; feature selection (SelectKBest); investigations of feature importances/coefficient values, seemingly systematic parameter tuning (GridSearchCV); and learning curve interpretation. Secondly, this reflection looks at some of the challenges encountered in prediction.

References

- Narayan, P. K., Sharma, S. S., & Thuraisamy, K. (2016). An analysis of price discovery from panel data models of stock returns and exchange rates. *Journal of International Financial Markets, Institutions and Money*, 42, 65–77.
- Nguyen, D., & Bellalah, M. (2008). Hedging effectiveness and multivariate GARCH models: Evidence from European stock markets and the euro/dollar exchange rate. *International Journal of Business*, 13(1), 1–17.
- Lahmiri, S. (2019). Performance of machine learning methods in predicting direction of stock price movement: Evidence from S&P 500 and Nasdaq 100 indices. *International Journal of Finance & Economics*, 24(4), 1586–1597.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.