BI

WISEflow
Europe/Oslo(CEST)
29 Jun 2023

Handelshøyskolen BI

GRA 19703 Master Thesis

Thesis Master of Science 100% - B

Predefinert informasjon

Startdato:	09-01-2023 09:00 CET	**Termin:**	202310
Sluttdato:	03-07-2023 12:00 CEST	**Vurderingsform:**	Norsk 6-trinns skala (A-F)
Eksamensform:	T		
Flowkode:	202310||11378||IN00||B||T		
Intern sensor:	(Anonymisert)		

Deltaker

Navn:	Philip Sigurd Risgaard Kanavin og William Gunnholt Hvalbye

Informasjon fra deltaker

Tittel *:	Forecasting Implied Volatility Returns for At-The-Money Currency Options Using Machine Learning
Navn på veileder *:	Stephen Walter Szaura

Inneholder besvarelsen konfidensielt materiale?:	Nei	**Kan besvarelsen offentliggjøres?:**	Ja

Gruppe

Gruppenavn:	(Anonymisert)
Gruppenummer:	9
Andre medlemmer i gruppen:	

Forecasting Implied Volatility Returns for At-The-Money Currency Options Using Machine Learning

Master Thesis

By

William Hvalbye
Philip Kanavin

MSc in Business – Major in Finance

Bergen, June 30, 2023

Supervisor:
Stephen Walter Szaura, Associate Professor
Department of Finance, BI Norwegian Business School

Acknowledgments

We would sincerely like to thank our supervisor Stephen Walter Szaura for his guidance and flexible supervision through quick and thorough feedback. We would also like to thank the library staff at BI campus Bergen for their support and kindness in our search for literature.

Abstract

This thesis contributes to the extensive and expanding literature on financial forecasting through machine learning techniques. Our investigation focuses on the predictive capacity of machine learning (ML) models in forecasting the returns of implied volatility (IV) for at-the-money (ATM) currency options, leveraging the established methodology outlined in Kelly et al. (2020, RFS). In contrast to prior empirical evidence pertaining to equity options prediction, our findings reveal that machine learning techniques do not exhibit superior performance when compared to the straightforward ordinary least squares (OLS) regressions. This research sheds light on the inherent limitations of machine learning models in effectively capturing the predictive power of relevant covariates or variables, thereby resulting in diminished forecasting accuracy. Consequently, our study underscores the significance of exploring alternative approaches and adopting meticulous model selection strategies to enhance the precision of financial forecasting for implied volatility returns in ATM options. Through the comparative analysis of machine learning techniques and traditional regression models, our study contributes to a more comprehensive understanding of the efficacy of diverse forecasting methodologies within the financial domain.

Contents

1. Introduction and Motivation

The returns of a financial asset are the most standard evaluation of its performance. The fluctuations in these returns are described by the asset's volatility (Medvedev & Wang, 2021, JFM). Volatility measures the dispersion of returns for a security or index (Brooks, 2019). This characteristic makes it an essential component of financial derivatives. There are two measures of volatility; historical volatilities and implied volatilities (IV) (Medvedev & Wang, 2021, JFM). Historical volatility tracks the market's volatility over time, while implied volatility is a prediction that traders use to determine future market volatility. Factors such as monetary policy, macroeconomic conditions, and market expectations drive future volatility (Medvedev & Wang, 2021, JFM). Forecasting changes in implied volatilities is, therefore, challenging yet highly rewarding. Implied volatilities play a crucial role in option valuation and are essential to the Black-Scholes options pricing model (Black-Scholes, 1973, JPE). This study aims to explore the predictability of implied volatility returns of at-the-money (ATM) currency options returns using machine learning (ML) models and to contribute to the existing literature on financial forecasting.

Given that the changes in IV returns for ATM currency options hold significant implications for the broader currency options market, these changes play an essential role in the global financial market, influencing risk management, investment decisions, and monetary policy. The ability to forecast changes in implied volatilities for ATM currency options returns is of great interest to investors and traders. Kelly et al. (2020, RFS) demonstrated the potential of ML models for forecasting asset prices and returns, inspiring the investigation of such models' applicability to the implied volatility of ATM options. Being capable of selecting variables that maximize out-of-sample performance and address issues related to multicollinearity among predictor variables (Kelly et al. 2020, RFS), ML techniques offer a good starting point. Additional papers also demonstrate the strength of machine learning regressions beating the ordinary least squares (OLS) regressions (Goyenko & Zhang, 2020, SSRN; Neuhierl et al., 2021, SSRN). However, the paper "Interacting Anomalies" (Müller & Schmickler, 2020, SSRN) displays that the method of double sorts can sometimes outperform ML strategies; hence ML is not always superior.

The primary objectives of this research are to determine the effectiveness of ML models in forecasting the implied volatility (IV) of ATM currency option returns, to compare the performance of different ML models in this context, and to identify the factors affecting the accuracy of these forecasts. The findings can directly aid traders or investors using these models for decision-making and indirectly benefit those intending to develop their own models. Additionally, understanding the impact of specific characteristics on forecast accuracy can provide insights into model optimization.

Not only can the research have practical implications in finance, but this research is also interesting from a theoretical perspective, as it could provide insights into the underlying relationships between different financial assets and how they are affected by various economic and market characteristics. By studying this topic, we can contribute to the existing literature on IV for ATM currency options and machine learning in finance.

2. Literature review

In this section, we will briefly discuss the currency option market along with the Black-Scholes model. We will also review economic and parametric methods that are relevant for volatility forecasting, along with the role that machine learning has in finance.

2.1 Currency option market

Currency options are financial derivatives, granting the holder the right, without the obligation, to purchase or sell a predetermined amount of one currency for another at a pre-established exchange rate, referred to as the strike price, either on or before a specified expiration date (Hull, 2018). These options, which originated in the early 1970s (Hull, 2018), are employed by a wide array of market participants, such as investors, corporations, and financial institutions, for a variety of reasons, such as hedging against foreign exchange risks, engaging in speculative activities, and risk management (Hull, 2018).

The currency options market constitutes a sizable part of the global foreign exchange market, renowned for being one of the world's largest and most liquid financial markets (King et al., 2013, JIMF). As reported by the Bank for International Settlements (BIS), the global foreign exchange market's average daily trading volume was around $6.6 trillion in April 2019 (BIS, 2019) and $7.5 trillion in April 2022 (BIS, 2022). While the currency options market contributes substantially to this volume, quantifying its precise share remains challenging due to many currency options transactions' over-the-counter (OTC) nature (Eichengreen et al., 2016).

There are two main categories of the currency options market: exchange-traded options and OTC options, as classified by Hull (2018). The exchange-traded options are standardized contracts listed and traded on formal exchanges like the Chicago Mercantile Exchange (CME) and the International Securities Exchange (ISE). These options feature uniform contract sizes, expiration dates, and strike prices and undergo clearing via a central clearinghouse, effectively reducing counterparty risk (Hull, 2018).

On the other hand, OTC currency options involve private agreements between two parties, usually financial institutions or their clients, and do not trade on formal exchanges (Levinson, 2005). OTC currency options provide increased flexibility regarding contract conditions and customization, enabling market participants to adapt the contracts to suit their individual requirements and risk preferences (Bekaert & Hodrick, 2018). Nevertheless, OTC options carry a higher degree of counterparty risk, as the absence of a central clearinghouse means there is no guarantee of contract realization (Bekaert & Hodrick, 2018).

Currency option pricing depends on a range of factors, including the current spot exchange rate, the strike price, the time until expiration, the volatility of the underlying currency pair, and the difference in interest rates between the two currencies involved (Black & Scholes, 1973; Merton, 1973, JSTOR). Multiple option pricing models, like the Garman-Kohlhagen model (Garman & Kohlhagen, 1983, JIMF), have been developed to determine the fair value of currency options by considering these influencing factors (Black & Scholes, 1973; Merton, 1973).

Over time, the currency options market had grown substantially, evolving from its origins in the early 1970s when the Chicago Mercantile Exchange (CME) introduced the first standardized currency futures contracts (Hull, 2018). Since then, the market has seen the development of new trading platforms, products, and risk management tools (Hull, 2018). Throughout the years, market participants have developed and implemented various currency options strategies tailored to diverse market conditions and risk-return preferences, such as straddles, strangles, risk reversals, and butterflies (Hull, 2018). Risk reversals and butterfly spreads are two prominent option strategies in the currency options market, utilized by market participants to manage risk exposure and optimize profit potential.

Risk reversals consist of simultaneously purchasing and selling out-of-the-money call and put options with exact expiration dates but varying strike prices (Hull & Sinclair, 2022, JIS). The risk-reversal skew, or the difference in implied volatility, offers valuable insights into market sentiment concerning the currency pair's direction because it provides a measure of the market's expectation for the future price direction of the underlying asset (Hull & Sinclair, 2022, JIS).
Conversely, butterfly spreads involve combining four options with the same expiration date but different strike prices, establishing a profit zone within a specified range of the underlying currency's spot rate (Hull, 2018). Investors typically employ this strategy when anticipating low volatility in the underlying currency (Hull, 2018).

Such strategies have evolved in conjunction with the currency options market attracting a diverse range of market participants, which has contributed to further enhancing the strategies (King et al., 2013, JIMF). In addition, technological advancements, such as high-speed cables, servers, and electronic trading, have improved the efficiency, transparency, and accessibility of the currency options market (Eichengreen et al., 2016; King et al., 2013, JIMF). Algorithmic trading has significantly changed how financial markets operate, including the currency options market, influencing trading strategies, liquidity, and market efficiency (Vega et al., 2014, JF). These improvements have attracted a broader range of market participants and intensified competition among liquidity providers, ultimately impacting exchange rate dynamics and contributing to the evolution of exchange rate regimes (Levinson, 2005).

2.2 Implied Volatility

Implied volatility plays a significant role in the currency options market, by providing a forward-looking measure of the expected fluctuations in currency prices. As a critical input into option pricing models, it reflects the market's collective estimation of how much the underlying currency pair is likely to move over the life of the option (Medvedev & Wang, 2021, JFM). Therefore, implied volatility is not just a technical concept but a vital tool for understanding and navigating the complexities of the currency options market.

In their paper "Multi-Step Forecasts of the Implied Volatility Surface Using Deep Learning," Medvedev and Wang (2021, JFE) explains how the Black-Scholes model (BS) (Black & Scholes, 1973) can be inverted to produce implied volatilities and implied volatility-smiles. All of these are assumed to be constant over time. The implied volatility in the Black-Scholes model is computed through an iterative technique that uses the BS-model backward, plugging in the known variables and iteratively adjusting the volatility until the model's output matches the observed market price of the option (Medvedev & Wang, 2021, JFM). The initial volatility (σ) is systematically adjusted using the market data until we get the optimal implied volatility (σ_{imp}) (Medvedev & Wang, 2021, JFM).

2.3 Machine learning in finance

In recent years, machine learning has gained significant attention in the field of finance. Due to its capacity to manage extensive datasets, reveal concealed patterns, account for the effects of nonlinearities, and address interactions among numerous option and stock characteristics, machine learning also helps to minimize the risk of in-sample model overfitting (Bali et al. 2021, SSRN). Several studies have demonstrated the effectiveness of machine learning techniques in predicting various financial variables, such as stock prices, exchange rates, and credit risk (Bao et al., 2017; Kelly et al., 2020; Krizhevsky et al., 2012). Machine learning models have further demonstrated promising results in predicting financial time series data, often outperforming traditional forecasting methods (Kelly et al., 2020). They have been applied to various aspects of finance, such as predicting stock prices, exchange rates, and option returns.

2.3.1 Machine Learning in financial forecasting

Some commonly used machine learning models in financial forecasting include Neural Networks, Decision Trees and Random Forests, and Ensemble Methods. Artificial Neural Networks (ANNs) are inspired by the human brain's structure and function, consisting of interconnected layers of nodes or neurons that process and transmit information through weighted connections (Zhang et al., 1998, IJF). Feedforward Neural Networks and Recurrent Neural Networks (RNNs) are popular types of neural networks used for time series forecasting (Zhang et al., 1998, IJF). Decision trees are hierarchical models that recursively partition the input space into regions based on a set of rules, leading to a decision or prediction at the terminal nodes or the "leaves" of the tree (Kara et al., 2011). Random Forests, however, consist of multiple decision trees combined through a bagging technique to reduce overfitting and improve prediction accuracy (Liaw & Wiener, 2002, R news). Ensemble methods, such as bagging and boosting, combine multiple weak learners to create a strong learner, improving the prediction accuracy and reducing the risk of overfitting (Dietterich, 2000, LNCS; Rokach, 2010, AIR). Combining individual classifiers to obtain a collective classifier that outperforms the individual ones emulates the human process of gathering and evaluating several opinions before making a decision. Humans weigh these decisions individually and reach a final decision by combining them (Polikar, 2006, IEEE).

Machine learning models have been applied to time series data for forecasting future values of financial variables, such as stock prices and exchange rates. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, have proven particularly useful in capturing temporal dependencies in time series data (Hochreiter & Schmidhuber, 1997, MIT; Bao et al., 2017, plos). Support vector regression (SVR) has also been utilized in time series forecasting, demonstrating promising results in comparison to traditional methods like autoregressive integrated moving average (ARIMA) models (Cao & Tay, 2001, 2003, IEEE).

2.3.2 Challenges and Limitations in financial forecasting

While machine learning models have shown potential in financial forecasting, there are some general challenges and limitations to consider:

• Overfitting: Machine learning models are prone to overfitting, mainly when dealing with noisy financial data. Overfitting occurs when a model learns the noise in the training data instead of the underlying patterns, resulting in poor generalization to new data (Hawkins, 2004, CI). Regularization techniques, such as Lasso and Ridge regression, have been employed to mitigate the risk of overfitting, but it remains a challenge in complex financial forecasting tasks (Tibshirani, 1996, JRSS).

• Interpretability: Many machine learning models, such as deep learning and ensemble methods, are considered "black boxes" because their internal workings are difficult to interpret. This lack of interpretability can be an issue in finance, where regulators and investors often require transparent decision-making processes (Guidotti et al., 2018, ACM). Some researchers have proposed explainable artificial intelligence (XAI) techniques to improve the interpretability of these models. However, the application of XAI in finance is still in the early phase of research (Adadi & Berrada, 2018, IEEE).

• Computational Complexity: Some machine learning models, especially deep learning models, require significant computational resources and time to train and optimize. This can be challenging when dealing with significant financial datasets or when real-time predictions are needed (Chen et al., 2018, arXiv). However, advancements in parallel computing and specialized hardware, such as GPUs, have helped alleviate some of these computational constraints (Raina et al., 2009, ACM).

Despite these challenges, machine learning continues to be a promising area of research in finance, with ongoing developments aimed at addressing these limitations and improving the performance and interpretability of machine learning models (Kelly et al., 2020, RFS).

Another area for improvement is our relatively novice experience with advanced machine learning techniques, such as neural networks, random forests, and Gradient Boosted Regression Trees (GBRT) may have constrained our ability to exploit the analytical advantages and possibilities of these methods fully. These limitations could have implications for the precision of our conclusions. We have to the best of our abilities, incorporated as many of the required nuances as possible to better align with our research objective. However, limitations arising from this may exist. We therefore urge readers to interpret our findings with these limitations in mind.

2.4 Existing studies on ML for forecasting

The application of machine learning (ML) techniques for forecasting currency option returns has been the focus of numerous studies in recent years. The pursuit of accurate financial forecasts has led to significant contributions in the field, as evidenced by the works of Hu et al. (2013, IJF), Feng et al. (2020, JF), Kelly et al. (2020, RFS), Bali et al. (2021, SSRN), Corte et al. (2021, JFE), and Fullwood et al. (2021, JFE).

Kelly et al. (2020, RFS) is an extensive empirical study that makes an important contribution to the literature on machine learning for financial forecasting in asset pricing and investment decision-making, making it a valuable resource for anyone interested in using machine learning in finance.

Bali et al. (2021, SSRN) is a paper that investigates the use of machine learning and big data to predict option returns. The authors find that these approaches can significantly improve the accuracy of option pricing models, particularly in capturing nonlinear relationships between option returns and their underlying factors.

Corte et al. (2021, JFE) makes an important contribution to the literature on currency volatility and risk premia and is a valuable resource for anyone interested in the factors that influence the demand for currency volatility risk. The findings of the paper have important implications for the pricing of currency-linked financial instruments and the management of currency risk in investment portfolios.

Fullwood et al. (2021, JFE) investigates the relationship between option returns and volatility in the foreign exchange (FX) market and concludes that option returns are positively related to volatility, with the relationship being more robust for options with longer maturities and lower liquidity. The paper's findings have important implications for using options as a risk management tool and for developing option pricing models.

In addition to these empirical studies, Hastie et al. (2009, Springer) provide the theoretical foundation for specific ML models, such as penalized regression, regression trees, and neural networks. While the machine learning literature typically focuses on prediction, Molnar (2014) emphasizes that model interpretation is crucial for understanding the underlying relationships in financial forecasting. This could identify areas for improvement and make models more accurate.

In summary, these studies collectively contribute to a growing body of literature on machine learning and currency option returns forecasting. They provide valuable insights and guidance for researchers and practitioners interested in exploring the potential benefits of incorporating various ML techniques into financial forecasting and investment decision-making processes.

3. Data

We obtain data for nine different currency pairs, consisting of 4066 observations spanning from 2006 to 2021. The quoting convention is standardized across dealers and exchanges. The currencies are the Japanese yen (JPY), Canadian dollar (CAD), Danish krone (DKK), Norwegian krone (NOK), Swedish Krona (SEK), Swiss franc (CHF), Pound sterling (GBP), Euro (EUR) and Australian dollar (AUD), all in relation to the United States Dollar (USD).

The change in implied volatility returns of ATM currency options (ΔIV_{ATM}) is computed as the difference between the implied volatility at time $t+1$ ($IV_{ATM,t+1}$) and the implied volatility in time t ($IV_{ATM,t}$), divided by the implied volatility at time t ($IV_{ATM,t}$). Multiplying by 100 gives us the percentage change in in implied volatility of ATM currency options returns.

$$\Delta IV_{ATM} = \frac{\left(IV_{ATM,t+1} - IV_{ATM,t}\right)}{IV_{ATM,t}} x\ 100$$

We will retrieve the most economically relevant data to see if changes in implied volatility returns for ATM currency options can be forecasted. In particular, we focus on key variables with substantial economic significance. All of our data is collected from the Bloomberg terminal. The data include the change in one-month ATM implied volatilities and Interbank Offered Rates. We also have the 1, 2, 3, 6, 6-3, 9, and 12-month data for the risk reversals, butterfly strategies, and forward rates.

The time period used is from January 2006 to August 2021, resulting in 4066 observations for all of our variables. Table 1 provides sample statistics on each instrument; the number of observations, the mean, the median, the standard deviation, and the maximum and minimum value for all currencies combined. See Appendix B for summary statistics for all currencies and each pair individually. The data is available from the authors upon request.

4. Methodology

In this section, we provide an overview of the machine learning methods used in our analysis. Our methodology is based on the work by Shihao Gu, Bryan Kelly, and Dacheng Xiu in their paper “Empirical Asset Pricing via Machine Learning” (Kelly et al., 2020, RFS). We use MATLAB code from Dacheng Xiu’s website (Kelly et al., 2019, GitHub), and we thank him for making this code publicly available. Each subsection focuses on a specific method and its three key components. Firstly, we have the statistical model, which defines the general form for predicting risk premiums (Kelly et al., 2020, RFS). Secondly, we discuss the objective function used to estimate the model parameters. We aim to minimize the mean squared prediction error (MSE) to ensure accurate predictions. To enhance out-of-sample performance and avoid overfitting, we introduce variations of the MSE objective. These variations would involve including penalties for parameterization and making the models more robust against outliers.

Additionally, even with a small number of predictors, nonlinear transformations can lead to a vast number of potential model combinations, especially considering our already high-dimensional predictor set. Lastly, we describe computational algorithms in each subsection that efficiently identify the best model specification from the available options. These algorithms are crucial in optimizing the models for our analysis (Kelly et al., 2020, RFS).

4.1 Sample splitting and tuning via validation

We align with the essential preliminary steps outlined in their paper. These include designing disjoint subsamples for estimation and testing and understanding the concept of "hyperparameter tuning" (Kelly et al., 2020). Hyperparameters, also called tuning parameters, manage machine learning model complexity and combat overfitting. However, determining the optimal hyperparameter values for out-of-sample performance could be challenging based on limited theoretical guidance.

Echoing Kelly et al.'s (2020, RFS) paper, we adopt an approach by adaptively selecting tuning parameters from a validation sample. Our sample is divided into three disjoint time periods: a training sample for model estimation, a validation sample for hyperparameter tuning, and a testing sample for evaluating predictive performance. This setup enables a simulated out-of-sample test, aiming for optimal model complexity. The tuning parameters are chosen from the validation sample, while the parameter estimates are derived solely from the training data (Kelly et al., 2020, RFS).

4.2 Simple Linear

To introduce our model analysis, we first present the simplest method - the ordinary least square (OLS) linear predictive regression model. Kelly at al. (2020, RFS) use it as a baseline to highlight the unique characteristics of more advanced techniques, since they anticipate that the OLS will underperform in contrast. The linear model assumes that future changes in implied volatility for ATM currency option returns can be approximated by a linear function of the raw predictor variables and a parameter vector, θ.

$$\Delta IV_{ATM\,(i,t+1)} = z_{i,t}'\theta$$

This model imposes a simple regression specification that does not allow for nonlinear effects and interactions with predictors. In our baseline estimation of the simple linear model, we employ a standard least squares approach with an objective function, L_2.

$$L_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}(\Delta IV_{ATM\,(i,t+1)} - z'_{i,t};\theta)^\wedge 2$$

Minimizing $L(\theta)$ yields the pooled OLS estimator. The advantage of using the baseline L_2 function is that it provides analytical estimates, eliminating the need for complex optimization and computation techniques (Kelly et al. 2020, RFS).

4.2.1 Simple Linear + Huber objective function

Kelly et al. (2020, RFS) states that replacing Equation (4) with a weighted least squares objective such as

$$L_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}w_{i,t}(\Delta IV_{ATM\,(i,t+1)} - z'_{i,t};\theta)^\wedge 2$$

The additional scale factor or weight, which is included in the fitting process, could potentially improve the fit (Rosopa et al. 2006). Financial returns and illiquidity/currency/stock predictor variables are known to exhibit heavy tails, indicating outliers (Bradley, B. O., & Taqqu, M. S. 2003). The least squares objective function (4) is convex, which means it strongly emphasizes errors. Accordingly, the OLS predictions could be vulnerable to influence of outliers (Choi, S. W, 2009). The statistics literature has developed modified least squares objective functions that tend to produce more stable forecasts than OLS in the presence of extreme observations. Huber robust objective function is an example of such variation, which is a common choice to counter the effect of heavy-tailed observations (Choi, S. W, 2009). The Huber robust objective function is defined as:

$$L_H(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}H\ (\Delta IV_{ATM\,(i,t+1)} - z'_{i,t};\theta,\xi),$$

Where,

$$H(x;\xi) = \begin{matrix} x^2 & if \quad |x| \le \xi \\ 2\xi|x| - \xi^2 & if \quad |x| > \xi \end{matrix}$$

The combination of the Huber loss function is determined by tuning parameter ξ, which can be optimized based on data (Fan et al., 2016). The function combines characteristics of squared loss for small errors and absolute loss for larger errors (Kelly et al., 2020). This incorporation of robust objective functions, enhance robustness of the simple linear model (Huber, P. J., 1964). In our empirical analysis, these functions would be applied for most of the methods, while we investigate predictive advantages of this incorporation in the machine learning techniques.

4.3 Penalized Linear (Lasso, Elastic net & Ridge)

In order to avoid overfitting, the objective of reducing number of estimated parameters is crucial. With many predictors, the simple linear model often fails based on number of predictors approaching number of observations. The simple linear would become inefficient, by overfitting noise rather than extracting signal. Avoiding overfitting in machine learning is commonly countered by appending a penalty to the objective function. To combat this, a “penalty” term is added to the linear regression. Echoing Kelly et al. (2020, RFS), we promote parameter simplicity by introducing this “penalty” term to encourage the selection of more concise model specifications. We wish to incorporate this “regularization” to eventually hope for improving the stability of the out-of-sample performance, ultimately reducing fit of noise, while preserving signal fit (Kelly et al., 2020, RFS). For our penalized linear model, the statistical model is the same as the simple linear model in equation (3). However, it differs by appending a penalty to the original loss function.

$$L(\theta; \cdot) = L(\theta) + \phi(\theta; \cdot)$$

Additionally, we focus on the penalty function known as “elastic net“ penalty which combines two popular types of penalties; the L1 penalty used in Lasso regression and the L2 penalty used in Ridge regression (Zou & Hastie, 2005). The “elastic net” takes this form.

$$\phi(\theta; \lambda, \rho) = \lambda(1-\rho)\sum_{j=1}^{P}|\theta_j| + \frac{1}{2}\lambda\rho\sum_{j=1}^{P}\theta_j^2$$

The $\rho = 0$ corresponds to the Lasso and uses an absolute value, l_1 as parameter penalization, while case of $\rho = 1$ corresponds to the ridge regression (Zou & Hastie, 2005). Ridge regression uses an l_2 parameter penalization, drawing all coefficient estimates closer to zero, but never exactly zeroes. In this regard, the ridge acts as a shrinkage technique, preventing coefficients from becoming excessively large in magnitude (Kelly et al., 2022, RFS). We optimize the tuning parameters λ and ρ, adaptively using the validation sample. To achieve this, we use the accelerated proximal gradient algorithm in implementation of penalized regression. This approach supports both least squares and Huber robust objective functions (Kelly et al., 2022, RFS).

4.4 PCR and PLS

Principal Component Regression (PCR) and Partial Least Squares (PLS) are two strategies for reducing the dimensionality of the data. When predictors in penalized linear models are highly correlated, the produced forecasts could be suboptimal, specifically when forcing coefficients on regressors near or exactly equal to zero. In this case, a subset of predictors via Lasso penalty could be less efficient in comparison to taking a simple average of predictors and using this as the sole predictor in a univariate regression. Moreover, linear combinations of predictors could help reduce noise and decorrelate otherwise highly reliant predictors (Kelly et al., 2020, RFS).

With PCR, we first execute the principal component analysis (PCA) which "… is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the table, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in maps" (Abdi & Williams, 2010).

One flaw of the PCR is its failure to incorporate the objective of implied volatility returns for ATM currency options in the dimension reduction step (Kelly et al., 2020, RFS). On the other hand, PLS directly exploits the covariation between predictors and the forecast target to perform dimension reduction (Kelly & Pruitt, 2015, JE). PLS then combines predictors into an aggregate component, prioritizing stronger predictors with higher weights. The process is iterated, were the target and predictors are orthogonalized to the previously constructed components until the desired number of PLS components is obtained (Kelly et al., 2020, RFS). Here we have:

$$\Delta IV_{ATM} = Z\theta + E$$

Where ΔIV_{ATM} is the NT x 1 vector of the percentage changes in implied volatility for each observation I at time t+1, Z is the NT x P matrix of stacked predictors z_(i,t), and E is a NT x 1 vector of residuals ϵ_(i,t+1). Both PCR and PLS seek reducing dimensionality by transforming the set of predictors from its original dimension P to a reduced set of K linear combinations of predictors. For both models, forecasting model is written:

$$R = (Z\Omega_k)\theta_k + \mathrm{E}\tilde{}$$

PCR seeks K linear combinations of Z that mimic the predictor set, while PLS, in contrast, seek K linear combinations.

4.5 Generalized Linear (Group Lasso)

Due to their simplicity and efficiency, linear models have become popular. This is because they essentially are a first-order approximation of real world data (White, 1980, IER). But data can be far more complex and nonlinear. A models forecasting error can be divided into three parts; approximation error, estimation error and intrinsic error (Kelly et al., 2020, RFS).

Approximation error arises from the model's inability to replicate in full the true data. This can be reduced by implementing more flexible specifications. However, this might increase the risk of overfitting and destabilizing the model out of sample (Kelly et al., 2020, RFS). Estimation error arises due to sampling variation. This is therefore determined by the data and can be reduced by adding new observations, but doing so might reduce the control for the econometrician. News or randomness in financial markets are classified as intrinsic error. It is simply unpredictable.

The generalized linear models offer a nonparametric approach by including transformation that is nonlinear of the predictors as additional terms in a linear model – adding flexibility levels while still being close counterparts to the approaches in (4.2.1) and (4.3). Our model adapts the simple linear form by adding a K-term splineseries expansion of the predictors (Kelly et al., 2020, RFS):

$$g(z;\theta,p(\cdot)) = \sum_{j=1}^{P} p(z_j)'\theta_j,$$

Where $p(\cdot) = \left(p_1(\cdot), p_2(\cdot), \dots, p_k(\cdot)\right)'$ $(\cdot))'$ is a vector of basic functions, and the parameters are now a K x N matrix $\theta = (\theta_1, \theta_2, \dots, \theta_N)$. We adopt a spline series of order two: $(1, z, (z-c_1)^2, (z-c_2)^2, \dots, (z-c_{K-2})^2$,, where $c_1, c_2, \dots, c_{K-2}$ are knots (Kelly et al., 2020, RFS).

We adopt the same least squares objective function as in linear models, with optimal modifications for robustness similar to the paper by Kelly et al. (2020, RFS) both with and without the Huber robustness modification. Similarly, the use of penalization is adopted to control for the number of model specifications. The penalization function is known as the group lasso and has the form:

$$\phi(\theta;\lambda,K) = \lambda \sum_{j=1}^{P} \left(\sum_{k=1}^{K} \theta_{j,k}^2 \right)^{\frac{1}{2}}$$

The group Lasso is compatible with both least squares and robust Huber objectives, and it employs the accelerated proximal gradient descent algorithm, similar to the elastic net (Kelly et al., 2020, RFS).

4.6 GBRT and Random forests

Gradient Boosted Regression Trees (GBRT) is a widely used learning algorithm in the field of machine learning today (Zheng et al., 2020). It constructs an additive regression model by utilizing decision trees as weak learners. One of the advantages of decision trees, including gradient boosted regression trees (GBRT), is their interpretability compared to other learning algorithms. Additionally, GBRT is highly adaptable as it allows for the use of various loss functions during the boosting process (Friedman, J.H., 2001).

We follow the exact same procedure and pattern as the paper of Kelly et al. (2020). Similarly, the procedure begins by fitting a shallow tree with a limited depth (e.g., L=1). This initial tree may provide a weak prediction with substantial bias in the training sample. Subsequently, a second shallow tree (also with depth L) is employed to model the prediction residuals from the first tree. The forecasts from these two trees are combined to form an ensemble prediction, but the contribution from the second tree is scaled down by a factor $\nu \in (0,1)$ to prevent overfitting of the residuals (Kelly et al., 2020, RFS).

In each subsequent step, a new shallow tree is fitted to the residuals from the model with b-1 trees, and its residual forecast is added to the ensemble prediction with a shrinkage weight of ν. This process continues iteratively until a total of B trees are included in the ensemble. Therefore, the final output is an additive model comprising shallow trees with three tuning parameters (L, ν, B), which we select adaptively through the validation step (Kelly et al., 2020, RFS).

Following the approach outlined in the paper by Kelly et al. (2020, RFS), we again adopt a similar methodology for random forests. Random forests consist of a combination of tree predictors, where each tree relies on the values of a random vector that is independently sampled from the same distribution for all trees in the forest (Breiman, 2001). This ensemble approach has demonstrated significant improvements in classification accuracy by aggregating the predictions of multiple trees and allowing them to vote for the most popular class (Breiman, 2001).
To construct these ensembles, random vectors are often generated to govern the growth of each tree within the ensemble. One notable example is bagging (Breiman, 1996), which involves selecting random subsets (without replacement) from the training set to grow each individual tree. This randomness in the selection process contributes to the diversity of the trees within the forest, enhancing the overall predictive performance (Breiman, 2001).

5.0 Results

Our training sample ranges from January 2006 to January 2013, our testing data ranges from February 2013 to January 2018 and our validation data uses the remaining time period stretching from February 2018 to August 2021. We conducted robustness tests for our training, testing, and validation models by employing different time periods, ensuring the reliability and validity of our findings.

5.1 Variable Importance and marginal relationships

Similarly to the paper by Kelly et al. (2020, RFS), we wish to identify covariates that have a significant impact on implied volatility returns of ATM options, in addition to accounting for other predictors. In our methodology, we adopt a similar metric to evaluate the importance of variables to try to represent the goodness of fit of the model to our observed data. When specific predictor values are set to zero, the metric measures the panel predictive R2 while holding the remaining model estimates constant (Kelly et al., 2019, JFE).

Moreover, we examine the marginal relationship between the implied volatility returns of ATM options and each characteristic through plots. This offers a visual presentation of the influence of each covariate in the machine learning models.

5.2 All currencies

In this section, we will present the results of our analysis and further discuss the impact and meaning of our findings.

5.2.1 Out of sample R-squared

The regressions used in this study aim to predict the implied volatility returns of the ATM currency options. Table 2 in Appendix B exhibits the out-of-sample R-squared for all currencies combined. Out of the thirteen model regressions, the simple OLS and simple OLS + H has an R-squared value of 0.0232, meaning that these models explain about 2,32% of the variance in the implied volatility. This indicates a relatively weak predictive power.

PCR and PLS have the lowest R-squared values of the models, close to zero and equal to zero. These regression models should handle dimensionality and multicollinearity in the data. However, the low R-squared values suggest that these models were unable to capture any meaningful relationship between the target variable and the input features.

Our regularization models, Lasso, Ridge, and Elastic net, which introduce penalties to prevent overfitting in the models, perform slightly better. The Lasso and its Huber loss function both yield an R-squared of 0,0176. The Ridge regression and its Huber loss function yield an R-squared of 0,0126 and 0,0127, respectively. The Elastic net and its Huber loss function yield an R-squared of 0,0155 and 0,0156, hence performing even worse.

The Oracle model yields a negative R-squared value, meaning that it actually performed worse compared to a model that only predicts the mean of the output variable.

Finally, the R-squared values for the group lasso regression and its Huber loss function are 0,0021 and 0,0020, respectively. Like the other models, these fail to predict and capture the relationship in high-dimensional data with the current feature grouping.

Overall, these R-squared yielded from our models suggest our models were largely unsuccessful in accurately predicting implied volatility returns of ATM currency options.

5.3 Individual currency pairs

In this section, we will look into the results of the regression models when run for each currency individually.

The simple OLS and its Huber loss function variant display some interesting variations. They both yield more robust R-squared values when applied to JPY, NOK, and GBP, returning values of 0,1381, 0, 1449, and 0,0628 for the simple OLS function, respectively. Compared to when all currencies were stacked together, the models explain a higher percentage of variation for these currencies individually.

Similar to all currencies combined, the PCR and PLS continue to perform poorly across all currencies individually, with R-squared close or equal to zero. This highlights the models' inability to capture the relationship between the target and input variables.

The regularization models, Lasso, ridge, and elastic net, exhibit varying performances across the corresponding currency pairs. The Lasso model yields an R-squared of 0,1607 for NOK, closely followed by the elastic net, yielding 0,1594. The regularized regressions perform best out of all the models for CAD, NOK, SEK, and AUD, while its performance was weaker for the other currencies. It is of interest how the Ridge regression for NOK have a relatively high R-squared (0,1393) but low values for all the other currency pairs.

Like the PCR and PLS, the Oracle model, the group lasso, and its Huber loss function generally performed poorly across all individual currency pairs, yielding R-squared close to zero. In the case of the CHF, the Oracle model yielded a negative R-squared, indicating the model's difficulty in providing accurate predictions.

RF and GBRT were applied to DKK, CHF, and AUD. Despite the relatively low results, both the RF and GBRT yielded higher R-squared values than the rest of the models, similar to the GBRT for DKK.

The predictive power of the models remains relatively low across the currency pairs. However, some currency pairs display higher performance, JPY, NOK, and AUD.

Table 3 displays the summary statistics for JPY in relation to USD. Starting with the IBOR and forward rates, the mean and median are close to each other, implying that there is a normal or near normal distribution with limited skewness in this data. Given a mean and a median close to zero, with a relatively low standard deviation, indicates low variations. This could possibly explain why Lasso, Ridge and Elastic net have such low R-squared values since these models can handle multicollinearity and selecting variables that are major contributors.

The forward rates have a higher standard deviation compared to the mean, suggesting high variability. If these rates significantly affect implied volatility, the variability in forward rates could contribute to the relatively high R-squared values for the simple OLS model.

The statistics for risk reversals show a strong skew towards negative values, highlighted by the 15 delta across all maturities, suggesting a market bias for protecting against depreciation of USD/JPY. There is far greater volatility in a 15 delta compared to a 35 delta, implying a larger market uncertainty for extreme movements in USD/JPY, aligning with the high negative values observed.

The statistics for the butterfly strategy display positive values across all maturities and deltas. This could suggest market anticipation with higher volatility in either direction of the USD/JPY forward rate. The volatility expectation increase as the horizon of the maturities increases. This makes sense due to increasing uncertainty over longer time horizons.

Linking these observations to the original regression results, models like Lasso, Ridge, and Elastic Net performed relatively well. These values are not particularly high, but these models are capturing some of the complexities and volatility in the USD/JPY forward rate, displayed in Figure C.1.1.

Figure C.1.1. also displays some context for why the PCR and PLS perform so poorly. There is a complex relationship and inherent volatility in the currency market, reflected in the low explanatory power of these models, similarly across all currency pairs.

In Table 2, the regression results for USD/NOK currency pair display a similar pattern as the USD/JPY. Overall, the R-squared values are low, indicating the models' relatively weak predictive power. However, some differences in the results are worth discussing.

The forward rates displayed in Table 1 exhibit a skew to the right, explained by a mean higher than the median. This could suggest an anticipation towards a future appreciation in the value of USD against the NOK. This could contribute to volatility in the exchange rate. This anticipation is strengthened by the positive risk reversal mean values across all maturities. The mean is also positive for the butterfly spread across, implying that markets add tail risk to their pricing. This could contribute to overall volatility.

5.4 Which characteristics matter

This subsection examines the significance of covariates and introduces two figures, namely Figure 1 and Figure 2. Using the importance measure discussed in Section 1.9, we assess the relative importance of individual covariates in relation to the performance of each model. Following the study by Kelly et al. (2020), we present Figure 1, which showcases the top 20 most influential covariates for each method. This presentation enables the interpretation of the relative importance of variables within each model, with the variable importance normalized to a sum of one.

Figure 1 presents the important characteristics of the generalized models and the dimension reduction model (PLS). The generalized model, specifically the Group Lasso + H, exhibits a moderate bias towards the 1-month tenor of the Interbank offered rate (IBOR). Furthermore, the Group Lasso models display a slight inclination towards the butterfly spread (15BF6_3M). Conversely, the PLS dimension reduction model highlights that the covariate IBOR1M holds the primary influence. In contrast, the other covariates appear to have negligible or zero impact on the predictions of the PLS models. With the exception of the PLS, all individual models demonstrate some degree of agreement regarding the most influential predictors, suggesting their utilization of predictive information from a broader range of predictors.

Across the models, the covariates exhibit variation, with a slight bias towards foreign exchange rates (FX_rates) for 1-3-month maturities observed in both the Simple OLS and Simple OLS + Huber. The higher R-squared values associated with the FX-rate variables indicate a stronger explanatory influence, thus accounting for a more significant portion of the variance in implied volatilities compared to the butterfly or risk reversal strategy variables. Considering the Simple OLS's treatment of variables on equal footing, without specific regularization or selection techniques, it is reasonable for foreign exchange rates to emerge as significant predictors for implied volatilities of ATM currency option returns.

Moreover, the Lasso models, with and without the Huber loss function, exhibit a slight bias towards the Risk reversal strategy for options trading (15RR6M) and the butterfly spread trading strategy (25BF9M). In contrast, the Ridge and Ridge models with Huber loss function demonstrate similarity in terms of the most influential predictors, although differentiated by expiration and maturity time horizons.

The Elastic net model (Enet + H) demonstrates a close alignment of predictors, with the foreign exchange rate (FX_rate) and the 3-month rate displaying the highest R-squared values. Additionally, the model exhibits a slight skewness towards the butterfly spread trading strategy across various time horizons. The penalized models exhibit a similar agreement among predictors, indicating that these models prioritize the same groups of variables and penalize those deemed less relevant.

Figure 1 demonstrates that, except for Oracle and the dimension reduction technique known as Partial Least Squares (PLS), there is a consensus among the models regarding the most influential predictive characteristics. These predictors can be categorized into four groups. The first group pertains to the risk reversal options strategy, which varies in terms of maturities and sensitivity to changes in the underlying assets (delta). The second group consists of the butterfly spread and the foreign exchange rate (FX_rate). Lastly, the Interbank offered rate (IBOR) serves as a predictor characteristic.

In Figure 2, the heatmap portrays the 51 characteristics based on their overall contribution to the models. Similar to the study conducted by Kelly et al. (2020), we rank the importance of each covariate for each method and subsequently sum their ranks. The color gradients in the heatmap reflect the model-specific ranking of covariates, ranging from least to most important, with lighter shades indicating lower importance and darker shades indicating higher importance.

The analysis of the regression models, accompanied by the heatmap visualization, reveals the significant impact of certain covariates on the predictive performance of the models. The heatmap exhibits a discernible pattern, wherein specific variables consistently exhibit higher R-squared values across various models, particularly the penalized models such as Lasso (Lasso + H), Ridge (Ridge + H), and Enet (Enet + H). The color gradient pattern observed in the heatmap indicates a consensus among these models regarding influential predictors, with certain variables consistently contributing more to the predictive power. Notably, the butterfly spread strategies (15BF6M, 15BF9M, 15BF12M) and the IBOR1M exhibit a strong influence, consistently appearing as darker shades across the penalized models.

5.4.1 Best Performers

In this subsection, our focus is on three specific currencies: Norwegian Krone (NOK), Japanese Yen (JPY), and Australian Dollar (AUD). The currencies are displayed in Figure C.2, C.1, and C.3, respectively, in Appendix B. These currencies have demonstrated superior out-of-sample performance compared to others. As in the previous analysis, we present the top 20 influential covariates for each method.

Examining the results for NOK, we find consistently low R-squared values across all models. The independent variables exhibit limited explanatory power in predicting performance. Although this pattern holds true across various models, specific predictors contribute more significantly to the out-of-sample predictive performance in specific models. Notably, the Generalized models (Group Lasso + H) exhibit a notable skew towards the FX_rate variable, while the remaining top variables demonstrate a high degree of agreement in their contributions. The penalized model (elastic net) gives slightly greater importance to the IBOR1M predictor. However, the R-squared value, approximately 17.5%, indicates a relatively modest level of explanatory power for predictive purposes.

In the case of the Japanese Yen (JPY), the penalized model (Elastic net + H) exhibits a distinct bias toward the 1-month foreign exchange rate. Furthermore, it incorporates predictive insights from a broader range of characteristics. The Ridge regression models, with and without the Huber loss function, demonstrate a pronounced inclination towards the foreign exchange rate and the 9-month butterfly trading strategy (25BF9M). Additionally, these models derive predictive information from the Interbank offered rates (IBOR1M) and (IBOR6-3M). The remaining variables, however, contribute little to the overall predictive performance of the models.

The Australian dollar (AUD/USD) exhibits negative R-squared values across multiple variables in the Simple OLS and OLS with the Huber extension. This indicates a potential violation of the assumption of linearity between the independent and dependent variables, as assumed by the Simple OLS regression. Consequently, the omission of certain variables may have resulted in a poor fit, failing to accurately capture the relationship's complexity. Conversely, the penalized models, particularly Elastic net and Lasso regressions, demonstrate positive R-squared values, indicating their effectiveness in forecasting. These models utilize information from all of the top 20 characteristics, suggesting a higher level of agreement in terms of predictive information. However, the Ridge regression displays a stronger bias towards the 1-month interbank offered rate (IBOR1M). Additionally, the Ridge regression and other penalized models incorporate predictive information from a broader range of the top 20 characteristics.

These results indicate that, while the performance of models improved for some individual currency pairs compared to all currencies combined, the predictive power of the models remains relatively low across all currencies. This suggests that the relationship between the independent variables and the implied volatility returns of ATM currency options may be complex, potentially nonlinear, and varies by currency. The difference in model performance for individual currencies indicate that there may be unique characteristics in each currency market that are not captured by the models. This calls for further investigation and potentially developing currency-specific models or including additional currency-specific predictors.

6. Conclusion

In our study, we investigate a range of machine learning methods – namely, Principal Component Regression (PCR), Partial Least Squares (PLS), Lasso, Lasso + Huber, Ridge, Ridge + Huber, Elastic Net, Elastic Net + Huber, Oracle, Group Lasso, Group Lasso + Huber, Random Forest (RF), and Gradient Boosted Regression Trees (GBRT) – and reveals that the comparatively simpler linear Ordinary Least Squares (OLS) regression outperforms these more advanced techniques. This outcome can be attributed to several key factors. Firstly, the simplicity of OLS lends itself to transparent and interpretable models, enabling a deeper understanding of coefficient effects on the response variable (James et al., 2013). This characteristic proves particularly beneficial in our research context, where interpretability and explainability are valued. Secondly, the assumptions inherent in OLS, including linearity, independence, and homoscedasticity, align favorably with the underlying structure of our data, leading to accurate and reliable results.

Nevertheless, this by no means diminishes the potential of machine learning techniques, which may outshine in different contexts or datasets (Müller & Guido, 2017). Considering the specific characteristics of different currency pairs, such as JPY/NOK and USD/AUD, where dimension reduction regressions such as Lasso and Elastic Net have exhibited relatively better performance compared to OLS, may offer valuable insights for developing tailored models for forecasting IV returns for ATM currency options.

Appendix: Data sources

A.1 One-month At the Money Implied Volatilities

We use nine different currencies in relation to the US dollar for the one-month at-the-money implied volatilities: Japanese yen (JPY), Canadian dollar (CAD), Danish krone (DKK), Norwegian krone (NOK), Swedish Krona (SEK), Swiss franc (CHF), Pound sterling (GBP), Euro (EUR) and Australian dollar (AUD). We retrieve the data on the one-month at-the-money implied volatilities from Bloomberg. The time period stretches from 01/04/2006 to 08/03/2021.

A.2 Risk reversals in FX options and butterfly strategies in currency options

Here, we have the 15 delta, 25 delta, and 35 delta risk reversals with a 1, 2, 3, 6-3, 6, 9, and 12-month expiry for the various currency pairs: USD/JPY, USD/CAD, USD/DKK, USD/NOK, USD/SEK, USD/CHF, GBP/USD, EUR/USD, and AUD/USD. The data was obtained from Bloomberg. The same time periods apply here as A.1

A.3 Interbank Offered Rates (IBOR)

We have the Interbank Offered Rates for the same currency pairs with period lengths of 1 month and 6-3 months. In certain economic conditions, negative IBOR rates can occur meaning that banks charge a storage fee for holding the money, instead of paying interest. The same time periods apply here as A.1.

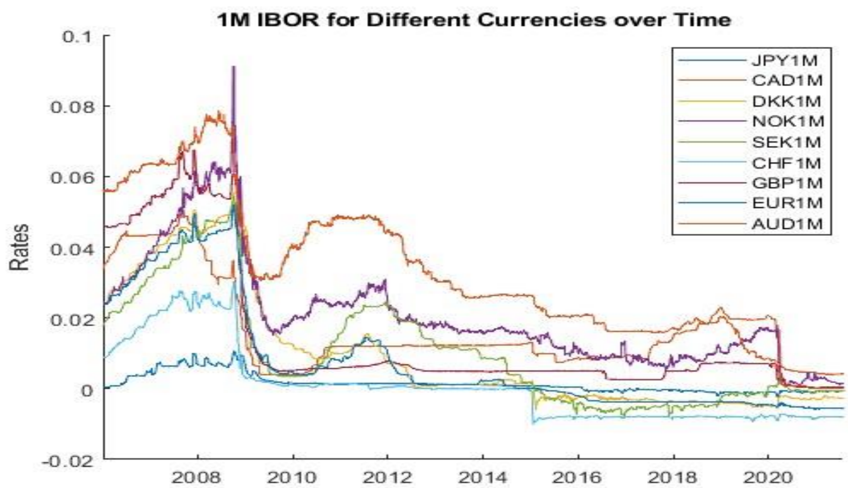


Figure A.1: 1-month IBOR for the various currencies.

Fig.A.1. This figure shows the Interbank offered rate for the different currencies over a sample period from January 2006 to August 2021.

A.4 Historical foreign exchange (FX) rates and currency forwards

For the corresponding currency pairs, we gather their historical FX rates. The data was obtained from Bloomberg. The same time periods apply here as A.1. We also have the 1, 2, 3, 6-3, 6, 9, and 12-month forward rates for these currency pairs.

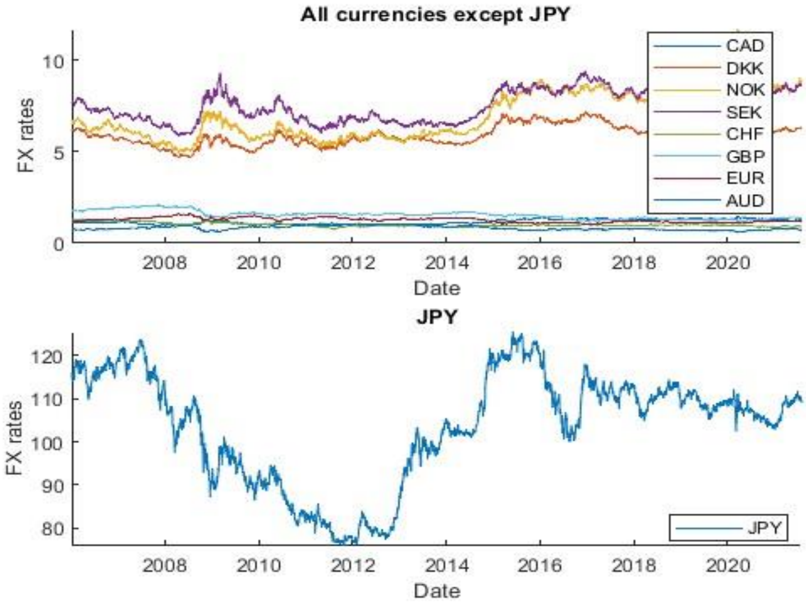


Figure A.4: FX rates for each currency

Fig.A.4. The top figure shows the FX rates for USD/CAD, USD/DKK, USD/NOK, USD/SEK, USD/CHF, GBP/USD, EUR/USD, and AUD/USD, while the bottom figure shows the USD/JPY FX rates over a sample period from January 2006 to August 2021.

Appendix B: Tables

B.1 All Currencies

	All currencies	N	Mean	Median	Std.Dev	Max	Min
	ATM Implied Volatility						
	ΔIV_{ATM}	4066	0.0008	-0.0011	0.0426	1.3665	-0.3585
	Interbank Offered Rate						
	1M	4066	0.0322	0.0265	0.0204	0.0785	0.0000
	6-3M	4066	0.0008	0.0009	0.0012	0.0049	-0.0058
	Forward Rates						
	1M	4066	0.8290	0.7840	0.1188	1.0973	0.5743
	2M	4066	0.8277	0.7833	0.1176	1.0931	0.5744
	3M	4066	0.8263	0.7829	0.1165	1.0888	0.5744
	6M	4066	0.8222	0.7812	0.1131	1.0765	0.5745
	9M	4066	0.8183	0.7794	0.1098	1.0650	0.5744
	12M	4066	0.8143	0.7778	0.1067	1.0535	0.5743
	Risk Reversal						
	15RR	4066	-0.2141	-0.1420	1.8588	11.1025	-15.7825
1M	25RR	4066	-0.1306	-0.0990	1.2198	7.3150	-10.2475
	35RR	4066	-0.0686	-0.0495	0.6780	4.1900	-5.6850
	15RR	4066	-0.2587	-0.1600	2.0696	11.0075	-15.8800
2M	25RR	4066	-0.1580	-0.1100	1.3570	7.2500	-10.2925
	35RR	4066	-0.0847	-0.0600	0.7510	4.1500	-5.7175
	15RR	4066	-0.3015	-0.1775	2.2618	10.8650	-16.0075
3M	25RR	4066	-0.1849	-0.1212	1.4794	7.1200	-10.3550
	35RR	4066	-0.1005	-0.0650	0.8134	4.0550	-5.7150
	15RR	4066	-0.3585	-0.1825	2.5282	10.5425	-16.1775
6M	25RR	4066	-0.2196	-0.1212	1.6455	6.7700	-10.3675
	35RR	4066	-0.1211	-0.0693	0.8940	3.8150	-5.6750
	15RR	4066	-0.4014	-0.1925	2.6431	9.6275	-16.2625
9M	25RR	4066	-0.2406	-0.1250	1.7120	6.2275	-10.3625
	35RR	4066	-0.1331	-0.0707	0.9247	3.4800	-5.6425
	15RR	4066	-0.4232	-0.2225	2.7549	10.1125	-16.3475
12M	25RR	4066	-0.2612	-0.1500	1.7891	6.4375	-10.3525
	35RR	4066	-0.1500	-0.0808	0.9697	3.5500	-5.6075
	Butterfly						
	15BF	4066	0.5331	0.4550	0.2975	3.6800	-0.0202
1M	25BF	4066	0.2398	0.2125	0.1196	1.4375	-0.1025
	35BF	4066	0.2388	0.2125	0.1207	1.4375	-0.1025
	15BF	4066	0.6230	0.5425	0.3198	3.6575	-0.0303
2M	25BF	4066	0.2767	0.2475	0.1289	1.4100	-0.0101
	35BF	4066	0.0896	0.0825	0.0425	0.5425	-0.2075
	15BF	4066	0.7176	0.6325	0.3540	3.5500	-0.0303
3M	25BF	4066	0.3149	0.2850	0.1423	1.3700	-0.0051
	35BF	4066	0.1015	0.0925	0.0461	0.5600	-0.1875
	15BF	4066	0.8642	0.7775	0.3936	3.5000	-0.0505
6M	25BF	4066	0.3712	0.3400	0.1576	1.5200	-0.0101
	35BF	4066	0.1182	0.1100	0.0506	0.5375	-0.1725
	15BF	4066	0.9373	0.8550	0.4125	3.6450	-0.0404
9M	25BF	4066	0.3999	0.3675	0.1647	1.6000	-0.0051
	35BF	4066	0.1267	0.1175	0.0532	0.6125	-0.1775
	15BF	4066	1.0055	0.9225	0.4333	3.8500	-0.0505
12M	25BF	4066	0.4253	0.3900	0.1751	1.7900	-0.0051
	35BF	4066	0.1347	0.1250	0.0578	0.5725	-0.1775

Table 1: Summary statistics for our data

Tab.1. The table displays summary statistics for each instrument in our data, reporting the mean, median, standard deviation, maximum and minimum value.

B.2 Out-of-sample R-squared

	All Currencies	JPY	CAD	DKK	NOK
Simple OLS	0.0232	0.1381	0.0564	-0.0756	0.1459
Simple OLS+H	0.0232	0.1371	0.0583	-0.0750	0.1459
PCR	0.0001	0.0000	0.0007	0.0007	0.0027
PLS	0.0000	0.0000	0.0000	0.0000	0.0082
Lasso	0.0176	0.0639	0.0887	-0.0253	0.1607
Lasso+H	0.0176	0.0544	0.0890	-0.0249	0.1606
Ridge	0.0126	0.0021	0.0211	0.0005	0.1393
Ridge+H	0.0127	0.0019	0.0228	0.0005	0.1393
Enet	0.0155	0.0817	0.0872	-0.0249	0.1594
Enet+H	0.0156	0.0850	0.0876	-0.0245	0.1594
Oracle	-0.0003	0.0007	0.0007	-0.0009	0.0036
Group Lasso	0.0021	0.0020	0.0000	-0.0002	0.0000
Group Lasso+H	0.0020	0.0021	0.0000	-0.0003	0.0000
RF	-	-	-	-0.0002	-
GBRT	-	-	-	0.0049	-

	SEK	CHF	GBP	EUR	AUD
Simple OLS	0.0208	-0.0298	0.0628	-0.0491	-0.0486
Simple OLS+H	0.0208	-0.0300	0.0628	-0.0492	-0.0498
PCR	0.0018	0.0017	0.0000	0.0015	0.0043
PLS	0.0000	0.0000	0.0000	-0.0157	0.0209
Lasso	0.0457	-0.0031	0.0560	-0.0132	0.1167
Lasso+H	0.0457	-0.0021	0.0560	-0.0133	0.1162
Ridge	0.0348	-0.0007	0.0568	-0.0231	0.0315
Ridge+H	0.0348	-0.0001	0.0568	-0.0232	0.0305
Enet	0.0461	-0.0031	0.0531	-0.0136	0.1095
Enet+H	0.0461	-0.0021	0.0531	-0.0137	0.1090
Oracle	0.0014	-0.0010	0.0046	0.0008	0.0042
Group Lasso	0.0006	0.0034	0.0016	0.0009	0.0113
Group Lasso+H	0.0007	0.0036	0.0017	0.0010	0.0102
RF	-	0.0070	-	-	0.0082
GBRT	-	0.0041	-	-	0.0279

Table 2: Out-of sample R-squared

Tab.2. The table shows the out-of sample R-squared yielded by the different models for each currency pair.

B.3 USD/JPY summary statistics

	USD/JPY	N	Mean	Median	Std.Dev	Max	Min
	ATM Implied Volatility						
	ΔIV_{ATM}	4066	0.0012	-0.0035	0.0526	0.8118	-0.2208
	Interbank Offered Rate						
	1M	4066	0.0014	0.0010	0.0026	0.0106	-0.0025
	6-3M	4066	0.0008	0.0007	0.0006	0.0025	-0.0001
	Forward Rates						
	1M	4066	103.3786	106.8769	12.7865	125.5683	75.7928
	2M	4066	103.2372	106.7658	12.7016	125.5254	75.7449
	3M	4066	103.0961	106.5840	12.6202	125.4822	75.7069
	6M	4066	102.6647	106.0534	12.3848	125.2664	75.5750
	9M	4066	102.2250	105.6318	12.1682	124.9230	75.4220
	12M	4066	101.7694	105.2067	11.9681	124.5545	75.2640
	Risk Reversal						
	15RR	4066	-1.7901	-1.3358	2.2506	1.5375	-15.7825
1M	25RR	4066	-1.1326	-0.8580	1.4273	1.0375	-10.2475
	35RR	4066	-0.6066	-0.4700	0.7671	0.5750	-5.6850
	15RR	4066	-2.0344	-1.6137	2.3403	1.9450	-15.8800
2M	25RR	4066	-1.2871	-1.0325	1.4857	1.2350	-10.2925
	35RR	4066	-0.6962	-0.5675	0.8041	0.6775	-5.7175
	15RR	4066	-2.2461	-1.8479	2.4332	2.3050	-16.0075
3M	25RR	4066	-1.4249	-1.1750	1.5507	1.4950	-10.3550
	35RR	4066	-0.7706	-0.6357	0.8427	0.8200	-5.7150
	15RR	4066	-2.5658	-2.1392	2.6094	2.4850	-16.1775
6M	25RR	4066	-1.6210	-1.3500	1.6593	1.6000	-10.3675
	35RR	4066	-0.8790	-0.7275	0.9025	0.8550	-5.6750
	15RR	4066	-2.7531	-2.3350	2.7097	2.3425	-16.2625
9M	25RR	4066	-1.7350	-1.4663	1.7163	1.5175	-10.3625
	35RR	4066	-0.9395	-0.7875	0.9361	0.7975	-5.6425
	15RR	4066	-2.9087	-2.5125	2.7930	2.3000	-16.3475
12M	25RR	4066	-1.8374	-1.5850	1.7748	1.4700	-10.3525
	35RR	4066	-1.0029	-0.8550	0.9665	0.7600	-5.6075
	Butterfly						
	15BF	4066	0.6585	0.5500	0.3253	2.8375	0.2875
1M	25BF	4066	0.2856	0.2575	0.1088	1.0825	0.1125
	35BF	4066	0.2849	0.2575	0.1100	1.0825	0.0526
	15BF	4066	0.7274	0.6113	0.3342	2.5625	0.2982
2M	25BF	4066	0.3093	0.2825	0.1103	0.8675	0.1250
	35BF	4066	0.0909	0.0925	0.0426	0.2675	-0.2075
	15BF	4066	0.8094	0.7025	0.3468	2.6025	0.2982
3M	25BF	4066	0.3389	0.3200	0.1175	0.9100	0.1250
	35BF	4066	0.0991	0.1000	0.0431	0.2900	-0.1875
	15BF	4066	0.9785	0.9050	0.3655	2.7050	0.3975
6M	25BF	4066	0.3924	0.3750	0.1336	0.9550	0.0875
	35BF	4066	0.1113	0.1125	0.0491	0.2925	-0.1725
	15BF	4066	1.0884	1.0225	0.3810	2.7850	0.4231
9M	25BF	4066	0.4294	0.4175	0.1462	0.9800	0.0725
	35BF	4066	0.1199	0.1200	0.0537	0.3375	-0.1775
	15BF	4066	1.1904	1.1275	0.4017	2.8650	0.4000
12M	25BF	4066	0.4665	0.4550	0.1600	0.9975	0.0375
	35BF	4066	0.1301	0.1275	0.0623	0.3500	-0.1775

Table 3: Summary statistics for our data

Tab.3. The table displays summary statistics for each instrument in our data, reporting the mean, median, standard deviation, maximum and minimum value for USD/JPY.

Appendix C: Individual currency pairs

C.1 Currency pair 1: Japanese yen (JPY) in relation to the US dollar (USD)

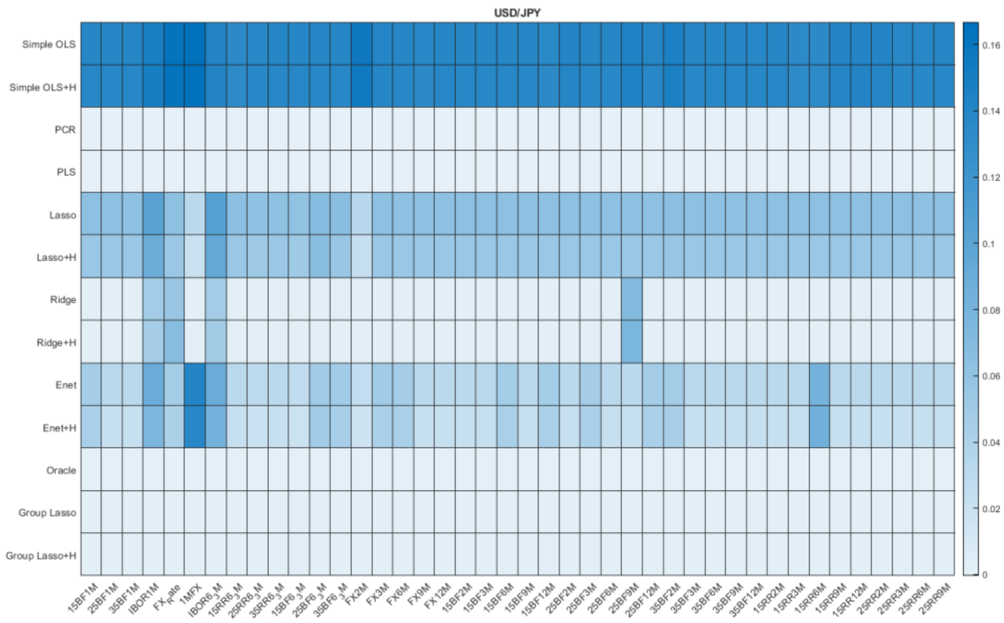
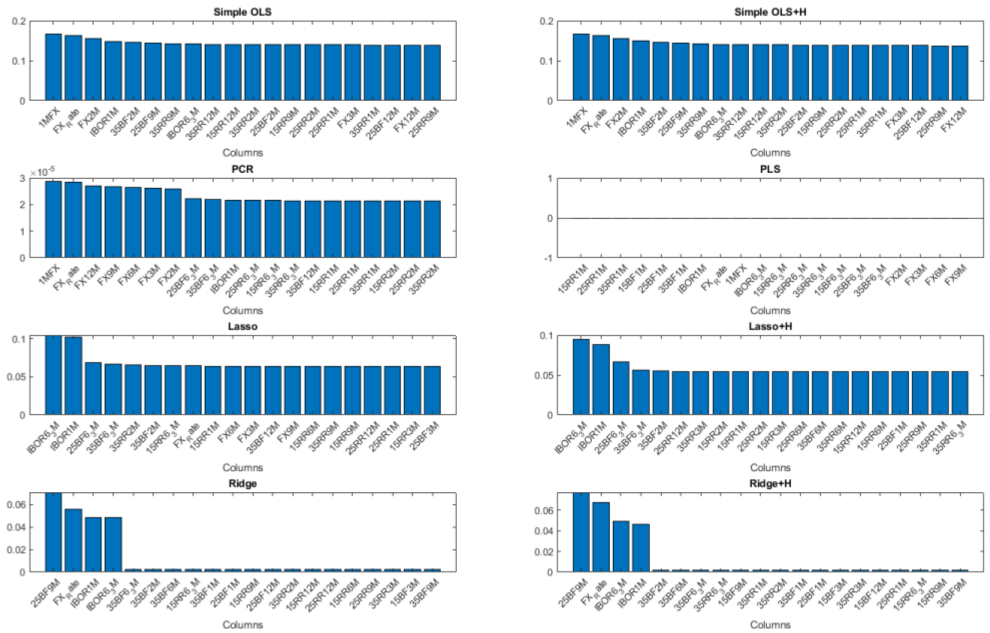


Figure C.1: Heatmap USD/JPY for 13 regression models.

Fig.C.1. This figure shows the heatmap for each of the 13 regression models, looping over each of the 51 variables.



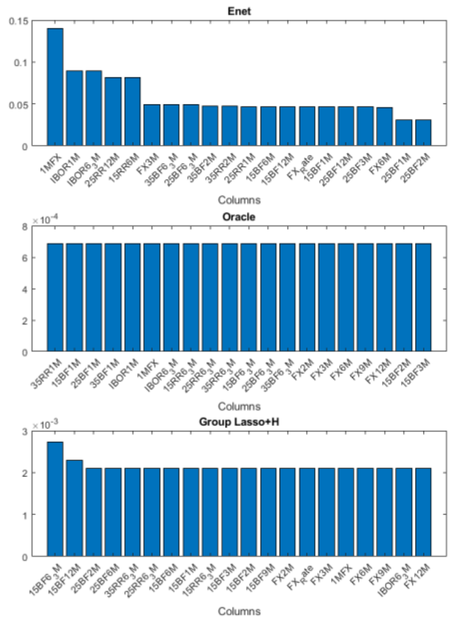


Figure C.1.1: Top 20 covariates for USD/JPY for 13 regression models.

Fig.C.1.1. This figure shows the top 20 covariates for each of the 13 regression models

C.2 Currency pair 2: Norwegian krone (NOK) in relation to the US dollar (USD)

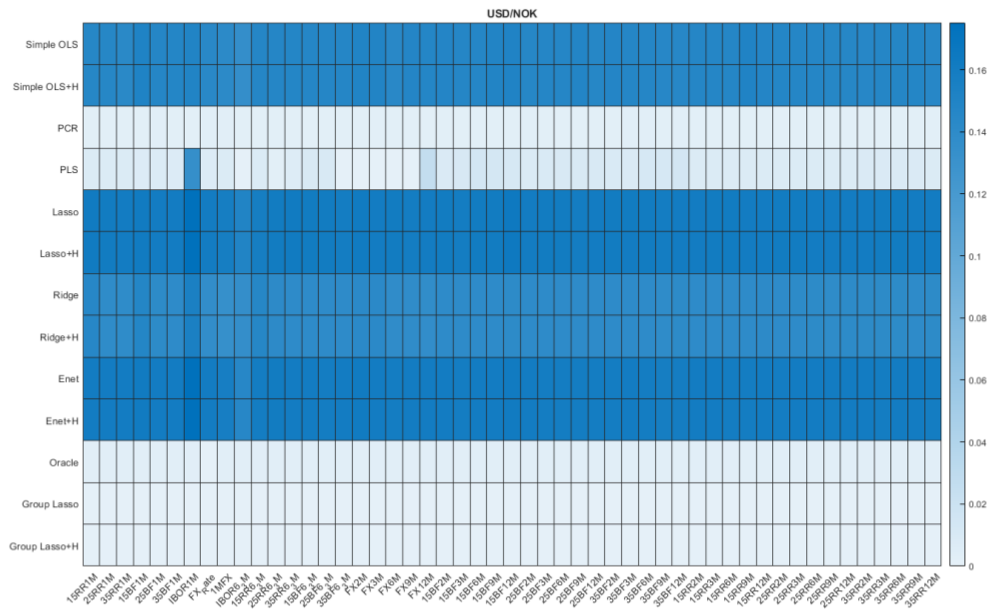


Figure C.2: Heatmap USD/NOK for 13 regression models.

Fig.C.2. This figure shows the heatmap for each of the 13 regression models, looping over each of the 51 variables.

Figure C.2.1: Top 20 covariates for USD/NOK for 13 regression models.

Fig.C.2.1. This figure shows the top 20 covariates for each of the 13 regression models

C.3 Currency pair 3: Australian dollar (AUD) in relation to the US dollar (USD)

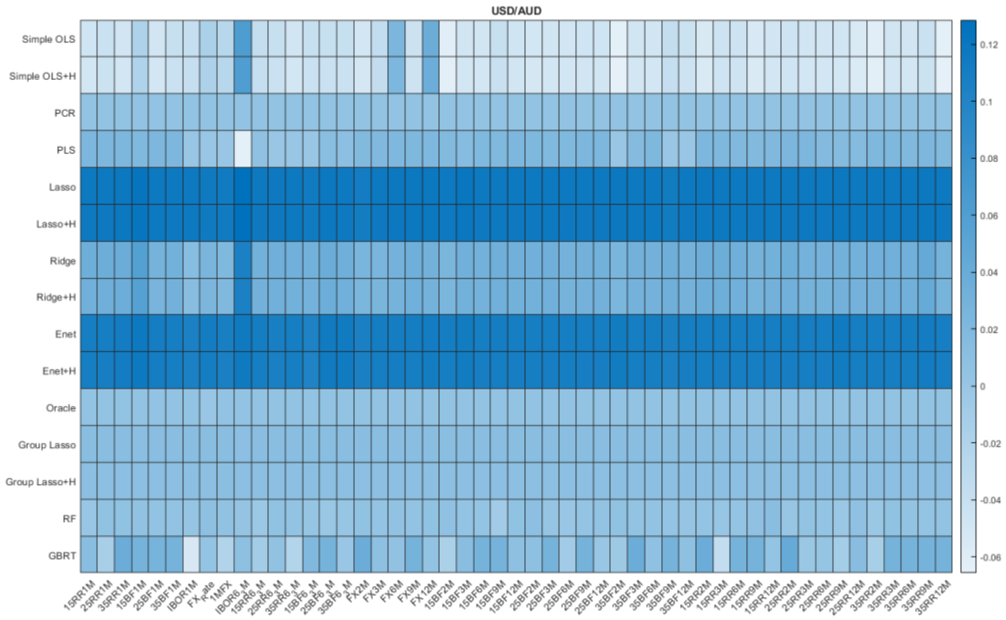


Figure C.3: Heatmap USD/AUD for 15 regression models.

Fig.C.3. This figure shows the heatmap for each of the 15 regression models, looping over each of the 51 variables.

Figure C.3.1: Top 20 covariates for USD/AUD for 13 regression models.

Fig.C.3.1. This figure shows the top 20 covariates for each of the 13 regression models

References

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

Ahmad, M. H., Adnan, R., & Adnan, N. (2006). A comparative study on some methods for handling multicollinearity problems. *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, 109-119. https://doi.org/10.11113/matematika.v22.n.179

Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques – Part II: Soft computing methods. *Expert Systems with Applications*, *36*(3), 5932-5941. https://doi.org/10.1016/j.eswa.2008.07.006

Bali, T. G., Beckmeyer, H., Moerke, M., & Weigert, F. (2021). Option return predictability with machine learning and big data. *Georgetown McDonough School of Business Research Paper*, (3895984).

Bank for International Settlements (BIS). (2019). *Triennial Central Bank Survey: Foreign exchange turnover in April 2019*. https://www.bis.org/statistics/rpfx19_fx.pdf

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one*, *12*(7), e0180944. https://doi.org/10.1371/journal.pone.0180944

Bekaert, G. and Hodrick, R.J. (2018) *International Financial Management.* (3rd ed.). Cambridge: Cambridge University Press.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637-654. https://doi.org/10.1086/260062

Cao, L. J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, *14*(6), 1506-1518. https://doi.org/10.1109/TNN.2003.820556

Carr, P., & Wu, L. (2007). Stochastic skew in currency options. *Journal of Financial Economics*, *86*(1), 213-247. https://doi.org/10.1016/j.jfineco.2006.03.010

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, *69*(5), 2045-2084. https://doi.org/10.1111/jofi.12186

Chen, S. H., Chang, C. L., & Du, Y. F. (2010). Agent-based economic models and econometrics. *The Knowledge Engineering Review*, *25*(4), 379-404. https://doi.org/10.1017/S0269888912000136

Chen, T., Ma, K., & Xiao, C. (2018). FastGCN: Fast learning with graph convolutional networks via importance sampling. In Proceedings of the 2018 World Wide Web Conference (pp. 1225-1234). https://doi.org/10.1145/3178876.3186046

Corte, P.D., Kozhan, R. and Neuberger, A. (2021). The cross-section of currency volatility premia. *Journal of Financial Economics, 139*(3): 950–970. https://doi.org/10.1016/j.jfineco.2020.08.010

Cusimano, J. (2023*) Forex vs stocks which is better?*. Statrys. https://statrys.com/blog/forex-vs-stock-market

Dietterich, T. G. (2000). *Ensemble methods in machine learning. In International workshop on multiple classifier systems* (pp. 1-15). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45014-9_1

Efron, B. and T. Hastie. (2021). *Computer Age Statistical Inference, Student Edition: Algorithms, Evidence, and Data Science*. Volume 6. Cambridge University Press

Eichengreen, B., Lafarguette, R., & Mehl, A. (2016). Cables, sharks and servers: Technology and the geography of the foreign exchange market. *National Bureau of Economic Research* [Working Paper] No. 21884.

Feng, G., Giglio, S. and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance, 75*(3): 1327–1370. https://doi.org/10.1111/jofi.12883

Fullwood, J., James, J. and Marsh, I.W. (2021). Volatility and the cross-section of returns on FX options. *Journal of Financial Economics, 141*(3): 1262–1284. https://doi.org/10.1016/j.jfineco.2021.04.030

Garman, M. B., & Kohlhagen, S. W. (1983). Foreign currency option values. *Journal of International Money and Finance, 2*(3), 231-237. https://doi.org/10.1016/S0261-5606(83)80001-1

Goyenko, R., & Zhang, C. (2020). The joint cross section of option and stock returns predictability with Big Data and machine learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3747238

Gu, S., Kelly, B., & Xiu, D. (2019). *ML_Codes*. GitHub. https://github.com/xiubooth/ML_Codes

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223-2273. https://doi.org/10.1093/rfs/hhaa009

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, *51*(5), 1-42. https://doi.org/10.1145/3236009

Hastie, T., R. Tibshirani, and J. Friedman. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (2nd ed.) Springer series in statistics. Springer.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences, 44*(1), 1-12. https://doi.org/10.1021/ci0342472

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hull, B. and Sinclair, E. (2022). The risk-reversal premium. *The Journal of Investment Strategies, 11*(1), pp. 59–75. https://doi.org/10.21314/JOIS.2022.003

Hull, J. C. (2017). *Options, Futures, and Other Derivatives* (10th ed.). Pearson.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: springer.

Jensen, T. I., B. T. Kelly, S. Malamud, and L. H. Pedersen. (2022). *Machine learning about optimal portfolios*. [Working paper]. Copenhagen Business School.

Jurek, J.W. (2014). Crash-neutral currency carry trades. *Journal of Financial Economics, 113*(3): 325–347. https://doi.org/10.1016/j.jfineco.2014.05.004

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications, 38*(6), 7533-7549. https://doi.org/10.1016/j.eswa.2008.01.003

King, M. R., Osler, C. L., & Rime, D. (2013). The market microstructure approach to foreign exchange: Looking back and looking forward. *Journal of International Money and Finance, 38*, 95-119. https://doi.org/10.1016/j.jimonfin.2013.05.004

Klein, M. W. (2013). *Exchange Rate Regimes: Choices and Consequences*. MIT Press

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems* (pp. 1097-1105).

Levinson, M. (2005). *Guide to Financial Markets* (4th ed.). The Economist.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*(3), 18-22.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77-91. https://doi.org/10.2307/2975974

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science, 4*(1), 141-183. https://doi.org/10.2307/3003143

Molnar, C. (2014). *Interpretable machine learning: A guide for making Black Box models explainable.* Munich: Christoph Molnar.

Moody, J., & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks, 12*(4), 875-889. https://doi.org/10.1109/72.935097

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. " O'Reilly Media, Inc.".

Müller, K., & Schmickler, S. (2020). Interacting anomalies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3646417

Neuhierl, A., Tang, X., Varneskov, R. T., & Zhou, G. (2021). Option Characteristics as Cross-Sectional Predictors. *SSRN Electronic Journal.* http://dx.doi.org/10.2139/ssrn.3795486

Ong, C. S., Huang, J. J., & Tzeng, G. H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications, 29*(1), 41-47. https://doi.org/10.1016/j.eswa.2005.01.003

Polikar R (2006) Ensemble based systems in decision making. *IEEE Circuits Syst Mag 6*(3):21–45. https://doi.org/10.1109/MCAS.2006.1688199

Raina, R., Madhavan, A., & Ng, A. Y. (2009). Large-scale Deep Unsupervised Learning using Graphics Processors. *Proceedings of the 26th Annual International Conference on Machine Learning,* 873-880. doi:10.1145/1553374.1553486

Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, *33*(1-2), 1-39.

Tay, F. E. H., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, *29*(4), 309-317.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society*: Series B (Methodological), 58(1), 267-288.

Bank for International Settlements. (2022). *Triennial Central Bank Survey of foreign exchange and Over-the-counter (OTC) derivatives markets in*

2022. Bank for International Settlements. https://www.bis.org/statistics/rpfx22.htm

Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, *36*(2), 2473-2480.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, *14*(1), 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7