BI

WISEflow
Europe/Oslo(CEST)
03 Jul 2023

# Handelshøyskolen BI

## GRA 19703 Master Thesis

### Thesis Master of Science 100% - W

#### 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\|\|11184\|\|IN00\|\|W\|\|T | | |
| **Intern sensor:** | (Anonymisert) | | |

#### Deltaker

| | |
|---|---|
| **Navn:** | **Anders Kristian Aasen** |

#### Informasjon fra deltaker

| | |
|---|---|
| **Tittel *:** | AI and traditional forecasting of currency return using Diebold Li factors from commodities |
| **Navn på veileder *:** | Jamie Cross |

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

#### Gruppe

| | |
|---|---|
| **Gruppenavn:** | (Anonymisert) |
| **Gruppenummer:** | 283 |
| **Andre medlemmer i gruppen:** | Deltakeren har innlevert i en enkeltmannsgruppe |

Programme: Master of science in business, major economics

AI and traditional forecasting of currency return using Diebold Li factors from commodities

Supervisor: Jamie Cross

# Contents

# Abstract

This research paper investigates out of sample forecasting AI compared to traditional models. The data used are oil crude futures prices and the currency pair NOK/USD. The crude oil futures prices is transformed to the commodity basis and termstructure of crude oil futures prices. The Diebold Li factors is extracted from the termstructure. These factors and the currency return are used with various models to forecast currency return. Encouraging results for the use of AI in out of sample forecasting is found.

# 1 Introduction

Research questions:

- Does the DieboldLi factors of commodities contain valuable information for forecasting currency return?
- Can forecast models beat the random walk at forecasting currency return?
- How does traditional models compare to AI, deep neural networks at forecasting currency return?

The ability to forecast exchange rates would be useful to governments, central banks, businesses, investors, and consumers. It would be especially important to a small open economy like Norway because of its dependency on international trade.

AI is used in many fields of science with great success. The field of AI as a science is rapidly expanding because of huge technological advances. Exploring the benefits and applications of AI in economics as in any other field is important. The unknown should always be explored. Maybe AI is the next big thing in economics, helping economists understand economic relationships and solve issues that has puzzled economists for decades, like the equity premium puzzle and exchange rate puzzles.

Frankish and Ramsey (2014) describes Artificial Intelligence (AI) as a cross-disciplinary approach to understanding, modelling, and replicating intelligence and cognitive processes using various computational, logical, mechanical, and biological principles and devices. Research on AI include exploring intelligence, consciousness, rationality, mental representation, perceptual experience, and human action. Historically AI practitioners comes from disciplines as logic, mathematics, engineering, philosophy, psychology, linguistics and computer science (Frankish & Ramsey 2014). Notice the absence of the discipline of economics. We should strive to add ourselves to the list of users and contributors in the field of AI.

Meese and Rogoff (1983a,b, 1988) finds no models to outperform the RW at forecasting exchange rates. The mushrooming literature on out-of-sample forecasting ability have not found economic models explaining exchange rate movements, even ex post (Ravazzolo et al, 2016).

This paper focus on out-of-sample forecasting of the Norwegian commodity currency. A commodity currency is a currency that co-move with commodity prices of a commodity that is important for that that country's export revenues. The NOK is considered a commodity currency. Crude brent oil is a commodity important for Norway's export revenue. An estimated weight of 30 % of Norway's export revenue comes from crude oil (Ravazzolo et al, 2016).

This paper builds on Ravazzolo et al (2016) who finds commodity prices to contain information to explain currency return. They use an asset pricing framework using the basis of commodity futures prices to forecast currency return out-of-sample. This paper extends on this with the termstructure of crude oil futures prices from the commodity basis and the extraction of the Diebold Li factors from it. This paper also extends with different forecasting models, including artificial intelligence to forecast currency return.

The currency return is forecasted out-of-sample using both the Diebold Li factors and currency return using the following models:  random walk, autoregressive models, linear model, autoregressive model with exogenous input, artificial intelligence in the form of a deep neural network and a model combination.

The root mean squared forecast errors of the models are compared. The Diebold Mariano test is used to check for statistical significance of the results. The cumulative sum of squared error differences is used to analyse the model's performance over time.

The paper is structured as follows:

Section 2 reviews earlier literature. Section 3 presents the theoretical framework. Section 4 describes the data. Section 5 describes the theoretical framework. Section 5 presents the forecasting results. Section 6 is a simple profitability analysis. Section 7 concludes.

# 2 Literature review

Meese and Rogoff (1983a,b, 1988) finds no models to outperform the RW at forecasting exchange rates. The mushrooming literature on out-of-sample forecasting ability have not found economic models explaining exchange rate movements, even ex post (Ravazzolo et al, 2016).

Engel and West (2005) explain the exchange rate puzzle with asset-pricing models, linking exchange rates and their fundamentals. Ravazzolo et al (2016) finds commodity prices to contain information in the commodity prices to explain currency return using an asset pricing framework. Chen and Rogoff (2003, 2012) find models based on commodity prices do not consistently outperform a random walk in out-of-sample forecasting.

McCulloch and Pitts (1943) creates a very simplified model of a functioning neuron that performs any Boolean operation. The subfield of AI called neural networks is created. Neural networks are intelligent systems of nodes in which each node contain a simplified model of a neuron, trying to replicate a nervous system (Franklin et al. 2014).

Arthur Samuel (1959) creates a program that not only plays checkers but learns and improvs. Samuel was supposedly not able to beat the program after a few months of learning.  The major subfield of AI called machine learning is created. Machine learning can be thought of as algorithms that train on training data to learn to perform tasks it was not explicitly programmed to do (Franklin et al. 2014).

Artificial neural networks are great at pattern recognition and are used in mutual fund-investing, fraud detection, credit scoring, real estate appraisal and many more (Franklin et al. 2014). Leung et al (2000), Huang et al (2004), Nag and Mitra (2004) amongst many find positive results in forecasting exchange rates using AI, neural networks. Adekoia et al (2021) and Datta et al (2021) finds positive results when using long short-term memory (LSTM), deep learning for predicting exchange rates.

# 3 Theoretical framework

## 3.1 Commodity currency

A commodity currency is a currency that co-move with commodity prices of a commodity that is important for that that country's export revenues. The NOK is considered a commodity currency. Crude brent oil is a commodity important for Norway's export revenue. An estimated weight of 30 % of Norway's export revenue comes from crude oil (Ravazzolo et al, 2016).

***Figure 3.1 Commodity Currency and FNB Crude Oil Futures contract price***

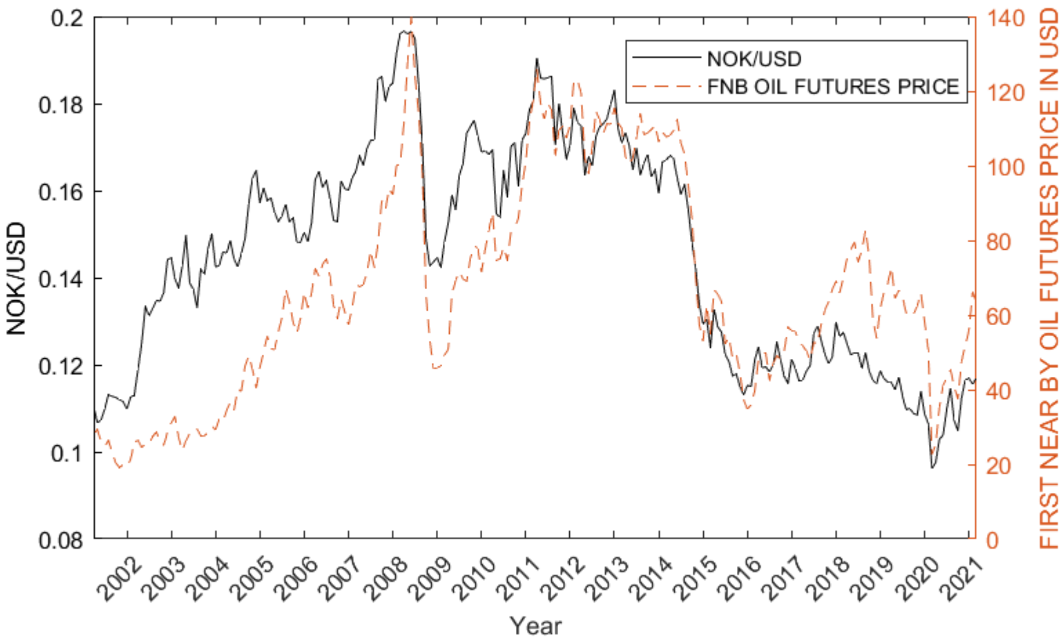



Figure 3.1 supports NOK and oil to be an exchange rate-commodity pair. It looks to be co-movement between the NOK/USD currency pair and the price of the first nearby (FNB) of crude oil futures contract. The FNB contract is the contract with the closest settlement date. Literature uses FNB futures contract price as a proxy for spot price (Ravazzolo et al, 2016).

A puzzling fact is there seems like the NOK/USD and oil price stop co-move from 2016. This implies a structural break. This could have implications for the performance of the models used for forecasting. One explanation is the Nordic exchange market are looked upon as one market by investors. Then the Scandinavian have the same stochastic discount factor and the currencies will be priced with the same stochastic discount factor. A market risk change will give a percent vice change with the same amount and in the same direction for all Nordic exchange rates. In a more unstable world the Swedish krone(SEK) is looked upon as a safe currency. Combined with Sweden's' low interest rates this make the SEK a lucrative carry trade currency. The SEK increases in value. If investors look upon the Nordic exchange market as one market, the value of the NOK increase as well. Therefore, one can argue the correlation with the SEK to be the reason the NOK and oil price stop correlate.

## 3.2 Theory of storage

The theory of storage was established by Kaldor (1939), Working (1949) and Brennan (1959) These equations explains the theoretical intuition behind the theory.

$$F(t,T) - S(t) = S(t)r(t,T) + W(t,T) - C(t,T)$$

Equivalently,

$$\frac{F(t,T) - S(t)}{S(t,T)} = r(t,T) + \frac{W(t,T) - C(t,T)}{S(t)}$$

Where F(t,T) is the commodity futures price at time t, maturing at time T. S(t) is the spot price of the commodity at time t. The return from purchasing the commodity at time t and selling it for delivery at time T is equal to the forgone interest, S(t)R(t,T) and the marginal storage cost, W(t,T), subtracted the marginal convenience yield from an additional unit of inventory, C(t,T). The left-hand side of the above equations is the basis. The basis in the futures market is the difference between the spot price and futures price of a commodity. A positive basis implies a higher futures contract price than the spot price. This occurs when the convenience yield net of storage costs is lower than the risk-free rate. A negative basis implies a higher spot price than future price because of higher risk-free rate than convenience yield net of storage costs. The marginal convenience yield can be explained as the value of holding a commodity opposed to holding a futures contract or the value of storing commodities to meet unexpected positive or negative demand or supply shocks, respectively (Ravazzolo et al, 2016)

Ravazzolo et al (2016) use the theory of storage in an asset pricing setup and ads a risk premium into the theory of storage.

$$b_{t,n} = E_t[cy_{t,n}] - sc_{t,n} - r_{t,n} - \mu^{cy}_{t,n}$$

Where $b_{t,n}$ is the relative commodity basis with maturity n, $cy_{t,n}$ is the commodity convenience yield, $sc_{t,n}$ is the commodity storage cost between periods t and t+1, $r_{t,n}$ is the risk-free rate and $\mu^{cy}_{t,n}$ is the risk premium associated with the stochastic nature of the convenience yield of the commodity.

They relate the basis and currency return to the price and quantity of risk and propose a negative relationship between currency return and the basis. An increase in risk aversion of investors, increase the price of risk. Investors will then demand higher compensation for taking on the risk. This includes all risky assets including the commodity portfolio. The price of the commodity portfolio, the commodity basis, should then fall. The currency return should increase.

### 3.3 Diebold Li factors

The Nelson Siegel factors are most commonly used in describing the yield curve. Empirical studies find the three factors level, slope and curvature to capture more than 99 % of treasury bond yield movements. Nelson and Siegel (1987) develop a parsimonious model for describing the monotonic, humped and S shaped nature of yield curves.

Nelson Siegel Model(original model):

$$i^j_t = \beta^{NS}_{1,t} + \beta^{NS}_{2,t}\left(\frac{1 - e^{-\lambda j}}{\lambda j}\right) - \beta^{NS}_{3,t}(e^{-\lambda j})$$

Diebold and Li (2006) modify this to:

$$i^j_t = \beta^{NS}_{1,t} + \beta^{NS}_{2,t}\left(\frac{1 - e^{-\lambda j}}{\lambda j}\right) - \beta^{NS}_{3,t}(\frac{1 - e^{-\lambda j}}{\lambda j} - e^{-\lambda j})$$

Where $i^j$ denotes the yield of the j-th maturity. The model is considered a dynamic latent factor model where the time varying $\beta$ parameters capture the value of the level slope and curvature of the yield curve. $\beta_1$, $\beta_2$ and $\beta_3$ are called the Level, Slope and Curvature factors respectively because they describe the level, slope and curvature of the yield curves. The level is considered a long-term factor. The Slope is considered a short-term factor. The curvature is considered a medium-term factor.

This paper extends on Ravazzolo et al (2016) with the termstructure of crude oil futures prices and the extraction of the Diebold Li factor from it. This paper proposes that the termstructure of commodities captures the information in the basis. This information from the termstructure is then captured by the Diebold Li factors. It is assumed that the Diebold Li factors accurately describe the term curve and its movements.

***Figure 3.2 Mesh Surface Plot of the Termstructure of Crude Oil Futures Prices***

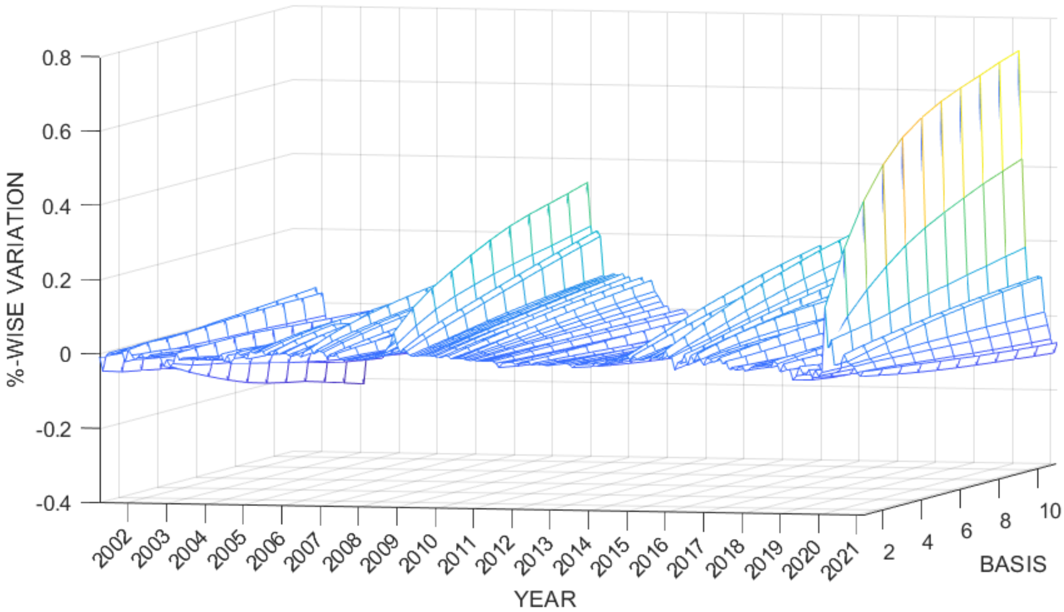
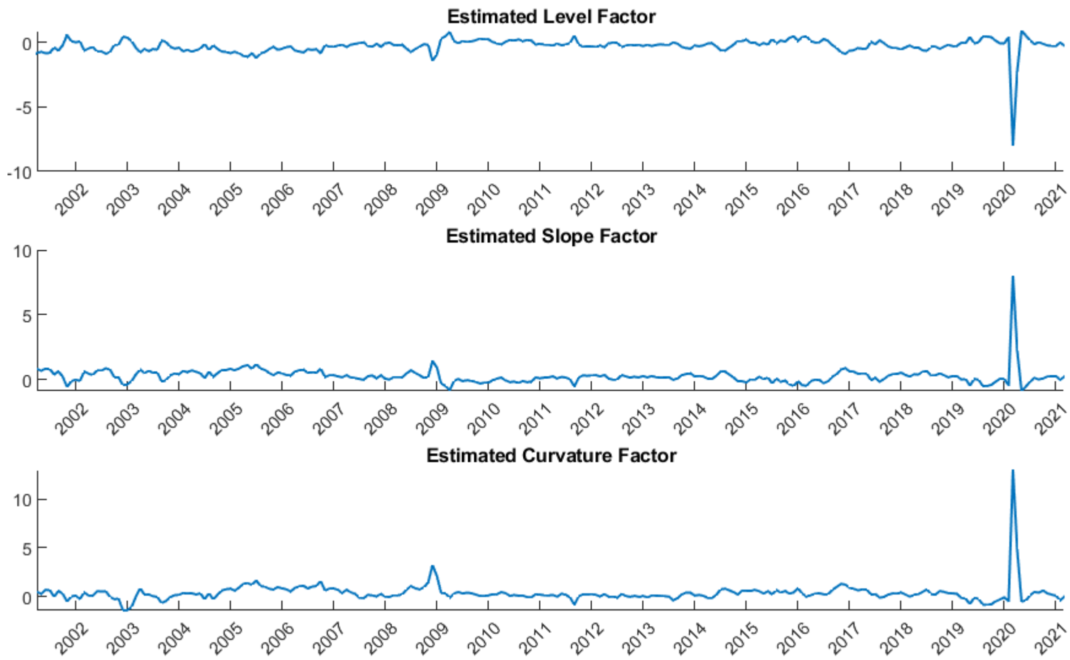



Figure 3.2 show an increase in basis variability for longer maturities. The termstructure is forward looking and assumed to contain information about the future. The information in the commodity basis is assumed to be captured by thetermstructure and its term curves.

Estimating the Diebold Li factors from the termstructure of crude oil futures prices using the standard $\lambda = 0.0609$ from the literature (Diebold and Li, 2006) gives figure 3.3.

***Figure 3.3 Estimated Diebold Li factors***




Notice the spikes of the Diebold Li factors in early 2020. This is when the corona virus hit globally. It will be interesting to see how the models perform during and after these spikes.

# 4 Data description

## 4.1 Data collecting

All data are monthly observations, collected from Datastream, Refinitiv Eikon. Datastream is a comprehensive economic and financial database. The exchange rate NOK/USD is from Global Trade Information Services (GTIS), FTID/TR. The crude oil futures contracts are traded on the Intercontinental Exchange (ICE). Futures contracts with 1 to 12 month maturity are used because of sufficient liquidity. Futures contracts with longer maturity lack liquidity and are considered not informative (Ravazzolo et al, 2016). All data are collected end-of-month on the last business day of the month. For the AI model using a big dataset a total of 67 timeseries are used. See appendix for description of the big dataset.

Sample period is April 2001 to March 2021. In March 2001, Norway went from a fixed to a floating exchange rate regime with inflation targeting. Preceding the monetary policy change was a transition period (Kleivset 2012, Gjedrem 2001).

The transition period could have unstable monetary policy behaviour. Data from transition period are therefore excluded.

Information gathered from Datastream platform:

The exchange rate data are mid-market rates. Mid-market rate is the average of bid and ask rates. The observations are monthly, last business day at 10 PM London time.

The ICE Brent Crude futures contract is a deliverable contract based on EFP delivery with an option to cash settle. The contract size is 1000 barrels and are traded in US Dollars and cents. Last trading day is the end of the designated settlement period on the last business day of the second month preceding the relevant contract month. The March contract will expire on the last business day of January. Daily settlements are the average price of trades during a two-minute settlement period from 19:28:00 to 19:30:00 London time. There is a daily margin where all open contracts are marked-to-market daily. The contracts can be traded as electronic futures, exchange of futures for physical (EFP), exchange of futures for swap (EFS) and block trades.

## 4.2 Data transformation

The basis is created as the relative change of the crude oil futures prices. The first-near-by(1 month maturity futures) is the reference. From each point in time the basis maturities create one term curve. The term curves make up the termstructure. From the termstructure, the level, slope and curvature are estimated following Diebold and Li (2006). The currency return is calculated as the percentage change from one period to the next. Log difference gives less variation and is not used as an approximation for any data as higher variation is better for forecasting. For the AI model using a big dataset, a total of 67 timeseries is used. This include different US treasury bonds, different exchange rates(SEK/USD, DKK/USD, EUR/USD), CPI, unemployment, housing prices and inflation in US, Denmark, Sweden, Finland and Norway.

*Table 4.1 Descriptive statistics on the basis of crude oil futures prices*

| *Variable* | *Mean* | *Median* | *Min* | *Max* | *SD* | *Var* |
|---|---|---|---|---|---|---|
| **1 month** | 0.0022 | 0.0008 | -0.0457 | 0.1588 | 0.0178 | 0.0003 |
| **2 month** | 0.0046 | 0.0008 | -0.0617 | 0.3043 | 0.0314 | 0.0010 |
| **3 month** | 0.0060 | 0.0000 | -0.0848 | 0.4178 | 0.0432 | 0.0019 |
| **4 month** | 0.0069 | -0.0026 | -0.1083 | 0.5009 | 0.0533 | 0.0028 |
| **5 month** | 0.0073 | -0.0063 | -0.1290 | 0.5607 | 0.0622 | 0.0039 |
| **6 month** | 0.0074 | -0.0076 | -0.1443 | 0.6007 | 0.0699 | 0.0049 |
| **7 month** | 0.0071 | -0.0087 | -0.1565 | 0.6302 | 0.0768 | 0.0059 |
| **8 month** | 0.0066 | -0.0105 | -0.1671 | 0.6526 | 0.0830 | 0.0069 |
| **9 month** | 0.0060 | -0.0121 | -0.1827 | 0.6724 | 0.0889 | 0.0079 |
| **10 month** | 0.0054 | -0.0143 | -0.1991 | 0.6935 | 0.0946 | 0.0090 |
| **11 month** | 0.0045 | -0.0178 | -0.2156 | 0.7111 | 0.1000 | 0.0100 |

Notice the increasing variance in maturity length.

## 4.3 Data evaluation

Measurement errors could come from the source of the data, provider of the data, the downloading of datafiles or the transformation of data. Examining the data, no missing datapoints or unusual values are found. The possibility of data errors is considered small.

Alternative data:

The natural gas futures prices are the other viable alternative. The cost of storage for gas is higher than the cost of storage for oil. Therefore, the basis varies more with a larger convenience yield. Ravazzolo et al. (2016) argues commodities with higher storage costs contain a lager risk premium/ convenience yield. Natural gas could therefore be more suitable for forecasting the NOKUSD. The downside of gas futures prices is that the market is not as developed. Gas futures prices are only available with limited amount of monthly maturity prices until 2007. Oil futures prices are chosen because number of datapoints are favoured. Ravazzolo et al. (2016) argues commodities with higher storage costs contain a lager risk premium/ convenience yield. Natural gas could therefore be more suitable for forecasting the NOKUSD. Crude oil futures prices is chosen because of more datapoints

Alternative data would be to use the NOK compared to a bundle of currencies. Good alternatives would be the trade weighted exchange rate, WTI, or the import-weighted krone index, I-44. The I-44 is a nominal effective exchange rate index based on the NOK exchange rate against a geometric weighted average of Norway's 44 most important trading partners exchange rates. For computational and intuitive simplicity the currency pair NOK-USD is used as the oil futures prices are in dollars.

One useful transformation of the exchange rate could have been to take out the interest rate differential (Ravazzolo et al, 2016). This is not done because they find no significant relationship between the basis and the interest rate.

# 5 Empirical framework

This section explains the econometric methodology applied for the out-of-sample forecasting exercise. Out-of-sample forecasting is important in economics and finance. Out-of-sample forecasting reduce the probability of model overfitting (Ashley et al. 1980). Swanson and White (1995) finds in-sample predictive ability not to be an indication of good out-of-sample forecasting ability. Selection of window estimation method and size will be discussed. Econometric models to be explained are the random walk model, autoregressive model, linear regression model, the ARX model and a forecast combination. The creation of the deep neural network is described. The two methods of RMSFE and CSSED to evaluate the forecasts are described. Lastly the Diebold Mariano test for significance is described.

**Ordinary Least Squares (OLS) and assumptions**

The Ordinary Least Square (OLS) method is used. The Gauss Markov assumptions are assumed to hold for all models, giving unbiased forecasts.

The relevant hypotheses to be tested to answer the research question:

| |
|---|
| Model j does not outperform model i at forecasting currency return. |
| Model j does outperform model i at forecasting currency return. |

With j and i being the different models applied.

## 5.1 Models

### 5.1.1 Random walk (RW)

The Random Walk (RW) is defined as a time series process depending only on past values of itself and Gaussian white noise errors. The best forecast of $y_{t+1}$ at time t is its past value (Bjørnland and Thorsrud, 2015):

$$E[y_{t+1}|y_t] = y_t$$

Adding a constant term to the RW model gives a deterministic trend, and a stochastic term. This is the Random walk with drift:

$$y_t = \propto t + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

### 5.1.2 Autoregressive (AR) process

The AR(p) model is a timeseries model that relates the value of a variable y to previous values of itself and a random disturbance $\varepsilon$, at time t.

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t$$

AR(1) to AR(10) are used to forecast currency return from past values of currency return. The best model is chosen. Stationarity is assumed to hold.

### 5.1.3 Multiple linear regression model

The multiple linear regression model can be expressed by:

$$y_t = B_0 + B_1 x_{1,t} + B_2\, x_{2,t} + B_i x_{i,t} + \varepsilon_t$$

Where y is de dependent variable to be forecasted and $x_i$ is the independent predictor variable. $B_i$ is the measured effect on y of variable $x_i$ (Hyndman and Athanasopoulos, 2018). The dependent variables used are the Diebold Li factors.

### 5.1.4 Autoregressive-exogenous (ARX) model

An ARX model combines autoregressive variables with exogenous variables. The ARX(n,m) model assumes the current system output is a function of previous n system outputs and previous m system inputs(Horner et al, 2019). An ARX(n,m) is used where n is the autoregressive variable currency return, and m are the exogenous variables; Level, Slope and Curvature. ARX(1,m) to ARX(10,m) are used to forecast currency return. The best model is chosen.

### 5.1.5 Artificial Intelligence

Description and intuition of a deep neural network:

Conventional programming tell the computer exactly what to do by breaking a larger problem into smaller precisely defined portions. You can think of this as a, somewhat advanced, recipe in a cookbook. As a result, depending on the input the result is deterministic (Nielsen, 2015).

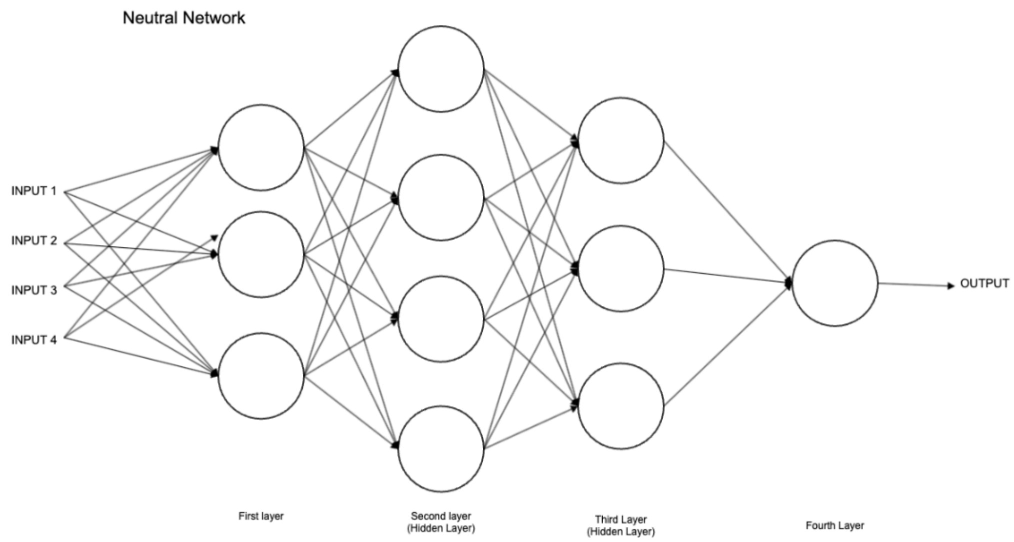
Neural networks an AI (Artificial Intelligence) tries to emulate or make decisions in similar way to the human brain. Humans make decisions based on multiple previous experiences, to some extent intuitive (Nielsen, 2015).

In neural networks the computer is not told what to do, but the neural network can be trained to make the right decisions. The training is done by providing input data with known results (output). The trained neural network can then provide answers to other values of input data. The hard part is to train the neural network to actually provide correct answers (Nielsen, 2015).

The smallest part in the neural network is the neuron/perceptron. A perceptron/neuron can take several inputs and provides a single output. By weighing the different inputs "correctly" the output can be tuned to match the known correct answer. The neuron calculates a result based on the input. Then based on a compare of the result with the neurons Treshold value, the output is either 0 or 1, in other words it is binary. Neurons can take input from other

neurons, consequently also the input values to the neurons are binary (Nielsen, 2015).
A neural network is multiple neurons in a network. Output from each neutron is used as input to several other neurons. The left most column of neurons (below) is called the first layer of perceptrons. The mid layers are hidden, as they have neither input or output. Each and every perceptron in the network is weighing the input. Building up this network of neurons makes is possible to make more complex decisions (Nielsen, 2015).

***Figure 5.1 Neural network***




A neural network with a large number of layers is a Deep neural network.
Figure 5.1 illustrates a feedforward network, where input can only travel from left to right. RNNs allow outputs from a set of neurons to be input to the same set of neurons (feedback). There are mechanisms within RNNs to avoid cyclic loops (Nielsen, 2015).
Long Short-term memory (LSTM) is a further development of RNN. It is suited for making predictions on time-series of data. It improves some of the weaknesses of plain RNN (Nielsen, 2015).

The AI model used in this paper:

A Long-Short-Term-Memory(LSTM), deep neural network is used to forecast currency return. The deep neural network consists of a sequence input layer, LSTM layer, fully connected layer and a regression layer. A sequence-to-sequence regression LSTM network is trained. At each time step of the training set, the

LSTM network learns to forecast the value of the next time step. The window of observations are partitioned into training and test data. The training data is the window size without the last observation. The test data is the window size without the first observation. The LSTM network learns by using the first observation of the training set to predict the next observation, the first observation of the test data. The model then uses the first two observations of the training data to predict the next observation, the second observation in the test data, and so on.  It then compares that prediction to the actual value in the test data. From the deviations between predicted and actual values the model learns as it goes. Threw the learning process the model forecasts the out-of-sample observation. The training data is standardized to have zero mean and unit variance for better fit and prevent diverging training.

### 5.1.6 Model combination (MC)

The model combination combines forecasts from multiple models. Model combinations offer diversification gains. This could give improved forecasting compared to each model individually. A combined forecast could be more robust to misspecification of individual models and structural breaks. (Bjørnland and Thorsrud, 2015):

$$y_t^c = \sum_i^n w_t^i \, \Delta y_t^i$$

The AR model, Linear model, ARX model and AI model using betas are used for the model combination forecast. A linear opinion pool with equal weights is used with equal weight on the individual models:

$$w^i = \frac{1}{4}$$

### 5.1.7 Window estimation method and size

Literature widely recognizes parameter instability as a crucial issue in forecasting (Inoue et al, 2017). There is widespread empirical evidence of parameter instability in financial forecasting (Goyal and Welch, 2003) and exchange rate forecasting (Schinasi and Swami, 1989). The rolling window estimation method produce a sequence of out-of-sample forecasts using a fixed number of the most recent data at each point of time (Inoue et al, 2014). The expanding estimation window always uses all data available.

With expanding window large deviations caused by rarities, such as special events, will affect the forecasts for the rest of the sample. Due to the ability of rolling window to take out these spikes in the data when the rolling window has moved past these data, and thereby eliminates the effect of the spikes on later forecasts when the rolling window is past the spike. On the other hand, rolling window might leave out valuable information from outside window.

Determining the best window method depends on the specific empirical application and the time series data properties. In general, determining the best window method is difficult (Bjørnland & Thorsrud, 2015). Due to the nature of AI, deep learning, the bigger test set gives more learning and is believed to increase forecast accuracy. Expanding window is expected to be optimal for AI forecasting.

Both window estimation methods are used in this paper. The selection of window size is important as forecasting performance is often sensitive to changes in window size (Inoue and Rossi, 2012). For window size there is a trade-off between omitting and including both valuable and not valuable information for forecasting. Inoue et al (2016) develops a method for selecting window size for forecasting. Ravazzolo et al (2016) finds optimistic results with a window size of 60 observations when forecasting currency return. A window size of 60 observations is therefore used as it has been proven to perform.

## 5.2 Evaluation methods

**Out-of-sample forecast evaluation**

### 5.2.1 Root mean squared forecast error (RMSFE)

The *Root Mean Squared Forecast Error* (RMSFE) is the standard statistic for evaluating forecast accuracy. It is a measure of the size of the forecast error, calculated as the square root of the mean squared error:

$$RMSFE = \sqrt{E[(e_{T+h})^2]} = \sqrt{E[(y_{T+h} - \hat{y}_{T+h})^2]}$$

The difference between the actual values, $y_t$, and the predicted values, $\hat{y}_t$, gives the forecast errors. The RMSFE is a symmetric loss function that puts equal weight on negative and positive forecast errors. Smaller forecast errors are considered better than larger ones. Lower RMSFE values indicates better forecasting performance (Bjørnland & Thorsrud, 2015).

### 5.2.2 Cumulative Sum of Squared forecast Error Difference (CSSED)

To evaluate the models forecasting performance over time, the *Cumulative Sum of Squared forecast Error Difference* (CSSED) introduced by Welch and Goyal (2008) is used:

$$CSSED_{m,\tau} = \sum_{\tau=R}^{T} (\hat{e}^2_{bm,\tau} - \hat{e}^2_{m,\tau})$$

This statistic compares the squared error of the benchmark model, denoted by $\hat{e}^2_{bm,\tau}$, to that of the alternative model, denoted by $\hat{e}^2_{m,\tau}$. Parameters R and T denote the beginning and end of the forecast evaluation period, respectively. An increase in the CSSED implies the alternative model to outperform the benchmark. A decrease in the CSSED implies the benchmark to outperform the alternative model.

### 5.2.3 Diebold-Mariano (West) test

The Diebold-Mariano (West) test by Diebold and Mariano (1995) and West (2006) tests for statistically significant difference in two models forecasting performance of the same variable.

Given two models with the following squared forecasting errors:

$$e^2_{T+h+i,1} = \left(y_{T+h+i} - \hat{y}_{T+h+i,1}\right)$$

$$e^2_{T+h+i,2} = \left(y_{T+h+i} - \hat{y}_{T+h+i,2}\right)$$

gives the loss differential:

$$d_{i,h} = e^2_{T+h+i,1} - e^2_{T+h+i,2}$$

A regression is run on:

$$d_{i,h} = \beta_0 + u_i$$

The hypothesis test:

$$H_0: \beta_0 = 0 \text{ vs. } H_1: \beta_0 \neq 0$$

The null hypothesis implies no significant difference in forecasting performance. A rejection of the null hypothesis means the forecast performance is significantly different (Bjørnland & Thorsrud, 2015).

# 6 Forecasting results

## 6.1 Full sample

***Table 6.1 RMSFE of forecasts***

| *Modell* | *RMSFE* |
|---|---|
| **AI Big dataset rolling** | 3.1445 |
| **AI Big dataset expanding** | 3.2730 |
| **AI Betas expanding** | 3.2870 |
| **Model Combination rolling** | 3.3136 |
| **AI Currency Return expanding** | 3.3572 |
| **Model Combination expanding** | 3.3979 |
| **RW without drift** | 3.4504 |
| **RW with drift rolling** | 3.4685 |
| **AR(1) expanding** | 3.4739 |
| **RW with drift expanding** | 3.4752 |
| **ARX(1) expanding** | 3.4973 |
| **AI Betas rolling** | 3.5208 |
| **AR(1) rolling** | 3.5316 |
| **ARX(1) rolling** | 3.5975 |
| **Linear model Betas expanding** | 3.6032 |
| **AI Currency Return rolling** | 3.6102 |
| **Linear model Betas rolling** | 3.7192 |

Table6.1 show the models in descending order on forecasting performance. Lower RMSFE is better forecasting performance The AI models tend to outperform the RW models. Six AI models outperform the RW models. Two AI models are outperformed by the RW models. AI models tend to outperform traditional models. Traditional models tend to be outperformed by the RW models. The RW without drift perform best of the RW models. The AI model using the big dataset perform best of all models. The linear model using betas perform amongst the worst. The AR model is the only traditional model that outperform the rolling RW model with drift. Including an AI model with traditional models in a forecast combination model gives predictive gains using rolling window, but predictive losses using expanding window.

Discussion

It is interesting that the traditional linear model with betas perform bad while the expanding AI model using betas perform good. This suggest the AI model captures information in the betas that the traditional model does not. The expanding AI models using betas and currency return outperform their rolling counterparts. This is as predicted as AI is believed to benefit from more data. The expanding AI model using currency return outperform the traditional AR models. This suggests the AI model captures more information than the traditional AR model. The expanding and rolling linear AI models outperform the expanding and rolling AI models using currency return. This suggest the AI model captures more information in the betas than in the currency return. The AI model using the big dataset perform best of all AI models, as expected, as more datapoints is considered an advantage. It is surprising that the rolling AI model using the big dataset, outperform its expanding counterpart. This suggest that the LSTM network is not as able to filter what data are useful and what data are not. This suggest that what window method and window size is used is not a given in forecasting using AI models. Window method and window size should be explored when using AI models in forecasting exercises.

***Table 6.2 Relative RMSFE***

| *Model* | *RW no drift* | *RW drift roll* | *RW drift exp* |
|---|---|---|---|
| **RW drift roll** | 1.005 | | |
| **RW drift exp** | 1.007 | 1.002 | |
| **AR roll** | 1.024 | 1.018 | 1.016 |
| **AR exp** | 1.007 | 1.002 | 1.000 |
| **Lin β roll** | 1.078** | 1.072** | 1.070** |
| **Lin β Exp** | 1.044* | 1.039 | 1.037 |
| **ARX roll** | 1.043 | 1.037 | 1.035 |
| **ARX exp** | 1.014 | 1.008 | 1.006 |
| **AI β roll** | 1.020 | 1.015 | 1.013 |
| **AI β exp** | 0.953 | 0.948 | 0.946 |
| **MC roll** | 0.960 | 0.955 | 0.953 |
| **MC exp** | 0.985 | 0.980 | 0.978 |
| **AI CR roll** | 1.046 | 1.041 | 1.039 |
| **AI CR exp** | 0.973 | 0.968 | 0.966 |
| **AI BD roll** | 0.911 | 0.907 | 0.905 |
| **AI BD exp** | 0.949 | 0.944 | 0.942 |

A relative RMSFE < 1 means the left hand side model outperform the right hand side model. A relative RMSFE > 1 means the left hand side model outperform the right hand side model.

Table 6.2 show the RW models to outperform the rolling linear model at a 5 % level. RW w/o drift outperform the exponential linear model at a 10 % level. No AI model is significantly different than the RW models.

| Model | AR roll | AR exp | Lin β roll | Lin β exp | ARX roll | ARX exp | AI β roll | AI β exp | MC roll | MC exp | AI CR roll | AI CR exp | AI BD roll |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AR exp | 0.984 | | | | | | | | | | | | |
| Lin β roll | 1.053 | 1.071** | | | | | | | | | | | |
| Lin β Exp | 1.020 | 1.037 | 0.969* | | | | | | | | | | |
| ARX roll | 1.019 | 1.036 | 0.967 | 0.998 | | | | | | | | | |
| ARX exp | 0.990 | 1.007 | 0.940* | 0.971 | 0.972 | | | | | | | | |
| AI β roll | 0.997 | 1.014 | 0.947 | 0.977 | 0.979 | 1.007 | | | | | | | |
| AI β exp | 0.931 | 0.946 | 0.884** | 0.912* | 0.914* | 0.940 | 0.934 | | | | | | |
| MC roll | 0.938** | 0.954* | 0.891*** | 0.920** | 0.921*** | 0.947** | 0.941 | 1.008 | | | | | |
| MC exp | 0.962** | 0.978 | 0.914*** | 0.943** | 0.945* | 0.972** | 0.965 | 1.034 | 1.025 | | | | |
| AI CR roll | 1.022 | 1.039 | 0.971 | 1.002 | 1.004 | 1.032 | 1.025 | 1.098 | 1.090 | 1.062 | | | |
| AI CR exp | 0.951 | 0.966 | 0.903* | 0.932 | 0.933 | 0.960 | 0.954 | 1.021 | 1.013 | 0.988 | 0.930 | | |
| AI BD roll | 0.890 | 0.905 | 0.845** | 0.873* | 0.874* | 0.899 | 0.893 | 0.957 | 0.949 | 0.925 | 0.871* | 0.937 | |
| AI BD exp | 0.927 | 0.942 | 0.880** | 0.908* | 0.910* | 0.936 | 0.930 | 0.996 | 0.988 | 0.963 | 0.907 | 0.975 | 1.041 |

Table 6.3 show the models relative RMSFE with significance testing using the Diebold Mariano test statistic. The model combination performs the best in terms of outperforming other models at the highest significance level. The linear model performs worst in terms of being outperformed by other models at the highest significance level.

Comparing traditional models finds the rolling linear model to be outperformed by the exponential AR model, exponential linear model and the exponential ARX model at the 5%, 10% and 10% level respectively.

Comparing AI models finds the rolling AI model using big dataset to outperform the rolling AI model using currency return at the 10% level.

Comparing AI models to traditional models finds the exponential AI model using betas to outperform the rolling and exponential linear models at the 5% and 10% level respectively. The model combination models outperform at different significance levels all but one of the traditional models. The model combinations perform especially well against the rolling and exponential linear models at the 1% and 5% level respectively.

**Discussion**

The exponential AI model using betas outperform the rolling and exponential linear model using betas at 5% and 10% level, respectively. This suggests the AI model captures information in the betas better than the linear model.

The model combination has the most significant results. It outperform most of the traditional models at the 10%, 5% and 1% level.

While no model outperforms any of the RW models at any significance level, significant results when comparing the other models to another is found.

All but one model outperforms the linear models, with many at different significant levels. The linear model looks to be suboptimal at this forecasting exercise using Diebold Li factors.

The AR models outperform the linear model with only AR expanding outperforming linear rolling significantly, at a 5% level. This indicates that past values of currency return contain more information about future values of currency return than betas do.

The MC model perform the best compared to many of the models with being significantly better than many of them at different significance levels.

The LSTM B roll outperform the linear B model with both window estimations, but not significantly. The AI B expanding outperform the linear rolling and linear expanding at a 5% and 10% significance level respectively. This indicates the LSTM network captures more information than the traditional linear model.

The model combination models with both window estimations outperform many of the traditional models with the highest significance level.

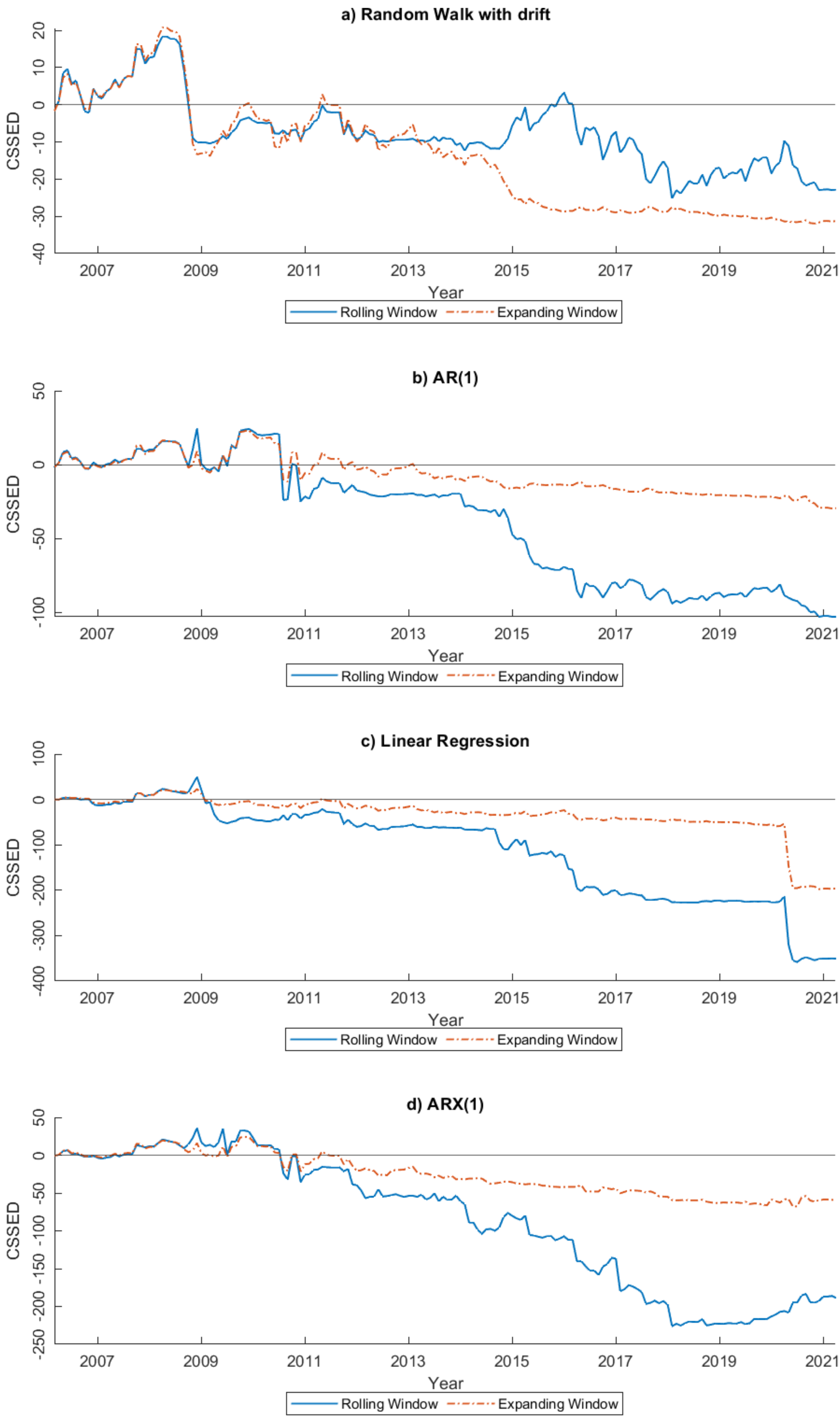
It is interesting to see that some models using LSTM with a rolling window outperforms the same model using an expanding window. This was not expected as LSTM favours more datapoints.

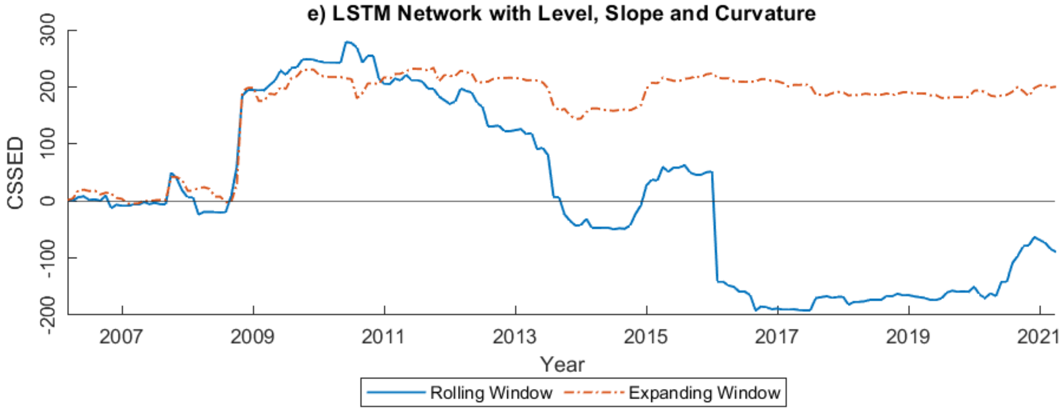
## 6.2 CSSED

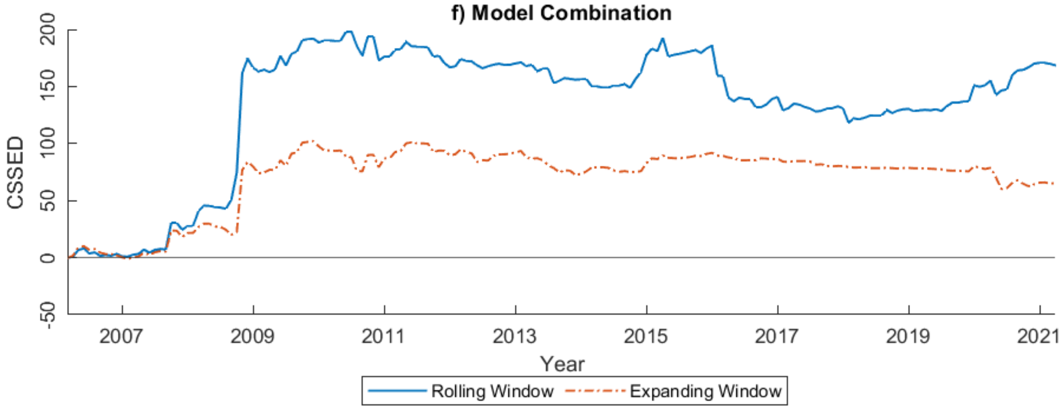
CSSED shows the difference in the squared forecast errors between two timeseries. In the above figure each timeseries forecast is compared to the RW forecast timeseries set as the benchmark.

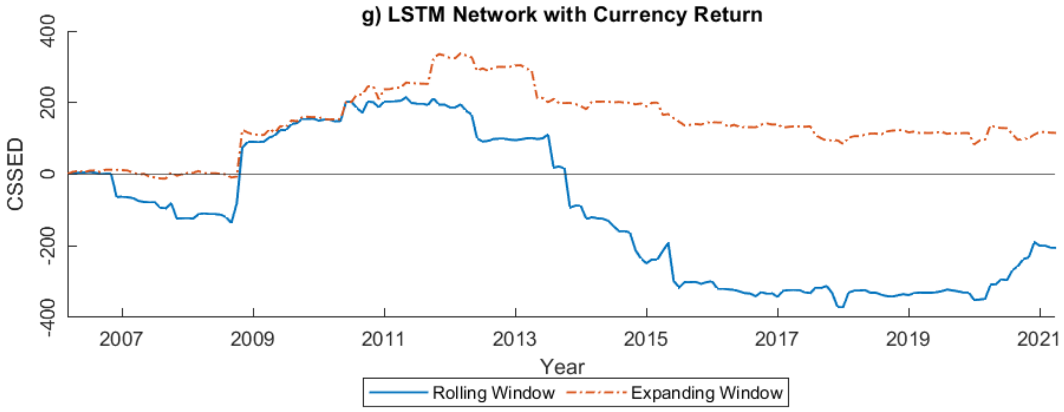
$$CSSED_{m,\tau} = \sum_{\tau=R}^{T} (\hat{e}_{bm,\tau}^2 - \hat{e}_{m,\tau}^2)$$

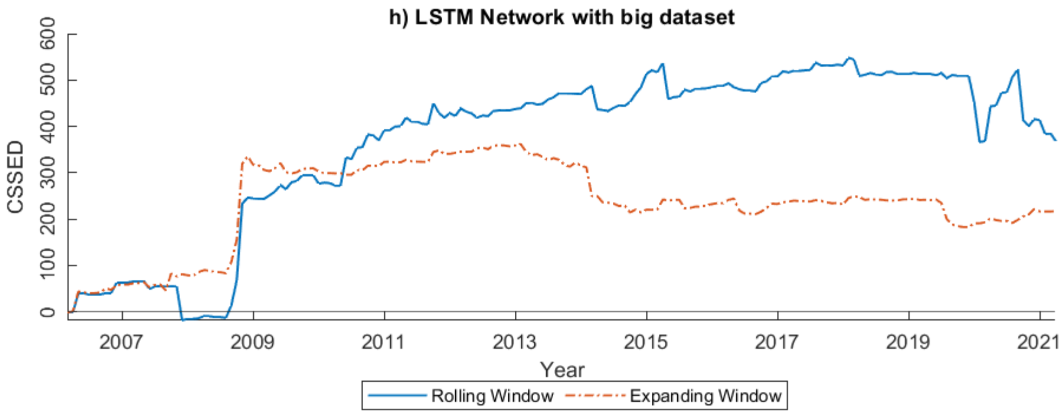
The formula shows that an increase in CSSED is the result of the models' outperformance of the benchmark. A decline of the CSSED shows the benchmark outperforms the model.


a) Random Walk with drift
CSSED
Year
Rolling Window
Expanding Window
b) AR(1)
CSSED
Year
Rolling Window
Expanding Window
c) Linear Regression
CSSED
Year
Rolling Window
Expanding Window
d) ARX(1)
CSSED
Year
Rolling Window
Expanding Window




a) shows an early outperformance of the benchmark the first 24 periods for moving average for both window options. Rolling window has periods of outperforming the RW. Over the timespan of forecasting, the BM outperforms both models. Rolling window outperforms expanding window.

b) Shows AR(1) to perform well for the first 48 months for both window options. The benchmark outperforms consistently, by a small margin, expanding window. Rolling window have shorter periods of worse forecasts, but performs equally good or better than the benchmark for some periods The BM systematically outperforms A(1) for both window options. The expanding window option outperforms the rolling window option.

c) Shows linear regression to perform well for the first 30 months. For the rest of the period expanding window outperforms rolling window. The BM outperforms linear regression for both window options. Expanding window outperforms rolling window.

d) Shows ARX(1) to perform well the first 60 months. Then the expanding window perform consistently, by a small margin, worse than BM. Rolling window outperform BM the last 40 months. The BM outperform both window options. Expanding window outperforms rolling window.

e) Shows the LSTM network finds information in the level, slope and curvature to outperform the benchmark in some periods. Expanding window outperforms the BM from mid 2008 to mid 2010. After mid 2010 expanding window and BM performs similarly. Rolling window performs slightly better than expanding window until mid 2010. Then it performs worse than benchmark and expanding window until 2014. From late 2014 rolling window performs better than BM and expanding window until 2016. From 2016 rolling window performs similarly to expanding window and BM until Rolling window performs better at the end of the forecasting timeframe. Expanding window outperforms the BM and rolling window. The BM outperforms the rolling window. Notice that rolling window outperform the BM and expanding window in some time intervals.

f) Shows the model combination perform best in the short interval before 2009 for both window options, but especially for rolling window. Rolling window perform better than the BM and expanding window in some periods and worse in others. Both window options outperform the BM over the forecasting timeframe.

g) Shows the LSTM network finds information in the currency return valuable for forecasting currency return. Expanding window outperforms the BM from late 2008 until 2012. Then it performs slightly worse for the rest of the forecasting period. Rolling window perform worse than BM and expanding window until late 2008 and between 2012 and mid 2015. From 2020, rolling window outperforms the BM and expanding window.

h) Shows the LSTM network finds information in the big dataset valuable for forecasting currency return. Both window options perform best in the short interval around mid 2008. While Expanding window falls off and perform similarly to the BM, the rolling window continue to outperform the BM and hence the expanding window. At the end of the forecasting timeframe the rolling window performance varies more.

Notice that some of the models have a shift in late 2014 and early 2020 where the performance between rolling window and expanding window go separate ways. This could imply a structural break. There looks to be short timeframes with large forecasting errors. This could come from volatility clustering, that large changes tend to be followed by large changes. Future research could look into this.

The dataset is split into pre-/post corona based on the results from the CSSED.

## 6.3 Pre/post pandemic

Table 6.3 relative RMSFE

| | pre korona | | | after korona | | |
|---|---|---|---|---|---|---|
| Model | RW no drift | RW drift roll | RW drift exp | RW no drift | RW drift roll | RW drift exp |
| RW drift roll | 1.004 | | | 1.012 | | |
| RW drift exp | 1.008 | 1.004 | | 1.001 | 0.989 | |
| AR roll | 1.023 | 1.019 | 1.014 | 1.028* | 1.015* | 1.027* |
| AR exp | 1.006 | 1.002 | 0.998 | 1.011 | 0.999 | 1.010 |
| Lin β roll | 1.060** | 1.056** | 1.051* | 1.170 | 1.155 | 1.168 |
| Lin β Exp | 1.015 | 1.011 | 1.007 | 1.187 | 1.172 | 1.186 |
| ARX roll | 1.058* | 1.054* | 1.049 | 0.958 | 0.947 | 0.958 |
| ARX exp | 1.018 | 1.014 | 1.010 | 0.991 | 0.979 | 0.990 |
| AI β roll | 1.043 | 1.039 | 1.034 | 0.894 | 0.883 | 0.893 |
| AI β exp | 0.949 | 0.945 | 0.941 | 0.973 | 0.961 | 0.972 |
| MC roll | 0.962 | 0.958 | 0.954 | 0.953 | 0.941 | 0.952 |
| MC exp | 0.979 | 0.975 | 0.971 | 1.015 | 1.003 | 1.014 |
| AI CR roll | 1.087 | 1.083 | 1.078 | 0.798* | 0.789* | 0.798* |
| AI CR exp | 0.968 | 0.964 | 0.960 | 0.999 | 0.987 | 0.999 |
| AI BD roll | 0.849** | 0.846** | 0.842** | 1.189 | 1.174 | 1.188 |
| AI BD exp | 0.949 | 0.945 | 0.941 | 0.949 | 0.937* | 0.948 |

Table 6.3 compare the different models performance against the RW models pre and post corona. When looking at pre corona, traditional models tend to be outperformed by RW models while models including AI tend to outperform RW models. The traditional linear model is significantly worse than the RW at 5% and 10% levels. The ARX is statistically worse at 10% level than the RW w/o drift and rolling drift. The ARX perform less worse than the linear model, hence past values of currency return could contain valuable information about future values of currency return. The rolling AI big data model is the most interesting result. It performs significantly better than the RWs at the 5% level. The AI rolling currency return model is the only AI model that is outperformed by the BM's, though not significantly. The RW w/o drift is best performing BM.

After corona, the best performing BM is also the RW w/o drift. Some traditional models perform better and some perform worse than the BM's. The rolling AR model perform significantly worse than the RW models at 10% level. The ARX models perform better than the BM's, but not significantly better. The most interesting result is the AI rolling currency return model perform significantly better than the BM's at the 10% level. The AI expanding big data model perform better than BM's and significantly better than rolling RW w/drift at 10 % level. The expanding MC and rolling AI big data model perform worse than the BM, but not significantly.

When comparing before and after corona the most interesting result is the change in performance and significance of some models. The rolling AI big data model is significantly better than BM's at 5% level before corona and worse than BM's after corona, though not significantly worse. This could imply a structural break. The rolling AI currency return model is worse than BM's before corona and significantly better than BM's after corona at 10% level, This could imply a structural break. The expanding AI big data model perform better than BM's both before and after corona and becomes significant after corona at 10% level against rolling RW w/drift. The rolling AR model perform worse than BM's both before and after corona and becomes significantly worse than BM's after corona at 10% level.

Some models perform significantly better or worse before and after corona. There seems to be a change in the models ability to find useful information in the data for forecasting. These results could imply a structural break.

Table 6.4

| Model pre | AR roll | AR exp | Lin β roll | Lin β exp | ARX roll | ARX exp | AI β roll | AI β exp | MC roll | MC exp | AI CR roll | AI CR exp | AI BD roll |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AR exp | 0.984 | | | | | | | | | | | | |
| Lin β roll | 1.036 | 1.053* | | | | | | | | | | | |
| Lin β Exp | 0.993 | 1.009 | 0.958** | | | | | | | | | | |
| ARX roll | 1.034 | 1.051* | 0.998 | 1.042 | | | | | | | | | |
| ARX exp | 0.995 | 1.012 | 0.961 | 1.003 | 0.962* | | | | | | | | |
| AI β roll | 1.020 | 1.037 | 0.984 | 1.027 | 0.986 | 1.024 | | | | | | | |
| AI β exp | 0.928 | 0.943 | 0.895** | 0.934 | 0.897* | 0.932 | 0.910 | | | | | | |
| MC roll | 0.940** | 0.956 | 0.908*** | 0.947* | 0.909*** | 0.945* | 0.922 | 1.014 | | | | | |
| MC exp | 0.957** | 0.973 | 0.924*** | 0.964* | 0.926*** | 0.962** | 0.939 | 1.032 | 1.018 | | | | |
| AI CR roll | 1.063 | 1.080 | 1.025 | 1.070 | 1.027 | 1.068 | 1.042 | 1.145** | 1.130* | 1.110 | | | |
| AI CR exp | 0.946 | 0.962 | 0.913 | 0.953 | 0.915 | 0.951 | 0.928 | 1.020 | 1.006 | 0.989 | 0.891* | | |
| AI BD roll | 0.830** | 0.844** | 0.801*** | 0.836** | 0.803*** | 0.834** | 0.814** | 0.895* | 0.883* | 0.867** | 0.781*** | 0.877* | |
| AI BD exp | 0.928 | 0.943 | 0.895** | 0.934 | 0.897* | 0.932 | 0.910 | 1.000 | 0.986 | 0.969 | 0.873* | 0.980 | 1.117* |

Table 6. 5

| Model after | AR roll | AR exp | Lin β roll | Lin β exp | ARX roll | ARX exp | AI β roll | AI β exp | MC roll | MC exp | AI CR roll | AI CR exp | AI BD roll |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AR exp | 0.984 | | | | | | | | | | | | |
| Lin β roll | 1.138 | 1.157 | | | | | | | | | | | |
| Lin β Exp | 1.154 | 1.174 | 1.015 | | | | | | | | | | |
| ARX roll | 0.932* | 0.948 | 0.820 | 0.808 | | | | | | | | | |
| ARX exp | 0.964 | 0.980 | 0.847 | 0.835 | 1.034 | | | | | | | | |
| AI β roll | 0.869 | 0.884 | 0.764 | 0.753* | 0.932 | 0.902 | | | | | | | |
| AI β exp | 0.947 | 0.962 | 0.832 | 0.820 | 1.015 | 0.982 | 1.089 | | | | | | |
| MC roll | 0.927* | 0.942 | 0.815 | 0.803* | 0.994 | 0.962 | 1.066 | 0.979 | | | | | |
| MC exp | 0.987 | 1.004 | 0.868 | 0.855 | 1.059 | 1.024 | 1.136 | 1.043 | 1.065** | | | | |
| AI CR roll | 0.777* | 0.789* | 0.683* | 0.673** | 0.833 | 0.806 | 0.893 | 0.820 | 0.838 | 0.787* | | | |
| AI CR exp | 0.972 | 0.988 | 0.855 | 0.842 | 1.043 | 1.009 | 1.118 | 1.027 | 1.049 | 0.985 | 1.252* | | |
| AI BD roll | 1.157 | 1.176 | 1.017 | 1.002 | 1.240 | 1.200 | 1.330 | 1.222 | 1.248 | 1.171 | 1.489 | 1.190 | |
| AI BD exp | 0.923* | 0.938* | 0.811 | 0.799* | 0.990 | 0.957 | 1.061 | 0.975 | 0.995 | 0.935* | 1.188 | 0.949 | 0.798 |

Note.

Table 6.4 and 6.5 show the relative RMSFE pre and post corona. The most interesting result is the increased performance of the model combination and the rolling AI model using the big dataset.

# 7 Profitability analysis

The investing strategy chosen is to buy a put option or a call option based on the predicted currency return.

When predicted value is positive, $\left(\frac{NOK}{USD}\right)_{t+1} > \left(\frac{NOK}{USD}\right)_{t}$:

A USD put option is bought at time t with expiration date t+1 with exercise price equals spot price at time t. At t+1, USD is bought according to the put option and the currency is sold immediately at spot price.

When predicted value is negative, $\left(\frac{NOK}{USD}\right)_{t+1} < \left(\frac{NOK}{USD}\right)_{t}$:

A USD call option is bought at time t with expiration date t+1 with exercise price equals spot price at time t. At t+1, USD is bought at spot price and immediately sold according to the call option.

A limitation to this investment strategy is it does not consider the precise forecast accuracy. It is only affected by the sign accuracy, that positive and negative predictions are followed by positive and negative true values. If the sign of the forecast equals true value a positive profit is made. If the sign of the forecast does not equal the true value, there is a negative profit.

We start with wealth of 1 at time t-1 and invest in either a call or put option. A positive or negative profit is made at time t if the sign of the prediction is correct or not to the true value. The new wealth at time t is then invested in a new put or call option, and so on. For simplicity, transaction costs are set to 0.

From graph 3 there is an increase in wealth by 19 times the original investment if the investment strategy was implemented in February 2001 using the LSTM network with the big dataset in a rolling scheme. This is huge profit. The other models are more moderate in revenue, but positive for the most part.

Figure 7.1 and table 7.1 show monthly average return, yearly average return and final wealth of the investing strategy.

***Figure 7.1 Profitability analysis***

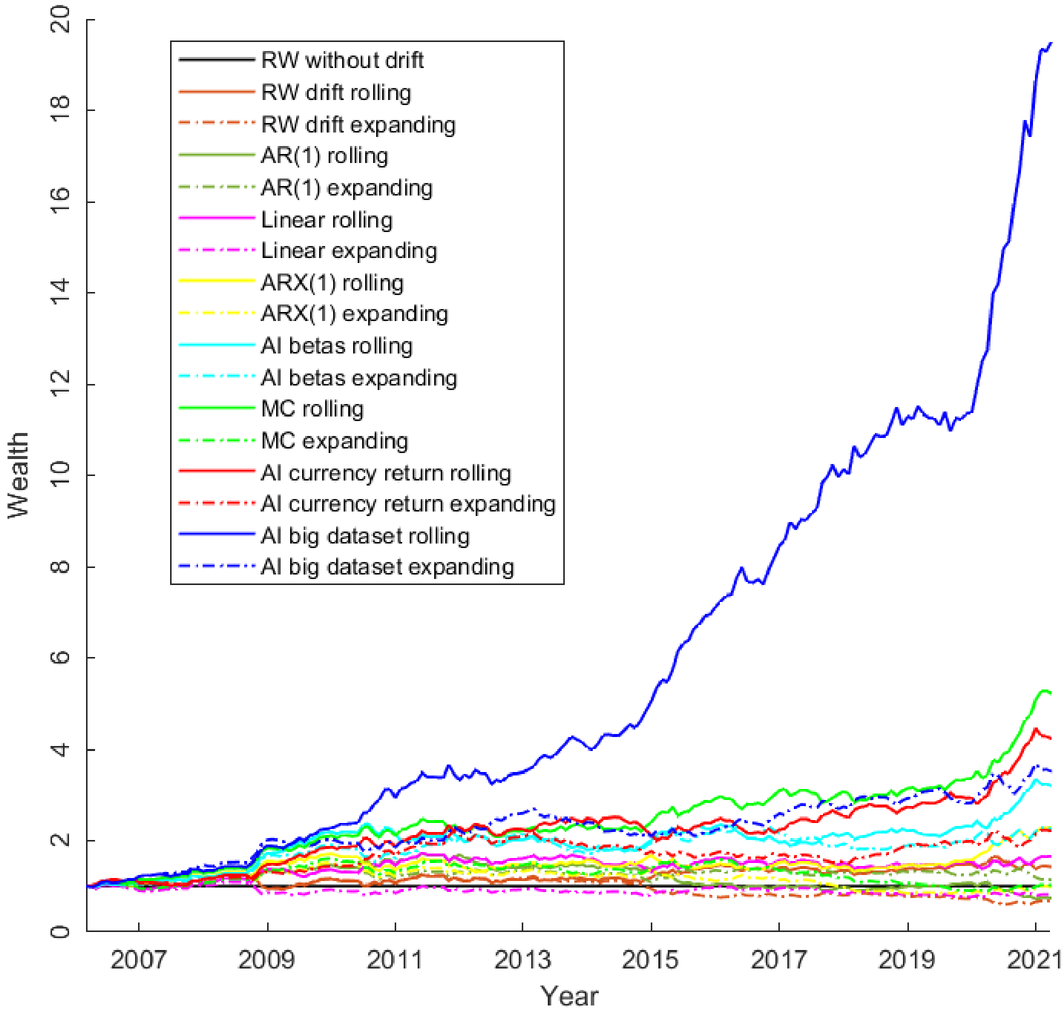


***Table 7.1 Profitability analysis***

| Model | Monthly return | Yearly return | Wealth |
|---|---|---|---|
| AI Big dataset rolling | 1.63 % | 21.41 % | 19.27 |
| Model Combination rolling | 0.90 % | 11.36 % | 5.16 |
| AI Currency Return rolling | 0.79 % | 9.80 % | 4.18 |
| AI Big dataset expanding | 0.70 % | 8.68 % | 3.56 |
| AI Betas rolling | 0.63 % | 7.84 % | 3.16 |
| AI Betas expanding | 0.45 % | 5.57 % | 2.29 |
| ARX rolling | 0.45 % | 5.53 % | 2.27 |
| AI Currency Return expanding | 0.43 % | 5.24 % | 2.18 |
| Linear model rolling | 0.27 % | 3.29 % | 1.64 |
| RW with drift rolling | 0.20 % | 2.43 % | 1.44 |
| AR expanding | 0.08 % | 0.92 % | 1.15 |
| Model Combination expanding | 0.02 % | 0.21 % | 1.03 |
| RW without drift | 0.00 % | 0.00 % | 1.00 |
| ARX expanding | -0.01 % | -0.14 % | 0.98 |
| Linear model Expanding | -0.11 % | -1.31 % | 0.82 |
| AR rolling | -0.17 % | -2.04 % | 0.73 |
| RW with drift expanding | -0.21 % | -2.53 % | 0.68 |

Monthly average return is calculated solving for r from the compound interest formula:

$$A = P\left(1 + \frac{r}{n}\right)^{nt}$$

where A = final wealth, P = original investment, r = annual rate of return, n = compounding frequency and t = time in years.

Yearly average return is calculated solving for r from the simple annual interest formula:

$$A = P(1 + r)^{t}$$

where A = final wealth, P = original investment, r = annual rate of return and t = time in years.

Table 7.1 and figure 7.1 show models outperformed by the RW when comparing RMSFE, now outperform the random walk models using the investment strategy. This indicates a sign test to be more appropriate for evaluating out-of-sample forecasts. This shows the models to perform good in real world investing.

## 7.1 Final remarks

The research question could have been more well defined and simpler. A weakness that I am doing research on both deep neural networks and DieboldLi factors as predictors for currency return.

Another approach to test deep neural networks would be to use data that has been proven to determine/forecast some variable to isolate the effect to the deep neural network.

The large number of forecasting model used could lead to p-hacking and overestimate the significance of the results.

The causal relationship between the Diebold Li factors is not explored and is subject for future research.

# 8 Conclusion

There is information in the commodity futures prices that is useful for forecasting commodity currencies. The Diebold Li factors from the commodity futures prices does contain valuable information for forecasting currency return.

The Diebold Li factors does capture information in the termstructure of crude oil futures prices valuable for forecasting currency return. This information is only captured by an AI model.

AI models is able to largely outperform several traditional forecasting models out-of-sample. Random walk models tend to outperform traditional models. AI models can outperform the random walk in forecasting currency return.

Future research could test forecasting other commodity currencies with different commodity futures prices.

**References:**

Adekoya, A. F., Nti, I. K., & Weyori, B. A. (2021). Long short-term memory network for predicting exchange rate of the Ghanaian cedi. *FinTech*, *1*(1), 25-43.

Ashley, R., Granger, C. W., & Schmalensee, R. (1980). Advertising and aggregate consumption: An analysis of causality. *Econometrica: Journal of the Econometric Society*, 1149-1167.

Brennan, M. J. (1976). The supply of storage. *The economics of futures trading*, 100-107.

Chen, Y. C., & Rogoff, K. (2003). Commodity currencies. *Journal of international Economics*, *60*(1), 133-160.

Datta, R. K., Sajid, S. W., Moon, M. H., & Abedin, M. Z. (2021). Foreign currency exchange rate prediction using bidirectional long short term memory. In *The big data-driven digital economy: Artificial and computational intelligence* (pp. 213-227). Cham: Springer International Publishing.

Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of econometrics*, *130*(2), 337-364.

Engel, C., & West, K. D. (2005). Exchange rates and fundamentals. *Journal of political Economy*, *113*(3), 485-517.

Frankish, K., & Ramsey, W. M. (Eds.). (2014). *The Cambridge handbook of artificial intelligence*. Cambridge University Press.

Goyal, A., & Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, *49*(5), 639-654.

Horner, M., Pakzad, S. N., & Gulgec, N. S. (2019). Parameter Estimation of Autoregressive-Exogenous and Autoregressive Models Subject to Missing Data Using Expectation Maximization. *Frontiers in Built Environment*, *5*, 109.

Huang, W., Lai, K. K., Nakamori, Y., & Wang, S. (2004). Forecasting foreign exchange rates with artificial neural networks: A review. *International Journal of Information Technology & Decision Making*, *3*(01), 145-165.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Inoue, A., Jin, L., & Rossi, B. (2014). Window selection for out-of-sample forecasting with time-varying parameters.

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of econometrics*, *196*(1), 55-67.

Kaldor, N. (1939). Speculation and economic stability. *The review of economic studies*, *7*(1), 1-27.

Kleivset, C. (2012). From a fixed exchange rate regime to inflation targeting.

Leung, M. T., Chen, A. S., & Daouk, H. (2000). Forecasting exchange rates using general regression neural networks. *Computers & Operations Research*, *27*(11-12), 1093-1110.

Mathworks. (n.d.). Train Network for Time Series Forecasting Using Deep N https://se.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-network-designer.html

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample?. *Journal of international economics*, *14*(1-2), 3-24.

Meese, R., & Rogoff, K. (1983). The out-of-sample failure of empirical exchange rate models: sampling error or misspecification?. In *Exchange rates and international macroeconomics* (pp. 67-112). University of Chicago Press.

Nag, A. K., & Mitra, A. (2002). Forecasting daily foreign exchange rates using genetically optimized neural networks. *Journal of Forecasting*, *21*(7), 501-511.

Nelson, C. R., & Siegel, A. F. (1987). Parsimonious modeling of yield curves. *Journal of business*, 473-489.

Ravazzolo, F., Sveen, T., & Zahiri, S. K. (2016). Commodity futures and forecasting commodity currencies.

Rossi, B., & Inoue, A. (2012). Out-of-sample forecast tests robust to the choice of window size. *Journal of Business & Economic Statistics*, *30*(3), 432-453.

Schinasi, G. J., & Swamy, P. A. V. B. (1989). The out-of-sample forecasting performance of exchange rate models when coefficients are allowed to change. *Journal of International Money and Finance*, *8*(3), 375-390.

Swanson, N. R., & White, H. (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, *13*(3), 265-275.

Working, H. (1949). The theory of price of storage. *The American economic review*, *39*(6), 1254-1262.

Yu-Chin, C., & Kenneth, R. (2012). Are the commodity currencies an exception to the rule?. *Global Journal of Economics*, *1*(01), 1250004.

Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25, pp. 15-24). San Francisco, CA, USA: Determination press.