Norwegian School of Economics
Bergen, Spring 2021

Density Forecasting of the EUR/NOK Exchange Rate

An Evaluation of Out-of-Sample Forecasting Performance of Empirical Exchange Rate Models

Sigurd Blom Breivik and Peder Vinje Samuelsen

Supervisor: Gernot Doppelhofer

Master thesis, Economics and Business Administration

Major: Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgements

This thesis is written as the final part of our Master of Science in Economics and Business Administration at Norwegian School of Economics, with majors in Financial Economics.

We would like to express our gratitude to our supervisor Professor Gernot Doppelhofer for the good cooperation throughout the semester. His guidance and constructive feedback have proven valuable for our progress throughout the writing process.

Norwegian School of Economics

Bergen, June 2021

Sigurd Blom Breivik

Peder Vinje Samuelsen

Abstract

This thesis investigates the predictive ability of fundamental economic and financial indicators on the EUR/NOK exchange rate. In doing so, we explore the emerging field of density forecasting, in addition to the standard point forecasting literature. Using a set of well-established empirical models, we construct short-term pseudo out-of-sample forecasts for the exchange rate. The results are benchmarked against a naïve random walk model, using a range of evaluation statistics grounded in the literature. The empirical analysis reveals that no models significantly outperform the random walk model using neither a point nor density forecast approach. However, we find evidence that fundamental models outperform in terms of forecasting appreciation tail risk at the one-month horizon. Furthermore, we find that a simple normal distribution is a better fit compared to an empirically backed skewed t-distribution derived from quantile regression. Our findings add to the growing strand of literature investigating the Meese & Rogoff puzzle from a density forecast perspective.

Contents

Figures

Tables

1. Introduction

The foreign exchange (FX) market plays a critical role within macroeconomics and international finance. Li (2016) highlights the facilitation of currency conversion, exchange rate risk management, and speculation as the three main functions of the market. The facilitation of currency conversion enables businesses and individuals to perform transactions outside their local currency, allowing them to make investments abroad and import and export goods and services. This allows countries to access goods, services, and markets that may otherwise have been unavailable to them. Usually, settlement of a trade deal will occur weeks to months after the deal was agreed. This exposes the transaction parties to currency risk. The facilitation of exchange rate risk management allows the parties to use the FX market to hedge their exposure and safeguard their interests. For this function to operate appropriately, the hedger must be able to pass the risk onto an entity that is better able to bear it. The facilitation of speculation provides a market for the hedger to do this. The FX market is the most liquid financial market globally, with a daily trading volume of $6.6 trillion as of April 2019 (BIS, 2019), and it is open Monday through Friday. This makes the FX market highly attractive to speculating entities that are willing to take on risk in an attempt to profit from currency fluctuations.

The mechanisms outlined above make exchange rate forecasting of great importance to investors, speculators, and policymakers alike. For investors and speculators, exchange rate forecasting can be an essential tool to design trading or hedging strategies that helps maximize return and minimize risk. It is vital to understand the uncertainty associated with the forecast to design effective strategies. Thus, investors and speculators are concerned not only with the expected outcome but the entire probability distribution. This includes the variance, any asymmetric 'leaning,' and the fatness of the outer tails. As stated shrewdly by Crnkovic & Drachman (1997, p. 47): *"At the heart of market risk measurement is the forecast of the probability density functions (PDF) of the relevant market variables… a forecast of a PDF is the central input into any decision model for asset allocation and/or hedging… therefore, the quality of risk management will be considered synonymous with the quality of PDF forecasts."* For policymakers like central banks, exchange rate forecasting is important because future exchange rate dynamics may impact interest rate policies and currency intervention decisions.

Given the importance of accurate exchange rate forecasts, knowing what drives exchange rates is essential to the forecaster. In the last several decades, economists have applied standard economic theory to develop a wide range of fundamental exchange rate determination models that incorporates macroeconomic and financial indicators as predictor variables. Yet, the performance of such models has been under scrutiny since Meese & Rogoff (1983) published their highly influential research paper '*Empirical exchange rate models of the seventies: Do they fit out of sample?'* . Using various fundamental exchange rate determination models to forecast several dollar-related currency pairs during the post-Bretton Woods era, the authors demonstrate that a naïve random walk model performs just as well in out-of-sample forecasting. The 'Meese & Rogoff puzzle' has since inspired an extensive body of literature that attempts to outcompete a driftless random walk model in out-of-sample analysis, with results varying greatly depending on econometric approach, sample period, currency pair, and forecast horizon. The results have led institutions like the Bank of Canada, the European Central Bank, and Statistics Norway to assume unchanged exchange rates in their predictions (Hungnes, 2020).

The vast majority of the Meese & Rogoff puzzle literature assesses the forecast accuracy of a simple point forecast. However, as previously outlined, understanding the uncertainty associated with a forecast can be vital. In this thesis, we therefore go beyond point forecasting to discuss a distinct part of the literature that deals with out-of-sample *density* forecasting. Prior research such as Wang & Wu (2012) applies a semiparametric method to generate out-of-sample exchange rate intervals for ten dollar-related OECD currency pairs. They find that, compared to a random walk model, fundamental models generate tighter forecast intervals which cover the realized exchange rates equally well. Gaglianone & Marins (2017) construct point and density forecasts at horizons of up to twelve months for the BRL/USD currency pair, using various statistical and economics-driven models. They find that fundamental economic indicators are useful when modeling exchange rate appreciation. In summary, the findings indicate that, while economic and financial indicators may have limited predictive value for point forecasting, they could have a predictive value when forecasting exchange rate probability distributions.

Several research papers have been written on the topic of the relationship between fundamental economic indicators and the Norwegian krone exchange rate. However, most of the literature approaches the topic from an ex-post analysis perspective, i.e., models are tested in-sample[1]

[1] In-sample prediction refers to predictions made inside the data sample in which the parameter estimates are obtained.

instead of out-of-sample (see, e.g., Akram, 2019 and Martinsen, 2017). Furthermore, the vast majority of the literature investigating the Meese & Rogoff puzzle on the Norwegian krone exchange rate approaches the puzzle from a point forecast perspective rather than density forecast perspective (see, e.g., Akram, 2004). In this thesis, we aim to apply the distinct body of literature investigating the Meese & Rogoff puzzle from a density forecast perspective to the krone-related currency pair that involves Norway's most important trading partner, the EU. This leads us to the following research question:

Can out-of-sample density forecasts based on fundamental exchange rate models outperform a naïve random walk model in forecasting the EUR/NOK exchange rate?

Specifically, we apply a series of empirically grounded exchange rate models to forecast the exchange rate at horizons ranging from one to twelve months. To assess whether density forecasts of the EUR/NOK exchange provide an informational advantage relative to standard point forecasts, we consider both approaches. The forecasting performance of the respective approaches are evaluated by use of a range of well-established statistical tests.

The EUR/NOK currency pair is interesting to consider for several reasons. First of all, the EU is Norway's most important trading partner. This makes the exchange rate particularly relevant for Norges Bank's decisions regarding interest rate policy and currency intervention. Furthermore, it makes the currency pair particularly relevant with regards to Norwegian import and export businesses' hedging strategies. As one of the most liquid krone-related currency pairs, the EUR/NOK is also suitable for investment and speculation purposes. Additionally, the NOK has depreciated significantly against the Euro and other currencies since 2013, puzzling economists across the country. This has implications for monetary policies and expectations of market participants. Thus, investigating the predictive value of fundamental economic and financial indicators on the EUR/NOK exchange rate may be of interest to various entities and for a wide range of applicational purposes.

The next chapter presents the theories that build the foundation of the empirical work. These theories are interest rate parity, purchasing power parity, the monetary model of exchange rate determination, the Taylor rule model of exchange rate determination, and the behavioral exchange rate model. In chapter 3, the econometric methodology is presented. Using ordinary least squares and quantile regression, we estimate point and density forecasts for the EUR/NOK exchange rate at the one-, three-, six- and twelve-month horizon. A series of test statistics are then applied to evaluate the forecast accuracy of the various model specifications. We evaluate the point forecast accuracy, the directional change accuracy, the performance of the full-density forecast, and the risks associated with the tails of the distributions. Chapter 4 presents the data variables included in the various models and explains how the data are transformed for analysis purposes. The aforementioned chapters culminate in an empirical analysis presented in chapter 5. Finally, the thesis is summarized and concluded in chapter 6.

2. Theory

This chapter presents the underlying theories that build the foundation of the empirical work. Although the forecasting performance of such theoretical models have come under scrutiny since Meese & Rogoff (1983) published their highly influential paper, they remain helpful in providing understanding and insight into the long-term and short-term drivers of exchange rates. Additionally, the literature investigating such models from a density forecast perspective remains limited. We will present the theories about interest rate parity, purchasing power parity, the monetary model, the Taylor rule model, and the behavioral exchange rate model. However, first, we give a brief introduction to the foreign exchange market.

2.1 The Foreign Exchange Market

The foreign exchange market is the biggest financial market in the world, with a daily volume of about $6.6 trillion as of April 2019 (BIS, 2019). The market operates around the clock from Monday through Friday and enables businesses to perform transactions outside their local currency, thus facilitating international trade. Contrary to, e.g., the equity market, trading does not take place in a central marketplace and is instead conducted over-the-counter. The trading occurs through a worldwide linkage of bank currency traders, non-bank dealers, and FX brokers that trade through telephones, computer terminals, and automated dealing systems (Kumar, 2014). Although trading desks are closed on the weekends, it is still possible to execute a transaction through, e.g., a bank. The bank will then supply the buyer from a stock obtained prior to the weekend at a rate where the bank makes a slight profit.

Foreign exchange transactions occur either in the spot market or the forward market. Transactions in the spot market occur at the prevailing exchange rate; the spot rate and deliveries are almost instant. In the forward market, foreign exchange is bought and sold for delivery at a future date. The forward rate of exchange is settled today but may deviate from the prevailing spot rate with quotes at either a premium or a discount.

2.2 Interest Rate Parity

Historically, one of the most popular trading strategies in the foreign exchange market has been the carry trade. A currency carry trade is carried out by borrowing a currency in a country with low interest rates to fund a currency in a country with high interest rates. This trade implies a fundamental relationship between exchange rates and interest rates and has given rise to several theories, one of which is the interest rate parity theory. Interest rate parity is a no-arbitrage condition that seeks to explain movements in the exchange rate on the back of the interest rates available on bank deposits in the two respective countries. This no-arbitrage condition exists in two forms elaborated on in the following subsections.

2.2.1 Covered Interest Rate Parity

The first form is covered interest rate parity. Interest rate parity is covered when the no-arbitrage condition is satisfied using a forward contract that hedges exchange rate risk. In this situation, an investor will be indifferent between investing domestically or abroad, as the forward exchange rate sustains equilibrium so that the return on domestic deposits is equal to the return on foreign deposits. The potential for arbitrage profits is thereby eliminated. The equation below represents covered interest rate parity, with i^d and i^f being the domestic and foreign money market rate, respectively, S_t the spot rate and, F_t the forward rate.

$$(1 + i^d) = \frac{F_t}{S_t}(1 + i^f)$$

The return on domestic deposits on the left side of the equation is equal to the return on foreign deposits on the right side of the equation, expressed in domestic currency. Note that here and in the following, we adopt a currency convention where the domestic currency is expressed per unit of foreign currency. That is, an increase (decrease) in the exchange rate implies a depreciation (appreciation) of the domestic currency.

2.2.2 Uncovered Interest Rate Parity

Uncovered interest rate parity is the second form of the no-arbitrage condition. Interest rate parity is uncovered when the no-arbitrage condition is satisfied without using a forward contract to hedge exchange rate risk. Instead, a risk-neutral investor will be indifferent between investing domestically or abroad because the spot rate is expected to adjust so that the return on domestic deposits is equal to the return on foreign deposits, measured in domestic terms at a future date. The potential for arbitrage profits is thereby eliminated also in this situation. The equation below represents uncovered interest rate parity, with $E_t(S_{t+k})$ being the expected future spot exchange rate a time $t + k$.

$$(1 + i^d) = \frac{E_t(S_{t+k})}{S_t}(1 + i^f)$$

Again, the return on domestic deposits on the left side of the equation is equal to the return on foreign deposits on the right side of the equation, expressed in domestic currency.

Covered and uncovered interest rate parity rests on the assumptions that capital flows freely between countries and that assets are perfectly interchangeable. Given these assumptions, investors would be expected to hold the assets that offer more significant returns, regardless of the assets being domestic or foreign. Thus, arbitrage opportunities are immediately traded away so that a single investor will expect to earn equivalent returns domestically and abroad.

2.2.3 Empirical Evidence of Interest Rate Parity

Empirically, covered interest rate parity generally holds for freely traded currencies. However, due to the presence of market imperfections such as transaction costs, tax implications, and counterparty risk, studies find evidence of short-term deviations, indicating that it does not hold with precision (Levich, 2011). As for uncovered interest rate parity, previous empirical studies largely reject its validity at short horizons. The evidence is more supportive at longer horizons with, e.g., Chinn & Meredith (2004) finding evidence supporting uncovered interest rate parity at horizons longer than one year. A large body of newer literature has suggested potential explanations for why the same is not observed at shorter horizons. These explanations include the potential presence of time-varying risk premia. Ismailov & Rossi (2017) provide empirical evidence that uncovered interest rate parity holds at short horizons when uncertainty is 'not exceptionally high' and breaks down during periods of high uncertainty.

2.3 Purchasing Power Parity

The concept of purchasing power parity has existed since the 16th century but was molded into its current form by Swedish economist Karl Gustav Cassel in 1916. Cassel was an advocate for restoring the gold standard and the system of fixed exchange rates in the aftermath of World War I as a means to restore international trade and further stable and balanced growth. In his writings around this time, he recommended fixing exchange rates at a level corresponding to purchasing power parity, arguing it would prevent trade imbalances between nations (Rogoff, 1996). Although not formulated by Cassel as a theory of exchange rate determination, the doctrine has proven integral to understanding the relationship between price levels and foreign exchange rates.

Purchasing power parity is based on the law of one price, which states that identical goods sold in different locations must sell at the same price under the assumptions of free competition, price flexibility, and no trade frictions. Given these assumptions, market participants would be expected to buy goods in cheap areas to profit in expensive areas. Thus, arbitrage opportunities should be traded away and result in equal prices at all locations. As with interest rate parity, purchasing power parity exists in two forms, both of which are detailed below.

2.3.1 Absolute Purchasing Power Parity

The first form is absolute purchasing power parity. This condition states that the foreign exchange rate is expected to adjust so that the price level of a basket of goods domestically is equal to the price level of an equivalent basket of goods abroad, measured in a common currency. Mathematically, absolute purchasing power parity can thus be expressed as:

$$P^d = S \times P^f,$$

where P^d is the domestic price index, P^f the foreign price index, and S the spot rate. The price level of a domestic basket of goods on the left side of the equation is equal to the price level of a foreign basket of goods on the right side of the equation.

2.3.2 Relative Purchasing Power Parity

The second form is relative purchasing power parity, which is a dynamic form of purchasing power parity that relates relative changes in inflation rates to the exchange rate. In other terms, relative purchasing power parity states that the foreign exchange rate is expected to adjust with the relative change in price level between two countries. Mathematically, the relationship can be expressed as:

$$\frac{S_t}{S_0} = \frac{\left(P_t^d / P_0^d\right)}{\left(P_t^f / P_0^f\right)}$$

The change in the spot rate from time 0 to t on the left side of the equation is equal to the relative change in price level between the domestic and foreign countries on the right side of the equation.

2.3.3 Empirical Evidence of Purchasing Power Parity

Absolute purchasing power parity does not have much empirical support, partly explained by difficulties in obtaining comparable baskets of goods across countries[2]. Support for relative purchasing power parity has also proven weak in the short term. In contrast, studies find evidence of its significance in the long term but suggest that real exchange rates adjust to the purchasing power parity level at a prolonged rate. According to Rogoff (1996), consensus tends to estimate a half-life of adjustment of three to five years. Dornbusch's (1976) theory of exchange rate overshooting, which explains short-term deviations from purchasing power parity by the stickiness of goods' prices compared to the flexible prices of financial instruments, has been one suggested explanation for this result. However, as shown by Rogoff, one would expect to observe a half-life of adjustment of one to two years if this was the case. Studies on Norwegian quarterly data find support for purchasing power parity in the long-term with a half-life of adjustment of around 18 months (Akram, 2000, 2002, as cited in Akram et al., 2003).

2.4 The Monetary Model

The monetary model of exchange rate determination appeared in its modern form after the collapse of the fixed exchange rate system in the 1970s, when proponents of the monetary approach to the balance of payments developed its parallel for floating exchange rates. The model exists in several forms, but all have in common that they can be regarded as asset market view models of exchange rate determination. This means that exchange rates are viewed as relative prices of assets priced in a forward-looking fashion, i.e., allowing for the inclusion of non-observable factors among the fundamentals. Today, the monetary model of exchange rate determination is a standard workhorse within fundamental exchange rate forecasting.

[2] Issues relate e.g. to differences in quality, purchasing patterns and labor costs, absence of international trade for certain goods, and country specific costs.

Following Zhang et al. (2007), the model rests on two distinct assumptions. The first one is that purchasing power parity holds continuously over time. This relationship can be expressed as:

$$s_t = p_t - p_t^* + c + \varepsilon_t, \tag{1}$$

where c is a constant, s_t the logarithm of the exchange rate expressed in units of domestic currency, p and p^* the domestic and foreign price levels, and ε the error term. The equation implies that absolute purchasing power parity holds if $c = 0$, while relative purchasing power holds if $c \neq 0$.

The second assumption is that of money market equilibrium. The money market is said to be in equilibrium at the given interest rate that balances the quantity of money demanded to the quantity of money supplied. The monetary model assumes a stable money demand function where the equilibrium condition depends on the logarithm of real income y, the logarithm of the price level p, and the nominal interest rate i. Domestically, money market equilibrium[3] is thus expressed as follows:

$$m_t = p_t + \beta_2 y_t - \beta_3 i_t + \mu_t, \tag{2}$$

with m being the logarithm of money demanded[4], β_2 the income elasticity of money demanded, β_3 the semi-elasticity of interest rates, and μ the error term. The same equation applies to foreign countries but with variables denoted by an asterisk:

$$m_t^* = p_t^* + \beta_2^* y_t^* - \beta_3^* i_t^* + \mu_t^* \tag{3}$$

From here, monetary models of exchange rate determination go in different directions. Two of the main types will be examined further in the following subsections.

---

[3] The equation is obtained by taking the natural logarithm of Cagan's (1956) semi-logarithmic demand for money function $M_t/P_t = Y_t^k e^{-\lambda i_t}$ and solving for m.

[4] In equilibrium, money demand is assumed to be equal to the respective money supply.

2.4.1 Flexible Price Monetary Model

By rearranging equations (2) and (3) to solve for the domestic and foreign price levels and substituting for the price levels in equation (1), one obtains the flexible price monetary model of exchange rate determination. Here, c is an arbitrary constant and ε_t^* a disturbance term.

$$s_t = m_t - m_t^* - \beta_2 y_t + \beta_2^* y_t^* + \beta_3 i_t - \beta_3^* i_t^* + c + \varepsilon_t^*$$ [5]

The equation above implies that a relative increase in domestic to foreign money supply will depreciate the exchange rate. The opposite is true for a relative increase in domestic to foreign real income. Excess demand for the domestic money stock reduces expenditure, causing prices to fall until money market equilibrium is achieved. This implies an appreciation of the exchange rate through the purchasing power parity mechanisms. A relative increase in domestic interest rates reduces domestic demand for the money stock, leading to depreciation.

The flexible price monetary model assumes that prices of goods behave in the same way as prices in financial markets, i.e., changing market conditions reflect goods' prices immediately. The model furthermore assumes that uncovered interest rate parity continuously holds.

2.4.2 Sticky-Price Monetary Model

After the transition to the floating exchange rate regime, real exchange rates experienced high volatility, sowing doubt over the assumption of continuous purchasing power parity. This led to the development of the sticky price monetary model of exchange rate determination by Dornbusch (1976). The model sought to explain features of exchange rate behavior that deviated from the predictions of the flexible price monetary model, including the unexpected occurrence of an immediate depreciation in the exchange rate following a monetary expansion.

Contrary to the flexible price model, Dornbusch's model centers around the concept of goods' prices being sticky in the short run. This is captured in a framework in which prices of domestic goods are sticky, while domestic currency prices of foreign goods move freely with the exchange rate. Thus, the model outputs a long-run equilibrium that the exchange rate is expected to adjust towards over time, but the exchange rate may overshoot its long-run

[5] Note that the monetary model is sometimes specified without the interest rate differential. It may also be specified with an inflation differential. For reference, see, e.g., Meese & Rogoff (1983) and Wang & Wu (2012).

equilibrium level in the short-run. Compared to the flexible price model in section 2.4.1, this implies that the sign of the interest rate differential is interpreted differently in the short run. In the sticky-price model, an increase in interest rates must offset a cut in the money supply for the money market to clear. Raised interest rates lead to capital inflow and an appreciating nominal exchange rate. Since prices are sticky, this also implies an appreciation of the real exchange rate. Domestic interest rates decline in accordance with the changing money market equilibrium as domestic prices begin to fall. Thus, the exchange rate converges on its long-run equilibrium.

2.4.3 Empirical Evidence of Monetary Models

Various techniques and research methodologies have been applied to test the empirical significance of monetary models in the last several decades. The results have generally been mixed, depending on the currency pair and sample period used. Meese & Rogoff (1983) show that a naïve random walk model outperforms the flexible and sticky-price monetary models in short-term out-of-sample prediction. Mark (1995), on the other hand, finds that monetary models outperform when the forecast horizon is longer. The evidence relating to monetary models are more supportive in short-term forecasting when advanced cointegration techniques and error-correction models[6] are applied. MacDonald & Taylor (1994), e.g., find that a dynamic monetary error correction model that allows for flexible short-run dynamics, outperforms random walk mechanisms on forecasting horizons of up two twelve months in out-of-sample prediction of the sterling-dollar exchange rate.

2.5 The Taylor Rule Model

Most research on out-of-sample exchange rate predictability until the mid-2000s was based on empirical exchange rate models akin to those presented in section 2.4. Thus, a disconnect had risen against literature on monetary policy evaluation, which had its basis in the Taylor rule framework. The Taylor rule was proposed by John B. Taylor (1993) as a targeting monetary policy technique to stabilize economic activity. Engel & West (2005) used the

[6] For a discussion of cointegration and error correction specifications, see sections 3.2.1 and 4.2.2.

framework to derive the exchange rate as a present value asset price. Since then, a growing strand of the literature has used the Taylor rule to model exchange rate predictability.

2.5.1 Asymmetric Taylor Rule Model with Smoothing

In Taylor's original formulation, the rule states that the central bank sets the nominal interest rate as a response to a divergence in observed inflation from the target inflation rate and in observed GDP from potential GDP, i.e., it can be specified as:

$$\tilde{i}_t = \pi_t + \phi(\pi_t - \tilde{\pi}) + \gamma y_t + \tilde{r},$$

where $\tilde{i}_t$ is the short-term nominal interest rate target, π_t the inflation rate, $\tilde{\pi}$ the inflation rate target, y_t the output gap, and $\tilde{r}$ the real interest equilibrium rate. Following Molodtsova & Papell (2009), $\tilde{\pi}$ and $\tilde{r}$ can be combined into a constant[7] term $\mu = \tilde{r}$ - $\phi\tilde{\pi}$, leading to

$$\tilde{i}_t = \mu + \lambda\pi_t + \gamma y_t, \tag{4}$$

where $\lambda = 1 + \phi$. For the foreign country, it is commonly assumed that the central bank also targets the exchange rate level that makes purchasing power parity hold. If the exchange rate depreciates from equilibrium, the foreign central bank increases the nominal interest rate and vice versa. Thus, the real exchange rate q_t is included in the Taylor rule for the foreign country.

$$\tilde{i}_t = \mu + \lambda\pi_t + \gamma y_t + \delta q_t \tag{5}$$

No distinction is made between the actual nominal interest rate and the target interest rate in the original Taylor rule. The target rate is assumed to be achieved within one period. However, according to Molodtsova & Papell (2009), it has become common practice to adjust this assumption so that the interest rate only partially adjusts within one period. The interest rate i_t is then assumed to adjust to the target rate as in the following formula:

$$i_t = (1 - \rho)\tilde{i}_t + \rho i_{t-1} + v_t \tag{6}$$

By substituting equation (5) into equation (6), the following is obtained:

[7] Note that rolling window regression allows the constant to be time-varying (Molodtsova & Papell, 2009).

$$i_t = (1-\rho)(\mu + \lambda\pi_t + \gamma y_t + \delta q_t) + \rho i_{t-1} + v_t \quad (7)$$

The interest rate differential function is constructed by subtracting equation (7) of the foreign country from equation (4) of the domestic country, yielding:

$$i_t - i_t^* = \alpha + \alpha_{d\pi}\pi_t - \alpha_{f\pi}\pi_t^* + \alpha_{dy}y_t - \alpha_{fy}y_t^* - \alpha_q q_t^* + \rho_d i_{t-1} - \rho_f i_{t-1}^* + \eta_t, \quad (8)$$

where * denotes foreign variables, d are domestic coefficients, f are foreign coefficients, α is a constant, $\alpha_\pi = \lambda(1-\rho)$, $\alpha_y = \gamma(1-\rho)$ and $\alpha_q = \delta(1-\rho)$.

Molodtsova & Papell (2009) show that equation (8) can be combined with a series of predictions to construct an exchange rate forecasting equation. The first prediction is that higher inflation in the domestic country will lead to exchange rate appreciation. Vice versa, higher inflation in the foreign country will lead to an exchange rate depreciation. Second, an increase in the output gap domestically will cause the central bank to raise interest rates, leading to an appreciation of the currency. If the output gap increases in the foreign country, raised foreign interest rates will depreciate the domestic currency. Third, an increase in the real exchange rate in the foreign country is predicted to lead to higher interest rates in the foreign country, causing a depreciation of the domestic currency. Finally, if the interest rate smoothing assumption hold, an increase in the lagged interest rate is predicted to lead to higher current and future interest rates. According to the uncovered interest parity, raised interest rates will result in an immediate appreciation of the domestic currency and forecasted depreciation. However, empirical evidence suggests that both an immediate and forecasted appreciation is a more reasonable assumption. Gourincha & Tornell (2004), e.g., provide survey evidence that investors tend to underestimate the persistence of interest rate shocks. This causes the currency to appreciate for longer than uncovered interest rate parity predicts, as investors gradually revise their beliefs about the persistence of the shock. The exchange rate forecasting equation can then be constructed as:

$$\Delta s_{t+1} = \omega - \omega_{d\pi}\pi_t + \omega_{f\pi}\pi_t^* + \omega_{dy}y_t - \omega_{fy}y_t^* - \omega_q q_t^* - \omega_{di}i_{t-1} - \omega_{fi}i_{t-1}^* + \eta_t, \quad (9)$$

where s_t is the log of the nominal exchange rate expressed in domestic currency terms.

2.5.2 Taylor Rule Model Variations

Equation (9) expresses an exchange rate determination model with asymmetric Taylor Rule fundamentals and interest rate smoothing. It can be altered to produce several different variations with slightly different interpretations. In our empirical analysis, four different variations are used: a model with asymmetric Taylor Rule fundamentals with and without smoothing and a model with symmetric Taylor Rule fundamentals with and without smoothing. The fundamentals are symmetric if the foreign central bank does not target the exchange rate level for which purchasing power parity holds, in which case $\delta = \alpha_q = 0$. If the actual nominal interest rate is assumed to adjust to the target rate within the same period, the model is specified without smoothing, in which case $\rho_d = \rho_f = 0$. Additionally, the model can be specified as homogenous or heterogeneous. The model is homogenous if the inflation-, output gap- and interest rate smoothing coefficients are the same domestically and abroad, i.e., $\alpha_{d\pi} = \alpha_{f\pi}, \alpha_{dy} = \alpha_{fy}$, and $\rho_d = \rho_f$, and heterogenous otherwise. Finally, the constant $\alpha = 0$ if the inflation-, inflation target-, interest rate smoothing-, and equilibrium real interest rate coefficients are the same between the domestic and foreign country.

2.5.3 Empirical Evidence of Taylor Rule Models

Compared to the models and theories explored earlier in this chapter, Taylor rule fundamentals have generally been found to improve forecasting ability in out-of-sample exchange rate forecasting in the short term. E.g., using 16 different Taylor rule variations, Molodtsova & Papell (2009) find evidence of short-term out-of-sample exchange rate predictability in eleven currency pairs. The evidence is more assertive with Taylor rule models than with conventional models. A symmetric model with heterogenous coefficients, smoothing, and a constant achieve the most robust results. However, beyond forecasting horizons of six months, the exchange rate predictability is generally found to be relatively poor. Furthermore, evidence of short-term exchange rate predictability is not without controversy. Rogoff & Stavrakeva (2008) argue that evidence of short-term exchange rate predictability in structural models is overstated due to misinterpretations of *"new out-of-sample tests for nested models, overreliance on asymptotic test statistics and failure to sufficiently check robustness to alternative time windows"* (p. 1).

2.6 Behavioural Equilibrium Exchange Rate Model

As pointed out in section 2.3.3, high volatility and slow mean reversions have raised questions about the usefulness of purchasing power parity as an isolated measure of the equilibrium exchange rate. Several approaches have been suggested to model the sources of these violations. One such approach is the behavioral equilibrium exchange rate (BEER) approach developed by Clark & MacDonald (1998). The principle of the BEER approach is that fundamental macroeconomic factors explain the slow mean reversion observed in empirical tests of purchasing power parity. The approach is not based on any specific model and, as such, various metrics may be employed during estimation.

Following Alstad (2010), the BEER approach is based on real uncovered interest rate parity:

$$E_k(q_{t+k}) - q_t = r_{t,k} - r^*_{t.k}, \tag{10}$$

where q is the logarithm of the real exchange rate, $r_{t,k}$ the k-period real domestic interest rate, and $r^*_{t.k}$ the real foreign interest rate. Under the assumption that there is a long-run linear relationship between the exchange rate and economic fundamentals, the expected real interest rate can be written as:

$$E(q) = E(\alpha + \beta' M), \tag{11}$$

where M is a vector of economic fundamentals, α a constant, and β' a vector of reduced-form coefficients. Thus, the model is equivalent to purchasing power parity if $\beta = 0$, as the long-run real exchange rate is expected to be constant. This is when the long-run economic variables are at their equilibrium levels. Deviations in the real exchange rate from the constant level are called total misalignments, while deviations from the equation at any point in time are referred to as current misalignments. Alstad shows that inserting equation (11) into (10) yields:

$$q = \alpha + \beta' E(M) - (r - r^*)$$

Under the BEER approach, the logarithm of the real exchange rate thus depends on the constant α, a vector of economic fundamentals M, and an interest differential $(r - r^*)$. As pointed out earlier, various metrics may be employed during estimation. What variables to include in the vector of economic fundamentals is primarily an empirical question. Section 4.1 presents the data variables we have chosen to include in this thesis and briefly reviews the empirical evidence associated with each variable in earlier analyses of the Norwegian krone exchange rate.

3. Methodology

This chapter presents the methodology applied in the empirical analysis and explains the underlying econometric theory. First, we present the general forecasting framework applied to estimate the point- and density forecasts, derived from ordinary least squares and quantile regression. Next, we provide a brief description of the point forecasting procedure and the underlying statistical assumptions of ordinary least squares. Following that, we introduce the concept of density forecasting and present a step-by-step approach to how the forecasted densities are generated. Furthermore, we present regression diagnostics applied to investigate whether the statistical assumptions of OLS and quantile regression hold. Finally, we give an overview of the methods and statistics used to evaluate the forecasts.

3.1 Forecasting Framework

This subsection discusses the forecasting framework applied in the empirical analysis. Consistent with the Meese & Rogoff puzzle literature, we apply pseudo-out-of-sample forecasting. The models are estimated using rolling-window estimation with a window size of 120 observations. The point and density forecasts are then forecasted using a direct forecasting method for horizons of one-, three-, six- and twelve months. Finally, the output for the structural models is benchmarked against a driftless random walk model. The forecasting methodology and the choices we have made are detailed further in the following.

3.1.1 Pseudo Out-of-Sample Forecasting

It is well established within the forecasting literature that a good in-sample fit of a forecasting model does not necessarily translate into good out-of-sample performance. A common cause for this is that data has been overfitted, i.e., the model is fitted with more parameters than what can be justified by the underlying structure of the data. Overfitting causes noise from the estimation period to be extracted to the fitted model, often leading to higher error rates out-of-sample. The out-of-sample forecasting method aims to address this issue. In true out-of-sample forecasting, forecasts for the future are constructed in real-time using a model where parameters are estimated by data available up until and including today. Given the time-consuming nature of this exercise, one typically relies on *pseudo* out-of-sample forecasting, where one simulates this exercise using an historical date $T_0 < T$ rather than today's date T as the starting point for the forecast period. The forecast is then computed for the date $T + h$ and repeated for all dates in the forecast period.

3.1.2 Rolling Window Estimation

There are a few different options to consider when specifying a model for pseudo out-sample forecasting. Among other things, this involves deciding on a forecasting scheme. Recursive- and rolling window estimation are the most common options in practice. In a recursive approach, sample data from $t = 1, \dots, T$ is used to estimate the forecast model, where T is the time period where the h-step ahead forecast is conducted. When the forecast from period T is made, the sample is increased by one observation, and the model is re-estimated for period $T + 1$, and so on. The starting point of the sample data is anchored to $t = 1$, meaning that the estimation window expands as T increases. This is where recursive window estimation differs from rolling window estimation. In rolling window estimation, the estimation window is fixed to a set number of sample observations. Thus, the forecast model for period $T + 1$ is estimated using sample data from $t = 2, \dots, T + 1$, rather than $t = 1, \dots, T + 1$.

Rolling window estimation is advantageous in cases where the independent variable's ability to forecast the dependent variable is time-varying, as it only uses the most recent observations to forecast the parameters. A growing strand of the literature concludes that the predictive content of financial and macroeconomic time series is time-varying (Rossi, 2013). In Norway, suggestions have been made that the weak Norwegian krone exchange rate observed in recent years results from increased climate transition risk (Kapfhammer et al., 2020), causing the currency to decouple from the oil price. Given the above, we opt to use rolling window estimation in this thesis.

There are no general guidelines for how many observations to include in the estimation window. In the literature on out-of-sample exchange rate forecasting, the number of observations ranges from a few dozen to several hundred. In general, fewer observations will allow the models to adapt more quickly to structural changes in the predictive content. However, too few observations could also cause the coefficient parameters to be estimated unreliably. Hence, we use 120 observations in this thesis, i.e., ten years of data. This is line with, e.g., Molodtsova & Papell (2009).

3.1.3 Direct Forecasting

In addition to the forecast estimation scheme, a forecaster also faces options concerning the forecasting method itself. In their influential paper, Meese & Rogoff (1983) use future realized fundamentals to prove that fundamental models do not forecasts exchange rates better than a random walk model. However, given that a pseudo out-of-sample model simulates out-of-sample forecasting, one would typically only use the information available at the time period the multiple step-ahead forecasts are made. The forecaster is then faced with the choice of making a direct forecast using lagged fundamental variables or using an iterative forecast method. In the direct forecast method, the multiple step-ahead forecast is made directly, without forecasting the intermediary horizons. Thus, the forecasted value is obtained using only lagged fundamentals that are available at the time the forecast is made as predictors. In the iterative forecast method, the predictors for the next period are estimated using an autoregressive process. The forecasted predictors are then used as inputs to forecast the dependent variable. The process is iterated for all time periods until the multiple step-ahead forecast for the given forecast horizon is obtained. Rossi (2013) notes that, for single equation linear models, the choice of independent variables matters more than the forecasting method. Furthermore, iterative forecasts are more susceptible to model misspecification (Marcellino et al., 2006). Hence, we use direct forecasting with lagged fundamentals in this thesis.

The forecast horizon in the literature ranges from short horizons of one month to long horizons of around five years. As described in section 2, the empirical evidence supporting fundamental model's predictability of exchange rates varies depending on the forecast horizon. It is generally advised to select a forecast horizon suited to the given model, which for a Taylor rule model could mean twelve months and for a purchasing power parity model could mean 60 months. However, given the relatively short estimation window and the use of first-differenced predictors at a monthly frequency, we opt to limit the forecast horizon to up to twelve months in this thesis.

3.1.4 Random Walk Model

Building on Meese & Rogoff's (1983) findings, it has become standard within the literature to use a driftless random walk model as a benchmark for testing exchange rate models. This is also considered the toughest benchmark to beat (Rossi, 2013). Hence, this is what we use as the benchmark in this thesis. A random walk is a time series process consisting of a succession of random steps. At time t, y is obtained by taking the previous value y_{t-1} and adding an independent random variable e_t with a zero mean.

$$y_t = y_{t-1} + e_t$$

By using repeated substitution and taking the expected value of both sides, it can be shown that $E(y_t) = E(y_0)$ for all $t \geq 1$, i.e., the best prediction of y_t is simply the previous value y_{t-1}.

$$y_t = e_t + e_{t-1} + \cdots + e_1 + y_0$$

$$E(y_t) = E(e_t) + E(e_{t-1}) + \cdots + E(e_1) + E(y_0) = \ E(y_t) = E(y_0) \text{ for all } t \geq 1$$

Unlike the mean, the variance of a random walk process depends on t. Following Wooldridge (2018), the variance can be computed by assuming y_0 is non-random so that $Var(y_0) = 0$. Assuming e_t is independent and identically distributed, it can be shown that the variance of a random walk model increases as a linear function of time.

$$Var(y_t) = Var(e_t) + Var(e_{t-1}) + \cdots Var(e_1) = \sigma_e^2 t$$

The above implies that a random walk process is non-stationary. It displays highly persistent behavior, as no matter how far into the future we try to predict y_{t+h}, today's value y_t will always be the best prediction. Importantly, as the variance increases with time, the confidence interval of a random walk model will grow larger as h increases.

3.2 Point Forecasting with OLS and Underlying Assumptions

The point forecast models in section 5 are estimated using ordinary least squares (OLS). This is a method for estimating the unknown parameters in a linear regression model that in its simplest form can be expressed as $y_{t+h} = \alpha + \beta x_t + \varepsilon_{t+h}$, where α and β are the true, unobserved parameters and ε_{t+h} is the error term. The method involves solving an optimization problem that minimizes the sum of the squared differences between the observed values and the predicted values. In other terms, OLS solves for the parameters $\hat{\alpha}$ and $\hat{\beta}$ that minimizes the sum of squared errors. Under a specific set of assumptions, the OLS estimator is, according to the Gauss-Markov theorem, the best unbiased linear estimator of the real values α and β (Wooldridge, 2018). In the following subsections, these assumptions will be presented. The statistics used to test the relevant OLS assumptions are outlined in section 3.4.

3.2.1 Assumption 1: Linearity, Stationarity and Weak Dependence

"*The stochastic process* $\{(x_{t1}, x_{t2}, \dots, x_{tk}, y_t): t = 1,2,\dots,n\}$ *follows the linear model*

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t$$

where $\{u_t: t = 1,2,\dots,n\}$ *is the sequences of errors or disturbances. Here,* n *is the number of observations*" (Wooldridge, 2018, p. 370).

The first assumption states that the time series process follows a model that is linear in its parameters. For some time-series problems, this assumption is not satisfied. In this case large sample properties of OLS must be applied. The above assumption is then extended to include that $\{(x_t, y_t): t = 1,2,\dots,n\}$ is stationary and weakly dependent. This implies that the law of large numbers and the central limit theorem can be applied to sample means[8].

Stationarity intuitively means that the statistical properties of a time series process do not change over time. In mathematical terms, (Wooldridge, 2018) defines the stochastic process $\{x_t: t = 1,2,\dots,n\}$ as stationary "*if for every collection of time indices* $1 \leq t_1 < t_2 < \cdots < t_m$, *the joint distribution of* $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ *is the same as the joint distribution of* $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ *for all integers* $h \geq 1$" (p. 367). If the stochastic process has a finite second moment, a weaker form of stationarity suffices. The stochastic process is then said to be covariance stationary if *"(i)* $E(x_t)$ *is constant, (ii)* $Var(x_t)$ *is constant and (iii) for any* $t, h \geq 1$, $Cov(x_t, x_{t+h})$ *depends only on* h *and not on* t" (p. 367). The use of non-stationary time series in OLS regression may lead to spurious results, as inference cannot be drawn from a time series if the statistical properties are time-variant. However, a regression model with two non-stationary variables will not generate spurious results if there is a linear relationship between the variables that in itself is stationary. If such a relationship is present, the variables are said to be cointegrated. We return to the topic of cointegration in section 4.2.2.

[8] This means that the sampling distribution of the sample means approximates a normal distribution as the sample size grows regardless of the variable's distribution in the population.

The concept of weak dependence deals with how strongly related the random variables x_t and x_{t+h} are allowed to be as the time distance h increases. According to (Wooldridge, 2018), a time series process $\{x_t: t = 1,2, ...\}$ is, loosely speaking, "*said to be weakly dependent if x_t and x_{t+h} are 'almost independent' as h increases without bound*" (p. 368). A covariance stationary time series is said to be weakly dependent if "*if the correlation between x_t and x_{t+h} goes to zero 'sufficiently quickly' as $h \to \infty$*" (p. 368). Most economic time series are highly persistent and must be transformed to satisfy the weak dependence criteria. Often, this requires first differencing the time series process. When a time-series process upon first differencing exhibits weak dependence, it is said to be integrated of order one, or I(1). When a time-series process naturally exhibits weak dependence, it is said to be integrated of order zero, or I(0).

3.2.2 Assumption 2: No Perfect Multicollinearity

"*In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others*" (Wooldridge, 2018, p. 340).

The above assumption implies that OLS cannot estimate a model that suffers from perfect multicollinearity. Perfect multicollinearity arises if two or more independent variables exhibit a perfect linear relationship, i.e., they are perfectly predictable and not random. When this occurs, regression coefficients cannot be determined, and standard errors become infinite. Thus, no inference is obtained from the regression model. Although independent variables cannot be perfectly correlated under this assumption, they can be near perfectly correlated.

3.2.3 Assumption 3: Zero Conditional Mean

"*For each t, the expected value of the error u_t, given the explanatory variables for all time periods, is zero. Mathematically, $E(u_t|X) = 0, t = 1,2, ..., n$*" (Wooldridge, 2018, p. 340).

This assumption implies that the error term u in any given time period t must be uncorrelated with each explanatory variable in all time periods. When this assumption is satisfied, explanatory variables are strictly exogenous. Under large sample properties of OLS, the assumption is relaxed to require only that the explanatory variables are contemporaneously exogenous. Explanatory variables are said to be contemporaneously exogenous when the error term u in the given time period t is uncorrelated with the explanatory variables in the same time period, i.e., no restrictions are placed on how u_t relates to the explanatory variables in other time periods. Contemporaneous exogeneity, in other words, requires u_t to have zero conditional mean and to be uncorrelated with x_{tj}. Mathematically, this can be expressed as $E(u_t) = 0, Cov(x_{tj}, u_t) = 0, j = 1, \ldots, k$.

3.2.4 Assumption 4: Homoskedasticity

"*Conditional on X, the variance of u_t is the same for all t: $Var(u_t|X) = Var(u_t) = \sigma^2, t = 1,2, \ldots, n$*" (Wooldridge, 2018, p. 342).

The fourth assumption means that the variance of the error term is consistent across all observations, i.e., the variance does not depend on X. This condition, which is known as homoskedasticity, is fulfilled when u_t and X are independent, and $Var(u_t)$ is constant over time. When the condition is not fulfilled, the errors are heteroskedastic. Heteroskedasticity causes the estimators of the variances to be biased. Thus, the standard errors, and the test statistics that utilize these standard errors, are invalidated. Under large sample properties of OLS, the assumption is relaxed to require only that the errors are contemporaneously homoscedastic. That is, conditioning is only on the current time period and not across all observations. Mathematically, this is expressed as $Var(u_t|x_t) = \sigma^2$.

3.2.5 Assumption 5: No Serial Correlation

"*Conditional on X, the errors in two different time periods are uncorrelated: $Corr(u_t, u_s|X) = 0$, for all $t \neq s$*" (Wooldridge, 2018, p. 342).

The above implies that conditional on X, one observation of the error term should not have predictive value over the following observation. When this assumption is not satisfied, the errors are said to suffer from serial correlation or autocorrelation. This causes incorrect standard errors and thereby invalidates t-statistics and F-statistics. The interpretation under large sample properties of OLS is nearly identical, except conditioning is only on the explanatory variables in the same time periods as u_t and u_s, i.e., for all $t \neq s, E(u_t, u_s|x_t, x_s) = 0$.

3.2.6 Assumption 6: Normality

"*The errors u_t are independent of X and are independently and identically distributed as $Normal(0, \sigma^2)$*" (Wooldridge, 2018, p. 344).

Under classical linear model assumptions, a final assumption is needed to validate OLS standard errors, t-statistics, and F-statistics. This assumption implies that the error terms are normally distributed, while the ratio of each coefficient to its standard error is t-distributed. Note that a similar assumption is not required under large sample properties of OLS, as assumption one through five allows the use of central limit theorems to show that the OLS estimators will approximate a normal distribution in large samples.

3.2.7 Inference

Under classical linear model assumptions, assumptions one through three implies that the OLS coefficient estimator is unbiased. When assumptions four and five also hold, the OLS estimator of the variance is unbiased, meaning standard tools for OLS inference can be used. Under assumptions one through five, the OLS estimators are the best linear unbiased estimators of the real values. Adding assumption six, t-statistics and F-statistics can also be used. Under large sample properties of OLS, assumptions one through three implies that the OLS estimator is consistent. Under assumptions one through five, the OLS estimators are asymptotically normally distributed. Furthermore, standard errors, t-statistics, F-statistics, and LM-statistics are asymptotically valid. Although the OLS estimators are consistent when large sample properties hold, they are not necessarily the best unbiased estimators of the real values.

3.3 Density Forecasting

In general, a density forecast can be defined as an estimate of the probability distribution of the possible future values of the dependent variable y. While a point forecast measures the best estimate of y, a density forecast provides a complete description of the uncertainty associated with this estimate (Rossi, 2014). The following subsection starts with a presentation of the general approaches to construct density forecasts. Next, we provide a brief summary of empirical evidence concerning the distribution of exchange rate returns. Finally, we present the chosen procedures to estimate the densities.

3.3.1 Parametric, Non-Parametric and Semi-Parametric Distributions

We can distinguish between two general approaches to construct density forecasts. One is a parametric approach, where parameter estimation of the densities is based on the assumed probability distribution (Zhang, et al., 2020). The normal distribution is a commonly used parametric model. This distribution requires two input parameters: the mean μ, and the standard deviation σ. The parameters are typically estimated through linear models, such as OLS. Here, the estimated mean $\hat{\mu}$ is the point forecast, and the estimated standard deviation $\hat{\sigma}$ is the standard deviation of the in-sample forecast errors. Parametric models are easy to work with, estimate, and interpret. A potential drawback of parametric models is that they are rigid and may be inadequate if the relationship between random variables does not follow the prespecified distribution (Mahmoud, 2021).

A second approach, which has received increasing attention, is to construct the densities by the use of non-parametric models. The idea behind non-parametric density estimation is to treat the data set as if it were drawn from some unspecific or unknown empirical distribution function. Non-parametric models have the advantage of being more flexible compared to parametric models (Mahmoud, 2021). A common method to construct a non-parametric density forecast is by the use of quantile regression, followed by a smoothing procedure to obtain the densities (van der Meer et al., 2018). This method is detailed in the next section.

A potential disadvantage of the non-parametric approach is that the constructed densities will be completely dependent on the data sample. This represents a drawback if the sample size is too small, or in general not representative of the entire population. In addition, the power of statistical tests based on non-parametric densities is known to decrease as the number of explanatory variables increases. This is known as the *"curse of dimensionality"* (Härdle et al., 2004).

Consequently, a third way of constructing density forecasts has emerged, which combines the features of the parametric and non-parametric approach. This is known as a semi-parametric approach. Semi-parametric models overcome some of the drawbacks of non-parametric models, while still providing more flexibility compared to parametric models (Mahmoud, 2021).

3.3.2 Density Estimation

The empirical literature on the distribution of exchange rate returns generally agrees that daily exchange rate returns are more peaked and exhibit more probability mass in the tails compared to a normal distribution. This is shown irrespective of the exchange rate regime (de Vries & Leuven, 1994). However, Coppes (1995) argues that monthly exchange rate returns are shown to be more normally distributed. In light of the, to some extent, inconclusive empirical literature regarding the distribution of monthly exchange rate returns, we consider two types of densities for each model.

The first density is a normal distribution, where the probability density function (PDF) can be defined as:

$$f(y;\ \mu, \sigma) = \frac{1}{\sigma}\Phi\left(\frac{y-\mu}{\sigma}\right),$$

where $\Phi(.)$ is the PDF of the standard normal distribution, μ the location parameter, and σ the scale parameter. The parameters are estimated using OLS, with the general equation:

$$s_{t+h} - s_t = \alpha + \beta x_t + \varepsilon_{t+h},$$

where the dependent variable $s_{t+\mathrm{h}} - s_t$ is the h-month-ahead exchange rate return, x_t a vector of explanatory variables, and $\varepsilon_{t+\mathrm{h}}$ the error term. The estimate of the location parameter is the point forecast $\hat{\mu}_{t+h} = \alpha + \beta x_t$, and the estimate of the scale parameter is the standard deviation of the in-sample errors, $\hat{\sigma}_{t+\mathrm{h}} = \sqrt{Var(}\varepsilon_{t+\mathrm{h}})$.

To account for the possibility of monthly exchange rate returns not being normally distributed, we consider a second approach to estimate the densities. Here, we follow the same procedure as in Adrian et al. (2019), where they use a two-step semi-parametrical approach to construct density forecasts for US output growth. The same method has also been employed by De Santis & Van der Veken (2020) for a risk assessment of US output growth and Yapi (2020) for exchange rates. In the following, we present a thorough explanation of the procedure.

The first step is to estimate the conditional quantiles. This is conducted by use of the quantile regression framework put forward by Keonker & Bassett (1978). Quantile regression differs from OLS, as it is based on asymmetric minimization of the weighted absolute errors. Koenker & Bassett argue that quantile regression estimators are more robust and efficient compared to OLS in a non-gaussian setting.

Let $y_{t+\mathrm{h}}$ denote the h-month-ahead exchange rate return $(s_{t+\mathrm{h}} - s_t)$, and x_t the vector of explanatory variables. We model the different quantiles of the exchange rate return's conditional distribution independently by estimating:

$$\hat{\beta}_{\tau} = \underset{\beta_{\tau}}{\text{argmin}} \sum_{t=1}^{T-\text{h}} \left(\tau * \mathbb{1}_{(y_{t+\text{h}} \geq x_t\beta)} |y_{t+\text{h}} - x_t\beta| + (1-\tau) * \mathbb{1}_{(y_{t+\text{h}} < x_t\beta)} |y_{t+\text{h}} - x_t\beta|\right),$$

where the indicator function $\mathbb{1}_{(y_{t+\text{h}} \geq x_t\beta)} = \begin{cases} 1, & if\ y_{t+\text{h}} \geq x_t\beta \\ 0, & if\ \ y_{t+\text{h}} < x_t\beta \end{cases}$

The predicted values from these quantile regressions $\hat{Q}_{y_{t+\text{h}|x_t}}(\tau|x_t) = x_t\hat{\beta}_{\tau}$ corresponds to the quantiles τ of the predictive distribution of $y_{t+\text{h}}$ conditional on x_t. For a low (high) level of τ, the τth conditional quantile of the exchange rate return ($x_t\hat{\beta}_{\tau}$) describes the behavior of the dependent variable $y_{t+\text{h}}$ at the left (right) tail of the distribution.

In the second step, the conditional quantiles $\hat{Q}_t(\tau|x_t)$ obtained from the quantile regressions are used to fit a skewed t-distribution. The PDF of the skewed t-distribution can be denoted:

$$f(y;\ \mu,\ \sigma,\ \alpha,\ \nu) = \frac{2}{\sigma} t\left(\frac{y-\mu}{\sigma}; \nu\right) T\left(\alpha \frac{y-\mu}{\sigma} \sqrt{\frac{\nu+1}{\nu+\left(\frac{y-\mu}{\sigma}\right)}}; \nu+1\right),$$

where $t(\cdot)$ and $T(\cdot)$ respectively denote the PDF and CDF of the student t-distribution. The skewed t-distribution has four parameters, where μ is the location parameter, σ the scale parameter, α the shape parameter, and ν the fatness parameter. The shape parameter α is where the skewed t-distribution differs from the student's t-distribution.

An advantage of using the skewed t-distribution is that it provides flexibility. For example, when $\alpha = 0$, the skewed t-distribution reduces to a t-distribution. When $\alpha = 0$ and $\nu = \infty$, it reduces to a normal distribution, and when $\nu = \infty$ and $\alpha \neq 0$, the distribution is skewed normal.

To obtain the parameter estimates of the skewed t-distribution, the method uses an optimization procedure where the estimates of the 5th, 25th, 75th, and 95th quantile from quantile regression serve as input variables. The optimization procedure minimizes the squared distance between the estimated conditional quantiles and the inverse cumulative density function of the skewed t-distribution. The equation is defined as:

$$\{\hat{\mu}_{\text{t+h}}, \hat{\sigma}_{t+h}, \hat{\alpha}_{t+h}, \hat{\nu}_{t+h}\} = \underset{\mu,\,\sigma,\,\alpha,\,\nu}{\text{argmin}} \sum_{\tau} \left(\hat{Q}_{\text{y}_{\text{t+h}}}(\tau|x_t) - F^{-1}(\tau,\ \mu,\ \sigma,\ \alpha,\ \nu)\right)^2,$$

where F is the cumulative probability distribution function of the skewed t-distribution and f is the associated probability density function.

3.4 Regression Diagnostics

In section 3.2, we presented both the classical linear model assumptions and the large sample assumptions of OLS. Due to the nature of the time-series data that underlies the empirical models, we appeal to large sample properties of OLS in this thesis. Furthermore, as we do not attempt to verify the causality of the explanatory variables, we are mainly concerned with the OLS estimators being consistent, i.e., that assumption one through three of large sample properties of OLS holds. Out of these three, it is primarily the stationarity assumption that is tangible to test. For this purpose, we apply the augmented Dickey-Fuller and KPSS tests.

3.4.1 Augmented Dickey-Fuller Test

A time series that contains a unit root displays a systematic, unpredictable pattern. Unit roots are, in other words, a cause of non-stationarity, meaning a time series variable can be tested for stationarity by assessing whether or not it possesses a unit root. Several tests are available for this purpose, with the most common one being the Dickey-Fuller test. The Dickey-Fuller test tests whether $\rho = 1$ in the model $y_t = \alpha + \rho y_{t-1} + e_t$, in which case y_t has a unit root. A convenient way of re-writing this equation is to detrend the model to the form

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t,$$

and define $\theta = \rho - 1$ (Wooldridge, 2018). Given $E(e_t | y_{t-1}, y_{t-2}, \dots, y_0) = 0$, the Dickey-Fuller test can then be applied to test $H_0\colon \theta = 0$ against $H_A\colon \theta \neq 0$. At the 5% significance level, the null hypothesis of a unit root presence is rejected in favor of the alternative hypothesis that the data is stationary at a critical value of $t_{\hat{\theta}} < -2{,}86$.

In models with more complicated dynamics, the Dickey-Fuller test can be extended to clean up serial correlation in Δy_t. The above formula is then augmented to include lags, e.g.:

$$\Delta y_t = \alpha + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + e_t,$$

where $|\gamma_1| < 1$. The null hypothesis of a unit root presence is tested by regressing Δy_t on $y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p}$. The number of lags included determines how many time-delayed expressions are required before the serial correlation in the error term is equal to zero. There are no hard rules to follow regarding how many lags to include. The data used in this thesis is monthly, in which case Wooldridge (2018) suggests one might include twelve lags.

3.4.2 Kwiatkowski-Phillips-Schmidt-Shin Test

Like the augmented Dickey-Fuller test, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test is a type of unit root test that tests a time series for stationarity. Unlike the augmented Dickey-Fuller test, the KPSS test can test for stationarity around a deterministic trend. This makes the test useful as a complementary diagnostic tool when testing for stationarity. Following Syczewska (2010), the KPSS model takes the form of:

$$y_t = \xi t + r_t + \varepsilon_t,$$

where $r_t = r_{t-1} + u_t$. In this equation, ξ is a deterministic trend and r_t a random walk process, while ε_t and u_t are error terms that by assumption are identically distributed independent random variables with an expected value of zero and constant variation.

The KPSS test is applied to test the $H_0\colon \sigma_u^2 = 0$ against $H_A\colon \sigma_u^2 \neq 0$. If $\xi = 0$, y_t is stationary around the constant r_0 under the null hypothesis, while $\xi \neq 0$ implies that y_t is stationary around a linear trend. Here, we test the latter null hypothesis, in which case, the LM-statistic is defined mathematically as follows:

$$LM = \sum_{i=1}^{t} S_t^2 \,/\sigma_\varepsilon^2,$$

where S_t is a partial sum of errors. At the 5% significance level, the null hypothesis that the time series process is stationary is rejected in favor of the alternative hypothesis that a unit root is present at a critical value of $LM > 0{,}463$.

3.5 Point Forecast Evaluation

The forecast evaluation in the empirical analysis in this thesis is two-fold. We start by conducting a standard point forecast evaluation, followed by a density forecast evaluation. In this section, we outline the point forecast evaluation methodology. The first step is to compute the root mean squared error for every model and compare that to the random walk model. We then use the Diebold-Mariano test to assess the statistical significance of the results.

Additionally, we use the Pesaran-Timmermann test to assess whether the models are able to predict the directional change of the EUR/NOK exchange rate. The test statistics are detailed further below.

3.5.1 Root Mean Square Error

The root mean square error (RMSE) is a relative performance measure of point forecasts. It is defined as the square root of the mean squared error, which measures the average squared difference between fitted and observed values. RMSE is defined mathematically as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}},$$

where $\hat{y}_t$ is the point forecast made in period $t - h$, and y_t is the observed value. Given that the errors are squared, equivalent errors with differentiated signs are equally weighted. This means that the RMSE is always positive. Thus, a lower RMSE also indicates a better fit.

3.5.2 Diebold-Mariano Test

Historically, assessments of forecast accuracy of point estimates revolved around comparing error measures like mean squared error and mean absolute errors between a forecast model and a benchmark model. However, limited attention was given to the issue of statistical significance. This was the basis of Diebold and Mariano's *Comparing Predictive Accuracy* (1995) paper, in which they proposed a general statistical test for comparing forecast accuracy. This test, known simply as the Diebold-Mariano test, has since gained wide traction in the forecasting literature.

The Diebold-Mariano test is based on a sample path of loss differentials $\{d_t\}_{t=1}^{T}$. If mean squared error is used as the loss function, then $d_t = e_t^2 - \breve{e}_t^2$, i.e., the loss differential is defined as the difference between the forecast error e_t^2 and the benchmark error $\breve{e}_t^2$. A key assumption is that this loss differential is a covariance stationary time series. The DM statistic used to test $H_0: E(e_t^2 - \breve{e}_t^2) = 0$ against $H_A: E(e_t^2 - \breve{e}_t^2) \neq 0$ is defined as follows:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi f_d(0)}{T}}},$$

where DM under the H_0 asymptotically converge to a standard normal distribution $N(0,1)$. The null hypothesis is rejected in favor of the alternative hypothesis that the forecast model has different levels of accuracy if the DM statistic falls outside the range of $\left(-z_{\frac{\alpha}{2}},\ z_{\frac{\alpha}{2}}\right)$. The z-value is the critical value from a standard normal distribution corresponding to half the desired significance level α. At a 5% significance level, this corresponds to $z = \pm 1{,}96$.

3.5.3 Pesaran-Timmermann Test

It is often of interest to predict the proper directional movement of a time series in a financial context. This might not always coincide with returning small forecast errors. Hence, we find it beneficial to include a directional forecast accuracy test in our empirical analysis. For this, we utilize the Pesaran-Timmermann (1992) test, which is prominent within the literature.

The test first defines y_t as the return series of the actual value at time t, $\hat{y}_t$ as the predictor value of y_t based on information available at time $t-h$, and n as the total number of observations in the forecast series. Then, the following dummy variables are calculated:

$$Y_t = 1 \text{ if } y_t > 0 \text{ and } Y_t = 0 \text{ if } y_t \leq 0$$
$$\hat{Y}_t = 1 \text{ if } \hat{y}_t > 0 \text{ and } \hat{Y}_t = 0 \text{ if } \hat{y}_t \leq 0$$
$$Z_t = 1 \text{ if } y_t\hat{y}_t > 0 \text{ and } Y_t = 0 \text{ if } y_t\hat{y}_t \leq 0$$

Next, the test denotes $P_y = \frac{1}{n}\sum_{t=1}^{n} Y_t$ and $P_{\hat{y}} = \frac{1}{n}\sum_{t=1}^{n} \hat{Y}_t$. This corresponds to the proportion of time that the actual dependent variable and its respective forecast is greater than zero. Under the assumption that y_t and $\hat{y}_t$ are independently distributed, the amount of correct sign predictions is binomially distributed with n trials. The success probability is equal to:

$$P_* = P_y P_{\hat{y}} + (1 - P_y)(1 - P_{\hat{y}})$$

The respective sample estimators equal:

$$\hat{P}_y = \frac{1}{n}\sum_{t=1}^{n} y_t,\ \hat{P}_{\hat{y}} = \frac{1}{n}\sum_{t=1}^{n} \hat{y}_t \text{ and } \ \hat{P}_* = \hat{P}_y \hat{P}_{\hat{y}} + (1 - \hat{P}_y)(1 - \hat{P}_{\hat{y}})$$

Under the H_0 that y_t and $\hat{y}_t$ are independently distributed, and thus that $\hat{y}_t$ has no power in forecasting y_t, the test statistic is:

$$PT = \frac{\hat{P} - \hat{P}_*}{\sqrt{var(\hat{P}) - var(\hat{P})}}$$

The null hypothesis is rejected in favor of the alternative hypothesis that $\hat{y}_t$ has power in forecasting y_t if $1 - z < a$, where z is the critical value from a standard normal distribution corresponding to the PT statistic and a is the desired significance level. At a 5% significance level, the critical z-value above which the null hypothesis is rejected is 1,64.

3.6 Density Forecast Evaluation

This section presents the density forecast evaluation framework adopted in the empirical analysis. Evaluating density forecasts might appear to be a difficult task, as the true sequence of density forecasts is never observed. However, there exist several reliable methods that are constructed for this purpose. We start by deriving the Probability Integral Transforms, which serves as input for the following calibration tests. These tests are the Anderson-Darling test and the Berkowitz test. We then outline the logarithmic score and a subsequent test statistic used to evaluate the relative performance of the forecasted densities. Finally, we describe risk measures to quantify appreciation and depreciation tail risk and a statistical test to evaluate these measures. The framework is detailed in the following.

3.6.1 The Probability Integral Transform

Diebold et al. (1998) put forward a widely used method to evaluate density forecasts by the use of the Probability Integral Transform (PIT). A PIT is the cumulative probability evaluated at the actual, realized value of the dependent variable. It measures the likelihood of observing a value less than the actual realized value, where the probability is measured by the density forecast (Rossi, 2014).

Let the realization of the dependent variable y be denoted by Y_t. The forecasted density for the dependent variable is denoted $\hat{f}_t$, where the forecast was made in period $t - h$. The PIT u_t is then given by:

$$u_t = \hat{F}_t(Y_t) = \int_{-\infty}^{Y_t} \hat{f}_t\,(q)dq,$$

where $\hat{F}_t$ denotes the forecasted cumulative distribution function associated with $\hat{f}_t$.

Theory suggests that if the series of PITs $\{u_t\}_{t=1}^{T}$ are independent, identically distributed, and uniform over the interval (0,1), the forecasted densities are correctly specified (Diebold et al., 1998). This can be denoted as $\{u_t\}_{t=1}^{T} \sim i.i.d.\ U(0,1)$. When this is the case, densities are referred to as perfectly calibrated, meaning that the density forecast coincides with the true sequence of densities associated with the predicted variable (Knüppel, 2015). For $h > 1$, it is standard practice in the literature to omit the independence assumption as it well-known that multi step-ahead forecasts produce serially correlated forecast errors, leading to serially correlated PITs (Hall & Mitchell, 2007). Hall & Mitchell argue that even in the presence of serial correlation, the multi-step ahead density forecasts can be correctly specified.

3.6.2 Anderson-Darling Test

The uniformity property means that no matter whether we consider high or low realizations of the variable we are forecasting, the probability that the value is higher (lower) than the forecasted value is, on average, the same over time (Rossi, 2014). It then follows that a histogram of the PIT series takes the shape of a rectangle, and the subsequent empirical CDF follows a straight 45-degree line. As uniformity is a common property that must be in place for all forecast horizons in order for the density forecasts to be evaluated as well-calibrated, we devote a test solely for this property.

The Anderson-Darling test assesses whether the PITs of the realized dependent variable with respect to the forecast densities follow $U(0,1)$. Simulation exercises have documented the Anderson-Darling test's power advantage compared to other statistical tests for uniformity, such as the Kolmogorov-Smirnov test (Rossi, 2014).

The test statistics is defined as:

$$A^2 = -T - \frac{1}{T}\sum_{t=1}^{T}(2t-1)\left[\ln U(u_t) + \ln\left(1 - U(u_{T+1-t})\right)\right],$$

where $U(\cdot)$ is the CDF of the uniform distribution. The null hypothesis that the data follow $U(0,1)$ is rejected in favor of the alternative hypothesis that the data do not follow $U(0,1)$ if the test statistic is greater than the critical value corresponding to the desired significance level of the theoretical uniform distribution.

3.6.3 Berkowitz Test

Berkowitz (2001) argues that a drawback of statistical tests for uniformity is that they are not sufficiently powerful for small samples and do not account for the $i.i.d$ property. He proposed a test based on the inverse normal of the PITs, which is a joint test for the uniformity and the

$i.i.d$ property. Let the inverse normal of the PITs be denoted $z_t = \Phi^{-1}(u_t) = \Phi^{-1}\left(\hat{F}_t(Y_t)\right)$, where Φ^{-1} is the inverse normal of the standard normal distribution function. Under the null hypothesis, it follows that the sequence $\{z_t\}_{t=1}^{T} \sim i.i.d.\ N(0,1)$.

The test is conducted using a first-order autoregressive model:

$$\mathrm{z_t} = \mu + \rho z_{t-1} + \varepsilon_t, \text{ with } t = 1,2,\dots,T \text{ and } \varepsilon_t \sim N(0,\sigma^2)$$

If the null hypothesis is true, we have that $\mu = 0,\ \rho = 0\ and\ var(\varepsilon_t) = 1$. The test statistic is based on the log-likelihood function, defined as:

$$\ln L = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{\left(z_1 - \left(\frac{\mu}{1-\rho}\right)\right)^2}{\frac{2\sigma^2}{1-\rho^2}} - \frac{T-1}{2}\log(2\pi) - \frac{T-1}{2}\log(\sigma^2) - \sum_{t=2}^{T}\left(\frac{(z_t-\mu-\rho z_{t-1})^2}{2\sigma^2}\right)$$

To test whether $\{z_t\}_{t=1}^{T} = \ \sim i.i.d.\ N(0,1)$, Berkowitz suggests using the joint test statistic:

$$\hat{\beta}^{ind} = (\ln L(\hat{\mu}, \hat{\sigma}, \hat{\rho}) - \ln L(0, 1, 0)),$$

which converges to a $\chi^2(3)$-distribution under the null hypothesis. For $\mathrm{h} > 1$, a modified version of the test is used by practitioners (Knüppel, 2015), which relaxes the independence property. Here, we use the test statistics:

$$\hat{\beta} = 2\left(\ln L(\widehat{\mu,}\ \hat{\sigma}, \hat{\rho}) - \ln L\left(0, \sqrt{1-\hat{\rho}^2}\ ,\ \hat{\rho}\right)\right),$$

which now converges to a $\chi^2(2)$-distribution under the null hypothesis.

3.6.4 Logarithmic Score

Another approach to evaluate density forecasts is by the use of scoring rules. Scoring rules measure the quality of the density forecasts by assigning a numerical score to the density forecast based on how well it forecasts the dependent variable. If the observed dependent variable falls within an area with high (low) predictive density, this gives a subsequent high (low) score. A benefit of using scoring rules is that it allows us to rank competing models (Gneiting et al., 2007).

One of the most used scoring rules to evaluate density forecasts is the logarithmic score, originally proposed by Good (1952). To evaluate a series of density forecasts, the logarithmic score is defined as:

$$S^l(\hat{f}_t(Y_t)) \ = \ln(\hat{f}_t(Y_t)),$$

where $\ln\left(\hat{f}_t(Y_t)\right)$ is the natural logarithm of the density forecast made at time $t-h$, evaluated at the observed value of the dependent variable. Based on a sequence of two competing density forecasts $\hat{f}_t$ and $\hat{g}_t$, we can evaluate their predictive performance by their average score (Diks et al., 2011):

$$\mathrm{T}^{-1}\sum_{t=1}^{T}\log\hat{f}_t\,(Y_t) \text{ and } T^{-1}\sum_{t=1}^{T}\log\hat{g}_t\,(Y_t)$$

A higher score implies a better model. To test if the predictive performance of the two competing models is significantly different, we can use a test statistic based on their average log score difference:

$$d_t^l = \log\hat{f}_t\,(Y_t) - \ \log\hat{g}_t\,(Y_t).$$

The null hypothesis is given by H_0: $\mathbb{E}\left(d_t^l\right) = 0$ for all $t,\ t+1, \dots, T$. Let $\overline{d}^l$ denote the sample average of the log score differences, that is $\overline{d}^l = n^{-1}\sum_{t=1}^{T} d_t^l$. In order to test H_0 against the alternative H_a: $\mathbb{E}\left(d_t^l\right) \neq 0$, we use a Diebold & Mariano (1995) type statistic:

$$\mathrm{t} = \frac{\overline{d}^l}{\sqrt{\frac{\hat{\sigma}^2}{n}}},$$

where $\hat{\sigma}^2$ is a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator of $\sigma^2 = Var(\sqrt{T}\,\overline{d}^l)$.

3.6.5 Expected Shortfall and Longrise

Risk measures such as Value-at-Risk are commonly applied to quantify the risk associated with extreme outcomes, so-called tail risk. Coverage tests such as the Kupiec test or Christoffersen test are often applied to assess the reliability of these risk measures. However, a drawback of such tests is their inability to account for extreme losses and some mathematical properties[9]. A more frequently used risk measure in recent years is Expected Shortfall (ES), defined as the average loss beyond the VaR level:

$$ES_{t+\mathrm{h}} = \frac{1}{\pi}\int_0^{\pi} \hat{F}^{-1}_{y_{t+\mathrm{h}|x_t}}(\tau|x_t)d\tau,$$

where π is the risk level and $\hat{F}^{-1}_{y_{t+\mathrm{h}|x_t}}(\tau|x_t)$ the estimated CDF. This corresponds to the appreciation tail risk of the currency. For exchange rates, we are also interested in the right tail of the distributions. Adrian et al. (2019) named this counterpart as Expected Longrise (EL), defined as:

$$EL_{t+\mathrm{h}} = \frac{1}{\pi}\int_{1-\pi}^{1} \hat{F}^{-1}_{y_{t+\mathrm{h}|x_t}}(\tau|x_t)d\tau$$

We consider 5% ES and EL. To evaluate whether the forecasted ES and EL are well specified, we use the conditional calibration test of Nolde & Ziegel (2017). For the pair VaR and ES at risk level $\tau \in (0,1)$, the test chooses the strict identification function:

$$V(Y,\, v,\, e) = \begin{pmatrix} \tau - \mathbb{1}_{\{Y \le v\}} \\ e - v + \mathbb{1}_{\{Y \le v\}}(v - Y)/\tau \end{pmatrix},$$

where e is the expected shortfall, v the Value-at-Risk, and Y the observed value. The test's expectation is zero, if and only if, v and e equal the true VaR and ES of the observed value Y. The two-sided conditional calibration backtest of the forecast for the VaR, $\hat{v}_t$, and for the ES, $\hat{e}_t$ is based on the hypotheses:

$$\mathbb{H}_0^{2s} = \mathbb{E}[V(Y_{t+\mathrm{h}}, \hat{v}_t, \hat{e}_t)] = 0 \text{ agains}t\ \mathbb{E}[V(Y_{t+\mathrm{h}}, \hat{v}_t, \hat{e}_t)] \neq 0, \text{ and}$$

Nolde & Ziegel (2017) propose to use the Wald-type test statistics:

$$T_{CC} = T\left(\frac{1}{T}\sum_{t=1}^{T} V(Y_t, \hat{v}_t, \hat{e}_t)\right)^T \hat{\Delta}_T^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} V(Y_t, \hat{v}_t, \hat{e}_t)\right),$$

[9] Value-at-Risk tests fail to account for the so-called subadditivity property. For further reading, see Chen (2018).

where $\hat{\Delta}_T^{-1} = \frac{1}{T}\sum_{t=1}^{T} V(Y_t, \hat{v}_t, \hat{e}_t))\left(V(Y_t, \hat{v}_t, \hat{e}_t)\right)^T$ is a consistent estimator of the covariance of the two-dimensional vector $V(Y_t, \hat{v}_t, \hat{e}_t)$. Under $\mathbb{H}_0$, the test statistic asymptotically follows a χ_2^2 distribution.

4. Data

This section presents the data variables included in our exchange rate models. We give a brief explanation on why the data is included, state their expected effect on the exchange rate, and explain how the data is transformed for analysis purposes. The data are summarized below. For plots c-f, Norwegian data is represented by a blue line and Eurozone data by a green line.

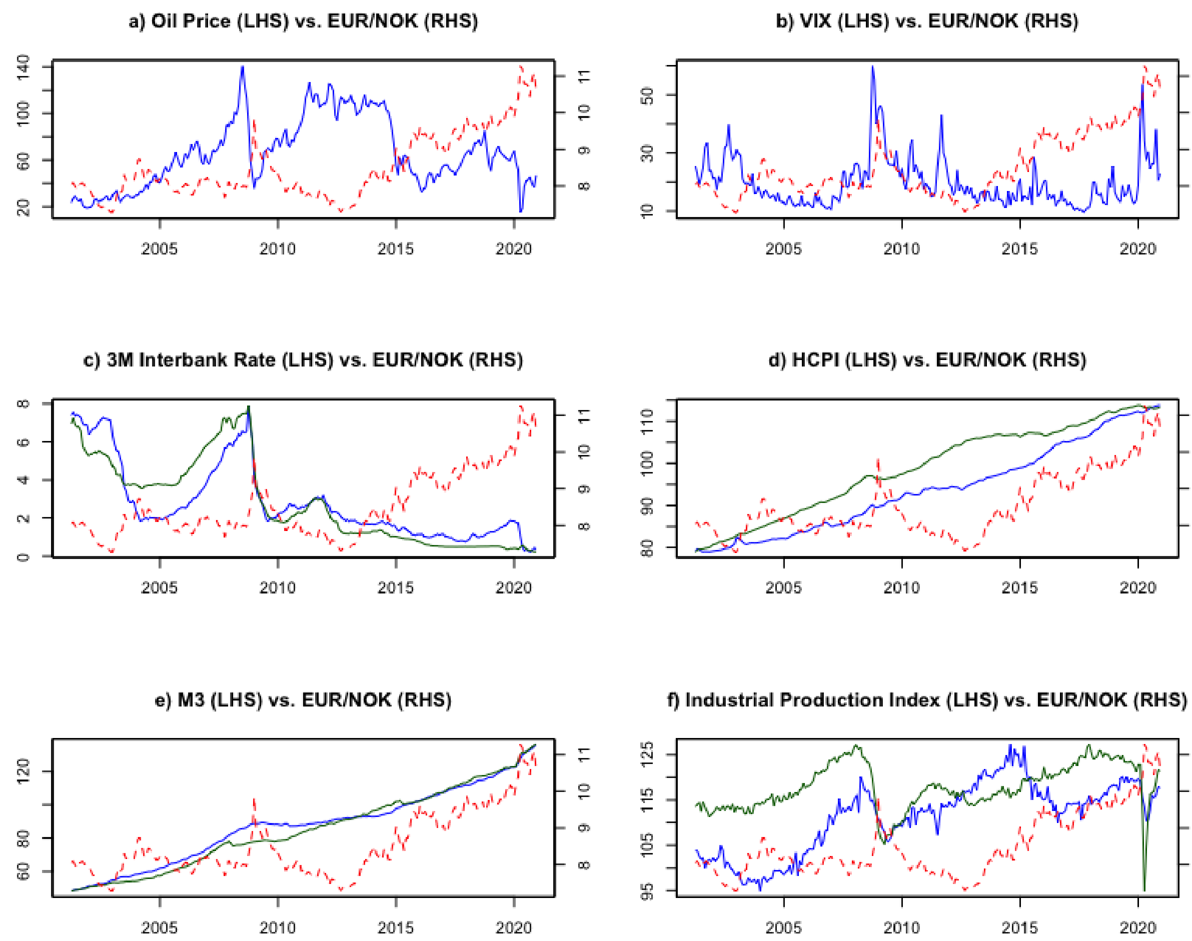


Figure 1: Summary of data used for independents variables (lhs) and the dependent variable (rhs).

4.1 Variables

The data is selected based on exchange rate theory and empirical findings. We have selected April 1st, 2001 as the starting point to avoid a structural break in the model associated with the introduction of the floating exchange rate regime in Norway on March 29th, 2001. The dataset includes monthly observations up to and including December 1st, 2020, yielding 237 observations. All data is secondary data originated for other initial purposes than this thesis. The data sources are presented in appendix I.

4.1.1 EUR/NOK

The dependent variable in all models in this thesis is the nominal EUR/NOK exchange rate. We apply a currency convention where NOK is expressed per unit of EUR. Bernhardsen & Røisland (2000) point out that the krone exchange rate is a non-unambiguous size that depends on the currency against which it is measured. The exchange rate should thus be measured against a basket of currencies rather than a single currency when quantifying its international value. However, in this thesis, the goal is not to quantify the international value but rather assess the forecasting ability of a tradeable currency pair. The EUR/NOK currency pair is chosen for the following reasons: (i) it is one of the most liquid currency pairs involving the Norwegian krone, (ii) the Euro area is Norway's largest and most important trading partner, and (iii) as Norway is part of the European economic area, comparable and consistent economic indicators are widely available. As stated by Rossi (2013), it is standard within the out-of-sample forecasting literature to use nominal exchange rates over real ones. In the following subsections, we adopt the variable name $eurnok$ for the EUR/NOK exchange rate.

4.1.2 Brent Crude Oil Price

According to economic theory, an increase in commodity prices should lead to a real appreciation of a commodity-exporting country's currency. As Torvik (2016) explains, rising commodity prices lead to higher revenues for the commodity-exporting country. Since production opportunities are fixed, consumption opportunities rise beyond the total production value of the shielded and exposed sectors. As such, the country will want to consume more of both shielded and exposed goods. However, as the country must produce everything it consumes of shielded goods, companies in sectors shielded from foreign competition must receive a signal to increase employment to facilitate the increased demand. This can only happen if the prices of shielded goods increase relative to the prices of exposed goods. An increase in the oil price should therefore be expected to lead to a real appreciation of the Norwegian krone. Alendal (2010) also points to an increased demand for oil-related stocks on Oslo Børs, expectations of rising interest rates associated with an expected increase in economic activity, and pure psychological effects as factors that could lead to higher demand for the Norwegian krone as the oil price rises, thereby causing an exchange rate appreciation.

The above discussion would suggest an inverse relationship between the NOK/EUR exchange rate and the oil price. The empirical evidence is also broadly supportive of commodity prices as predictors of exchange rates in in-sample estimation, see, e.g., Chen & Rogoff (2003). While this also holds for Norwegian data, Akram (2004) finds that the relationship is non-linear (2004), i.e., the strength of the correlation depends on the level and the trend in oil prices. Akram furthermore finds that an equilibrium correction model incorporating the non-linear relationship outperforms a random walk model in out-of-sample forecasting at a twelve-quarter horizon. The above findings lead us to include the oil price in the BEER model specification. We use the variable name oil to reference the Brent crude oil price in following subsections.

4.1.3 CBOE Volatility Index

Uncertainty in financial markets is another factor that is well documented to affect currencies, see, e.g., De Bock & de Carvalho Filho (2013). Investors will typically flee to assets that serve as a reliable and stable store of value during market turbulence. Within the foreign exchange market, currencies exhibiting such properties are referred to as safe-haven currencies. In contrast, currencies that tend to fluctuate or depreciate erratically against other currencies are referred to as soft currencies. Previous empirical work finds that the Norwegian krone tends to depreciate during times of increased volatility in the foreign exchange market, see, e.g., Bernhardsen and Røisland (2000). While not usually included amongst the major safe-haven currencies, the Euro is generally considered a safe haven alternative. Ronaldo and Söderlind (2010), e.g., find that the Euro exhibits safe haven characteristics during crises. As such, we expect to see a positive relationship between the EUR/NOK exchange rate and volatility in financial markets. Thus, we include a volatility measure in the BEER model specification.

Several indices are constructed to measure volatility in various currency pairs and broader financial markets. Indices commonly used in exchange rate models of the Norwegian krone are, e.g., the S&P 500 options-based implied volatility index (VIX), the global risk indicator (GRI), and various options-based indices measuring implied volatility in major currency pairs, such as Deutsche Bank's currency volatility index (CVIX). As pointed out by Akram (2019), the krone may have more in common with emerging market currencies than the world's major currencies. Akram furthermore points out that the implied volatility of emerging market currencies covaries with overall economic uncertainty as measured by equity options. As a result, we choose to use the VIX as the uncertainty measure in the BEER model specification. In later subsections, the CBOE volatility index is referenced by the variable name vix.

4.1.4 Three-Month Interbank Rates

We have previously elaborated on how interest rates are theorized to affect exchange rates in our review of uncovered interest parity and how its positioned as a central assumption in several theories of exchange rate determination. Thus, we include an interest rate differential in several of the models specified in the empirical analysis. As Bernhardsen and Røisland (2000) point out, it is vital to be aware that the interest rate differential is an endogenous variable. This is because the central bank takes the exchange rate into account when determining the policy rate. As such, the coefficient must be interpreted with caution.

Earlier empirical work primarily utilizes three-month or twelve-month interbank rates to construct interest rate differentials, with Bernhardsen (2012) finding a broad impact from the policy rate to money market rates both in Norway and abroad. Due to comparably better data access for three-month interbank rates, we have chosen to use the former in our exchange rate model, i.e., three-month NIBOR for Norway and three-month EURIBOR for the Eurozone. The $(i - i^{*})$ variable is constructed by subtracting the EURIBOR rate from the NIBOR rate.

4.1.5 Harmonized Consumer Price Indices

As elaborated on in the theory section, relative purchasing power parity, which underlies several theories of exchange rate determination, states that the foreign exchange rate is expected to adjust with the relative change in price level between two countries. Additionally, models using Taylor rule fundamentals and certain monetary model specifications incorporate an inflation differential. Thus, measures of the price levels in Norway and the Eurozone are required to construct variables for price differences and inflation differences. This thesis utilizes the harmonized consumer price indices for Norway and the Eurozone for this purpose. The harmonized consumer price index (HCPI) is a statistic designed for consistent comparisons of price developments between European countries (Statistics Norway, 2021).

The price differential variable $(p - p^{*})$ is constructed by subtracting the log-transformed HCPI for the Eurozone from Norway's log-transformed HCPI. The inflation differential variable $(\pi - \pi^{*})$ is constructed by subtracting the annual percentage change in the HCPI for the Eurozone from the annual percentage change in Norway's HCPI. Note that both indices have been seasonally adjusted.

4.1.6 M3 Money Supply Indices

Referring back to section 2.4, the monetary model of exchange rate determination includes a money demand differential variable. Money demand is assumed to be equal to the respective money supply in equilibrium. Hence, we need a measure for the money supply in Norway and the Eurozone. The most common money supply measures are the M1, M2, and M3 monetary aggregates. Both the Norwegian and European Central Bank follow the principles of "Manual on MFI Balance Sheet Statistics," making the metrics comparable for cross-country analysis purposes. The ECB defines M1 as the "*sum of currency in circulation and overnight deposits*" (n.d.). M2 additionally includes "*deposits with an agreed maturity of up to two years and deposits redeemable at notice of up to three months*." Finally, "*M3 is the sum of M2, repurchase agreements, money market fund shares/units and debt securities with a maturity of up to two years*". The literature offers no specific guidelines on which definition to use, but we have opted to use M3 in this thesis due to comparably better data access. The $(m - m^*)$ variable is constructed by subtracting the log-transformed M3 index for the Eurozone Norway's log-transformed M3 index.

4.1.7 Manufacturing Indices

Finally, the monetary exchange rate model includes a variable for differences in real income domestically and abroad. GDP differences could function as a proxy for this variable; however, lack of data at monthly intervals and otherwise short time series complicates this. Instead, we utilize an index of industrial production, in line with, e.g., Reese & Rogoff (1983). In Norway, mainland GDP is often used to assess the economic situation, as revenues from oil, gas, and shipping companies can vary greatly without impacting unemployment significantly. Given this, we use a manufacturing index that excludes oil and gas extraction, mining, and power supply. The resulting index correlates closely with GDP and should thus be an adequate proxy for real income. The $(y - y^*)$ variable is constructed by subtracting the log-transformed manufacturing index for the Eurozone from Norway's log-transformed manufacturing index. The indices have been seasonally adjusted.

We face similar problems when attempting to use GDP measures to construct the output gap differential variable included in models with Taylor rule fundamentals. Hence, the manufacturing index is also used to construct the $(y_{gap} - y_{gap}^{*})$ variable. For this, we use a Hodrick-Prescott filter. The Hodrick-Prescott (HP) filter is a data-smoothing technique that is commonly used to separate short-term fluctuations from a long-term trend. The long-term trend's sensitivity to short-term fluctuations is decided by the multiplier λ. For data at a monthly frequency, the lambda is commonly set to $\lambda = 14\ 400$. Hence, this is also what is used in this thesis. A problem with the HP-filter is that the short-term fluctuations converge on the long-term trend at the end of the sample, which can reduce the validity of the output gap measure. A common way of dealing with this, is to use an ARIMA forecast model to extend the time series. However, the usefulness of such an extension is limited by uncertainty over how many forecasts are needed and sensitivity to this number (Apel et al., 1996). Molodtsova & Papell (2009) presented one of the most prominent research papers on monthly out-of-sample forecasting of exchange rates with Taylor rule fundamentals. They find evidence that Taylor rule models outperform a random walk model in forecasts of several dollar-related currency pairs without addressing the endpoint issue of HP-filters. Given the above, we use the manufacturing indices without extending the data. By subtracting the observed data points from the long-term trend, we get a measure of the output gap. The differential is obtained by subtracting the output gap of the Eurozone from that of Norway.

4.2 Data Transformations

As discussed in section 3.2.1, using non-stationary time series may lead to spurious results. Hence, the concept of stationarity in time series is essential. This section presents the results of the augmented Dickey-Fuller and KPSS tests outlined in section 3.4 and discusses the choices we have made with regards to data transformations and model specifications.

4.2.1 Stationarity Tests

Recall that for the augmented Dickey-Fuller test, the null hypothesis of a unit root presence is rejected in favor of the alternative hypothesis that the data is stationary at a critical value of $t_{\hat{\theta}} < -2{,}86$. The number of lags included in the test follows Wooldridge's (2018) suggestion that one might include twelve lags on monthly data. For the KPSS test, the null hypothesis of stationarity is rejected in favor of the alternative hypothesis that a unit root is present at the critical value of $LM > 0{,}463$. In other words, to conclude that a time series is stationary, we wish to see a p-value below 0,05 for the augmented Dickey-Fuller test and above 0,05 for the

KPSS test. The critical values of the tests are summarized in the table below, with p-values presented in parenthesis beneath the critical value. As a reminder, $(i - i^*)$ is the interest rate differential, $(p - p^*)$ the price differential, $(\pi - \pi^*)$ the inflation differential, $(m - m^*)$ the money supply differential, $(y - y^*)$ the proxy real income differential, and $(y_{gap} - y^*_{gap})$ the output gap differential. 'Data' denotes absolute level data, while 'FD' denotes first-differenced data. Grey cells indicate results that are statistically significant at the 5% level.

Stationarity Tests

	ADF		KPSS	
Variable	Data	FD	Data	FD
$eurnok$	-1.74 (-0.68)	-4.47 (0.01)	1.06 (0.01)	0.21 (0.10)
$eurnok$ $lag = 3$		-4.48 (0.01)		0.28 (0.10)
$eurnok$ $lag = 6$		-4.71 (0.01)		0.27 (0.10)
$eurnok$ $lag = 12$		-3.97 (0.01)		0.34 (0.10)
oil	-1.70 (0.70)	-4.98 (0.01)	0.45 (0.06)	0.16 (0.10)
vix	-2.14 (0.52)	-6.00 (0.01)	0.16 (0.10)	0.06 (0.10)
$(i - i^*)$	-2.17 (0.50)	-4.11 (0.01)	0.12 (0.10)	0.08 (0.10)
$(p - p^*)$	-1.13 (0.92)	-3.84 (0.02)	1.56 (0.01)	0.06 (0.10)
$(\pi - \pi^*)$	-2.17 (0.50)	-5.01 (0.01)	1.56 (0.01)	0.06 (0.10)
$(m - m^*)$	-2.96 (0.17)	-3.40 (0.06)	1.50 (0.01)	0.09 (0.10)
$(y - y^*)$	-2.43 (0.39)	-3.57 (0.04)	0.63 (0.02)	0.06 (0.10)
$(y_{gap} - y^*_{gap})$	-3.85 (0.02)	-4.71 (0.01)	0.03 (0.10)	0.06 (0.10)

Table 1: Summary of the augmented Dickey-Fuller and KPSS tests. A p-value below 0,05 for the ADF test and above 0,05 for the KPSS test indicates that the time-series are stationary.

Although the KPSS test cannot reject the null hypothesis of stationarity in the raw data for the $oil, vix,$ and $(i - i^*)$ variables, we cannot reject the null hypothesis of unit root presence in the same variables using the augmented Dickey-Fuller test. These results indicate the time-series are trend stationary but not strict stationary. A simple way of removing a trend is to first-difference the data. Upon doing this, the results of both the augmented Dickey-Fuller tests and KPSS tests indicate that the mentioned variables are stationary. Thus, the first-difference transformations allow us to avoid problems with spurious results in OLS regression.

Meanwhile, the $eurnok,$ $(\pi - \pi^*), (m - m^*), (y - y^*),$ and $(p - p^*)$ variables tests negatively for stationarity in both the augmented Dickey-Fuller test and the KPSS tests. All the mentioned variables are thus first-differenced and re-tested. Since we forecast the EUR/NOK exchange rate for multiple horizons, the $eurnok$ variable is tested across all corresponding lag lengths. Upon first-differencing, all the above variables reject the augmented Dickey-Fuller test's null hypothesis of a unit root presence and fail to reject the KPSS test's null hypothesis of stationary. We thus conclude that these variables are I(1) and exhibit weak dependence upon being first-differenced. However, note that the $(m - m^*)$ variable is only significant at the 10% level. Contrary to the other variables, $(y_{gap} - y^*_{gap})$ is also stationary in its original form.

4.2.2 Model Specifications

Most empirical literature concerning in-sample prediction of exchange rates includes an error-correction term. A prerequisite for the use of error-correction models is that cointegration is present. As explained in section 3.2.1, cointegration is present when a linear combination of non-stationary variables is stationary. This allows for models to be estimated in levels, allowing inference of long-run relationships. Rossi (2013) summarizes the prominent literature within the out-of-sample forecasting literature pre-2013. Here, the choice of model specification is more mixed between error-correction and first-difference specifications. Moreover, some simply assume stationarity based on the underlying theory of exchange rate models. The common trait is that no model specification significantly outperforms the others.

In light of these findings and to achieve consistency across all models, we solely rely on first-difference specifications in this thesis.

5. Empirical Analysis

This section presents the results of the empirical analysis. We start by summarizing the selected models before moving onto the point forecast evaluation. Here, the root-mean-square error calculations are presented, along with the Diebold-Mariano and Pesaran-Timmermann test statistics. Following that, we elaborate on the results of the density forecasting evaluation.

5.1 Summary of Model Specifications

Building on the discussion in the theory-, methodology- and data section, we specify twelve distinct models subject to the empirical analysis. These models are summarized in the table below. The dependent variable for all models can be denoted Δs_{t+h}, where s is the log of the $eurnok$ exchange rate and $h = 1, 3, 6, 12$ months.

Models of the Exchange Rate

Model	Explanatory variables
Driftless random walk model	
Taylor rule model (symmetric w/o smoothing)	$\Delta(\pi - \pi^*)$; $\Delta(y_{gap} - y^*_{gap})$
Taylor rule model (asymmetric w/o smoothing)	$\Delta(\pi - \pi^*)$; $\Delta(y_{gap} - y^*_{gap})$; Δq
Taylor rule model (symmetric w/ smoothing)	$\Delta(\pi - \pi^*)$; $\Delta(y_{gap} - y^*_{gap})$; $(i_{t-1} - i^*_{t-1})$
Taylor rule model (asymmetric w smoothing)	$\Delta(\pi - \pi^*)$; $\Delta(y_{gap} - y^*_{gap})$; Δq; $(i_{t-1} - i^*_{t-1})$
Relative purchasing power parity	$\Delta(p - p^*)$
Uncovered interest rate parity	$\Delta(i - i^*)$
BEER model	$\Delta(i - i^*)$; $\Delta brentp$; Δvix
Monetary model (w/o interest rates, w/o inflation)	$\Delta(m - m^*)$; $\Delta(y - y^*)$
Monetary model (w/ interest rate, w/o inflation)	$\Delta(m - m^*)$; $\Delta(y - y^*)$; $\Delta(i - i^*)$
Monetary model (w/o interest rates, w/ inflation)	$\Delta(m - m^*)$; $\Delta(y - y^*)$; $\Delta(\pi - \pi^*)$
Monetary model (w/ interest rates, w/ inflation)	$\Delta(m - m^*)$; $\Delta(y - y^*)$; $\Delta(i - i^*)$; $\Delta(\pi - \pi^*)$

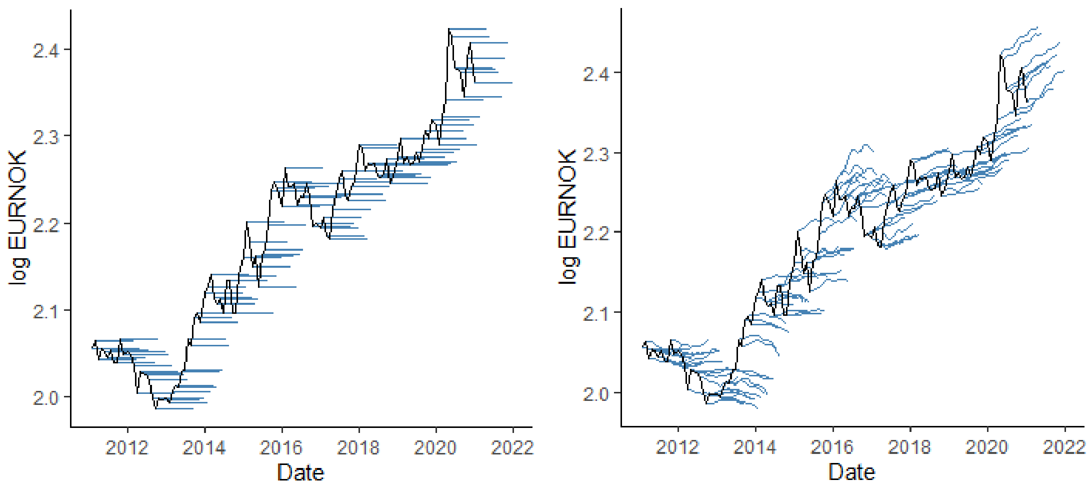
Table 2: Summary of model specifications.

The first model is the driftless random walk model, the benchmark of the empirical analysis. In levels, the predicted exchange rate $\hat{s}_{t+h} = s_t$. By subtracting s_t from both sides of the equation, we arrive at the first-difference specification which is simply $\Delta\hat{s}_{t+h} = \hat{s}_{t+h} - s_t = 0$. Thus, the random walk will always forecast an unchanged exchange rate. Next, we consider four different models with Taylor-rule fundamentals. We test the model with asymmetric Taylor rule fundamentals and smoothing presented in section 2.5 and also include variations with and without the real exchange rate and the smoothing factor. Furthermore, we specify models for relative purchasing power parity and uncovered interest rate parity, respectively. Referring back to section 2.6, we also construct a BEER model. Building on the discussion of earlier empirical evidence of factors affecting the Norwegian krone exchange rate, we include the oil price and VIX as variables in addition to the interest rate differential. Papers such as Akram (2019) and Martinsen (2017) include more explanatory variables in their in-sample estimations. However, in addition to the overfitting issue described in section 3.1.1, Rossi (2013) finds that out-of-sample predictability is most apparent when only a small number of predictors are considered. As such, we only include the three aforementioned variables. Finally, we consider the monetary model derived in section 2.4.1, and three additional variations with and without interest rate and inflation differentials.

5.2 Point Forecast Evaluation

The model evaluation begins by assessing the standard point forecasts estimated by the above models. The point forecast is obtained from the mean response of the OLS-estimated parameters across all models, time periods, and forecast horizons. The figure below presents example outputs of the point forecasts from the PPP model and the random walk model.

a) Random Walk   b) PPP



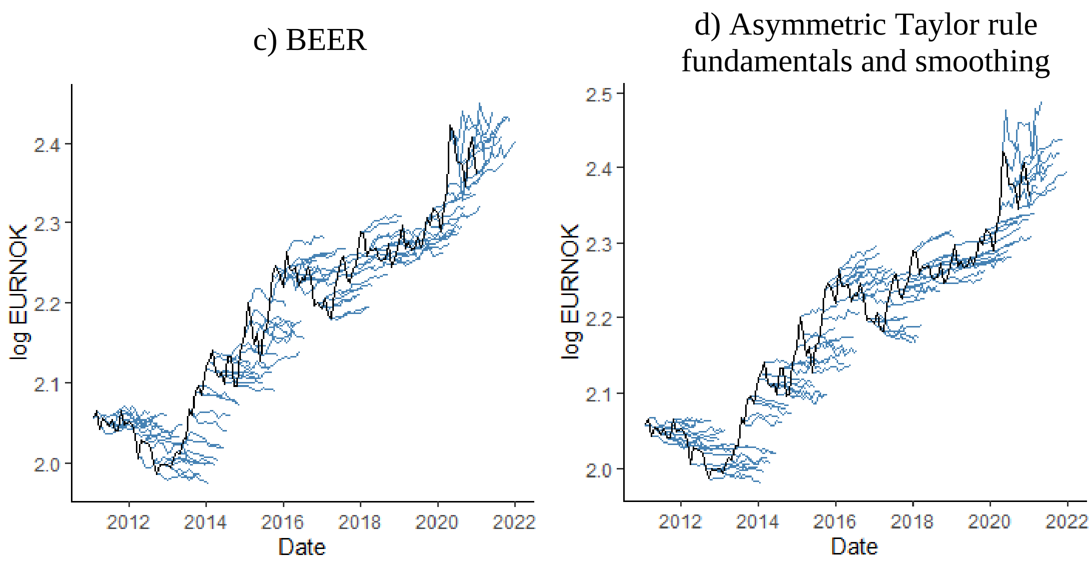


Figure 2: Point forecasts of selected exchange rate models. The graphs are constructed by adding the forecasted exchange returns to the level of the EUR/NOK at the time period the forecast is made.

5.2.1 Diebold-Mariano Test

To assess whether the models can beat the random walk model, we return to the Diebold-Mariano test statistic. Recall that the Diebold-Mariano test is based on a loss differential that must first be computed. Here, we utilize the root mean square error (RMSE). The resulting computations are presented in the table below for the forecast horizons of one, three, six, and twelve months. Grey cells denote models that achieve a lower RMSE than the benchmark.

Root Mean Square Error

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RW	0.0214	0.0338	0.0427	0.0641
Taylor rule *symmetric* *no smoothing*	0.0219	0.0341	0.0439	0.0651
Taylor rule *asymmetric* *no smoothing*	0.0217	0.034	0.0437	0.0655
Taylor rule *symmetric* *smoothing*	0.022	0.0343	0.044	0.0655
Taylor rule *asymmetric* *smoothing*	0.0218	0.0341	0.0439	0.0664

PPP	0.0215	0.0338	0.0423	0.0634
UIRP	0.0213	0.0346	0.0423	0.0632
BEER	0.0206	0.035	0.0432	0.0638
Monetary *no interest rates no inflation*	0.0219	0.0338	0.0412	0.0625
Monetary *interest rates no inflation*	0.0215	0.0349	0.0417	0.063
Monetary *no interest rates inflation*	0.0219	0.0343	0.0428	0.0648
Monetary *interest rates inflation*	0.0215	0.0355	0.0434	0.0654

Table 3: Summary of RMSE results. A lower RMSE indicates a better fit.

In total, five models generate forecasts that achieve a lower value than the random walk model, and the results vary notably across different forecast horizons. Moreover, the deviations from the benchmark are generally minor, which raises the question of whether the results are statistically significant. Thus, we calculate the Diebold-Mariano test statistic, the results of which are presented in the table below. Recall that the null hypothesis that the forecast errors and benchmarks errors are the same is rejected in favor of the alternative hypothesis that the forecast model has a different level of accuracy at a critical value of $z = \pm 1{,}96$, for the 5% significance level. The p-values are presented in parenthesis beneath the z-values.

Diebold-Mariano Test

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Taylor rule *symmetric no smoothing*	-0.52 (0.60)	-0.69 (0.49)	-1.12 (0.26)	-0.70 (0.49)
Taylor rule *asymmetric no smoothing*	-0.34 (0.74)	-0.24 (0.81)	-0.98 (0.33)	-0.94 (0.35)
Taylor rule *symmetric smoothing*	-0.62 (0.53)	-0.96 (0.34)	-1.35 (0.18)	-1.01 (0.32)
Taylor rule *asymmetric smoothing*	-0.52 (0.60)	-0.41 (0.68)	-1.17 (0.24)	-1.42 (0.16)

PPP	-0.28	0.21	0.39	0.64
	(0.78)	(0.84)	(0.70)	(0.52)
UIRP	0.41	-1.23	0.82	0.92
	(0.68)	(0.22)	(0.42)	(0.36)
BEER	1.70	-0.96	-0.43	0.25
	(0.09)	(0.34)	(0.67)	(0.80)
Monetary *no interest rates*	-0.57	0.15	1.39	0.77
no inflation	(0.57)	(0.88)	(0.17)	(0.44)
Monetary *interest rates*	-0.16	-1.43	1.01	0.52
no inflation	(0.87)	(0.15)	(0.32)	(0.60)
Monetary *no interest rates*	-0.62	-0.88	-0.07	-0.30
inflation	(0.54)	(0.38)	(0.95)	(0.76)
Monetary *interest rates*	-0.16	-1.94	-0.62	-0.61
inflation	(0.88)	(0.05)	(0.53)	(0.54)

Table 4: Summary of Diebold-Mariano test statistics. At the 5% significance level, the null hypothesis that a model has the same forecast accuracy as the naïve random walk model is rejected at a critical value of $z = \pm 1{,}96$.

As shown in the table, no model specification is significant at the 5% level. Even at the 10% level, only one specification is significant. The results are detailed in the following.

The Taylor rule models generally perform quite poorly. No model achieves a lower RMSE than the driftless random walk model at any forecast horizon. The discrepancy is especially large at the six- and twelve-month horizons. Furthermore, the Diebold-Mariano test statistics show that we cannot conclude that any model incorporating Taylor rule fundamentals has a different level of forecast accuracy than the benchmark model. The result contrasts Molodtsova and Papell (2009), who find evidence of short-term exchange rate predictability using Taylor rule fundamentals, and more so than for other structural specifications. However, as noted in section 2.5.3, the finding has been criticized by Rogoff & Stavrakeva (2008).

Table 3 shows that the relative purchasing power parity and uncovered interest rate parity models perform slightly better than the models with Taylor rule fundamentals. The PPP model achieves a lower RMSE than the benchmark at the six- and twelve-month horizons. The same is true for the UIRP model, which additionally achieves a lower RMSE at the one-month horizon. However, the Diebold-Mariano test statistics show that no result is statistically significant. Thus, we are again unable to conclude that the models have different levels of forecast accuracy than the random walk model at any forecast horizon. The results align with the large body of existing literature that does not find support for relative purchasing power parity and uncovered interest rate parity at forecast horizons of twelve months and shorter.

The BEER model achieves a lower RMSE than the benchmark model at the one- and twelve-month forecast horizons. From the Diebold-Mariano test statistics, we cannot conclude that the model has a different level of forecast accuracy than the random walk model at horizons longer than three months. However, at the one-month horizon, the model achieves a test result that is significant at the 10% level. The result suggests that a model incorporating the oil price, volatility in financial markets, and an interest rate differential has out-of-sample predictability over the EUR/NOK exchange at a one-month horizon. However, the evidence is weak, as the result is not statistically significant at the 5% level. Unlike Akram (2004), we do not find clear evidence that a BEER model incorporating the oil price outperforms a random walk model at the twelve-month horizon. Differences in the forecasting method may explain this.

The individual monetary models vary significantly in their predictive performance relative to the benchmark. Without an interest rate and inflation differential, the monetary model achieves a lower RMSE than the driftless random walk model at the six- and twelve-month horizons. The corresponding model with the interest rate differential also achieves a lower RMSE at the six-month horizon. Other specifications of the monetary model are unable to achieve a lower RMSE. Regardless, we cannot conclude that *any* monetary model has better forecast accuracy than the driftless random walk model from the Diebold-Mariano test statistics. The results are in line with Meese and Rogoff (1983), who do not find evidence of short-term predictability in monetary models at horizons of up to one year.

In summary, the results are mainly in line with the extensive body of literature, beginning with Meese and Rogoff (1983), which concludes that it is practically difficult to outperform a driftless random walk model in out-of-sample forecasting. Several of the classic exchange rate models achieve a lower RMSE than the benchmark at various forecast horizons but none of the results are statistically significant at the 5% level. We do, however, find weak evidence supporting the theory that a BEER model that incorporates commodity price and currency volatility dynamics performs better than a driftless random walk model at a one-month forecast horizon. The finding lends support to the body of literature that concludes that there is a link between the oil price, uncertainty, and the Norwegian krone exchange rate (see, e.g., Akram, 2019).

5.2.2 Pesaran-Timmermann Test

Next, we turn to the question of whether the models accurately forecast the direction of change for the EUR/NOK exchange rate. For this, we use the Pesaran-Timmermann test, introduced in section 3.5.3. Table 5 below presents the directional accuracy for each model specification, i.e., the percentage of exchange rate changes forecasted in the same direction as the observed change. Furthermore, the p-value is presented in parenthesis below the directional forecast accuracy value. Recall that the null hypothesis of the Pesaran-Timmermann test is that the independent variables do not forecast the sign of the dependent variable, while the alternative hypothesis is that the independent variables *do* forecast the sign of the dependent variable. Grey-colored cells indicate results that are statistically significant at the 5% level.

Pesaran-Timmermann Test

Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Taylor rule *symmetric no smoothing*	0.54 (0.19)	0.54 (0.29)	0.56 (0.31)	0.63 (0.17)
Taylor rule *asymmetric no smoothing*	0.53 (0.22)	0.55 (0.19)	0.54 (0.43)	0.62 (0.15)
Taylor rule *symmetric smoothing*	0.49 (0.60)	0.51 (0.51)	0.59 (0.11)	0.61 (0.24)
Taylor rule *asymmetric smoothing*	0.51 (0.42)	0.50 (0.52)	0.55 (0.24)	0.57 (0.34)
PPP	0.48 (0.71)	0.56 (0.16)	0.64 (0.02)	0.63 (0.30)
UIRP	0.54 (0.20)	0.51 (0.51)	0.56 (0.31)	0.56 (0.85)
BEER	0.56 (0.10)	0.49 (0.67)	0.58 (0.16)	0.57 (0.59)
Monetary *no interest rates no inflation*	0.48 (0.73)	0.53 (0.41)	0.59 (0.11)	0.61 (0.24)

Monetary	0.52	0.53	0.59	0.61
interest rates *no inflation*	(0.38)	(0.35)	(0.10)	(0.24)
Monetary	0.48	0.51	0.59	0.61
no interest rates *inflation*	(0.74)	(0.51)	(0.11)	(0.24)
Monetary	0.52	0.51	0.56	0.61
interest rates *inflation*	(0.36)	(0.47)	(0.25)	(0.13)

Table 5: Summary of directional forecast accuracy. At the 5% significance level, the null hypothesis that the independent variables do not forecast the sign of the dependent variable is rejected at a p-value below 0,05.

As illustrated by the above table, all models achieve a directional forecast accuracy above 50% on at least one horizon. However, only one result is statistically significant at the 5% level.

The relative purchasing power parity model achieves the highest directional forecast accuracy at the three-, six- and twelve-month horizons. However, it is only at the six-month horizon that we can reject the null hypothesis that the independent variable does not forecast the sign of the dependent variable. The result indicates that a price differential has predictive value for forecasting the direction of change in the EUR/NOK exchange rate at a six-month horizon.

The monetary model specified with an interest rate differential and without an inflation differential also achieves a statistically significant result at the six-month horizon. However, we can only reject the null hypothesis that the independent variable does not forecast the sign of the dependent variable at the 10% significance level. Additionally, we fail to reject the null hypothesis at the same significance level for other monetary model specifications. Thus, we only find weak evidence supporting a model incorporating monetary fundamentals being able to forecast the direction of change in the EUR/NOK exchange rate at a six-month horizon.

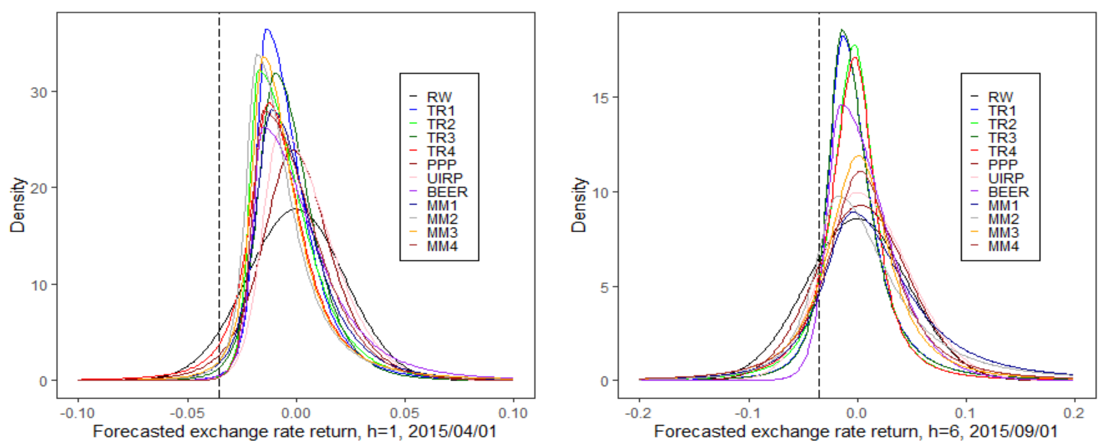
No other models achieve a statistically significant test result at any forecast horizon. The directional forecast accuracy varies between 48 and 56% percent at the one-month horizon. Here, the BEER model is closest to being statistically significant at the 10% level with a p-value just above 0,10. The results are not materially different at the three-month horizon, where the directional forecast accuracy fluctuates between 49 and 56%. However, at the six-month horizon, the accuracy generally improves. The Taylor rule model specified with symmetric fundamentals and interest rate smoothing achieves a directional forecast accuracy of 59%. With a p-value of 0,11, the result narrowly falls short of being statistically significant at the 10% level. Generally, the directional forecast accuracy improves further at the twelve-month horizon. However, none of the models achieve a statistically significant test result.

How do the results compare against other empirical evidence? Cheung et al. (2005) run 216 direction of change statistics on five dollar-related currency pairs, using a set of economics- and productivity-based models. They find that 50 of the statistics achieve a significantly better directional forecast accuracy than 50% at the 10% level. Furthermore, they note that a model's ability to predict the direction of change correctly seems currency-specific and that the models are generally more successful at longer horizons. Compared to Cheung et al., we achieve a notably smaller share of statistics with a significantly better directional forecast accuracy than 50% at the 10% level. On the other hand, similar to Cheung et al., we observe that the results to some extent achieve greater directional accuracy at longer horizons. Additionally, we find that a purchasing power parity model achieves the best results out of all the models tested and that a monetary model performs better than an interest rate parity model. These results are consistent with Cheung et al. However, note that Cheung et al. does not test Taylor rule models in their paper.

5.3 Density Forecast Evaluation

We introduce the density forecast evaluation section by illustrating graphically some of the forecasted densities. Next, we turn to the Probability Integral Transforms and Anderson-Darling Test, followed by the Berkowitz test and logarithmic scores. Finally, we present the tail risk measures and the corresponding conditional calibration tests.

Skewed t-Distribution



Normal Distribution

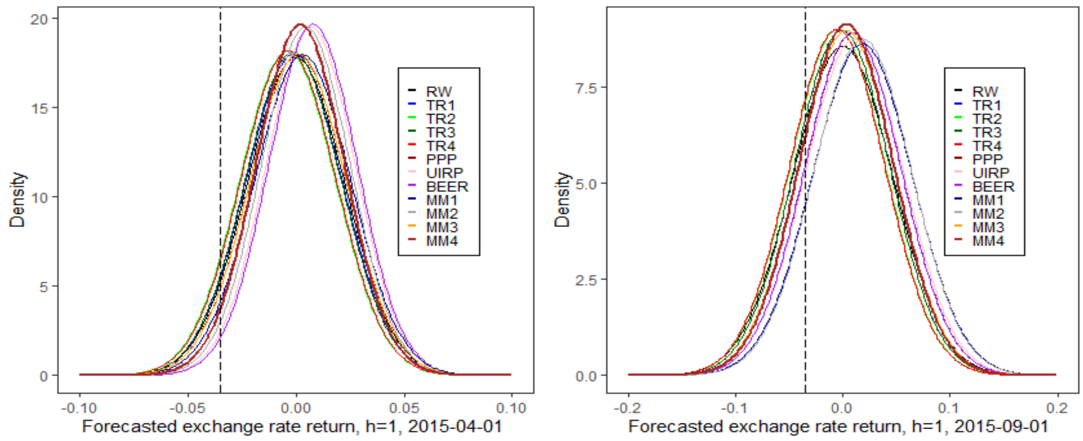


Figure 3: Visual representation of forecasted densities.

Figure 3 illustrates the estimated densities for some arbitrary selected months for $h = 1$ and $h = 6$. The dashed lined represent the observed value. From the skewed t-distribution derived from quantile regression, we can observe that most models are right-skewed, while some models are more symmetric. The peak of the skewed t-distributions corresponds to the outcome with the highest probability mass. If the skewness parameter α of the forecasted density is greater (smaller) than zero, this indicates more uncertainty associated with the right (left) side of the distribution relative to the peak. In the case where the peak corresponds to a forecasted exchange rate return of zero, a positive (negative) skew corresponds to a higher probability of depreciation (appreciation) of the exchange rate. Furthermore, a common trait of the skewed t-distributions is that they are leptokurtic, i.e., their peaks are higher and tails "fatter"[10] compared to the normally distributed densities. This is in line with the description of de Vries, Leuven (1994) of daily exchange rate returns. From the width of the distributions, we can also observe that for $h = 6$, the uncertainty concerning the future value of the exchange rate return is higher compared to $h = 1$. This corresponds to a larger scale parameter $\hat{\sigma}$. We are able to visually inspect the densities across time by the construction of a three-dimensional plot.

[10] "Fatter" tails indicate a higher probability of extremal outcomes.

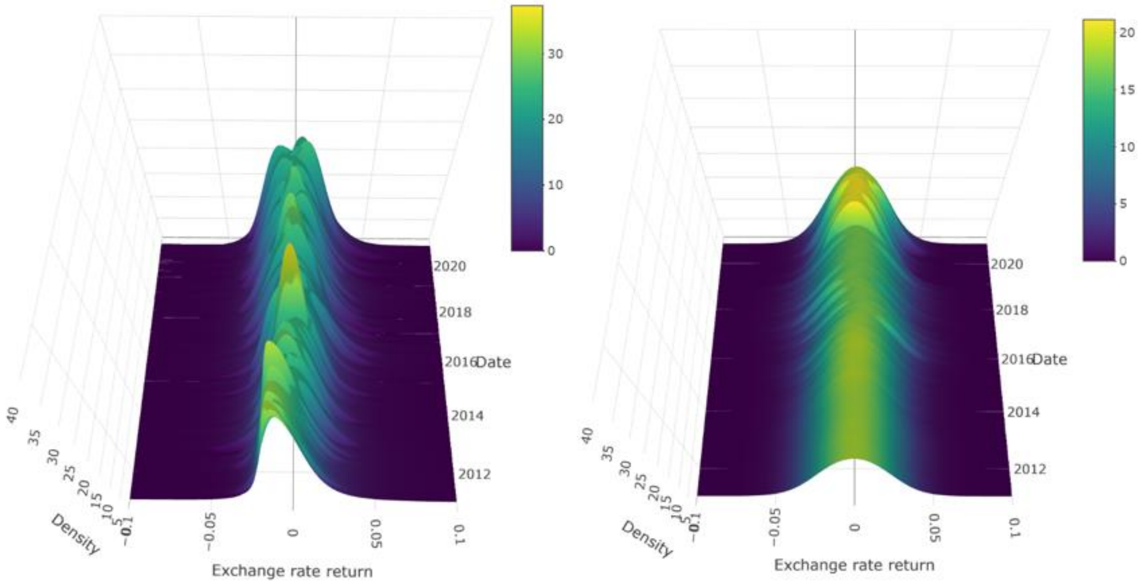


Figure 4: 3D plot of forecasted densities for the UIRP model at the one-month horizon.

The 3D plot shows the forecasted densities from the UIRP model for $h = 1$. We can observe that the aforementioned properties of the densities in general are very similar across the whole out-of-sample period. The skewed t-distribution often exhibits a right-skew, meaning that there is more uncertainty concerning outcomes associated with a weaker Norwegian krone relative to what is forecasted by the peak of the distribution. The peak is also in general higher for the skewed t-distribution compared the normal distribution. The normal distribution shows small variations across time, with exception of the higher peak starting from the middle of 2018. This is a result of smaller in-sample standard deviations of the errors, as the volatile period during the global financial crisis no is longer a part of the in-sample estimation window.

5.3.1 Probability Integral Transforms and Anderson-Darling Test

Figure 5 presents histograms of the probability integral transforms (PIT) of the monetary model without interest rate and inflation differentials for $h = 1$ and $h = 6$. The red horizontal lines specify the confidence intervals for the Anderson-Darling (AD) test. If the top of the histogram bins falls within the confidence interval, the AD-test suggests that we cannot reject the null hypothesis that the PIT-series follow a uniform distribution.

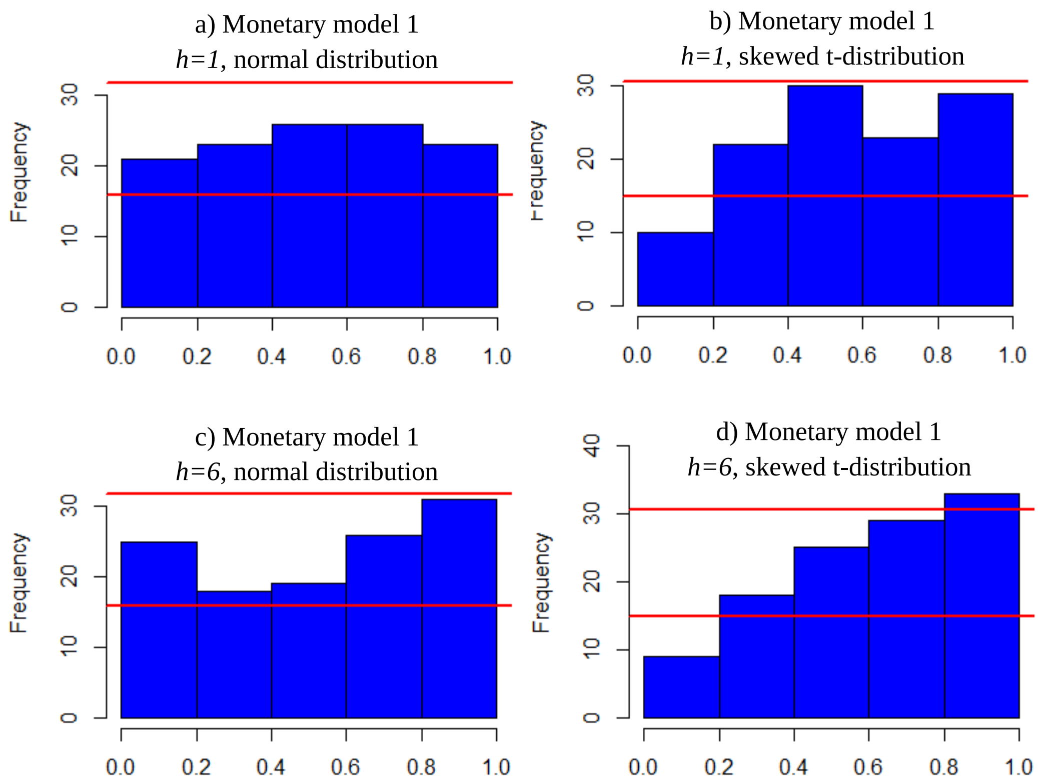


Figure 5: Histogram of the PITs of the monetary model without interest rates and inflation for the one- and six-month horizon. A perfectly calibrated histogram of a PIT-series takes the shape of a rectangle.

For $h = 1$, the height of all the histogram bins falls within the confidence interval for both the skewed t-distribution and the normal distribution, meaning that we cannot reject the AD test null hypothesis of uniformity. However, recall that we cannot conclude that the forecasted distributions from these models are well calibrated for the one-month horizon before the $i.i.d.$ assumption is tested. For $h = 6$, there is a lacking number of observations that fall within the lower part of the distribution for both models. The AD test thus rejects the null hypothesis of uniformity. This indicates that the forecasted distributions from this model is not well calibrated for the six-month horizon. The following table shows the AD test results. The values presented are the associated p-values. The null hypothesis that the histogram of a sequence of PITs follow a uniform distribution is rejected if $p \leq 0{,}05$. Grey cells indicate results that are statistically significant.

Anderson-Darling Test

	Normal Distribution				Skewed t-Distribution			
Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RW	0.38	0.02	0.00	0.00				
Taylor rule *symmetric no smoothing*	0.51	0.06	0.00	0.00	0.09	0.02	0.00	0.00
Taylor rule *asymmetric no smoothing*	0.40	0.05	0.00	0.00	0.13	0.01	0.00	0.00
Taylor rule *symmetric smoothing*	0.46	0.07	0.00	0.00	0.10	0.01	0.00	0.00
Taylor rule *asymmetric smoothing*	0.34	0.04	0.00	0.00	0.10	0.01	0.00	0.00
PPP	0.72	0.12	0.01	0.00	0.29	0.05	0.00	0.00
UIRP	0.54	0.06	0.00	0.00	0.12	0.02	0.00	0.00
BEER	0.48	0.07	0.00	0.00	0.33	0.01	0.00	0.00
Monetary *no interest rates no inflation*	0.64	0.07	0.00	0.00	0.12	0.03	0.00	0.00
Monetary *interest rates no inflation*	0.43	0.06	0.00	0.00	0.24	0.02	0.00	0.00
Monetary *no interest rates inflation*	0.52	0.07	0.00	0.00	0.11	0.01	0.00	0.00
Monetary *interest rates inflation*	0.35	0.06	0.00	0.00	0.19	0.02	0.00	0.00

Table 6: Summary of Anderson-Darling test results. A p-value greater than 0,05 indicates that we cannot reject the null hypothesis of uniformity.

The Anderson-Darling (AD) test suggest that for $h = 1$, the test cannot reject the null hypothesis of uniformity for any of the models, irrespective of the density type. For $h = 3$ with the normal distribution, no model is rejected, with the exception of the asymmetric Taylor rule model with smoothing and the random walk model. For this forecast horizon, the skewed t-distribution performs worse compared to the normal distribution. The only model not rejected is the PPP model. For $h = 6$ and $h = 12$, the test rejects all of the models for both distributions, indicating that these densities are not well calibrated. Referring back to section 3.6.1, the literature commonly omits the independence assumption for $h > 1$. Thus, the AD test suggests that most of the models for $h = 3$ with a normal distribution are, in fact, well calibrated. This means that, according to the AD test, we cannot reject that the density forecasts coincide with the true sequence of densities for the predicted variable.

5.3.2 Berkowitz Test

Berkowitz (2001) argues that his test has higher power compared to uniformity tests such as the AD-test. In addition, for $h = 1$, it also tests the $i.i.d.$ property. The table below presents the results of the test. Similar to the AD test, the cells show the p-values of the test. For $h = 1$, the null hypothesis that the histogram of a given PIT sequence follows a uniform distribution and is $i.i.d.$ is rejected if $p \leq 0{,}05$. For $h > 1$, recall that we use a modified version of the test where the independence property does not need to be satisfied in order for the test to be significant.

Berkowitz Test

	Normal Distribution				Skewed t-Distribution			
Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RW	0.43	0.00	0.00	0.00				
Taylor rule *symmetric no smoothing*	0.26	0.02	0.00	0.00	0.06	0.01	0.00	0.00
Taylor rule *asymmetric no smoothing*	0.60	0.01	0.00	0.00	0.04	0.01	0.00	0.00
Taylor rule *symmetric smoothing*	0.27	0.02	0.00	0.00	0.07	0.00	0.00	0.00
Taylor rule *asymmetric smoothing*	0.51	0.01	0.00	0.00	0.11	0.00	0.00	0.00
PPP	0.66	0.07	0.00	0.00	0.23	0.04	0.00	0.00
UIRP	0.57	0.01	0.00	0.00	0.17	0.01	0.00	0.00
BEER	0.46	0.01	0.00	0.00	0.23	0.00	0.00	0.00

Monetary *no interest rates no inflation*	0.26	0.03	0.00	0.00	0.10	0.02	0.00	0.00
Monetary *interest rates no inflation*	0.25	0.01	0.00	0.00	0.15	0.02	0.00	0.00
Monetary *no interest rates inflation*	0.27	0.02	0.00	0.00	0.07	0.01	0.00	0.00
Monetary *interest rates inflation*	0.24	0.00	0.00	0.00	0.09	0.00	0.00	0.00

Table 7: Summary of Berkowitz test results. A p-value greater than 0,05 indicates that we cannot reject the null hypothesis.

The test results suggest that all of the models coupled with a normal distribution are well calibrated for $h = 1$. For $h > 1$, all models are rejected except for the PPP model at the three-month horizon. For the skewed t-distribution, only the asymmetric Taylor rule model without smoothing can be rejected by the test for $h = 1$. For $h > 1$, all models are rejected.

In summary, the PIT-based AD and Berkowitz tests suggest that for $h = 1$, the density forecasts are generally well calibrated for both the normal and skewed t-distributions, as both the uniformity and $i.i.d.$ assumptions are satisfied. Thus, we cannot reject that these density forecasts coincide with the true sequence of predicted densities. This implicates that all models provide a good description of the uncertainty surrounding the one-month exchange rate return. For the normal distribution at the three-month horizon, the results are more ambiguous. The PPP model is the only model not rejected in either test. Most other models are rejected by the Berkowitz test, but not the AD-test, meaning we only find weak evidence that models are well calibrated at this horizon. For the skewed t-distribution, all models are rejected in both tests, except for the PPP model in the AD test. The above results may indicate that a PPP-model provides a more accurate description of the uncertainty associated with the future outcome of EUR/NOK exchange rate return at this horizon compared to the other models. For $h > 3$, both the tests suggest that all models perform poorly in terms of accurately describing this uncertainty.

5.3.3 Logarithmic Scores

This section presents the logarithmic score for the predictive densities. As the AD test and the Berkowitz test are 'absolute' tests, we cannot conclude whether or not one model performs better than the other if not rejected. The logarithmic score allows us to evaluate the relative performance of the models. The scores are presented in the table below. Grey cells represent logarithmic scores that are higher than that of the random walk model.

Logarithmic Scores

	Normal Distribution				Skewed t-Distribution			
Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RW	2.41	1.95	1.71	1.30				
Taylor rule *symmetric* *no smoothing*	2.38	1.92	1.67	1.28	2.32	1.87	1.57	1.24
Taylor rule *asymmetric* *no smoothing*	2.39	1.93	1.68	1.27	2.27	1.88	1.59	1.13
Taylor rule *symmetric* *smoothing*	2.37	1.92	1.67	1.27	2.32	1.87	1.56	1.22
Taylor rule *asymmetric* *smoothing*	2.38	1.92	1.67	1.25	2.29	1.86	1.60	1.10
PPP	2.40	1.93	1.71	1.30	2.38	1.99	1.67	1.29
UIRP	2.41	1.90	1.70	1.30	2.38	1.96	1.65	1.25
BEER	2.44	1.89	1.68	1.29	2.41	1.92	1.62	1.27
Monetary *no interest rates* *no inflation*	2.38	1.93	1.73	1.29	2.31	1.95	1.65	1.27
Monetary *interest rates* *no inflation*	2.40	1.90	1.72	1.28	2.37	1.92	1.68	1.27
Monetary *no interest rates* *inflation*	2.38	1.92	1.70	1.26	2.29	1.91	1.62	1.23
Monetary *interest rates* *inflation*	2.40	1.87	1.68	1.24	2.33	1.86	1.62	1.19

Table 8: Summary of logarithmic scores. A higher score implicates a better model.

At first glance, we can observe that the logarithmic scores between the models show small variations. Among the Taylor rule models, the asymmetric Taylor rule model without smoothing with a normal distribution performs the best for the one-, three- and six-month forecast horizons. The other Taylor rule models generally show very similar results across all forecast horizons. However, none of these models achieve log scores higher than the random walk. We can clearly observe that the skewed t-distributions achieve lower logarithmic scores compared to the normal distributions.

Among the PPP, UIRP and BEER model with a normal distribution, the latter performs the best at a one-month forecasting horizon and is the only model which achieves a higher score compared to the random walk. For the three- and six-month horizons, the PPP model performs best of the three models, while the results are more similar for the twelve-month forecast horizon. Interestingly, for the three-month forecasting horizon, several of the models achieve a higher score with the skewed t-distribution.

Concerning the monetary models, the most noteworthy are the two first specifications, which both show a higher score at the six-month forecast horizon compared to the random walk. Furthermore, at the three-month horizon, these specifications achieve a higher score with the skewed t-distribution than the normal distribution. However, for every other horizon and model specification, the normal distribution performs the best.

In general, the above results argue in favor of the normal distribution providing a more accurate description of the predictive densities for the monthly EUR/NOK exchange rate returns. However, the results are not completely unambiguous, as some models achieve a higher score with the skewed t-distribution. Referring back to the literature on the distribution of exchange rate returns, Coppes (1995) argues that monthly exchange rate returns are more normally distributed, while that of daily exchange rate returns are more leptokurtic. Thus, our results largely support Coppes' findings on the distribution of monthly exchange rate returns.

To evaluate the significance of the logarithmic scores, we return to the test statistics outlined in section 3.6.4. The table below presents the resulting test statistics and the associated p-values. The test statistics represents the average log-score difference between the respective fundamental models and the random walk. A positive value (grey cells) means that the given fundamental model achieves a higher log-score, while the p-values indicate whether the average log-score difference is significantly different from zero.

Test Statistics for Logarithmic Scores

	Normal Distribution				Skewed t-Distribution			
Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
Taylor rule *symmetric no smoothing*	-0.04 (0.14)	-0.03 (0.01)	-0.03 (0.28)	-0.03 (0.94)	-0.09 (0.14)	-0.09 (0.14)	-0.14 (0.09)	-0.06 (0.36)
Taylor rule *asymmetric no smoothing*	-0.03 (0.21)	-0.03 (0.03)	-0.03 (0.33)	-0.03 (0.85)	-0.14 (0.11)	-0.07 (0.32)	-0.11 (0.17)	-0.17 (0.07)
Taylor rule *symmetric smoothing*	-0.04 (0.18)	-0.04 (0.01)	-0.04 (0.15)	-0.03 (0.76)	-0.09 (0.06)	-0.08 (0.21)	-0.15 (0.06)	-0.09 (0.13)
Taylor rule *asymmetric smoothing*	-0.03 (0.17)	-0.03 (0.03)	-0.03 (0.22)	-0.05 (0.58)	-0.12 (0.01)	-0.09 (0.25)	-0.11 (0.13)	-0.20 (0.04)
PPP	-0.02 (0.22)	-0.02 (0.16)	0.00 (0.99)	0.00 (0.99)	-0.03 (0.28)	0.04 (0.52)	-0.04 (0.61)	-0.02 (0.84)
UIRP	-0.01 (0.73)	-0.05 (0.04)	-0.01 (0.97)	-0.01 (0.94)	-0.03 (0.42)	0.01 (0.91)	-0.06 (0.42)	-0.05 (0.54)
BEER	0.03 (0.17)	-0.06 (0.14)	-0.02 (0.43)	-0.02 (0.73)	0.00 (0.98)	-0.03 (0.74)	-0.08 (0.23)	-0.03 (0.70)
Monetary *no interest rates no inflation*	-0.03 (0.06)	-0.02 (0.09)	0.02 (0.62)	-0.01 (0.87)	-0.11 (0.04)	0.00 (0.98)	-0.06 (0.45)	-0.04 (0.74)
Monetary *interest rates no inflation*	-0.02 (0.30)	-0.06 (0.07)	0.01 (0.81)	-0.02 (0.72)	-0.05 (0.25)	-0.03 (0.52)	-0.03 (0.68)	-0.04 (0.76)
Monetary *no interest rates inflation*	-0.04 (0.10)	-0.03 (0.01)	-0.01 (0.89)	-0.04 (0.74)	-0.13 (0.03)	-0.04 (0.52)	-0.09 (0.28)	-0.07 (0.33)
Monetary *interest rates inflation*	-0.02 (0.36)	-0.07 (0.03)	-0.02 (0.54)	-0.06 (0.58)	-0.08 (0.09)	-0.09 (0.09)	-0.08 (0.21)	-0.11 (0.20)

Table 9: Summary of test statistics for logarithmic scores. A p-value below 0,05 indicates that the average log-score difference between a given model and the random walk is significantly different from zero.

Our findings suggest that no model significantly outperform the random walk model. The BEER model achieves the lowest p-value of the models with a positive average log-score difference. However, the p-value of 0,17 is not a statistically significant at conventional significance levels. The results are in line with Gaglianone & Marins (2016), who find that none of the tested models significantly outperform the random walk in density forecasting of the BRL/USD exchange rate. As such, these findings suggest that the Meese & Rogoff puzzle extends to sphere of density forecasting.

5.3.4 Expected Shortfall Conditional Calibration Test

Although the evaluation of the full densities reveal that no models can beat the random walk, we consider the possibility that they are better at forecasting the risk associated with the tails of the distributions. To do so, we first consider the left tail of the distributions by estimating the 5% Expected Shortfall (ES) and Value-at-Risk, and apply the joint conditional calibration test (Nolde & Ziegel, 2017) to evaluate whether they are well specified. Note that this corresponds to appreciation tail risk. The table below presents the output of the test. A p-value greater than 0,05 (grey cells) indicates that the models forecast of the left tail of the distributions are well specified.

Expected Shortfall Conditional Calibration Test

	Normal Distribution				Skewed t-Distribution			
Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RW	0.00	0.00	0.00	0.00				
Taylor rule *symmetric* *no smoothing*	0.56	0.00	0.00	0.00	0.00	0.05	0.06	0.00
Taylor rule *asymmetric* *no smoothing*	0.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Taylor rule *symmetric* *smoothing*	0.55	0.00	0.00	0.00	0.00	0.01	0.03	0.00
Taylor rule *asymmetric* *smoothing*	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PPP	0.08	0.00	0.00	0.00	0.01	0.00	0.00	0.00
UIRP	0.56	0.00	0.00	0.00	0.73	0.00	0.00	0.00
BEER	0.00	0.63	0.00	0.00	0.00	0.44	0.01	0.00
Monetary *no interest rates* *no inflation*	0.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Monetary *interest rates no inflation*	0.72	0.17	0.00	0.00	0.00	0.00	0.00	0.00
Monetary *no interest rates inflation*	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Monetary *interest rates inflation*	0.56	0.09	0.00	0.00	0.03	0.52	0.00	0.00

Table 10: Summary of Expected Shortfall conditional calibration test. A p-value below 0,05 indicates that the left tails are well specified.

The results show that, for $h = 1$ with a normal distribution, the only models that are rejected are the BEER model and the random walk model. Thus, most models outperform the random walk model in terms of forecasting appreciation tail risk at this horizon. For $h = 3$, the BEER model and the second and fourth monetary model, cannot be rejected by the test, while the other models are rejected. At the six- and twelve-month horizons, all models are rejected.

For the skewed t-distribution, the only model not rejected for $h = 1$ is the UIRP model. For $h = 3$, the BEER model, the monetary model with interest rates inflation, and the symmetric Taylor rule model without smoothing are not rejected by the test. The latter is also the only model not rejected for $h = 6$. For the twelve-month horizon, all models are rejected.

Overall, the test reveals that the fundamental models with a normal distribution consistently outperform the random walk in forecasting the appreciation tail risk for $h = 1$, while this is only the case for a few models for $h > 1$. With the skewed t-distribution, only a few models at selected forecast horizons are able to outperform the benchmark. Also here, the results argue in favor of a normal distribution as compared to a skewed t-distribution at a one-month horizon.

5.3.5 Expected Longrise Conditional Calibration Test

We perform the same test procedure for the right tail of the distribution. Here we estimate the Expected Longrise (EL) outlined in section 3.6.5, and the 95% Value-at-Risk. We apply the joint conditional calibration test (Nolde & Ziegel, 2017) to test whether the models are able to forecast the right tail of the distribution, which corresponds to depreciation tail risk. The results of the test are presented in the table below.

Expected Longrise Conditional Calibration Test

	Normal Distribution				Skewed t-Distribution			
Model	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RW	0.28	0.31	0.11	0.32				
Taylor rule *symmetric no smoothing*	0.27	0.22	0.10	0.14	0.04	0.08	0.03	0.17
Taylor rule *asymmetric no smoothing*	0.57	0.21	0.10	0.05	0.01	0.14	0.03	0.07
Taylor rule *symmetric smoothing*	0.41	0.22	0.10	0.08	0.03	0.07	0.04	0.13
Taylor rule *asymmetric smoothing*	0.75	0.22	0.10	0.06	0.00	0.10	0.03	0.06
PPP	0.57	0.25	0.16	0.30	0.13	0.26	0.06	0.29
UIRP	0.19	0.15	0.24	0.30	0.16	0.07	0.00	0.01
BEER	0.41	0.18	0.10	0.13	0.14	0.03	0.00	0.07
Monetary *no interest rates no inflation*	0.27	0.24	0.25	0.01	0.24	0.01	0.00	0.02
Monetary *interest rates no inflation*	0.08	0.10	0.17	0.00	0.30	0.07	0.00	0.01
Monetary *no interest rates inflation*	0.18	0.16	0.17	0.05	0.15	0.02	0.03	0.02
Monetary *interest rates inflation*	0.08	0.09	0.15	0.01	0.26	0.01	0.11	0.00

Table 11: Summary of Expected Longrise conditional calibration test. A p-value below 0,05 indicates that the right tails are well specified.

The results suggest that that, in general, most of the models do well in forecasting depreciation tail risk for all forecasting horizons with the normal distribution. Only three of the monetary models are rejected by the test. For the skewed t-distribution, the results are mixed depending on model and forecast horizon. Note that for $h = 1$ with a normal distribution, the fundamental models appear to perform well in forecasting both tails of the distribution, while the random walk model only performs well in forecasting the right tail. Thus, fundamental models appear to provide a better measure of the risk associated with extremal outcomes of the EUR/NOK exchange rate return compared to a random walk model at the one-month horizon. For $h > 1$, most of the models perform better in forecasting the depreciation tail risk as compared to the appreciation tail risk.

5.3.6 Point Versus Density Forecasts – What Have We Learned?

We devoted the first part of the empirical analysis to an evaluation of the point forecasts generated by the fundamental models. This analysis revealed that, in accordance with the empirical literature on the Meese & Rogoff puzzle, no model was able to significantly outperform the random walk model in pseudo out-of-sample forecasting. Furthermore, the models were generally unable to achieve a significantly better directional forecast accuracy than 50%. In light of these findings, we return to the question of whether density forecasts provide an informational advantage relative to standard point forecasts. The test statistics for the logarithmic scores shows that fundamental models do not have significantly better forecast performance compared to the random walk. However, the Anderson-Darling and Berkowitz tests reveal that, for the one-month forecast horizon, most of the densities provided a good description of the uncertainty associated with the forecasted exchange rate return. Furthermore, the fundamental models appear to provide a more accurate description of the risks associated with appreciation and depreciation tail risk compared to the random walk model at the same horizon. Decision makers such as investors, risk managers and central banks are often more concerned with the uncertainty associated with the exchange rate rather than the central estimate, as this has proven a particularly difficult exercise to forecast. Thus, it can be argued based on the findings in this paper, that density forecasting of exchange rates provide an informational advantage relative to point forecasting for these decisionmakers at the one-month horizon.

6. Conclusion

This thesis investigated the predictability of fundamental economic and financial indicators on the EUR/NOK exchange rate from a pseudo out-of-sample perspective. The thesis deviates from the vast body of literature investigating the predictability of the Norwegian krone exchange rate in that it utilizes both a point and density forecasting methodology. In total, forecasts have been made for eleven exchange rate models selected on the grounds of standard economic theory, at forecast horizons of one, three, six, and twelve months. The thesis sought to evaluate whether fundamental exchange rate models could outperform a naïve random walk model in out-of-sample density forecasting of the EUR/NOK exchange rate. To do so, we benchmarked the fundamental models against the naïve random walk model using a wide range of empirically grounded statistics. Furthermore, to assess whether density forecasts of the EUR/NOK exchange provide an informational advantage relative to standard point forecasts, we considered both approaches.

We first generated point forecasts for the selected models using ordinary least squares. A Diebold-Mariano test statistic was applied to investigate whether the empirical exchange models could achieve more accurate point forecasts than the naïve random walk model. No model achieved a forecast accuracy that was significantly better than the benchmark at the 5% significance level. Furthermore, a Pesaran-Timmermann test statistic was applied to test whether the empirical exchange rate models could accurately forecast the direction of change for the exchange rate. The models were generally found to be unable of achieving a significantly better directional forecast accuracy than 50%. Overall, the results are largely in line with the extensive body of Meese & Rogoff puzzle literature that concludes on the difficulty of outperforming a driftless random walk model in out-of-sample forecasting.

Density forecasts were generated using two sets of distributions, the first one being a normal distribution. To account for the possibility of exchange rate returns not being normally distributed, we also estimated a skewed t-distribution derived from quantile regression. An evaluation of the density forecasts revealed that, in terms of calibration, most of the theoretically grounded models were well calibrated at the one-month forecast horizon, while they in general showed poor performance at the remaining horizons. The logarithmic score test statistics displayed that no model could significantly outperform the random walk. However, an analysis of the tail-risk of the distributions revealed that the fundamental models were significantly better at forecasting appreciation tail risk compared to the random walk model. Most models presented fairly good results in relation to forecasting depreciation tail risk. This suggest that, although no model could outperform the random walk in a full-density evaluation, they can perform better in distinct parts of the distribution. In terms of the selected densities, the normal distribution generally provided a better description of the uncertainty associated with exchange rate returns compared to the skewed t-distribution.

Based on the findings in the empirical analysis, a discussion was presented on whether density forecasting provide an informational advantage relative to standard point forecasting. The discussion highlighted that certain decisionmakers are more concerned with the uncertainty associated with an exchange rate forecast rather than a central estimate. The findings in this paper suggests that density forecasting provide an informational advantage relative to point forecasting for these decisionmakers at a one-month horizon.

Possible extensions of this research include: (i) extending the conventional exchange rate models with additional predictors, (ii) using an alternative forecasting framework, and (iii) applying the methodology to alternative currency pairs and forecast horizons. First, several papers investigate the impact of incorporating additional fundamentals to conventional exchange rate models. Gloria (2010), e.g., provides evidence that incorporating commodity prices into a Taylor rule-based model improves forecasting performance for commodity-exporting countries' currencies. Additionally, the BEER model might be specified with a variety of alternative or additional predictors. It could, e.g., be interesting to investigate the out-of-sample predictability of decomposed oil price dynamics (Akram, 2019) and climate transition risk (Kapfhammer et al., 2020). Second, as seen in section 3.1, a range of different options are available with regards to, e.g., forecasting methodology, model specification, and forecast horizon. Thus, a possible extension of this research is to test whether an alternative approach to the forecasting framework verifies the conclusions of this thesis. Third, the body of literature investigating the out-of-sample predictability of krone-related currency pairs from a density forecasting perspective remains very limited. Consequently, another possible extension of this research is to apply the methodology to alternative currency pairs.

References

Adrian, T., Boyarchenko, N., & Giannone, D. (2019, April). Vulnerable Growth. *American Economic Review*, 109(4), 1263-1289.

Akram, Q. F. (2004, November 25). Oil prices and exchange rates: Norwegian evidence. *The Econometrics Journal*, 7(2), 476-504.

Akram, Q. F. (2019, September 13). Oil price drivers, geopolitical uncertainty and oil exporters' currencies. *Norges Bank Working Paper*, 15, 1-29.

Alendal, L. A. (2010). Risikopremien på norske kroner. *Norges Bank Staff Memo*, 3, 1-12.

Alstad, G. E. (2010, October 15). The long-run exchange rate for NOK: a BEER approach. *Norges Bank Working Paper*, 19, 1-21.

Apel, M., Hansen, J., & Lindbergh, H. (1996). Potential Output and Output Gap. *Quarterly Review of the Bank of Sweden*, 3, 24-35.

Berkowitz, J. (2001). Testing Density Forecasts, With Applications to Risk Management. *Journal of Business & Economic Statistics*, 19, 465-474.

Bernharden, T. (2012). Sammenhengen mellom styringsrenten og pengemarkedsrentene: 2007-2012. *Norges Bank Aktuell Kommentar*, 2, 1-8.

Bernhardsen, T., & Røisland, Ø. (2000). Factors that influence the krone exchange rate. *Economic Bulletin*, 4, 143-151.

BIS. (2019, September 16). *Triennial Central Bank Survey: Foreign exchange turnover in April 2019.* Retrieved May 2021, from Bank for International Settlements: https://www.bis.org/statistics/rpfx19_fx.pdf

Cagan, P. D. (1956). The monetary dynamics of hyperinflation. In M. Friedman, *Studies in the Quantity Theory of Money.* Chicago: University of Chicago Press.

Chen, J. M. (2018, Juni 1). On Exactitude in Financial Regulation: Value-at-Risk, Expected Shortfall, and Expectiles. *Risks*, 61, 1-29.

Chen, Y.-c., & Rogoff, K. (2003). Commodity Currencies. *Journal of International Economics*, 60, 133-160.

Cheung, Y.-W., Chinn, M. D., & Pascual, A. G. (2005). Empirical exchange rate models of the nineties: Are any fit to survive? *Journal of International Money and Finance*, 24, 1150-1175.

Chinn, M. D., & Meredith, G. (2004). Monetary Policy and Long-Horizon Uncovered Interest Parity. *IMF Staff Papers*, 51(3), 409-430.

Clark, P. B., & MacDonald, R. (1998, May). Exchange Rates and Economic Fundamentals: A Methodological Comparison of BEERs and FEERs. *IMF Working Paper*, 67, 1-38.

Coppes, R. C. (1995, February). Are exchange rate changes normally distributed? *Economics Letters*, 47(2), 117-121.

Crnkovic, C., & Drachman, J. (1997). Quality Control. I S. Grayling, *VAR: Understanding and applying value-at-risk.* Risk Publications.

De Bock, R., & de Carvalho Filho, I. (2013, January). The Behavior of Currencies During Risk-Off Episodes. *IMF Working Paper*, 8, 1-33.

De Santis, R. A., & Van der Veken, W. (2020, December). Macroeconomic Risks Across the Globe due to the Spanish Flu. *ECB Working Paper*, 2466, 1-39.

de Vries, C. G., & Leuven, K. U. (1994). Stylized Facts of Nominal Exchange Rate Returns. *Purdue CIBER Working Paper*, 2.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13, 253-265.

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating Density Forecasts, with Applications to Financial Risk Management. *International Economic Review*, 39, 863-883.

Diks, C., Panchenko, V., & van Dijk, D. (2011, August). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, 163(2), 215-230.

Dornbusch, R. (1976, December). Expectations and Exchange Rate Dynamics. *The Journal of Political Economy*, 85, 1161-1176.

Engel, C., & West, K. D. (2005). Exchange Rates and Fundamentals. *Journal of Political Economy*, 113(3), 485-517.

European Central Bank. (n.d.). *Monetary Aggregates*. Retrieved May 2021, from European Central Bank: https://www.ecb.europa.eu/stats/money_credit_banking/monetary_aggregates/html/index.en.html

Gaglianone, W. P., & Marins, J. T. (2017). Evaluation of exchange rate point and density forecasts: An application to Brazil. *International Journal of Forecasting*, 33, 707-728.

Gloria, C. M. (2010, April 29). *Exchange rate predictability: Taylor rule fundamentals and commodity prices.* Retrieved May 2021, from Université de Montréal: https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/3829/a1.1%20g%201142_Gloria_Cleide%20Mara_2010_rapport.pdf?sequence=1&isAllowed=y

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007, April). Probabilistic forecasts, calibration and sharpness. *Statistical Methodology*, 69(2), 243-268.

Good, I. J. (1952, January). Rational Decisions. *Journal of the Royal Statistical Society*, 14, 107-114.

Gourinchas, P.-O., & Tornell, A. (2004, December). Exchange rate puzzles and distorted beliefs. *Journal of International Economics*, 64(2), 303-333.

Hall, S. G., & Mitchell, J. (2007). Combining density forecasts. *International Journal of Forecasting*, 23, 1-13.

Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and Semiparametric Models.* New York City: Springer.

Hungnes, H. (2020, June). Predicting the exchange rate path: The importance of using up-to-date observations in the forecasts. *Statistics Norway Discussion Papers*, 934.

Ismailov, A., & Rossi, B. (2017, March). *Uncertainty and Deviations from Uncovered Interest Rate Parity*. Retrieved April 2021, from Centre de Recerca en Economia Internacional: https://crei.cat/wp-content/uploads/2021/02/UDUI-1.pdf

Kapfhammer, F., Larsen, V. H., & Thorsrud, L. A. (2020, December 16). Climate risk and commodity currencies. *Norges Bank Working Paper*, 18, 1-47.

Knüppel, M. (2015). Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments. *Journal of Business & Economic Statistics*, 33, 270-281.

Koenker, R., & Bassett, G. (1978, Januar). Regression Quantiles. *Econometrica*, 46, 33-50.

Kumar, R. (2014). *Strategies of Banks and Other Financial Institutions: Theories and Cases.* Amsterdam: Elsevier.

Levich, R. M. (2011, August 31). *Evidence on Financial Globalization and Crises: Interest Rate Parity*. Retrieved April 2021, from CORE: https://core.ac.uk/download/pdf/162461167.pdf

Li, S. (2016). *East Asian Business in the New World: Helping Old Economies Revitalize*. Amsterdam: Elsevier.

MacDonald, R., & Taylor, M. P. (1994, June). The monetary model of the exchange rate: long-run relationships, short-run dynamics and how to beat a random walk. *Journal of International Money and Finance*, 13(3), 276-290.

Mahmoud, H. F. (2021). Parametric Versus Semi and Nonparametric Regression Models. *International Journal of Statistics and Probability*, 10(2), 1-90.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135, 499-526.

Mark, N. C. (1995, March). Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability. *The American Economic Review*, 85, 201-218.

Martinsen, K. (2017, September 4). Norges Bank’s BEER models for the Norwegian effective exchange rate. *Norges Bank Staff Memo*, 7, 2-15.

Meese, R. A., & Rogoff, K. (1983, February). Empirical Exchange Rate Models of the Seventies: Do they fit out of sample? *Journal of International Economics*, 14, 3-24.

Molodtsova, T., & Papell, D. H. (2009, April). Out-of-sample exchange rate predictability with Taylor rule fundamentals. *Journal of International Economics*, 77(2), 167-180.

Nolde, N., & Ziegel, J. F. (2017, December). Elicitability and backtesting: Perspectives for banking regulation. *Ann. Appl. Stat.*, 11(4), 1833-1874.

Pesaran, M. H., & Timmermann, A. (1992, October). A Simple Nonparametric Test of Predictive Performance. *Journal of Business & Economic Statistics*, 10(4), 461-465.

Ranaldo, A., & Söderlind, P. (2010). Safe Haven Currencies. *Review of Finance*, 14(3), 385-407.

Rogoff, K. (1996, June). The Purchasing Power Parity Puzzle. *Journal of Economic Literature*, 34, 647-668.

Rogoff, K., & Stavrakeva, V. (2008, July 30). *The Continuing Puzzle of Short Horizon Exchange Rate Forecasting.* Retrieved April 2021, from Scholars at Harvard: https://scholar.harvard.edu/files/rogoff/files/exchange_rate_forecasting.pdf

Rossi, B. (2013). Chapter 21 - Advances in Forecasting under Instability. I G. Elliott, & A. Timmermann, *Handbook of Economic Forecasting.* Amsterdam: Elsevier.

Rossi, B. (2013, December). Exchange Rate Predictability. *Journal of Economic Literature*, 51(4), 1063-1119.

Rossi, B. (2014, September). Density forecasts in economics and policymaking. *Els Opuscles del CREI*, 37, 3-31.

Statistics Norway. (2021, May 10). *Harmonized index of consumer prices*. Retrieved May 2021, from Statistics Norway: https://www.ssb.no/en/priser-og-prisindekser/konsumpriser/statistikk/harmonisert-konsumprisindeks

Syczewska, E. M. (2010). Empirical power of the Kwiatkowski-Phillips-Schmidt-Shin test. *Warsaw School of Economics Working Paper*, 3, 1-20.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

Torvik, R. (2016). Realvalutakurs og oljeprisfall. *Samfunnsøkonomen*, 3, 41-49.

van der Meer, D. W., Munkhammar, J., & Widén, J. (2018, September 1). Probabilistic forecasting of solar power, electricity consumption and net load: Investigating the effect of seasons, aggregation and penetration on prediction intervals. *Solar Energy*, 171, 397-413.

Wang, J., & Wu, J. J. (2012, February). The Taylor Rule and Forecast Intervals for Exchange Rates. *Journal of Money, Credit and Banking*, 44, 104-144.

Wooldridge, J. M. (2018). *Introductory Econometrics: A Modern Approach (7e).* Boston: Cengage.

Yapi, N. A. (2020). Exchange rate predictive densities and currency risks: A quantile regression approach. *EconomiX Working Papers*, 16.

Zhang, S., Lowinger, T. C., & Tang, J. (2007, June). The Monetary Exchange Rate Model: Long-run, Short-run, and Forecasting Performance. *Journal of Economic Integration*, 22(2), 397-406.

Zhang, S., Wang, Y., Zhang, Y., Wang, Dan, & Zhang, N. (2020, November 1). Load probability density forecasting by transforming and combining quantile forecasts. *Applied Energy*, 277.

Appendix I: Data Sources

Eurostat. (2021). *HCPI for the Euro Area.* Retrieved February 2021 from Eurostat: https://ec.europa.eu/eurostat/databrowser/view/prc_hicp_midx/default/table?lang=en

Eurostat. (2021). *Manufacturing Index for the Euro Area.* Retrieved April 2021 from Eurostat: https://ec.europa.eu/eurostat/databrowser/view/sts_inpr_m/default/table?lang=en

Federal Reserve Bank of St. Louis. (2021). *CBOE Volatility Index: VIX*. Retrieved April 2021 from FRED: https://fred.stlouisfed.org/series/VIXCLS

Macrobond. (2021). *EURIBOR, 3 Month*. Retrieved April 2021.

Macrobond. (2021). *NIBOR, 3 Month*. Retrieved February 2021.

Norges Bank. (2021). *EUR/NOK Exchange Rate*. Retrieved April 2021 from Norges Bank: https://www.norges-bank.no/en/topics/Statistics/exchange_rates/?tab=currency&id=EUR

OECD. (2021). *M3 Index for Norway and the Euro Area.* Retrieved April 2021 from OECD: https://data.oecd.org/money/broad-money-m3.htm

Statistics Norway. (2021). *HCPI for Norway*. Retrieved February 2021 from Statistics Norway: https://www.ssb.no/en/statbank/table/03700/

Statistics Norway. (2021). *Manufacturing Index for Norway.* Retrieved February 2021 from Statistics Norway: https://www.ssb.no/en/statbank/table/07095/

U.S. Energy Information Administration. (2021). *Europe Brent Spot Price FOB*. Retrieved March 2021 from EIA: https://www.eia.gov/dnav/pet/hist/rbrteD.htm