Strategies for High Frequency FX Trading
- The choice of bucket size

Malin Lunsjö and Malin Riddarström

Department of Mathematical Statistics
Faculty of Engineering at Lund University

June 2017

Abstract

This thesis aims at developing and evaluating a model for high frequency foreign exchange data, that beats the TWAP benchmark the majority of the time. This is done by dividing the total order time into smaller time buckets and trading a smaller quantity of the total order volume in each bucket. The second purpose of the thesis is to determine if there is an optimal bucket size in which to trade in order to achieve the best results. Four different models were developed and it was found that the model that traded both passively and aggressively without a set order time performed the best. It was discovered that this model always beats the TWAP benchmark on an average day on the market. The best performing model also took the prevailing market conditions, modelled as market risk and spread risk, into account. The market risk was modelled using a prediction of the volatility during the time interval of the order and the spread risk was modelled by using a prediction of the spread. The purpose of the risk factors was to get an indication of how to choose the level at which to trade passively and aggressively in the buckets, which will be explained further in this thesis.

It was concluded that an optimal bucket size does not exist. Instead, it was decided that the client's preferences regarding potential risks and profits should be the deciding factor in determining optimal bucket size for an order. This is achieved by allowing the client to choose a certain probability of succeeding with a passive trade in a bucket and calculating the bucket size based on this probability. Prior to making the choice, the client is presented with the potential profit, market risk and spread risk for each probability. A low probability results in shorter bucket sizes and thus a shorter order time. This in turn results in a low market risk but a high spread risk. A high probability on the other hand, results in longer bucket sizes and a longer order time which implies a low spread risk but a high market risk. This means that a risk averse client chooses the low probability with less risk of market changes at the expense of loosing the spread, and vice versa for a less risk avert client.

The three currency pairs that were considered in this thesis are EUR/SEK, EUR/NOK and EUR/USD. High frequency was in this thesis defined as second-by-second up to minute-by-minute observations.

Keywords: High frequency, FX trading, Shifted geometric distribution, Monte Carlo, Time Series Analysis, EWMA, Bucket size, TWAP.

Acknowledgement

Contents

Chapter 1

Introduction

1.1 Purpose and aim of the thesis

In FX trading an order can be placed to buy or sell a set quantity of a certain currency pair during a specific time interval. One way of evaluating the performance of an order is by comparing it to the time weighted average price, TWAP, benchmark. The TWAP benchmark is simply the average execution price of the observed aggressive prices over a specified time interval. There are trading algorithms that try to match the TWAP benchmark. A disadvantage of these algorithms is that the trading patterns are usually uniform and independent of prices and volumes. Furthermore, the approach is not very flexible. Another approach would be to allow trading with different frequencies as well as to attempt to trade passively.

The aim of this thesis is to develop a model that beats the TWAP benchmark the majority of the time. This is investigated by dividing the total order time into smaller time intervals, called time buckets, and trading in each bucket instead of at every second. The trade in each bucket can be either passive or aggressive depending on the prevailing market conditions. In conjunction with this, the purpose of the thesis is also to determine if there is an optimal bucket size in which to trade to achieve the best results.

As mentioned, the study takes into account the prevailing market conditions. More specifically, the market risk and spread risk will be investigated. The market risk is modelled by using a prediction of the volatility during the time interval of the order and the spread risk by using a prediction of the spread in combination with other factors determined by the algorithm. The purpose of the risk factors is to get an indication of how to choose a level of probability of trading passively in the buckets, which will be explained further in the thesis.

1.2   The Foreign Exchange Market

The Foreign Exchange Market, also known as the FX Market, is said to be the world's largest financial market with a daily turnover of approximately 5 trillion US dollars [1]. The FX market is primarily an Over-The-Counter, OTC, traded market which means that the market participants connect to each other directly or via different brokers. Within the FX market there is a broad range of different types of market participant, for example commercial banks, investment banks, large corporation and hedge funds [2].

Electronic dealing systems were first introduced in the 1980's. The systems collected market data across multiple dealers and exchanges which allowed market participants to trade with each other at the best prices available on the systems. The development of electronic trading systems has lead to a market which is dominated by high-frequency trading. One difference between traditional FX trading and high frequency trading is that many traditional FX traders hold their trading positions for long periods, that is to say weeks, days or minutes, whilst high frequency trades can be done at a millisecond level. Other differences between traditional and high frequency trading include that high frequency trading is done automatically and that the frequent transactions have a relatively low average gain per trade compared to the traditional way [3].

1.3   Delimitations

The thesis will only cover three currency pairs; EUR/SEK, EUR/NOK and EUR/USD, as these are the most frequently traded pairs by Swedish banks. Additionally, the smallest spaced data is second-by-second observations. Throughout the thesis all examples, figures and tables will represent EUR/SEK and the other currencies can be found in the appendix. It is not certain that the conclusions will hold for other currencies or time horizons than those evaluated in the thesis.

It is assumed that the market is always liquid when trades are executed and that the trades never result in any essential market impact. This is to ensure that all orders can be executed without having to develop models for the market liquidity or market impact.

Furthermore, the number of trades executed in each bucket is limited to one throughout the thesis and the volume traded in each bucket is constant. This simplification is due to the time limit of the thesis.

Finally, only sell orders are evaluated in the method and results since sell and buy orders are symmetric.

Chapter 2

Theory

2.1 Terminology

Below follows some trading terminology that is used throughout the thesis.

2.1.1 Currency pair

When trading currency pairs, one currency is sold to buy the other. A currency pair is for example denoted EUR/SEK, where EUR is the base currency and SEK is the quote currency. This ratio is one unit - the *currency pair* is traded and not just EUR or SEK. The exchange rate indicates the quantity of SEK needed to buy one EUR. The ratio EUR/SEK is greater than one since EUR is a stronger currency than SEK.

2.1.2 Bid and offer price

A *bid price* is the price that a buyer is willing to pay for a certain amount of a currency and the *offer price* is the price a seller is willing to sell a certain amount of the currency for. The bid price is lower than the offer price because people want to buy at a low price and sell at a high price. Since numerous platforms are being traded at simultaneously, there are always several bid and offer prices on the market. The lowest offer price in the market is referred to as the *best offer price*, since this is the best price at which to buy a certain currency. The highest bid price is referred to as the *best bid price*. The difference between the best bid and offer prices creates a spread which is referred to as a *bid-offer spread*, or simply just *spread*, s_i, and it is calculated according to equation (2.1). The price found in the middle of the spread is referred to as the *mid-price*, p_i, and calculated according to equation (2.2). In this thesis only the best bid and offer prices will be considered.

$$s_i = o_i - b_i, \tag{2.1}$$

$$p_i = \frac{o_i + b_i}{2}, \qquad\qquad (2.2)$$

where o_i is the best offer price and b_i is the best bid price.

2.1.3   Buckets

The time interval during which the algorithm is to be executed is split into smaller sized intervals, referred to as *buckets*. A part of the order is executed in each bucket and how it should be traded is evaluated in each separate bucket. The bucket size is the total number of seconds in a bucket.

2.1.4   Passive and aggressive trading

A trade in each bucket can be done *passively* or *aggressively*. Trading passively means placing a certain price to sell or buy a currency pair, and waiting for someone to accept the price. Trading aggressively means accepting a price that someone else has placed. The advantage of trading passively is that a more favourable price is received compared to when trading aggressively. A market participant cannot expect to always succeed with a passive trade since there are many market participants attempting to trade passively at the same time. Therefore, a passive trade will be successful with a certain probability. This probability is an important factor to be considered in the execution algorithm. In this thesis succeeding through passive trading means receiving the offer price when executing a sell order and the bid price for a buy order.

2.2   Time Weighted Average Price

Benchmark comparison is arguably the most used tool for analyzing algorithmic performance. It is done by selecting an appropriate benchmark and comparing this to the average execution price. Matching or beating the benchmark would indicate a good performance [4].

The Time Weighted Average Price, TWAP, benchmark is the average of all observed prices over a set period. The average of the prices received with a trading algorithm is compared to the TWAP benchmark. TWAP reflects how the market price has changed over time. Equal weights are given to all trades and market conditions are not taken into account. Trading algorithms that try to match this benchmark are usually based on a uniform time-based schedule and are unaffected by any other factors, such as market price or volume. The advantages of these algorithms are that they are easy to implement and execute, and market impact is avoided by splitting the order into smaller pieces. The disadvantage of this strategy is its predictability, it can signal other market participants when the trades will take place. Furthermore, the strict time schedule can lead to poor execution as it does not take unfavourable prices or liquidity drops into consideration. Below is the equation for calculating TWAP for a number n prices.

$$\text{TWAP} = \frac{1}{n} \sum_{i=1}^{n} P_i,$$

where P represents the price [4].

2.3 Shifted Geometric Distribution

The shifted geometric distribution is a version of the geometric distribution which represents the probability distribution of the number X needed to get one success in a sequence of Bernoulli trials. A Bernoulli trial process is a random process in probability. Each trial has two possible outcomes, success or failure, and each trial is independent of the outcome of the other trials. The probability of success in any independent trial is $p \in [0, 1]$, also referred to as the success parameter of the process. The success parameter is the only controlling parameter of the shifted geometric distribution [5]. With the probability p of success for each trial, the probability of first success on the nth trial is given by equation (2.3) [6].

$$P(X = n) = (1-p)^{n-1}p, \quad n = 1, 2, 3.... \tag{2.3}$$

2.4 Monte Carlo Integration

Monte Carlo integration makes it possible to approximate an integral of an average value that may not be done analytically. To estimate the value of the integral, a set of points is drawn randomly from a distribution with support over the range of integration. In order to approximate the integral for some expectation, the expression below is used:

$$\tau \overset{\text{def}}{=} \mathbb{E}(\phi(X)) = \int_A \phi(x)f(x)dx,$$

where

X: is a random variable with possible values in A$\subseteq \mathbb{R}^d$ where $d \in \mathbb{N}$ may be very large,

f: $A \to \mathbb{R}_+$ is the probability density of X referred to as the target density and

$\phi : A \to \mathbb{R}$ is a function referred to as the objective function such that the above expectation is finite.

Let $X_1, ..., X_N$ be independent random variables with target density f, then by the law of large numbers

$$\tau_N = \frac{1}{N} \sum_{i=1}^{N} \phi(X_i) \to \tau = \mathbb{E}(\phi(X)),$$

when N tends to infinity [7].

2.4.1   Estimating quantiles

Monte Carlo methods can also be used to estimate quantiles of a distribution. A quantile is a cutoff point that tells how much can be lost in the worst case scenario. The p-quantile Q^p, $p \in [0,1]$, is defined by

$$P(Y \leq Q^p) = p, \tag{2.4}$$

where Y is the data set considered. This says that a proportion p of the evaluated values are below the value Q^p. By simluating N samples and sorting them by size, the p-quantile can be estimated to satisfy equation (2.4). Similarly, this can be done to compute the best case scenario [8].

2.5   Time Series Modelling

A time series is a sequence of values of a variable collected at regular intervals over a period of time in successive order. A time series is useful to understand the underlying forces and structures that produce the observed data in order to fit a model to the data and make forecasts and predictions. The underlying forces and structures of data points taken over time can be autocorrelation, trend and seasonal variation. Autocorrelation is a measure of the correlation within a time series at different lags, trend is the long run evolution and seasonal variation is periodic fluctuation. Data with this type of non-stationary behaviour is unpredictable and cannot be modeled or forecasted as this could produce spurious results of dependence between variables that do not exist. Non-stationary data therefore needs to be transformed into covariance stationary data.

2.5.1   Time Series Properties

A process y_t is weakly or covariance stationary if:

1. $E(y_t^2) < \infty$, $\forall$ integers t.

2. $E(y_t) = m$, $\forall$ integers t.

3. $Cov(y_t, y_{t+h}) = \gamma(h)$, $\forall$ integers t and h.

In other words the process has constant mean and covariance over time as well as finite variance. For a covariance stationary process the autocorrelation function, ACF, gives the correlation of observations at different lags k:

$$\rho_k = \frac{Cov[y_t, y_{t-k}]}{\sqrt{Var[y_t]Var[y_{t-k}]}} = \frac{\gamma_k}{\gamma_0},$$

where γ_i is the autocovariance. The partial autocorrelation function, PACF, gives the partial correlation of a time serie at different lags. It gives the direct correlation between a time series y_t and the lagged series y_{t-k}, with the linear dependence of the intermediate variables y_s, $y_{t-h} < y_s < y_t$, removed [9].

White noise is a series of uncorrelated, identically distributed random variables [10]. A white noise error term has the following properties:

1. $E[\epsilon_t] = 0$.

2. $E[\epsilon_t^2] = \sigma^2$.

3. $E[\epsilon_t \epsilon_s] = 0, t \neq s$.

Some of the commonly used processes for modelling time series are presented below.

2.5.2 Moving-average model

The moving-average, MA, process of order q is given by:

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q},$$

where $\theta_1, ..., \theta_q$ are the parameters of the model and $\epsilon_{t-1}, ..., \epsilon_{t-q}$ are the white noise error terms. Therefore, a MA process is a linear combination of current and previous error terms. The ACF for the MA model is non-zero for lags $k = 0, 1..., q$ and the PACF is non-zero for all lags but decays as the lag increases [9].

2.5.3 Autoregressive model

The autoregressive, AR, process of order p is given by:

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t, \tag{2.5}$$

where $\phi_1, ..\phi_p$ are the parameters of the model and ϵ_t is the white noise error term. The current value for the AR process is linearly related to the past values plus an error term. The ACF for the AR model is non-zero for all lags but decays for increase in lag and the PACF is zero for all lags $k > p$ [9].

2.5.4 Autoregressive Moving-Average model

The Autoregressive Moving-Average, ARMA, process is a combination of the AR and MA processes. An ARMA process of order (p, q) is given by:

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}$$
$$= \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}.$$

This model accounts for both previous values of the process as well as previous error terms. The ACF and PACF for the ARMA model are non-zero for all lags but both decay gradually with increase in lag [9].

2.5.5   Autoregressive Conditional Heteroscedasticity model

A common finding in studies of financial data is that volatility is not constant over time. There are periods of uncertainty with high volatility and calm periods with low volatility. This is known as volatility clustering and in order to capture this, the Autoregressive Conditional Heteroscedasticity, ARCH, model was introduced.

Consider the AR(p) model given by equation (2.5) but with an error term η_t that is uncorrelated and zero mean noise with changing variance instead of the white noise error term ϵ_t. Assume that the error term can be represented by

$$\eta_t = \sigma_t \xi_t,$$

$$\xi_t \sim N(0, 1),$$

where all ξ_t are independent and identically distributed. Further assume that

$$\sigma_t^2 = \omega + \alpha_1 \eta_{t-1}^2 + \alpha_2 \eta_{t-2}^2 + ... + \alpha_q \eta_{t-q}^2.$$

Then the conditional variance of η_t can be calculated as

$$\begin{aligned} Var[\eta_t | \eta_{t-1}, ..., \eta_{t-q}] &= E[\eta_t^2 | \eta_{t-1}^2, ..., \eta_{t-q}^2] \\ &= \sigma_t^2 \\ &= \omega + \alpha_1 \eta_{t-1}^2 + \alpha_2 \eta_{t-2}^2 + ... + \alpha_q \eta_{t-q}^2. \end{aligned}$$

This is defined as an ARCH(q) model where the conditional variance depends on past values of squared errors [9]. The parameters ω and α_i are required to be non-negative to ensure positive values of the variance and $\sum_{i=1}^{q} \alpha_i < 1$ to preserve stability [10].

2.5.6   Generalized Autoregressive Conditional Heteroscedasticity model

The Generalized Autoregressive Conditional Heteroscedasticity, GARCH, model describes the conditional variance as a function of past squared errors *and* past conditional variances:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \eta_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2,$$

where ω and all α_i, β_i must be non-negative to ensure positive variances, and $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$ to preserve stability [10]. It can be noted that the model resembles an ARMA model apart from a white noise error term with constant variance from the MA part [9].

2.5.7   Exponentially Weighted Moving Average

The Exponentially Weighted Moving Average, EWMA, is a volatility weighted historical model and a special case of the GARCH(1,1)-model. The EWMA model is defined as

$$\sigma_{t+1}^2 = (1 - \lambda)x_t^2 + \lambda \sigma_t^2, \tag{2.6}$$

where $\lambda \in [0, 1]$ is a forgetting constant, σ_t^2 is the squared volatility and x_t is the time series value at time t. The EWMA model assigns weights to the past conditional volatility where more recent volatilities are assigned higher weights and older volatilities get lower weights. The smaller the value of λ, the larger the relative weight given to the most recent sample. This allows for volatility clustering in the model, in the sense that the volatility at the current time step $t + 1$ is positively correlated with the volatility at the previous time step t. When using high frequency data this is favourable since older volatilities will have less impact on the volatility at the current time step [11].

2.6 Maximum Likelihood Estimation

The Maximum Likelihood, ML, method can be used to estimate the parameters of a specific time series model if the distribution of the observations is known. The ML estimator is defined as the argument that maximizes the joint likelihood

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} L(\theta),$$

where $L(\theta)$ is called the likelihood function

$$L(\theta) = p(x_0, ..., x_N | \theta),$$

$$= \left(\prod_{n=1}^{N} p(x_n | x_{n-1}, ..., x_0, \theta) \right) p(x_0 | \theta).$$

The ML method says that the likelihood function should be maximized with respect to the unknown parameters in the model, which can be viewed as maximizing the likelihood that the observations were generated by the model.

Since the argument maximizing $L(\theta)$ is not affected by a logarithmic transformation $l(\theta) = \log L(\theta)$, the optimization problem can be rewritten as

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} \log p(x_0 | \theta) + \sum_{n=1}^{N} \log p(x_n | x_1, ..., x_{n-1}, \theta).$$

The estimates are asymptotically Gaussian converging according to

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightarrow N(0, I_F^{-1}),$$

where θ_0 is the true parameter. I_F is the Fisher Information matrix and is defined as

$$I_F = \mathbb{V}[\nabla_\theta \log p(X | \theta_0)],$$

where ∇_θ is the gradient. It can also be expressed as

$$I_F = \mathbb{E}[(\nabla_\theta \log p(X | \theta_0))(\nabla_\theta \log p(X | \theta_0))^T],$$

and

$$I_F = -\mathbb{E}[\nabla_\theta \nabla_\theta \log p(X | \theta_0)],$$

where $\nabla_\theta \nabla_\theta$ is the Hessian with respect to the parameters [10].

2.7   Data distributions

The skewness and kurtosis are important properties to consider in order to determine the distribution of data. A simple t-test can be applied to a data set to check if it is normally distributed with zero-mean values.

2.7.1   Skewness

Skewness is a measure of asymmetry of the distribution and can be used to test for a standardized distribution. The skewness of the observations X is measured according to

$$Skewness = \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3},$$

where $\mathbb{E}$ is the expected value, μ is the mean and σ is the standard deviation of the observations. The distribution is said to be symmetric if the skewness is zero. If the skewness is positive, the distribution is skewed to the right and similarly the distribution is skewed to the left if the skewness is negative [11].

2.7.2   Kurtosis

Another approach to test for a standardized distribution is by examining the kurtosis of a distribution. The kurtosis is how peaked a distribution is and how fat its tails are. The kurtosis of the observations X is measured according to

$$Kurtosis = \frac{\mathbb{E}[(X-\mu)^4]}{\sigma^4},$$

where $\mathbb{E}$ is the expected value, μ is the mean and σ is the standard deviation of the observations. The normal distribution has a kurtosis of 3, which means that if the kurtosis is higher than 3 the distribution considered is more peaked than the normal distribution and similarly if it is lower than 3 the distribution is more flat than the normal distribution [11].

2.7.3   One sample t-test

The one-sample t-test is a parametric test of the location parameter when the population standard deviation is unknown. The t-test returns a test decision for the null hypothesis that the data in a sample x comes from a normal distribution with mean μ equal to zero and unknown variance σ^2 at a significance level $1 - \alpha$. The test statistics is measured as

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}},$$

where $\bar{x}$ is the sample mean, μ is the hypothesized population mean, σ is the sample standard deviation and n is the sample size [12].

2.8   Random walk

It is possible to model the logarithmic price returns according to the random walk hypothesis. This means that the logarithmic price returns r_t follow a random walk defined as

$$r_t = \epsilon_t,$$

$$\epsilon_t \sim N(0, \sigma),$$

where all r_t are independent and identically distributed [13].

2.9   Value-at-Risk

Value-at-Risk, VaR, is defined as the quantile of the loss distribution of the stochastic variable X. VaR is measured according to

$$VaR_\alpha(X) = \inf\{x : P(X > x) \leq 1 - \alpha\},$$

where α is the confidence level for the quantile of the loss distribution, P is the probability function and x is the smallest loss such that the probability of a future loss X is larger than x, is less than or equal to $1 - \alpha$ [14].

2.10   Box plot
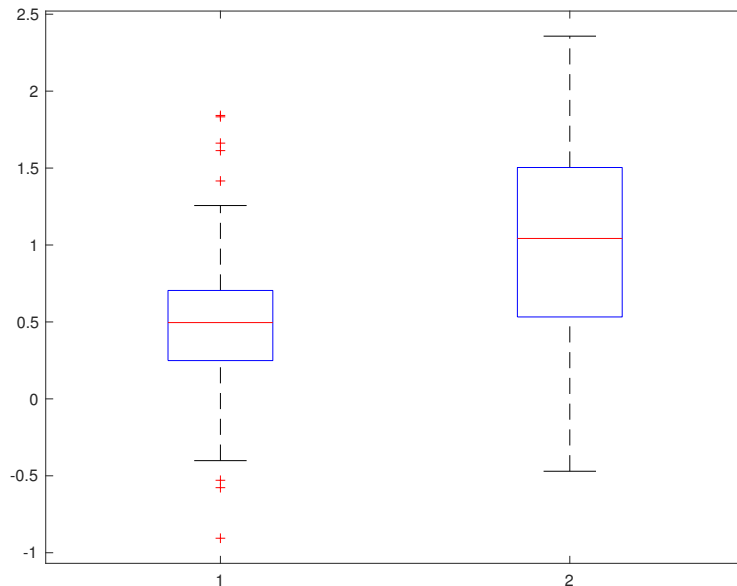
Figure 2.1: Example of box plot

Quartiles are a way of dividing data into four equal groups by using three points in the data; the lower quartile, the median and the upper quartile. The median is the middle

point in the data. The lower quartile is the middle point between the median and the minimum value in the data. The upper quartile is the middle point between the median and the maximum value in the data. In the box plot in figure (2.1) the rectangular box stretches between the lower and upper quartile. The median is the line drawn through the box. The vertical lines are called whiskers and display the variability outside the quartile. The red points are called outliers and represent the extreme values in the data. The results of the box plot can be used to interpret the distribution of the data. A short box implies low variation in the data in the quartile and a long box means there is more variation. The same interpretation can be used for the data outside the quartile, for instance outliers that are far apart indicate that there is a large variation in the extreme points in the data [15].

Chapter 3

Data

3.1 Description

Three currency pairs were used; EUR/SEK, EUR/NOK and EUR/USD and each pair was observed during a one month period, from 2016.02.01 to 2016.02.28, sampled at every second. The data consisted of the date, time, best bid price and best offer price. The data for each pair was captured in the same month to ensure that the results could be accurately compared. Table (3.1) illustrates the data and as demonstrated by the table, not all seconds were observed and represented in the data set. Additionally, some values were represented by zeroes. This is due to absence of price changes or some technical failure at those particular seconds. The aim of the thesis is based on second-by-second data and therefore it is essential to capture all seconds during the observed period. For this reason the missing data was handled by filling in the missing seconds and the corresponding prices with the most recently observed price. Another method that could have been used is interpolation, but since this requires the use of information available at time $t+1$, it would imply the use of information that is unknown at the current time t of the missing data. The valid price in reality at time t, given that there has been no new observation, is the most recent observed price. The missing values represented by zeroes were also replaced by the previous value. An example of the data after the missing values were added can be seen in Table (3.2).

Date	Time	Best Bid	Best Offer
2016.02.01	00:00:00	9.4017	9.4097
2016.02.01	00:00:01	9.4017	9.4097
⋮	⋮	⋮	⋮
2016.02.01	00:00:07	9.402	9.4097
2016.02.01	00:00:12	9.4018	9.4097
⋮	⋮	⋮	⋮
2016.02.01	21:59:56	0	0
⋮	⋮	⋮	⋮
2016.02.28	23:59:59	9.5428	9.548

Table 3.1: Example of data with both missing seconds and missing values, EUR/SEK

Date	Time	Best Bid	Best Offer
2016.02.01	00:00:00	9.4017	9.4097
2016.02.01	00:00:01	9.4017	9.4097
⋮	⋮	⋮	⋮
2016.02.01	00:00:07	9.402	9.4097
2016.02.01	00:00:08	9.402	9.4097
2016.02.01	00:00:09	9.402	9.4097
2016.02.01	00:00:10	9.402	9.4097
2016.02.01	00:00:11	9.402	9.4097
2016.02.01	00:00:12	9.4018	9.4097
⋮	⋮	⋮	⋮
2016.02.01	21:59:56	9.4449	9.4754
⋮	⋮	⋮	⋮
2016.02.28	23:59:59	9.5428	9.548

Table 3.2: Example of data after filling in missing seconds and values, EUR/SEK

The data was further modified by removing all observations during the weekend, as it was assumed that no trading took place at this time. Furthermore, only the hours 07:00-18:00 (GMT) were studied based on the fact that during the hours when the market is closed, there is very little movement in the prices resulting in many values of the data remaining constant for a long time. This is not favourable when modelling the data, which is explained further when transforming the data into log-returns. Figure (3.1) illustrates the final data for EUR/SEK for three different days; February 1st, February 8th and February 16th. As mentioned in section 2.1.2, the offer price lies above the bid price.
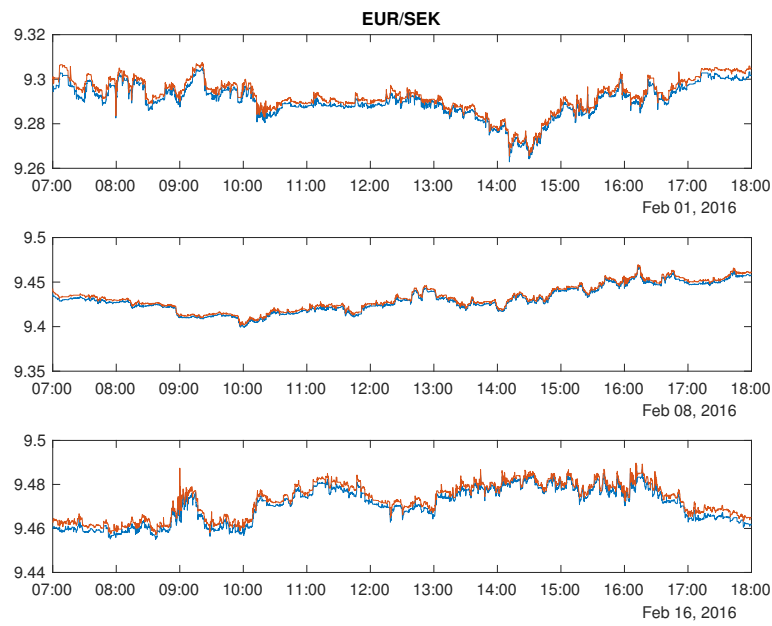
Figure 3.1: EUR/SEK bid and offer prices for three different days

The first five days of the data were used for model estimation whilst the next five days were used as comparison days. This is explained further in the volatility chapter. The last ten days were used for testing the models. An *average day* was calculated in order to achieve results that were accurate for any given day as opposed to a specific day. This was done by taking the average of the observed prices at every second between the hours 07:00-18:00 for the ten days used for testing. This resulted in a day that did not follow a specific, local trend but rather a general trend for an average day on the market. The bid and offer prices for the average day for EUR/SEK are shown in figure (3.2).
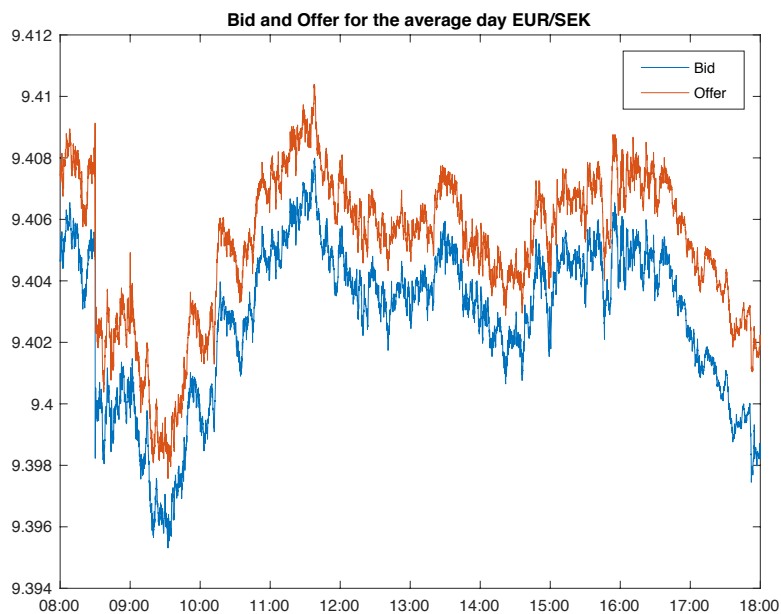
Figure 3.2: Bid and Offer During the Average Day EUR/SEK

As demonstrated in figure (3.2), there was a jump in the exchange rate around 08:30. This is due to the information release from the Swedish National Bank about the repo rate on February 11th.

In some cases there were large deviations in the data, for example peaks in the spread where the offer and bid prices were a lot further apart than commonly observed. These deviations could have been caused by one participant executing a large order, resulting in a sudden change in the price and a peak in the spread. An example is shown in figure (3.3), where there is a large peak in the spread between 08:00-09:00 for EUR/SEK. These peaks often disappear quickly since the spread is regulated by market makers who place prices until the spread has stabilized. Since modelling the extreme data values is out of scope for this thesis, the extreme values were removed for all currencies. The result after removing the extreme values in figure (3.3) is shown in figure (3.4).
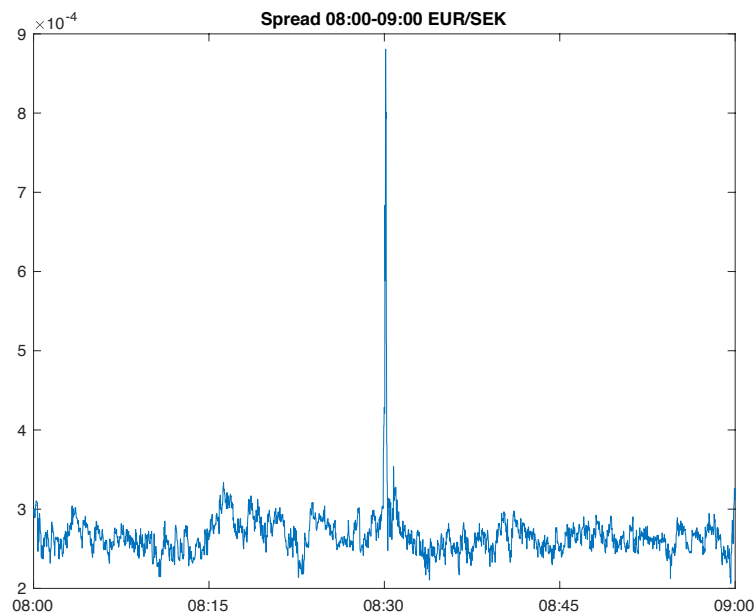
Figure 3.3: Example of an extreme value in the EUR/SEK spread between 08:00-09:00
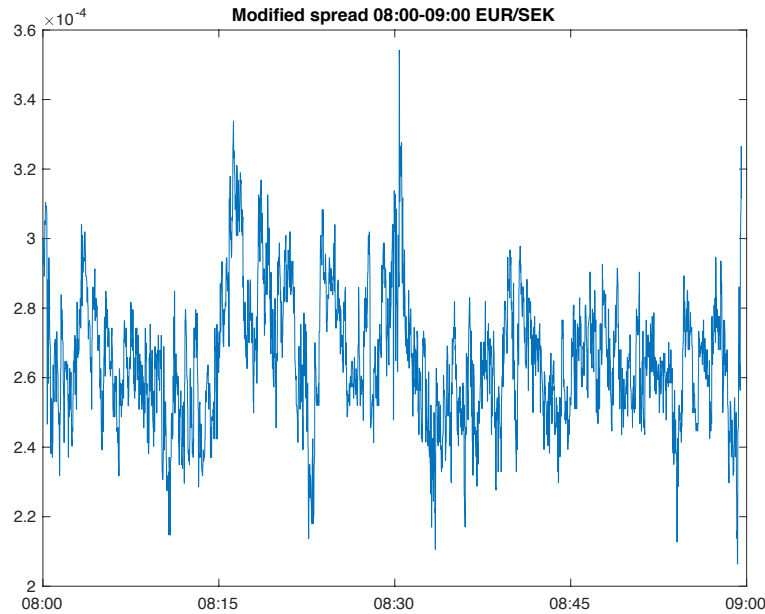
Figure 3.4: EUR/SEK spread between 08:00-09:00 after removing extreme values

3.2 Volatility modelling

A log-return transformation of the data was performed in order to calculate the volatility in a way that was comparable between the data sets. The volatility was estimated in order to get an indication of the currency exchange rate movements in general and not the specific movements of bid and offer prices. Therefore, it was sufficient to study the movement of the mid-price rather than the bid and offer prices separately. The log-returns were calculated as:

$$r_t = \log\left(\frac{p_t}{p_{t-1}}\right) = \log(p_t) - \log(p_{t-1}).$$

Long sections of constant data can prove problematic when calculating log-returns. This is because two consecutive prices of the same value result in a log-return of zero, meaning a long section of prices of the same value result in long sections of zeroes. This in return results in long sections with a realized volatility of zero. To mitigate this problem, the data was limited to the hours where there is most activity and therefore most price changes in the market, as mentioned in section 3.1. However, there were still large parts of the data that consisted of zeroes after limiting the hours. To solve this problem, minute-by-minute data was used instead of second-by-second when estimating the volatility, as this allowed more time for price fluctuations.

3.3   Average turnover

Each day experiences significant variation in the activity levels on the market between certain hours and even minutes. This means that there are periods during a day when more trades occur and there is a higher probability of succeeding with a passive trade, and vice versa. Data consisting of the average number of executed trades per half hour was evaluated over a three month period in order to calculate this probability. The data was first divided by two based on the assumption that half of the trades were sold and half were bought. It was further divided by the number of trading days in the three month period to get the average number of executed passive trades per half hour during one day. Based on this information the probability of success needed to receive the number of passive trades was calculated. This resulted in a probability of succeeding with passive trades for each half hour. In figure (3.5) this probability is shown for EUR/SEK. The curve also reflects the amount of activity on the market during different times of the day.
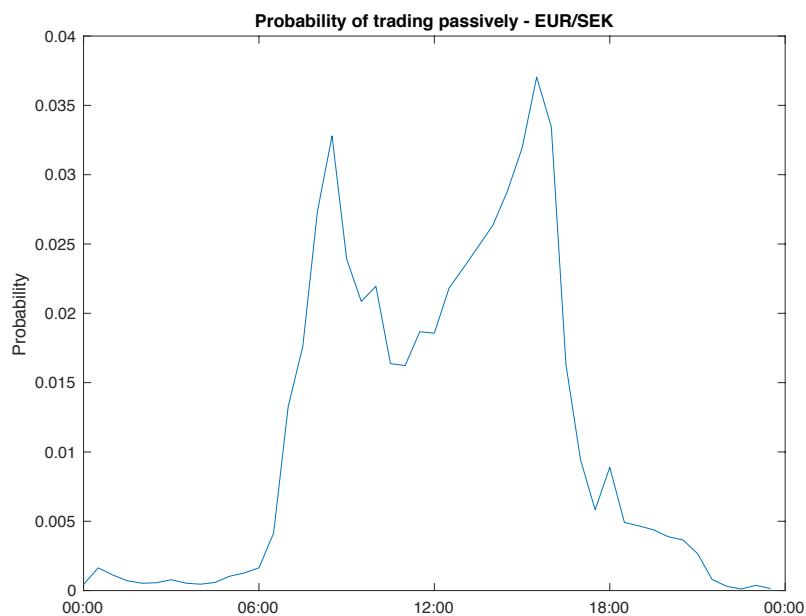
Figure 3.5: Probability of trading passively EUR/SEK, p

The data was also used to calculate the average volume of each currency pair traded during each half hour. The average traded volume in the market is close to 1 million, hence it was approximated to 1 million. Thus, the average volume was simply determined by taking the number of trades executed per half hour in one day and multiplying by 1 million. The average volume was then used to set a constraint for the recommended maximum volume to be traded at a certain time during the day. It was assumed that 50% of the whole volume traded on the market was a reasonable constraint in order to prevent market impact. The constraint resulted in a recommended maximum order size during a certain time presented in table (3.3).

Time	Recommended max order size, millions		
	EUR/SEK	EUR/NOK	EUR/USD
07:00:00	12	13	161
07:30:00	16	12	169
08:00:00	25	20	288
08:30:00	30	16	221
09:00:00	22	21	220
09:30:00	19	16	228
10:00:00	20	14	196
10:30:00	15	11	160
11:00:00	15	14	178
11:30:00	17	14	189
12:00:00	17	15	217
12:30:00	20	22	241
13:00:00	21	20	305
13:30:00	22	20	383
14:00:00	24	21	342
14:30:00	26	22	370
15:00:00	29	19	404
15:30:00	33	24	431
16:00:00	30	21	357
16:30:00	15	11	233
17:00:00	8	7	158
17:30:00	5	4	167

Table 3.3: Recommended max order size for each currency

Chapter 4

Model Development

Four models were developed and evaluated in an attempt to beat the TWAP benchmark. The TWAP benchmark was calculated as the average of all observed prices from the second an order was initiated until the last second of the order. Furthermore, the TWAP benchmark only captured the prices that entailed aggressive trading, meaning the bid prices when executing a sell order. All models were simulated and executed in MATLAB.

4.1 Model 1 - Local Greedy

Initially, the bucket sizes that were evaluated and the time during which the order should run were predetermined. The time interval was then split equally into a number of buckets based on the bucket size. The decision regarding which bucket sizes to evaluate was based on the currency pair; for EUR/SEK and EUR/NOK trades are not as high frequent as for EUR/USD. Based on the liquidity of the currency pairs it was concluded that reasonable bucket sizes lie between 30 seconds to 10 minutes for EUR/SEK and EUR/NOK, and 5 to 30 seconds for EUR/USD.

A model called *local greedy* was created as a first attempt to beat the TWAP benchmark. This model sought to trade in each bucket the first time a price was observed that was better than the TWAP benchmark. A better price means a higher price when executing a sell order and a lower price when executing a buy order. The algorithm traded at the last second in the bucket if no such price had been observed by the end of the bucket. The TWAP benchmark was updated continuously during the time interval of the order.

4.2 Model 2 - Trading passively and aggressively with a set order time interval

A key limitation of the local greedy model is that it only considers trading aggressively and never attempts to trade passively. To address this a new model was produced that allowed for both passive and aggressive trading.

4.2.1 Generating a time series of trades

To generate a time series of trades for an order it was assumed that all buckets were the same and independent of each other. It was further assumed that the times when passive trades occurred were randomly and independently distributed. The assumptions were necessary in order to model the passive trades with a shifted geometric distribution, where $P(X = n)$ in equation (2.3) was the probability of succeeding with a passive trade after n seconds. The formula below was used to model the time series of trades using the shifted geometric distribution:

$$\int_k^{k+1} \lambda e^{-\lambda x} dx = e^{-k\lambda}(1 - e^{-\lambda}). \tag{4.1}$$

Set

$$p = (1 - e^{-\lambda}),$$

then λ can be expressed as

$$\lambda = -\ln(1 - p),$$

and (4.1) can be written as

$$e^{-k\lambda}(1 - e^{-\lambda}) = (1 - p)^k p.$$

Let P be defined as

$$P = 1 - e^{-\lambda x}, \tag{4.2}$$

then x can be expressed as

$$x = \frac{\ln(1 - P)}{\ln(1 - p)},$$

where p is the probability of succeeding with a passive trade described in section 3.3. The number of seconds in each bucket was generated through $n = x$. A passive trade was attempted at every second in each bucket apart from the last second. If no passive trade was successful in a bucket, an aggressive trade was made at the last second. Each order was replicated N number of times in order to ensure a reliable result, thus generating N possible outcomes.

4.2.2 Trading size restrictions

The volume traded at each trade had to be restricted in order to reflect realistic trades. The administrative costs when trading are the settlement cost and the brokerage cost. The settlement cost was set to two dollars per trade and the brokerage cost was set to three dollars per million dollars traded. The income per million dollar traded was estimated to eight dollars. These assumptions are based on actual costs and incomes provided by SEB. Furthermore, the costs are assumed to be the same regardless of the currency pair. This resulted in the following equation, which was set up in order to calculate the minimum value that can be traded at each trade without resulting in a loss:

$$P_f = 8 \cdot M - 3 \cdot M - 2 \cdot N, \tag{4.3}$$

where P_f is the profit, M is the quantity traded in millions, and N is the number of trades.

4.2.3 Algorithm

A new model could be developed using the method for generating a time series of trades and the trading size restrictions. The parameters of the algorithm were; bucket size, start time, end time, order volume, side (buy/sell), passive probability and number of replicates. The number of buckets was determined by dividing the order time by the bucket size and then rounded down to the nearest integer. A consequence of rounding down was that leftover seconds that did not fit into a bucket were not used for generating the time series of trades. The time series was generated half hour-wise as the probability p of being passive changes during the day, as mentioned in section 3.3. The generated time series was then mapped against the corresponding passive or aggressive prices to determine which prices were received.

The results of the study demonstrated that client preference is the most important factor in determining the optimal bucket size. One client may prefer the chance of receiving a higher profit compared to the TWAP at the expense of the risk of negative market change. In comparison, another client may be willing to pass up the chance of receiving a higher profit in preference to a lower risk of a negative market change. Therefore, model 3 was developed which takes the client's preferences into account.

4.3 Model 3 - Trading passively and aggressively without a set order time interval

To be able to evaluate different ways of executing an order the following factors were predetermined; the currency pair, the total volume to trade and what time between 07:00-18:00 to start executing the order. The time series of trades and trading sizes were generated in the same way as in model 2. Furthermore, the model was mainly based on the client's willingness to trade passively.

4.3.1 Choosing bucket sizes

To be able to ensure that a client had *at least* the probability P of trading passively in each bucket, the bucket size was modelled as follows:

Let P be the level at which the client wishes to succeed at being passive in each bucket and let p be the probability of succeeding with a passive trade for each second in the bucket. The shifted geometric distribution can then be used to calculate the bucket size for each P as

$$n = 1 + \frac{\ln(1 - P)}{\ln(1 - p)}, \tag{4.4}$$

where n is the total number of seconds in each bucket, rounded up to the nearest whole second. The number 1 is one second which is added to the bucket size in order to have one second in each bucket at which an aggressive trade can be made. This means that

the bucket size will always be at least two seconds. Since the bucket size depends on P, different bucket sizes were examined. The probability of being passive for each second varies over the day and therefore the bucket sizes were calculated half hour-wise.

4.3.2   Order time interval

The time interval during which an algorithm is executed depends on the order volume, bucket size and trading size. The total number of trades equals the total number of buckets, n_b, since one trade is executed per bucket and these are determined by the order volume and the trading size according to

$$n_b = \frac{\text{order volume}}{\text{trading size}}.$$

The total order time T is then determined by the bucket sizes and the number of buckets. The order time was restricted such that the order has to be completed before 18:00.

4.3.3   Spread risk

Two risk measures were defined in order to take the risk of the different choices of P into account: (1) spread risk and (2) market risk.

The spread risk can be seen as a measure of the cost of trading aggressively as opposed to passively for the executed trades. A high spread will result in a large cost when trading aggressively, whilst a low spread means there are minimal differences between aggressive and passive trades. The spread risk is a measure of the total cost of all the aggressive trades executed in an order.

In order to calculate the spread risk the spread needed to be predicted for the period during which the execution took place. Since it is impossible to determine exactly what will happen in the future, the spread was predicted for the point in time where the execution starts and then assumed constant during the entire time period of the order. The assumption was based on a prediction of little spread change during the limited time period that was examined. The observed spread was calculated according to equation (2.1).

The distribution of the observed spread was investigated in order to be able to predict the spread, but it was found that it did not follow any standard distribution. Therefore, in order to predict the spread without making assumptions about the distribution, the spread was calculated by taking the mean of the observed spreads during the hour previous to the start time of the order. The previous hour was used since it is high frequency data and the spread is mostly affected by recent values. In order to make the spread comparable for all currencies, it was normalized by division of the mid-price in the following way:

$$\hat{s}_i = \frac{s_i}{p_i}.$$

Once the spread was predicted, the spread risk was calculated according to

$$r_s = s \cdot n_b \cdot (1 - P),$$

where s is the predicted spread, n_b is the number of buckets and $(1 - P)$ is the probability of trading aggressively in each bucket.

4.3.4 Market Risk

The second risk measure defined was the market risk.

The market risk can be seen as a measure of market volatility, meaning how much the market prices may change in the future. A high market risk may result in large price fluctuations, which means that there is a risk for both high and low prices in the future. The favorability of these price fluctuations is hard to predict as they can result in either a big profit or a big loss. Since a stable outcome with small variance is created by keeping the market risk down and avoiding large fluctuations it was assumed that a high market risk is unfavorable when trading.

In order to predict the market risk, the volatility σ_{t+1} for the next time step $t + 1$ was predicted and used to estimate future price changes. The volatility σ_{t+1} was estimated with the EWMA model. The EWMA model was chosen as it assigns relatively large weights to recent samples, which is favorable when working with high frequency data. Since the model assumes a normal distribution with zero-mean values, the null hypothesis of zero-mean was first tested for the minute-by-minute data of the log-returns using a simple t-test.

The ML method was used on the log-returns for the first five days of the data set to estimate the EWMA parameters. Thereafter volatilities were calculated according to equation (2.6) with the estimated parameters and σ_{t+1} was set to the volatility for the entire time period of the order. Possible future log-returns were simulated with the random walk hypothesis for the entire order period. To get the corresponding price p_i equation (4.5) was applied.

$$p_i = p_0 \cdot e^{\sum_{k=1}^{i} r_k}, \tag{4.5}$$

where p_0 is the initial price and r_k are the log-returns.

N number of possible future price paths were simulated in order to calculate VaR_α, which corresponds to the biggest possible loss such that the probability of the loss is less than or equal to the confidence level α. The market risk was modelled as VaR_α of the largest possible loss.

4.4 Model 4 - Strategy in trending markets with low probability of passive trading

The probability of succeeding with a passive trade in model 3 was based on an average from the empirical data and assumed the same for all days. However, in reality there are days during which the market experiences negative trends, resulting in a very low probability of succeeding with passive trades. The reason for this is that it is hard to find buyers in an upward trending market, and vice versa. In some cases orders are placed

on these less favourable days due to the need to convert between currencies as opposed to only trading speculatively with the objective of earning money. A way of beating the TWAP benchmark during a negative trend was therefore investigated.

This model applied the same way of generating the time series of trades as in model 3. However, if no passive trade was successful in three consecutive buckets the new strategy was to over hedge at the end of the third bucket such that no trades were needed in the following bucket. This was done by trading twice the trading size aggressively at the last second in the bucket. The reason for this was that the trend was assumed likely to continue into the subsequent bucket. The algorithm then reset for the buckets that followed.

Model 4 was compared to model 3 by executing both algorithms during a time interval experiencing a negative trend. The same time interval, low passive probability, order volume, trading size and bucket size was used for both models. Both model's profit were then calculated and compared.

Chapter 5

Results

This chapter highlights the key findings of the research, whilst supporting figures are captured in the appendix. Profit was defined as the difference between the TWAP benchmark and the average executed price for each order. The profits in model 3 and 4 were normalized by the total order volume. A negative profit meant that there was a loss in comparison to the TWAP benchmark. The total order volume was set to 250 millions for all three currency pairs and all orders, as this was assumed to be a reasonable order size based on common orders from clients.

5.1 Model 1

A number of fictive orders were executed in order to evaluate the performance of the local greedy model. 1 million was traded in each bucket, resulting in 250 buckets. EUR/SEK was investigated during a 5 hour time interval due to the trading restrictions for the currency pair in order to avoid market impact. The results are shown in table (5.1). EUR/NOK was investigated during a 6 hour time interval and EUR/USD during a 1 hour time interval.

Time interval	TWAP	Local greedy	Profit
08:00-13:00	9.40244	9.40245	0.00001
09:00-14:00	9.40276	9.40273	-0.00003
10:00-15:00	9.40368	9.40370	0.00002
11:00-16:00	9.40411	9.40412	0.00001
12:00-17:00	9.40385	9.40390	0.00005
13:00-18:00	9.40317	9.40319	0.00002

Table 5.1: EUR/SEK for various time intervals

The findings demonstrated that the local greedy model beat the TWAP benchmark the majority of the time for EUR/SEK and EUR/NOK. However, for EUR/USD the results were less definite as the profits were very small as well as negative on more occasions

than for the other currency pairs. The bucket sizes were calculated by dividing the total order time by the number of buckets. This resulted in significantly shorter buckets for EUR/USD compared to the other currency pairs, as the order time for EUR/USD was set to 1 hour compared to 5 and 6 hours for the other pairs. The model resembles the TWAP benchmark for shorter bucket sizes, hence why the profits were smaller for EUR/USD and the model only beat the TWAP benchmark half of the time.

5.2   Model 2

In order to trade in every bucket, the total order volume needed to be divided into smaller trading sizes. Equation (4.3) was used for different combinations of trading sizes and order quantities, resulting in table (5.2). The costs described in the earlier chapter were assumed to be minimum fees in the calculations such that they were the same for all trading sizes less than or equal to 1 million. From this table it was clear that the minimum size traded at each trade should be 0.4 millions in order to avoid a loss.

Order size, Millions	Trade size, Millions				
	0.2	0.3	0.4	0.5	1
1	-5	-2	0	1	3
10	-50	-17	0	10	30
50	-250	-83	0	50	150
100	-500	-167	0	100	300
250	-1250	-417	0	250	750

Table 5.2: Trading profit for various order and trading sizes

The maximum trading size of each trade was also restricted in order to avoid market impact. The maximum size traded at each trade was set to 1 million, as this is the lowest volume allowed on many platforms and it is close to the average size of one trade.

The average executed prices across all buckets were calculated for each of the N replicates of the orders evaluated for model 2. Different bucket sizes were evaluated and box plots were produced to illustrate how the mean and standard deviation of the average executed prices varied across different bucket sizes. The average executed prices were first normalized by division of the mean of each replicate in order to facilitate the comparison of the standard deviation. The prices were also normalized by the TWAP benchmark in order to compare the mean more easily. The prices were then multiplied by 100 to ensure stable results. The algorithm was executed for each hour and the maximum volume allowance was restricted as per table (3.3). The results for EUR/SEK between 13:00-14:00 are shown in figure (5.1).
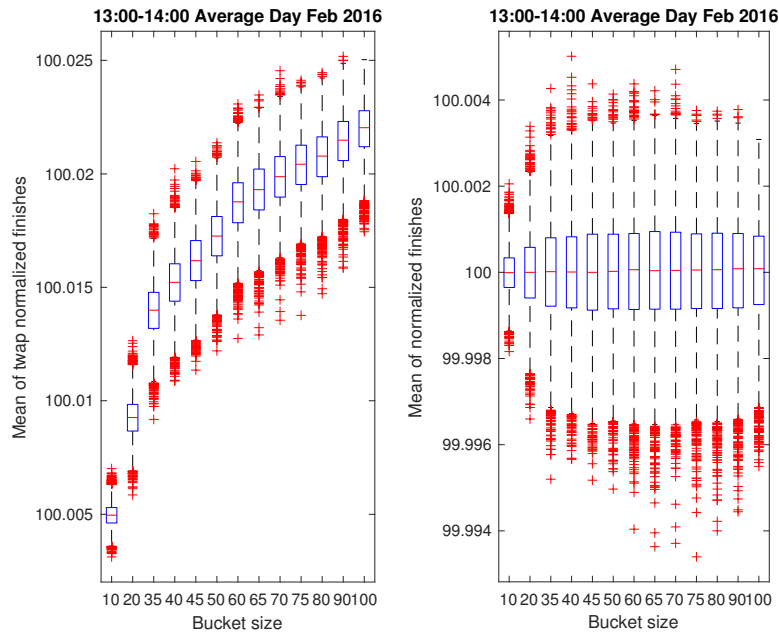
Figure 5.1: Box plot EUR/SEK 13:00-14:00

The average executed price increased for longer bucket sizes, which was arguably expected as this ensured more passive trades. The standard deviation on the other hand did not follow the same pattern. It initially increased and then fell as the bucket size became longer. This could also be expected as a short bucket size results in mostly aggressive trades and a long bucket size results in mostly passive trades, both of which result in lower standard deviations. For the bucket sizes that lie between these two cases the trades were both passive and aggressive resulting in a higher standard deviation.

The EUR/NOK results were the same as for EUR/SEK. In contrast, much shorter buckets were evaluated for EUR/USD and thus the standard deviation appeared to follow a different pattern. This was driven by the fact that for EUR/USD the passive probability for each bucket was significantly higher. The passive probability was in fact so high that the trades were rarely aggressive, even for very short buckets. This meant that for a short bucket there was a variety of passive and aggressive trades whilst all trades became passive for longer buckets, thus accounting for the difference in the pattern between the currencies.

These results indicate that the optimal bucket size is one long enough to ensure that all trades are passive. However, the market conditions also have to be taken into account when deciding on a bucket size. Trading in longer buckets implies trading for a longer time period, which in turn implies more risk that the market will change. The profit needs to be weighed against the risk in order to decide on a bucket size. This was investigated in model 3.

5.3   Model 3

5.3.1   Spread distribution

To determine a way to predict the spread at a certain time the distribution of the spread was investigated. The first 5 days of observations were used to examine the distribution of the data. This was done by taking all observed spreads between each hour for the first 5 days of observations and storing them in one vector, resulting in 18000 data points. The data for each hour was then examined using a histogram, QQ-plot, skewness and kurtosis. The results for the hours 08:00-09:00 for EUR/SEK are shown in figures (5.2) and (5.3). In table (5.3), the skewness and kurtosis for the same hour are presented for each currency.
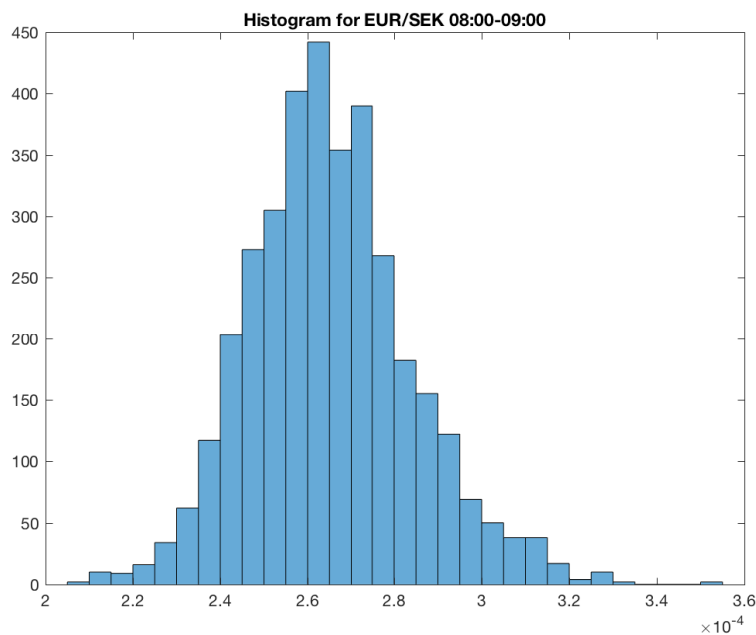
Figure 5.2: Histogram for EUR/SEK spread 08:00-09:00

Currency	Skewness	Kurtosis
EUR/SEK	0.3965	3.5902
EUR/NOK	0.0407	3.3587
EUR/USD	0.0993	3.2375

Table 5.3: Skewness and Kurtosis between 08:00-09:00

These illustrations show that the observed spread for all currencies was skewed to the right and had a high kurtosis. This meant that neither data was normally distributed. Furthermore, from the results of the QQ-plot and histogram, the EUR/USD spread appeared to be discrete. This is due to smaller price changes for EUR/USD compared to the other currency pairs, and that the bid and offer prices followed each other more closely. The data of all currency pairs did not appear to follow any standard distribution and therefore, as previously stated, in order to predict the spread without making assumptions
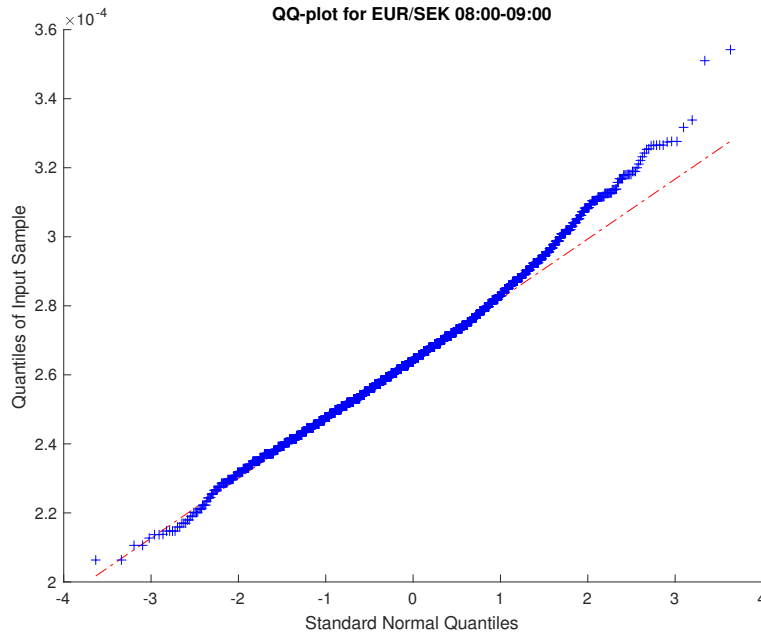
Figure 5.3: QQ-plot for EUR/SEK spread 08:00-09:00

about the distribution the spread was calculated by taking the mean of the 3600 observed spreads during the hour prior to the start time of the order.

5.3.2 Market Risk

The EWMA parameters needed to be estimated in order to calculate the market risk. This was done with the ML-method for the first five days of data during the time of 07:00-18:00, using log-returns on a minute basis. The lambdas for each currency pair can be found in Table (5.4). The results showed that the λ's for EUR/SEK and EUR/NOK were slightly higher than the λ for EUR/USD. A higher λ means that the volatility at the current time step t is more dependent upon recent volatilities than older volatilities.

Currency	λ
EUR/SEK	0.8973
EUR/NOK	0.8985
EUR/USD	0.8830

Table 5.4: Estimated λ for each currency pair

The minutes prior to the start time of the order were used to estimate the volatility. This meant that for a start time of 08:00 only 59 minutes of data could be used, potentially resulting in an inaccurate approximation of the volatility. In order to investigate if 59 minutes were sufficient, the volatility estimated for one of the days 6-10 using 59 minutes was compared to the volatility estimated using the whole forenoon consisting of 299 minutes of the previous day. For example, the estimated volatility using 59 minutes

on day 7 was compared to the estimated volatility for the whole forenoon of day 6 using 299 minutes. The difference between the estimated volatilities was very small and it was therefore concluded that 59 minutes of data was sufficient.

The volatility of each order was estimated using minute-by-minute data from 07:00 to the start time of the order. The volatility was then assumed to be constant during the whole order period. The volatility estimated for the start time of the order made it possible to simulate new log-returns and new prices with the random walk hypothesis. 10000 possible log-returns were simulated in order to receive a reliable result. $VaR_{0.05}$ was then calculated, using the last known price from the minute before the start time as the initial price and the simulated possible prices.

5.3.3   Executing the algorithm

A number of fictive orders were executed in order to evaluate the performance of the model. Six different probabilities of succeeding with a passive trade in a bucket, P, were investigated; $1\%, 5\%, 25\%, 50\%, 80\%$ and 90%. The purpose of the different P's was to present the client with a number of choices regarding the trade-off between passive and aggressive trading, and what each choice entailed regarding potential profits and risks. The level P was used to determine the bucket size, order time and risk factors, thus controlling the outcome of the algorithm. The trading sizes ranged between the lowest and highest allowed, 0.4 and 1 million. The number of replicates for each execution was set to $N = 10000$ and an average execution price was calculated for each replicate using the Monte Carlo method.

To calculate the profit, the N average execution prices were subtracted by the TWAP benchmark and normalized by division of the order volume, as per equation (5.1). This was done in order to allow the comparison of the models average execution prices and the TWAP average execution price in a dimensionless way. The upper and lower quantiles of 5% and 95% were then calculated. The results for one executed order are summarized in table (5.5), whilst table (5.6) shows the corresponding volumes and bucket sizes for each half hour.

$$\text{Profit} = \frac{\text{Average execution price} - \text{TWAP}}{\text{Total order volume}} \qquad (5.1)$$

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.01	[0.1184, 0.4478]	2.569	0.001	0.077	08:08
0.05	[0.3142, 0.7760]	3.525	0.002	0.074	08:12
0.25	[2.174, 3.123]	7.256	0.003	0.058	08:46
0.5	[4.444, 5.517]	8.182	0.005	0.039	09:52
0.8	[7.183, 8.015]	6.351	0.008	0.016	13:12
0.9	[7.924, 8.557]	4.786	0.009	0.008	15:07

Table 5.5: Results for EUR/SEK with start time 08:00 and trading size 1

P	Order volumes	Bucket sizes
0.01	250	2
0.05	250	3
0.25	150, 100	12, 10
0.5	69, 81, 60, 40	26, 22, 30, 34
0.8	30, 36, 26, 23, 24, 18, 18, 20, 20, 24, 11	60, 50, 68, 78, 74, 99, 100, 87, 87, 75, 70
0.9	21, 25, 18, 16, 17, 12, 12, 14, 14, 16, 18, 19, 20, 22, 6	85, 71, 97, 111, 105, 141, 142, 124, 124, 106, 99, 93, 88, 80, 72

Table 5.6: Example of order volumes and bucket sizes for different orders

To determine the actual return of the order, the non-normalized profit was multiplied by 250 million. One example of the profits converted to SEK is shown in table (5.7).

P	Profit (10^{-6})	Profit in SEK
0.01	[0.1184, 0.4478]	[7400, 27 988]
0.05	[0.3142, 0.7760]	[19 638, 48 500]
0.25	[2.174, 3.123]	[135 875, 195 190]
0.5	[4.444, 5.517]	[277 750, 344 810]
0.8	[7.183, 8.015]	[448 940, 500 940]
0.9	[7.924, 8.557]	[495 250, 534 810]

Table 5.7: Actual returns for EUR/SEK with start time 08:00 and trading size 1

When executing the orders it was found that for the lower probabilities, the bucket size was very small and thus the order time was very short. For instance, in the case of a probability of 1%, the whole order of 250 million was executed in 8 minutes. From table (3.3) it was clear that this was much more than the recommended amount and would imply market impact. A probability of 1% was therefore not recommended in this case. The highlighted rows in the tables indicate which orders were within or close to the recommended restrictions. There were fewer allowed probabilities for EUR/SEK and EUR/NOK as these markets are less liquid and there is a higher risk of market impact compared to EUR/USD. Furthermore, in the instance of high probabilities for smaller trading sizes the orders could not be completed within one day. For example, the

probabilities $P = 0.8$ and $P = 0.9$ for the trading size 0.4, found in table (A.3), could not be completed before 00:00 and are marked with " − " in the table. All the rows marked with " − " in the tables are orders that could not be completed before 00:00.

To ensure a sufficient quantity of data for evaluation, additional trading sizes were examined for EUR/SEK and EUR/NOK to compensate for the lack of possible choices in comparison to EUR/USD. Also, since the EUR/SEK and EUR/NOK markets are less liquid, a longer order time was required in order to complete an order compared to EUR/USD. As a result, fewer starting times were examined for EUR/SEK and EUR/NOK as the executions that were initiated later in the day took too long to finish.

The results showed that the profit increased monotonously when the probability of being passive increased. This was because a higher passive probability meant receiving favorable prices more often. When comparing the currency pairs to each other it was found that EUR/NOK and EUR/SEK reported higher profits than EUR/USD. This is because generally the spread for EUR/USD is smaller than the other currency pairs, and a smaller spread means a smaller profit when trading passively compared to aggressively.

It was also found that the standard deviation first increased with the probability of being passive and then decreased around the probability 0.5. This is due to the fact that around 0.5 half of the trades were passive and half aggressive, meaning there was a larger dispersion of the prices for the trades. A high standard deviation meant a greater uncertainty of the possible profits and not necessarily a higher risk, which is why the market risk and spread risk were calculated.

It was found that the market risk increased for longer order periods whereas the spread risk decreased. For smaller trading sizes the spread risk increased since a larger number of buckets/trades was required. The market risk varied depending on the start time as some hours experienced a higher estimated volatility, generating a larger spread of the simulated prices. When comparing the market risk for the different currency pairs to each other, the results showed that the market risk was lower for EUR/USD compared to EUR/SEK and EUR/NOK. The reason for this is that the orders for EUR/USD had a much shorter order time, which meant the market had less time to change.

When comparing the different trading hours that were examined for EUR/USD, the most significant difference was that the number of probability choices P varied during the day. This was driven by the variation in market liquidity throughout the day and during certain half hours it was not possible to sell the whole order volume. Notably, whilst predicted profits differed slightly for different times of the day there were no significant deviations. As suspected, the order took longer time to execute during the hours where there was a smaller probability of selling passively per second and hence the market risk was higher for those orders.

Orders with different trading sizes were executed in order to examine if there would be an increase in estimated profit when changing the trading size. As presented in the tables the estimated profits did not differ remarkably for the different trading sizes for all the currency pairs.

The results demonstrated that order time and probability were positively correlated, as

order time increased when probability increased. For instance, when P changed from 0.25 to 0.5 in (A.3) it took 3 hours and 30 minutes longer to execute the order. The reason that orders with high probabilities and small trading sizes could not be completed is that they continued to run at night time. Around 22:00 the probability of succeeding with passive trades for each second becomes one tenth smaller compared to during day time, resulting in longer bucket sizes compared to the day time bucket sizes. This is another reason as to why the trading hours for the algorithms were limited to 07:00-18:00 GMT.

For EUR/USD it was found that the end times were the same for the orders with the lower probabilities P. This is because the bucket sizes were rounded up so that the smallest bucket size was always 2 seconds. This in turn meant that the probability of trading passively in a bucket actually occurred with a higher probability than the one examined. For those cases the true P^* was calculated for the rounded bucket size by using equation (4.2) for $x = n - 1$. The true spread risk* was then calculated using P^*. One example of the difference in P and P^* is presented in table (5.8). For the orders that were completed during the same order time the P^* and spread risk were the same. The minimal probability of trading passively in a bucket for the order presented in table (5.8) was 32% as this corresponded to a bucket size of 2 seconds. This is a rather significant difference to for instance a probability of 1%. On the other hand, the P's and spread risks for the orders with larger bucket sizes were not as significant based on the results in table (5.8). Furthermore, for orders with longer order time over numerous half hours, with varying bucket sizes and therefore varying P^* values, an exact P^* could not be determined for the order as a whole. Therefore, the true P^*'s and spread risks were calculated for the orders that had the minimum bucket size of 2 seconds whereas the rest were considered accurate enough.

P	Spread risk	P*	Spread risk*	End time	Bucket size
0.01	0.046	0.32	0.031	08:08	2
0.05	0.044	0.32	0.031	08:08	2
0.25	0.035	0.32	0.031	08:08	2
0.5	0.023	0.54	0.021	08:12	3
0.8	0.009	0.85	0.007	08:25	6
0.9	0.005	0.90	0.005	08:29	7

Table 5.8: True P and spread risk for EUR/USD with start time 08:00 and trading size 1

5.4 Model 4

In order to accurately compare model 3 and model 4, they were run using the same factors; passive probability, order volume, bucket size and time interval. The days for the whole month were studied in order to choose the ones that had a clear trend. As few days had a clear downward sloping trend for EUR/SEK and EUR/NOK, buy orders were also executed over days with upward sloping trends. This was done to allow for the comparison of the models multiple times for each currency.

The idea behind the comparison was to evaluate the results when the probability of succeeding with passive trades p was lower than expected. In fact, the probability was

so low that hardly any passive trades were successful. The scenario can be seen as a stress test of model 3 compared to an alternative approach, model 4. The order volumes and bucket sizes were chosen based on the results from executing model 3. The level of willingness to trade passively was set to $P = 0.8$ for each order. Furthermore, each order was run for a number of probabilities of succeeding with a passive trade within the range $p \in [0.0001, 0.001]$. The probabilities were set this low in order to reflect the negative market trend.

As illustrated in figure (5.4) February 2nd experienced a clear upward sloping trend between 08:00-13:12, and was therefore studied for EUR/SEK. The order volumes and bucket sizes were chosen based on the results of the starting time 08:00 and trading size 1 for model 3, which can be found in table (5.6). The results of the two models profit are presented in table (5.9).

Bid and offer 2 February EUR/SEK
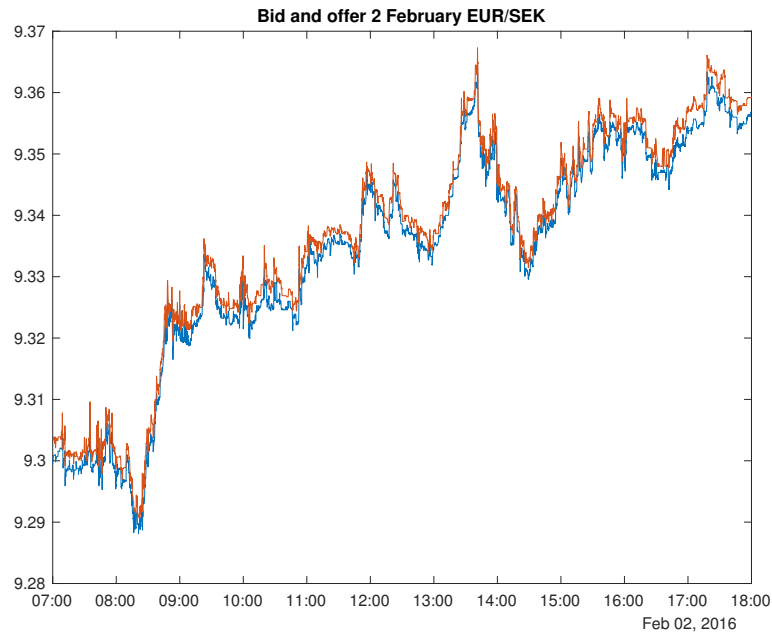
Figure 5.4: Day with an upward sloping trend EUR/SEK

p	Model 3 Profit(10^{-7})	Model 4 Profit(10^{-7})
0.0001	[-3.539, -1.640]	[-0.7197, 1.574]
0.0003	[-2.896, 0.5658]	[-0.7427, 3.195]
0.0005	[-1.957, 2.468]	[-0.4609, 4.579]
0.0007	[-0.9549, 4.217]	[0.09257, 5.959]
0.001	[0.5302, 6.728]	[1.163, 7.797]

Table 5.9: February 2nd with start time 08:00 EUR/SEK

The exact same execution was performed for EUR/SEK on February 25th, which showed an upward sloping trend.

The comparison for EUR/NOK was executed for February 17th, which showed a downward sloping trend, as well as on the 24th, which showed an upward sloping trend. The orders were run from 08:00-14:06, with corresponding order volumes and bucket sizes for each half hour. For EUR/USD, the 9th showed a clear upward sloping trend and an order was investigated for different start times during this day.

The results presented in the tables for EUR/SEK and EUR/NOK imply that model 4 always had a higher profit than model 3, when the probability of trading passively was low. It was also found that for EUR/USD, the results were not as coherent as for EUR/SEK and EUR/NOK. For the day investigated, it was discovered that for the start time 10:00, model 4 always had a higher profit than model 3. For the start time 14:00 it was found that model 3 had a higher profit apart from the upper quantile for $p = 0.001$. When analyzing the result from the start time 16:00, the lower quantile was higher for model 3 whereas the upper quantile was mostly higher for model 4. The reason for the high variation in the results is that the orders were executed in much shorter time intervals for EUR/USD. There was a negative overall trend during the day when the order was executed, yet the result of the execution was more dependent upon the local trend during the time interval of the order. Figure (A.10) illustrates that during the order time interval 14:00-14:20 the prices moved both up and down. It is therefore difficult to decide which model to use as it is impossible to predict exactly how the market will move during such a short time interval. For EUR/SEK and EUR/NOK the orders are executed over several hours and the choice of model is therefore easier. The overall result was that model 4 generally performed better for longer order periods.

Chapter 6

Discussion

6.1 Conclusions

This study found that both model 2 and model 3 outperform model 1, because the latter always trades aggressively. Model 3 is an improvement of model 2, as it both quantifies the risks of the choices regarding how to execute the order and takes the client's preferences into account. Thus, a conclusion can be made that model 3 is the favourable choice out of the first three models.

It is clear from the results that to determine one optimal bucket size is not possible. Instead it was found that the client's preferences regarding potential risks and profit should be the deciding factor in determining the optimal bucket size. It was also found that the bucket sizes should be allowed to vary during a day. If a client is risk averse, the chosen bucket size will be shorter as it was found that there is less market risk for shorter time intervals. However, this results in a higher spread risk as there is a lower probability of succeeding with passive trades. If a client on the other hand is less risk averse, the chosen bucket size will be longer, as it was found that a higher profit will be received when choosing a higher probability of being passive. However, since longer buckets imply a longer execution time a higher market risk will occur. In other words, a risk averse client looses the spread in favor of less market risk, whilst the less risk averse client earns the spread at the risk of market changes. The execution resembles the TWAP benchmark for short bucket sizes. Furthermore, it was discovered that model 3 always beats the TWAP benchmark on an average day with the average probability of succeeding with passive trades. This study therefore argues that the profits in actual returns for each of the currencies are significant enough to conclude that model 3 is a good strategy.

When comparing the currency pairs, the choices are more limited when using model 3 for EUR/SEK and EUR/NOK than for EUR/USD. This is due to the EUR/SEK and EUR/NOK markets being less liquid than the EUR/USD market. One restriction that affects EUR/SEK and EUR/NOK is that the order has to be completed before 18:00, which means that for late start times there are only a few possible ways to execute the order. In this thesis a limited amount of probability levels P were examined but by allowing other levels it is possible to extend the choices for all the currency pairs. Another limitation

to the possible choices is the allowed order volume, which affects all of the currencies. Overall there were more choices allowed for every order executed for EUR/USD compared to the other currency pairs. This means that clients that wish to trade EUR/USD will also have a choice about whether they want to trade during hours that are more or less liquid. It was found that for second-by-second data the smallest bucket size was 2 seconds and for EUR/USD this means that the lowest choice of the actual probability is $P^* = 0.22$. This in turn means that if the client wishes to have an even smaller probability of trading passively, the bucket sizes and the data need to be in microseconds.

The market risk and spread risk are higher for smaller trading sizes across all three currency pairs. This is due to the fact that smaller trading sizes imply a larger number of trades and thereby a larger number of buckets, which will lead to a longer order time. For this reason we do not recommend smaller trading sizes as it will result in more risk for the client. A larger amount of trades also implies a higher settlement cost, which although not included in the calculations of the profit or further examined in this thesis is another reason that we do not recommend the smaller trading sizes.

Model 4 can be seen as an extension of model 3, to be used in instances of a negative trend on the market. It was found that model 4 outperforms model 3 during such a time for EUR/SEK and EUR/NOK, whereas for EUR/USD the results are incoherent regarding which model has the best performance. This means that model 4 should be used instead of model 3 when there is a clear negative trend for EUR/SEK and EUR/NOK, whereas further research is required to determine if model 4 should be used for EUR/USD. In the results of model 4 it can also be seen that neither model 3 or 4 always outperform the TWAP benchmark for low probabilities. The low probabilities imply that hardly any passive trades are successful, meaning that the models trade similarly to the TWAP only less frequently. The potential reason for the TWAP performing better include that it captures all observed prices and among them all of the better prices, whereas model 3 and model 4 have less chance of receiving the best observed prices as they trade less frequently. However, this is not further investigated in this thesis.

The results and conclusions in this thesis are based on the most active trading hours in GMT-time. Further investigation would be required in order to establish if the models beat the TWAP benchmark during the remaining trading hours.

6.2   Method and Errors

The spread is used to predict the spread risk and is estimated by taking the average of the observed spreads during the hour prior to the start time of the order. The spread is then assumed constant over the entire time interval of the order. In reality the spread will vary during the time the order is executed to an extent, particularly for long time intervals such as 6 hours. This means that the spread risk could be over- or underestimated depending on how the market changes during the order. However, as previously mentioned it is impossible to predict how the market will change and therefore an assumption is necessary to enable the spread risk being taken into account when making a choice regarding how to execute an order.

Similarly, the volatility that is estimated with the EWMA method is also set constant for the whole order period. To assume that the volatility is constant the entire order period is not a realistic reflection of the market, especially not for EUR/SEK and EUR/NOK as they have order periods of several hours. However, as with the spread this is a necessary assumption in order to enable the client to choose how to execute the order with the estimated market risk in mind. The volatility estimation is slightly better for EUR/USD, since it has shorter order periods. A more accurate volatility estimation would be to update the volatility while executing the order, but as the volatility was used to get the sense of how the market risk differs for different levels of P, it was determined that the method applied was sufficient. If one would like to further develop model 3, it is possible to update the volatility for every second and have the option to alter P as the volatility changes.

Another uncertainty with the choice of method when modelling the market risk, is that the estimation of the λ's is done using the days 1-5, whilst the order is executed for the average day. The average day is less volatile since the local trends of each day is smoothed when taking the mean of the 10 days considered. Therefore, the estimation of the λ's leads to an overestimation of the market risk. However, since the market risk is estimated in order to give an indication how it varies for different choices of P and it is overestimated at all level's of P, the overestimation will not cause a less informed choice of P.

The results show that when the market risk is high, the spread risk is low and vice versa. In reality there are cases where the market liquidity decreases at the same time as the market volatility increases, causing both an increase in the spread risk and market risk. In this case it would be advisable to choose a short bucket size to avoid the risk associated with a longer order time. In addition, these cases imply a dependence between the spread risk and the market risk that is not modelled in the thesis.

Furthermore, the bucket sizes are calculated based on the probability P and then rounded up to the nearest whole second. This results in a true value of P^* that slightly differs from the P examined in this thesis and the same applies for the spread risk as it is calculated based on P. For the orders with the minimal bucket size of 2 seconds the true values of P^* were calculated as well as the true spread risk. However, for orders with longer bucket sizes the P and spread risk were considered accurate enough. In reality, for orders with several half hours and different bucket sizes, the P and spread risk will not remain the same throughout the entire order. This means that in some cases the P and spread risk may be under- or overestimated.

The probability p of succeeding with a passive trade each second for different half hours of the day is estimated using the average of empirical data and assumed the same for all days. In reality this probability differs for different days based on the current market trend. Therefore, calculations based on p may not always match the true values. Also, it is arguably optimistic to assume a very high passive probability P for an entire order period of several hours. This means that the estimated profits for the higher values of P might be slightly overestimated.

As mentioned in Chapter 4 the last seconds of the time interval that do not make up a whole bucket are not used for generating the trading times. As the trading times are generated in half-hour slots this means that instead of allowing a bucket to overlap between

two half-hours some seconds at the end of each half hour are not used. Therefore there are several seconds in each order where no passive trades are attempted and no aggressive trades are carried out that may in reality have resulted in favorable prices. By further developing model 3 so that there are no such breaks in an execution algorithm and by including these seconds, the orders would have been completed earlier than calculated in this thesis. This means that the orders could have been executed with less market risk.

6.3   Further Development

To further develop model 3 we suggest investigating if the model can be improved by updating the risk factors while the order is executed. This is arguably more relevant for EUR/NOK and EUR/SEK, as the orders run for several hours and updating the risks provides the possibility of changing strategy during the execution of the order to achieve better results.

Another theory that was never pursued in this thesis due to time constraints, was to create a market risk to spread risk ratio in order to quantify the comparison of the two risk factors. The aim would then be to investigate this ratio in order to find a balance of market risk and spread risk that is satisfiable for the client.

As mentioned earlier, model 4 needs to be further examined in order to determine when it should be in preference of model 3 for EUR/USD. Both model 3 and model 4 also need to be investigated in order to determine why they are outperformed by the TWAP benchmark for low probabilities. The purpose of this would be to develop the models so that they always beat the TWAP benchmark. We also think it would be valuable to look into whether model 3 and model 4 can be combined by switching between the two during one order if it is found that the probability is too low for model 3 to be effective.

Additonally, one way to further develop model 3 would be to allow more than one trade per bucket. If a passive trade is successful in a bucket it could be favorable to try for another passive trade within the same bucket as the current trend on the market is clearly positive. However, this requires another method for calculating the total probability of succeeding with a passive trade in the bucket and is therefore a rather extensive development of the model. Another way to develop model 3 would be to immediately close a bucket when a passive trade has been received and move on to the next bucket. With this approach it would be possible to shorten the total order time and reduce the market risk.

Finally, the model can be developed by not making the delimitations of this thesis described in chapter 1. This implies creating models for market liquidity and market impact in order to take these into account as well as allowing the volume traded in each bucket to change.

Appendix A

Appendix

A.1   Model 1

A.1.1   EUR/NOK

Time interval	TWAP	Local greedy	Profit
08:00-14:00	9.547306	9.547305	-0.0000001
09:00-15:00	9.54732	9.54734	0.00002
10:00-16:00	9.546542	9.546543	0.000001
11:00-17:00	9,54527	9,54532	0.00005
12:00-18:00	9.54315	9.543221	0.00007

Table A.1: EUR/NOK for various time intervals

A.1.2   EUR/USD

Time interval	TWAP	Local greedy	Profit
08:00-09:00	1.109600	1.109599	-0.000001
10:00-11:00	1.108112	1.108115	0.000003
12:00-13:00	1,107516	1,107517	0.000001
14:00-15:00	1,106174	1,106173	-0.000001
16:00-17:00	1,10676	1,10677	0.000001

Table A.2: EUR/USD for various time intervals

A.2   Model 2
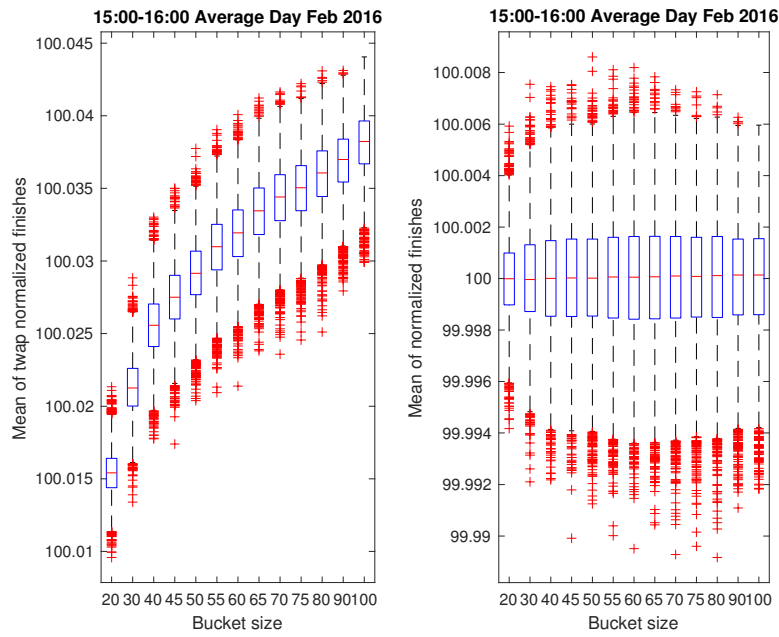
Figure A.1: Box plot EUR/NOK 15:00-16:00
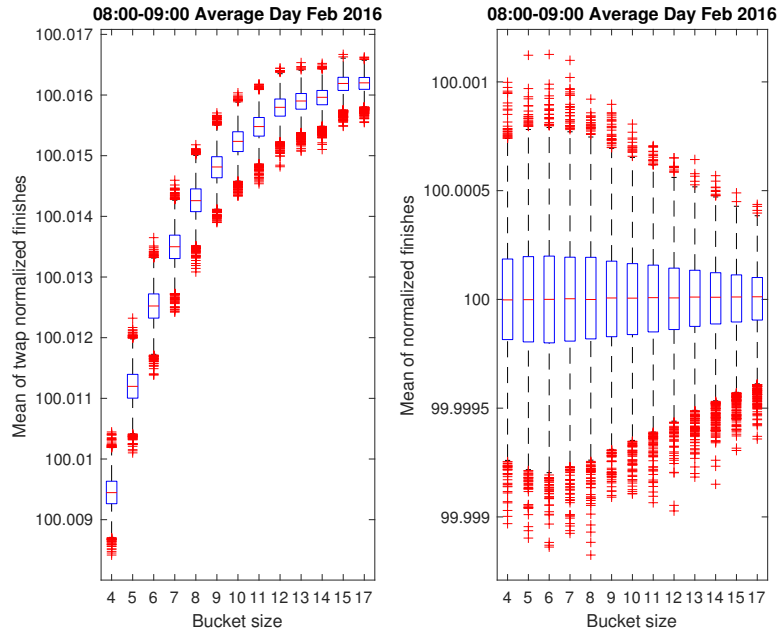
Figure A.2: Box plot EUR/USD 13:00-14:00

A.3   Model 3

A.3.1   Spread distribution
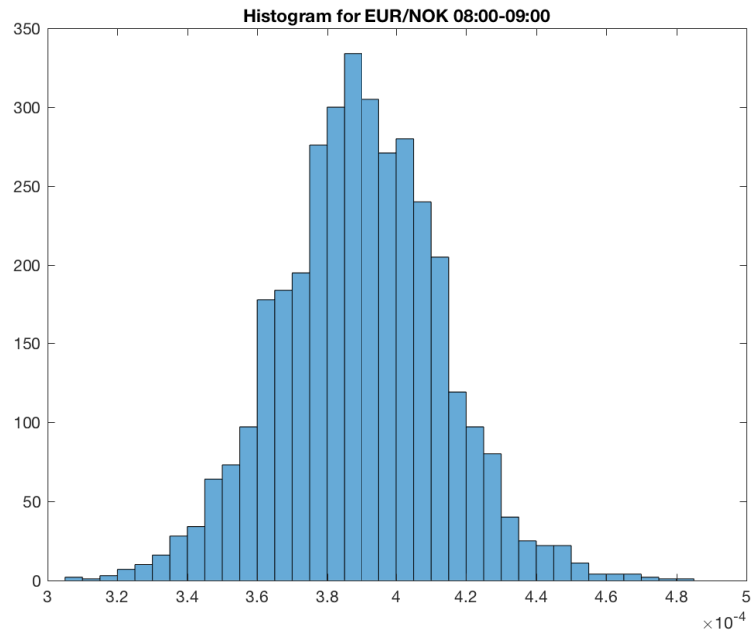
Figure A.3: Histogram for EUR/NOK spread 08:00-09:00
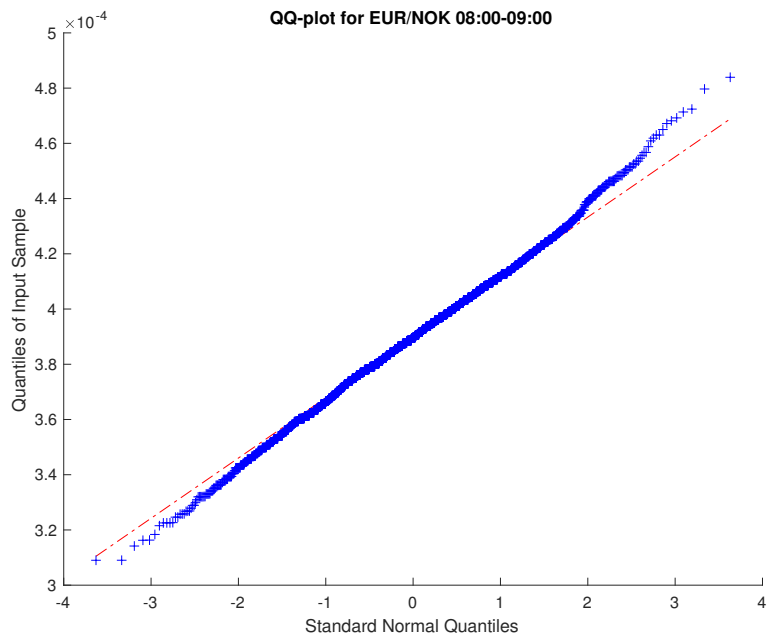
Figure A.4: QQ-plot for EUR/NOK spread 08:00-09:00
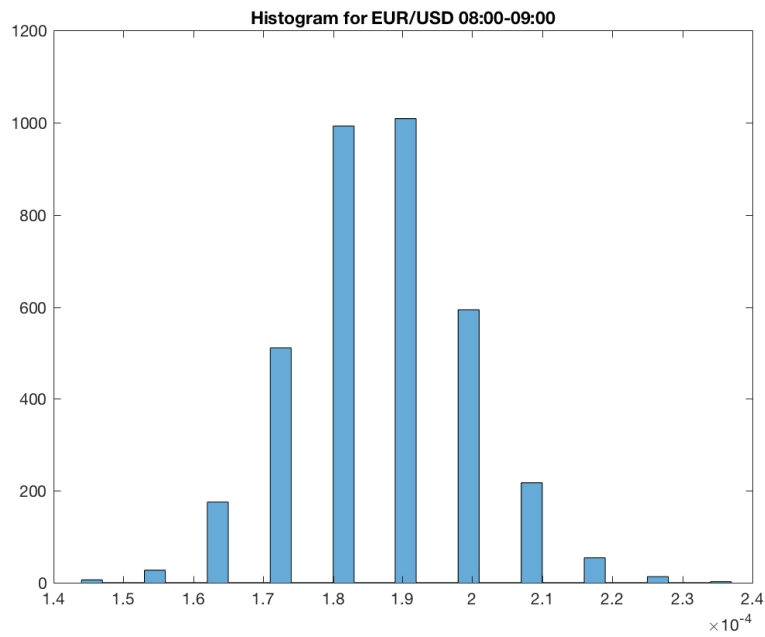
Figure A.5: Histogram for EUR/USD spread 08:00-09:00
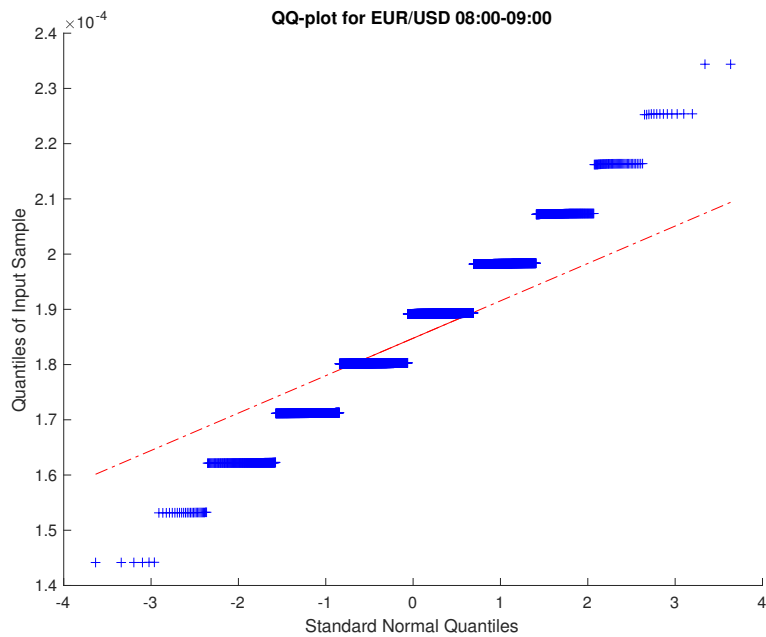
Figure A.6: QQ-plot for EUR/USD spread 08:00-09:00

A.3.2   EUR/SEK

When executing the orders it was found that some orders take too long to finish. The orders that are not done before 00:00 are therefore left blank in the tables of results.

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.01	[0.1808, 0.3964]	1.653	0.002	0.192	08:26
0.05	[0.0701, 1.224]	9.082	0.003	0.184	08:48
0.25	[2.256, 2.942]	5.240	0.005	0.146	10:21
0.5	[4.399, 5.056]	4.999	0.009	0.097	13:52
0.8	-	-	-	-	-
0.9	-	-	-	-	-

Table A.3: EUR/SEK with start time 08:00 and trading size 0.4

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.5	[4.564, 5.508]	7.169	0.005	0.048	10:28
0.8	[7.077, 7.808]	5.609	0.009	0.019	14:23
0.9	[7.968, 8.532]	4.302	0.010	0.010	16:20

Table A.4: EUR/SEK with start time 08:00 and trading size 0.8

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.5	[4.490, 5.703]	9.226	0.005	0.043	10:10
0.8	[7.138, 7.925]	5.992	0.009	0.017	14:46
0.9	[7.911, 8.514]	4.578	0.010	0.009	15:41

Table A.5: EUR/SEK with start time 08:00 and trading size 0.9

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.8	[6.909, 7.715]	6.128	0.013	0.013	14:25
0.9	[7.773, 8.389]	4.673	0.015	0.007	15:58

Table A.6: EUR/SEK with start time 09:00 and trading size 1

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.5	[4.272, 4.882]	4.665	0.013	0.083	14:55
0.8	-	-	-	-	-

Table A.7: EUR/SEK with start time 09:00 and trading size 0.4

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.8	[6.880, 7.587]	5.385	0.014	0.017	15:24
0.9	-	-	-	-	-

Table A.8: EUR/SEK with start time 09:00 and trading size 0.8

P	Profit (10^{-6})	Std (10^{-7})	Market risk	Spread risk	End time
0.8	[6.897, 7.646]	5.720	0.013	0.015	14:53
0.9	[7.861, 8.448]	4.456	0.015	0.007	16:34

Table A.9: EUR/SEK with start time 09:00 and trading size 0.9

A.3.3   EUR/NOK

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.01	[0.1382, 0.6199]	3.580	0.002	0.107	08:08
0.05	[0.6445, 1.423]	5.908	0.003	0.103	08:16
0.25	[2.922, 5.022]	15.98	0.007	0.081	09:03
0.5	[6.351, 8.100]	13.18	0.010	0.054	10:38
0.8	[10.43, 11.67]	9.560	0.016	0.022	14:06
0.9	[11.93, 12.91]	7.492	0.018	0.011	16:10

Table A.10: EUR/NOK with start time 08:00 and trading size 1

P	Profit (10^{-6})	Profit in NOK
0.01	[0.1382, 0.6199]	[8638, 38 744]
0.05	[0.6445, 1.423]	[40 281, 88 938]
0.25	[2.922, 5.022]	[182 625, 313 875]
0.5	[6.351, 8.100]	[396 940, 506 250]
0.8	[10.43, 11.67]	[651 875, 729 375]
0.9	[11.93, 12.91]	[745 625, 806 875]

Table A.11: Actual returns for EUR/NOK with start time 08:00 and trading size 1

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.01	[0.1926, 0.4820]	2.190	0.004	0.268	08:26
0.05	[0.6440, 1.149]	3.848	0.006	0.257	08:52
0.25	[3.268, 4.091]	6.249	0.011	0.203	10:59
0.5	[6.262, 7.678]	10.62	0.017	0.135	14:58
0.8	-	-	-	-	-
0.9	-	-	-	-	-

Table A.12: EUR/NOK with start time 08:00 and trading size 0.4

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.5	[5.891, 8.540]	23.34	0.012	0.068	11:36
0.8	[10.45, 11.55]	8.323	0.018	0.027	15:21
0.9	-	-	-	-	-

Table A.13: EUR/NOK with start time 08:00 and trading size 0.8

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.5	[6.089, 8.185]	15.95	0.0113	0.060	11:06
0.8	[10.45, 11.66]	9.187	0.017	0.024	14:40
0.9	-	-	-	-	-

Table A.14: EUR/NOK with start time 08:00 and trading size 0.9

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.8	[10.27, 11.50]	9.308	0.018	0.019	14:56
0.9	-	-	-	-	-

Table A.15: EUR/NOK with start time 09:00 and trading size 1

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.5	[6.369, 7.296]	7.034	0.019	0.122	15:26
0.8	-	-	-	-	-

Table A.16: EUR/NOK with start time 09:00 and trading size 0.4

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.8	[10.30, 11.48]	8.952	0.020	0.024	16:11
0.9	-	-	-	-	-

Table A.17: EUR/NOK with start time 09:00 and trading size 0.8

P	Profit (10^{-6})	Std (10^{-5})	Market risk	Spread risk	End time
0.8	[10.25, 11.42]	8.856	0.019	0.022	15:33
0.9	-	-	-	-	-

Table A.18: EUR/NOK with start time 09:00 and trading size 0.9

A.3.4 EUR/USD

In the tables, a * indicates spread risk calculated using P*.

P	P*	Profit (10^{-7})	Deviation (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.32	[2.170, 2.961]	5.965	1.552	0.031*	08:08
0.05	0.32	[2.170, 2.961]	5.965	1.552	0.031*	08:08
0.25	0.32	[2.170, 2.961]	5.965	1.552	0.031*	08:08
0.5		[3.929, 4.771]	6.407	1.922	0.023	08:12
0.8		[6.631, 7.237]	4.579	2.853	0.009	08:25
0.9		[7.068, 7.581]	3.903	3.027	0.005	08:29

Table A.19: EUR/USD with start time 08:00 and trading size 1

P	Profit (10^{-7})	Profit in USD
0.01	[2.170, 2.961]	[13 562, 18 506]
0.05	[2.170, 2.961]	[13 562, 18 506]
0.25	[2.170, 2.961]	[13 562, 18 506]
0.5	[3.929, 4.771]	[24 556, 29 819]
0.8	[6.631, 7.237]	[41 444, 45 231]
0.9	[7.068, 7.581]	[44 175, 47 381]

Table A.20: Actual returns for EUR/USD with start time 08:00 and trading size 1

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.32	[2.353, 2.847]	3.767	2.926	0.079*	08:26
0.05	0.32	[2.353, 2.847]	3.767	2.926	0.079*	08:26
0.25	0.32	[2.353, 2.847]	3.767	2.926	0.079*	08:26
0.5		[3.787, 5.182]	10.55	3.983	0.058	08:48
0.8		[6.612, 7.141]	4.031	5.211	0.023	09:21
0.9		[7.132, 7.945]	6.587	6.073	0.012	09:48

Table A.21: EUR/USD with start time 08:00 and trading size 0.4

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.22	[1.444, 2.158]	5.417	2.744	0.037*	10:08
0.05	0.22	[1.444, 2.158]	5.417	2.744	0.037*	10:08
0.25		[2.801, 3.637]	6.371	3.395	0.035	10:12
0.5		[3.860, 4.721]	6.526	4.021	0.023	10:16
0.8		[6.164, 7.209]	7.957	5.960	0.009	10:34
0.9		[7.113, 7.634]	3.983	6.944	0.005	10:48

Table A.22: EUR/USD with start time 10:00 and trading size 1

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.22	[1.572, 2.019]	3.379	5.090	0.091*	10:26
0.05	0.22	[1.572, 2.019]	3.379	5.090	0.091*	10:26
0.25		[2.292, 3.531]	9.487	6.944	0.088	10:48
0.5		[4.033, 4.629]	4.524	7.400	0.058	10:53
0.8		[6.439, 7.126]	5.228	11.00	0.023	11:49
0.9		[6.973, 7.428]	3.456	12.00	0.012	12:18

Table A.23: EUR/USD with start time 10:00 and trading size 0.4

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.24	[1.502, 2.176]	5.113	1.638	0.035*	12:08
0.05	0.24	[1.502, 2.176]	5.113	1.638	0.035*	12:08
0.25		[2.825, 3.600]	5.877	2.060	0.035	12:12
0.5		[3.878, 4.663]	5.947	2.411	0.023	12:16
0.8		[5.814, 6.452]	4.831	3.273	0.009	12:29
0.9		[6.752, 7.253]	3.840	3.917	0.005	12:40

Table A.24: EUR/USD with start time 12:00 and trading size 1

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.24	[1.623, 2.053]	3.289	3.119	0.088*	12:26
0.05	0.24	[1.623, 2.053]	3.289	3.119	0.088*	12:26
0.25		[2.098, 3.238]	8.608	4.219	0.087	12:48
0.5		[4.210, 4.759]	4.179	4.436	0.058	12:52
0.8		[6.010, 6.425]	3.150	5.618	0.023	13:21
0.9		[6.847, 7.131]	2.152	5.841	0.012	13:29

Table A.25: EUR/USD with start time 12:00 and trading size 0.4

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.38	[2.565, 3.364]	6.071	2.394	0.026*	14:08
0.05	0.38	[2.565, 3.364]	6.071	2.394	0.026*	14:08
0.25	0.38	[2.565, 3.364]	6.071	2.394	0.026*	14:08
0.5		[4.418, 5.219]	6.132	2.885	0.021	14:12
0.8		[6.418, 7.010]	4.497	3.888	0.009	14:20
0.9		[6.916, 7.407]	3.757	4.295	0.004	14:25

Table A.26: EUR/USD with start time 14:00 and trading size 1

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.38	[2.740, 3.244]	3.837	4.385	0.066*	14:26
0.05	0.38	[2.740, 3.244]	3.837	4.385	0.066*	14:26
0.25	0.38	[2.740, 3.244]	3.837	4.385	0.066*	14:26
0.5		[4.281, 5.545]	9.572	6.013	0.053	14:48
0.8		[6.536, 6.892]	2.688	6.444	0.021	14:56
0.9		[6.978, 7.539]	4.298	7.752	0.011	15:18

Table A.27: EUR/USD with start time 14:00 and trading size 0.4

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.40	[2.459, 3.208]	5.658	2.667	0.026*	16:08
0.05	0.40	[2.459, 3.208]	5.658	2.667	0.026*	16:08
0.25	0.40	[2.459, 3.208]	5.658	2.667	0.026*	16:08
0.5		[4.219, 4.952]	5.542	3.292	0.022	16:12
0.8		[6.108, 6.641]	4.039	4.406	0.009	16:20
0.9		[6.626, 7.054]	3.241	4.915	0.004	16:25

Table A.28: EUR/USD with start time 16:00 and trading size 1

P	P*	Profit (10^{-7})	Std (10^{-6})	Market risk (10^{-4})	Spread risk	End time
0.01	0.40	[2.462, 3.198]	5.530	5.041	0.065*	16:26
0.05	0.40	[2.462, 3.198]	5.530	5.041	0.065*	16:26
0.25	0.40	[2.462, 3.198]	5.530	5.041	0.065*	16:26
0.5		[4.222, 4.952]	5.544	6.884	0.054	16:48
0.8		[6.111, 6.633]	3.974	9.013	0.022	17:18
0.9		[6.623, 7.052]	3.273	9.351	0.011	17:28

Table A.29: EUR/USD with start time 16:00 and trading size 0.4
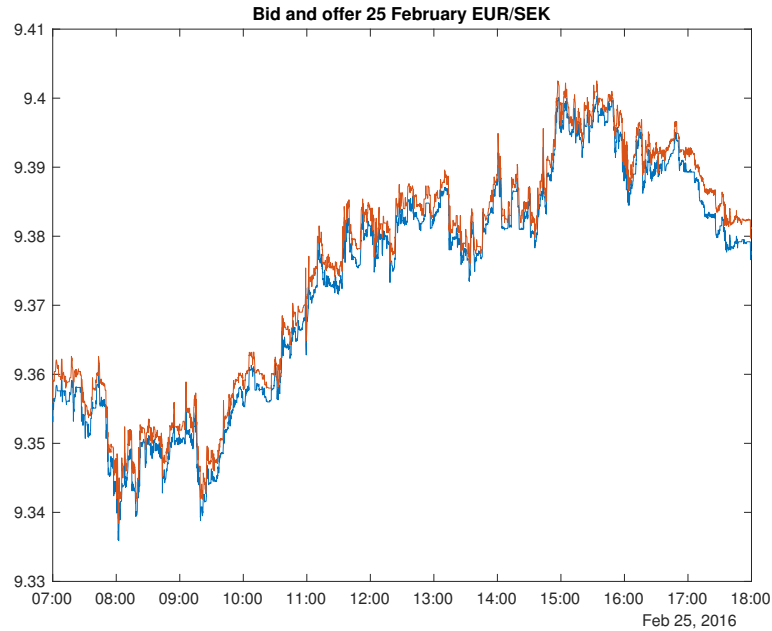
A.4 Model 4

A.4.1 EUR/SEK

Figure A.7: Second day with an upward sloping trend EUR/SEK

p	Model 3 Profit(10^{-7})	Model 4 Profit(10^{-7})
0.0001	[-4.471, -2.517]	[-1.329, 1.094]
0.0003	[-3.850, -0.2609]	[-1.339, 2.617]
0.0005	[-2.942, 1.714]	[-1.071, 3.977]
0.0007	[-1.876, 3.440]	[-0.5598, 5.214]
0.001	[-0.2446, 6.030]	[0.3064, 7.162]

Table A.30: 25 February with start time 08:00 EUR/SEK

A.4.2 EUR/NOK

For EUR/NOK the 17th and 24th are studied. The orders are executed from 08:00 to 14:06 and the corresponding order volumes each half hour are (24,20,25,19,16,13,16,17,18,27,25,24,6) and the bucket sizes are (73,90,71,94,106,129,107,101,95,65,72,74,69).
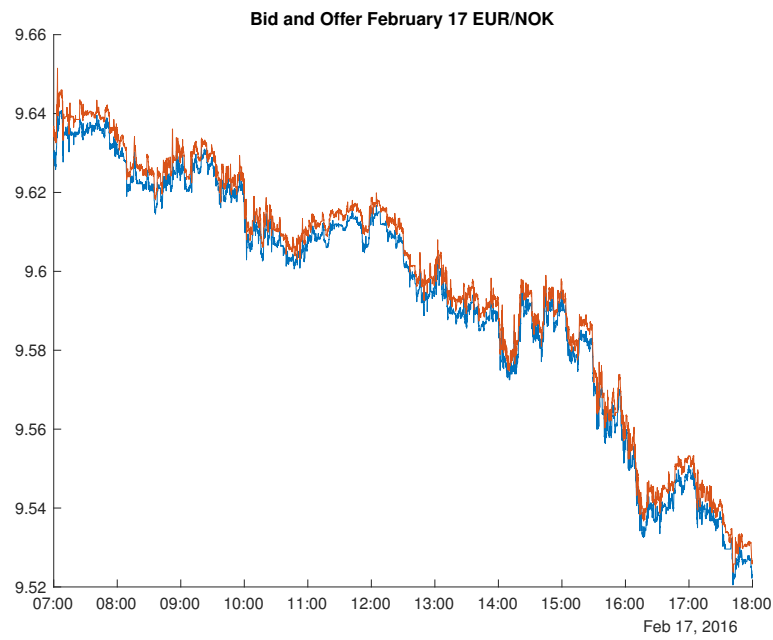
Figure A.8: Day with a downward sloping trend EUR/NOK

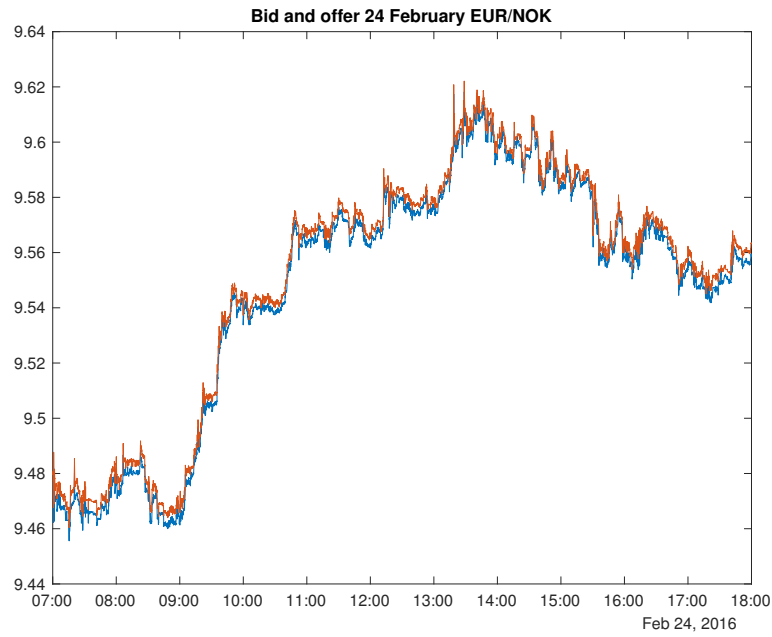|	Model 3 Profit(10^{-7})	Model 4 Profit(10^{-7})
p		
0.0001	[-5.228, -2.033]	[0.9932, 4.879]
0.0003	[-4.046, 1.845]	[1.055, 7.630]
0.0005	[-2.476, 5.164]	[1.907, 10.20]
0.0007	[-0.7446, 8.257]	[2.705, 12.40]
0.001	[1.975, 12.53]	[4.491, 15.42]

Table A.31: 17 February with start time 08:00 EUR/NOK

Figure A.9: Day with an upward sloping trend EUR/NOK

p	Model 3 Profit(10^{-7})	Model 4 Profit(10^{-7})
0.0001	[-10.07, -6.678]	[-0.1810, 6.311]
0.0003	[-8.938, -2.665]	[-0.8374, 8.980]
0.0005	[-7.278, 0.7214]	[-0.5060, 10.88]
0.0007	[-5.553, 3.877]	[0.1751, 12.64]
0.001	[-2.797, 8.282]	[1.610, 15.27]

Table A.32: 24 February with start time 08:00 EUR/NOK

A.4.3 EUR/USD

The time intervals, order volumes and bucket sizes of the orders executed are shown in the table below.

Time interval	Order volume	Bucket size
10:00-10:34	225, 25	8,10
14:00-14:20	250	5
16:00-16:20	250	5
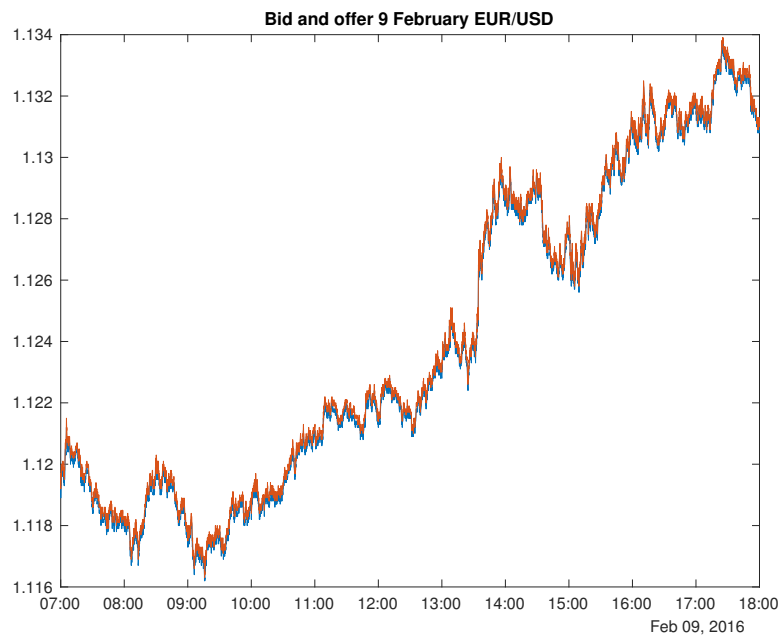
Figure A.10: Day with an upward sloping trend EUR/USD

p	Model 3 Profit(10^{-7})	Model 4 Profit(10^{-7})
0.0001	[-0.4221, -0.3954]	[-0.2265, -0.1288]
0.0003	[-0.4221, -0.3243]	[-0.2888, -0.1110]
0.0005	[-0.4221, -0.2532]	[-0.2976, -0.07542]
0.0007	[-0.4221, -0.2265]	[-0.2976, -0.04876]
0.001	[-0.4221, -0.1732]	[-0.3065, -0.03098]

Table A.33: 9 February with start time 10:00 EUR/USD

p	Model 3 Profit(10^{-7})	Model 4 Profit(10^{-7})
0.0001	[0.1521, 0.18413]	[-0.1199, -0.003987]
0.0003	[0.1521, 0.2001]	[-0.1199, 0.1201]
0.0005	[0.1521, 0.2161]	[-0.1199, 0.1681]
0.0007	[0.1521, 0.2161]	[-0.1199, 0.2161]
0.001	[0.1521, 0.2321]	[-0.1199, 0.2641]

Table A.34: 9 February with start time 14:00 EUR/USD

p	Model 3 Profit(10^{-7})	Model 4 Profit(10^{-7})
0.0001	[-0.1520, -0.1200]	[-0.2000, -0.2000]
0.0003	[-0.1520, -0.1040]	[-0.2640, -0.1360]
0.0005	[-0.1520, -0.08804]	[-0.3120, -0.04004]
0.0007	[-0.1520, -0.07204]	[-0.3280, 0.05596]
0.001	[-0.1520, -0.05604]	[-0.3440, 0.1040]

Table A.35: 9 February with start time 16:00 EUR/USD

Bibliography

[1] Bank for International Settlements. Triennial Central Bank Survey of foreign exchange and OTC derivatives markets in 2016; `http://www.bis.org/publ/rpfx16.html` (Retrieved 2017-05-10).

[2] Tim Weithers. *Trading Money, in Foreign Exhange: A Practical Guide to the FX Markets.* John Wiley & Sons, Inc., Hoboken, NJ, USA. 2012.

[3] Irene Aldrige. *High Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems.* John Wiley & Sons, Inc., Hoboken, NJ, USA. 2010.

[4] Barry Johnson. *Algoritmic Trading and DMA: An introduction to direct access trading strategies.* 4 Myeloma Press, London, Great Britian. 2010.

[5] University of Alabama in Huntsville; The Geometric Distribution; Introduction; `http://www.math.uah.edu/stat/bernoulli/Introduction.html` (Retrieved 2017-04-25).

[6] University of Alabama in Huntsville; The Geometric Distribution; Random; `http://www.math.uah.edu/stat/bernoulli/Geometric.html` (Retrieved 2017-04-25).

[7] Geof H. Givens, Jennifer A. Hoeting *Computational Statistics* John Wiley & Sons, Inc., Hoboken, NJ, USA. 2013.

[8] Art B. Owen *Monte Carlo theory, methods and examples* `http://statweb.stanford.edu/~owen/mc/Ch-intro.pdf` (Retrieved 2017-04-25).

[9] Peter J. Brockwell, Richard A. Davis *Introduction to Time Series and Forecasting* Springer-Verlag New York, USA. 2010.

[10] Erik Lindström, Henrik Madsen, Jan Nygaard Nielsen *Statistics for Finance* CRC Press, Taylor & Francis Group, Boca Raton, FL, USA. 2015.

[11] Kevin Dowd *Measuring Market Risk* John Wiley & Sons, Ltd., Hoboken, NJ, USA. 2005.

[12] Kent State University. One Sample t-test; `http://libguides.library.kent.edu/SPSS/OneSampletTest` (Retrieved 2017-05-10).

[13] Roel Oomen *Execution in an aggregator, Quantitative Finance* vol. 16 (2016). DOI: 10.1080/14697688.2016.1201589

[14] John C. Hull *Risk Management and Financial Institutions* John Wiley & Sons, Inc., Hoboken, NJ, USA. 2012.

[15] Minnesota State University; Five-Number Summary and Box-and-Whisker Plots; `http://web.mnstate.edu/peil/MDEV102/U4/S36/S365.html#Box` (Retrieved 2017-05-10).