Lund University
Department of Economics

Filtering techniques for asset allocation using a Discrete Time Micro-structure model: a comparative study

Henning Zakrisson

Master thesis, one year

supervised by
Birger Nilsson

June 5, 2017

Abstract

This paper is a comparative study of different approaches to using a Discrete Time Micro-structure model. By using the three filtering techniques Extended Kalman, Unscented Kalman and Bootstrap Particle, the hidden variables; excess demand and market liquidity, were estimated and used in an asset allocation strategy that invested in the asset when the excess demand as estimated as positive, due to the assumption that positive excess demand would make the price go up. Two different strategies were used—one based on threshold values of excess demand and one binary approach simply using the sign of the excess demand—to try to outperform a passive allocation strategy on 12 different stock indices. They were then evaluated in terms of average daily returns and market timing. The results showed favourable average daily returns for the Extended and Unscented Kalman filtering techniques using both kinds of strategies, though none of the results were statistically significant at the 5% confidence level. The Bootstrap Particle was deemed generally unreliable. The market timing tests rejected the null hypothesis of no market ability for most data sets using all three filtering techniques, with the two Kalman filters yielding the best results. Nothing was concluded about which filtering technique was superior, though the study indicates that Kalman filtering techniques can be used successfully in many cases while the Bootstrap Particle filter as used in this thesis is not reliable. The threshold-based strategy got slightly worse results in general than those of the binary approach, but this was tested without taking transaction costs into account.

Contents

1 Introduction

The dynamics behind asset prices has been studied for centuries, with the concept of technical analysis, or price prediction, being an ancient field of study (Lo and Hasanhodzic, 2011, p. 2). One common conception is that asset prices follow random walk processes (Fama, 1965). This is in line with the *efficient market hypothesis*, which basically stipulates that any inequality in information between traders is already incorporated into the price, and therefore cannot be predicted (Malkiel, 2003). Numerous attempts have been made to challenge this, for example by modeling asset prices as dynamic models dependent on different factors, or as augoregressive processes (Brooks, 2014, p. 614). A lot of traders today use technical analysis to predict the future prices and beat the market by buying when they believe the price will go up and selling when they think it is going down. They often base the price speculation on mathematical calculations of historical data such as the "head and shoulders"-method that looks for specific shapes of the price curve, or predicting the price as moving average processes (Lo and Hasanhodzic, 2011, p. viii). All these models are, in one way or another, challenging the random walk perspective.

With the increasing popularity of models from mathematical statistics, physics and engineering in the field of economics and finance, models working with filtering techniques from signal processing have become more common in the practical field of portfolio management (Drakakis, 2009). These methods, such as the Kalman filter or the particle filter, have been shown to yield promising results when applied in financial fields such as asset allocation strategies (Peng et al., 2016).

Another common conception is that of supply and demand as determinants of asset price—the notion that the market price can be derived from the intersection between the *supply* and *demand* curves, representing how much the market is ready to supply a good given a specific price and how much the market demands to buy the good at that price (Varian, 2010, p. 293). According to this theory, all deviations from this equllibrium, i.e. when the *excess* demand is non-zero, will lead to a price change until a new equilibrium is reached. Since market supply and demand are both generally unobservable variables for a traded asset, these models are mostly used as a theoretical framework rather than applied practically in trading strategies.

Based on a model proposed by Bouchaud and Cont (1998), Peng et al. (2003) created the Discrete Time Micro-structure model meant to closely model the price P as a nonlinear function. The model proposes that the price change depends on the unobserved variables *excess demand* ϕ and (inverse of) *marked depth* λ, a variable first introduced by Kyle (1985) meant to represent how much the price changes as an effect of the excess demand. The variable ϕ decides the sign of the price change and, together with the strictly positive λ, its magnitude. Peng et al. (2003) used this model and its state space representation (see appendix A for a short introduction of state space representation), with the hidden variables estimated using a Kalman filter and model parameters estimated using maximum likelihood parameters, in a asset allocation strategy, or portfolio strategy. The strategy based its prediction of the future price on its estimate of the excess demand variable ϕ—investing more when the excess demand was estimated positive and vice versa. Since the Discrete Time Micro-structure model—more thoroughly described in section 2.1—is non-linear, an Extended Kalman filter was used as a mean of linearization. The asset in this case was the currency pair USD/JPY, and Peng et al. (2003) managed to outperform a passive strategy in terms of gross return, even though no test was conducted for the significance of the portfolio performance.

Another student at Lund university, Strömberg (2006), later managed to reconstruct the results of Peng et al. (2003) on the two currency pairs SEK/USD and DKK/USD. A few other following studies have also been made to evaluate and develop the micro structure model, such as using local linearization to estimate the continuous version of the model, denoted the *Continous* Time Micro-structure model (Peng et al., 2005) or adding stochastic jumps (Peng et al., 2015).

Peng et al. (2016) recently argued that there were possible bias problems in using maximum likelihood parameter estimation on a discrete time model with estimated state variables, and therefore reformulated the Discrete Time Micro-structure model as a Self-organizing State Space type model—meaning a model able to estimate its own parameters—and applied it to an equivalent portfolio management strategy on Chinese index stocks using a Bootstrap Particle filter for estimation.

All articles Peng et al. (2003), Peng et al. (2005), Peng et al. (2015) and Peng et al. (2016) as well as the student paper Strömberg (2006) concluded that a portfolio strategy based on the different formulations of the model as introduced by Peng et al. (2003), using different filtering techniques are able to beat a passive strategy in terms of gross return, and the strategies are shown to accurately avoid price declines by allocating less money in the studied asset. This shows great promise in hidden state-form models for asset price prediction, and these kinds of models deserve more testing and evaluation. Peng et al. (2016) did not, however, explore any additional portfolio strategies for the same model formulation, nor did they compare the strategies application in different data sets. They also didn't test the excess returns in terms of statistical significance, nor did any tests of market timing, i.e. the ability to successfully enter and exit the market based on forecasts of returns—basically what the portfolio strategies aim to do. This kind of comparative study is performed in this thesis.

Unobserved variables in nonlinear state space models, such as the ones in the Discrete Time Micro-structure model described by Peng et al. (2003), have to be estimated using some kind of filtering technique to be useful in practice. *Linear* models are optimally recursively reconstructed, in a mean square error sense, using a linear Kalman filter (Jakobsson, 2013, p. 291), but nonlinear models like the ones mentioned above call for some kind of approximate filter. These filters all come with different strengths and weaknesses. Peng et al. (2016) discusses this and points out some problems using the Extended Kalman filter as done by Peng et al. (2003), such as a possibly biased maximum likelihood parameter estimation. To estimate the parameters and states, they consequently use a Bootstrap Particle filter instead, using a Self-organizing State Space formulation of the model. It is, however, not proven that this is the optimal filter technique for the model, even though Peng et al. (2016) got good results on one data set. There is reason to believe that other filters might also perform well or even better in portfolio strategy applications. One problem with the Self-organizing space state formulation of the Discrete Time Micro-structure model approach described by Peng et al. (2016) is that while it claims to estimate the model parameters using the particle filter, it relies on good initial distributions. Peng et al. (2016) suggests retrieving these from maximum likelihood estimation, why it could therefore be argued that the method doesn't really get around the problem of possibly erroneous initial parameter values.

One filter that Peng et al. (2016) has not tested is the *Unscented* Kalman filter, meant to be computationally less demanding than the particle filter, while retaining the ability to estimate the transition density of the hidden state rather than the transition function, by using a number of carefully chosen sample points (Wan and van der Merwe, 2000). While this model would also require parameters to be estimated using maximum likelihood as done in Peng et al. (2003), it could still be argued that the filter could suit this model well. Hence, the Unscented Kalman filter has been tested as an alternative filtering technique in this thesis.

This thesis is built mainly on the research of Peng et al. (2003) and the Discrete Time Micro-structure model, and applies some of the theory and method from the mentioned studies to make a comparative study of different filtering techniques and portfolio strategies, and search for more evidence regarding the application strength of these types of models. The two previously studied filters Extended Kalman and Bootstrap Particle were examined as well as the previously unstudied Unscented Kalman filter. The filters were used to estimate the current state of excess demand ϕ and take strategic portfolio actions accordingly, in the same manner as Peng et al. (2003). Also a new, simpler, kind of portfolio strategy without threshold values for ϕ was used based on the same estimated data of excess demand ϕ, but simply allocating all the money in the asset when $\phi > 0$ and none when $\phi < 0$, to more aggressively react on predicted future price declines. The filters and strategies were applied to price data from a number of stock indices, meant to represent market portfolios as suggested by Peng et al. (2016), and evaluated in order to determinate which filter, and which strategy, yielded the highest average daily return as well as which one has the best *market timing*, i.e. ability to accurately forecast when to invest in the market and when not to, using predictions of future returns. While these portfolio strategies' strength lies in timing their asset allocation with the market, no such tests seems to have been performed earlier.

The thesis means to contribute to this field of study by more thoroughly evaluating the different approaches of applying the Discrete Time Micro-structure model to portfolio strategies, by applying a number of filtering techniques and strategies to a series of data sets. The data sets in question are 12 stock indices from around the world. The main purpose is to test if the DTMS framework can be of use in portfolio management by testing the null hypotheses that portfolios built on the model assumptions and application do not yield significantly higher returns than a passive portfolio, and that they lack ability to time their asset allocation with market returns. Another purpose is to deduce whether or not there is one superior strategy when it comes to filtering technique and asset allocation approach, as well as to examine in what kind of markets the strategies work well and what markets they do not. Since the thesis does not claim to prove whether or not the real dynamics behind price changes really follow the Discrete Time Micro-structure model, its goal is rather to test if such model assumptions can be of use in portfolio strategy applications. It is difficult, if not impossible, to actually prove what dynamics determine price changes in an asset, but trying to estimate variables that can not be observed may still prove useful in trading, risk management and crisis prevention. In a broader perspective, these kinds of models can be applied in many areas of finance—for both private investors wanting a successful trading strategy, and for evaluating future risk by predicting price declines.

The rest of the paper is outlined as follows: Section 2, Theory, describes the previous research and theory used in the study. First, the Discrete Time Micro-structure model as described by Peng et al. (2003) is introduced, derived and explained, followed by the derivation of the Self-organizing space state-formulation of the same model from Peng et al. (2016). Then, the different filtering technique applications are described using the methodology of Jakobsson (2013), Wan and van der Merwe (2000) and Doucet et al. (2001) in order to to explain how Peng's model formulations can actually be estimated. Section 2.3 deals with the two different parameter estimation approaches that can be applied on the DTMS model using the above mentioned filtering techniques. Then, the portfolio strategy introduced by (Peng et al., 2003) is described, followed by measures of portfolio performance to be used in the comparative study. Section 3, Method, describes how the study was performed in terms of algorithms for model estimation, as well as which portfolio strategies was used. The data sets used for testing portfolio performance are also introduced here, as well as the comparison procedure used to evaluate the different strategies' performance. Section 4, Results, contains the results from the tests described in 3 Method. Section 5, Discussion, concludes the report by evaluating the results seen in 4, Results, discussing what these imply and suggesting future research on the topic.

2 Theory

2.1 The Discrete Time Micro-structure model

The Discrete Time Micro-structure model, henceforth denoted DTMS, was introduced by Peng et al. (2003). By arguing that it is necessary to deal with the dynamics of financial markets in many perspectives, a phenomenological model proposed by Bouchaud and Cont (1998) is used to follow the asset price dynamics. In that model,, P_t is the asset price with change described by

$$\mathrm{d}P_t = \lambda \phi_t \, \mathrm{d}t, \tag{2.1}$$

λ is the *inverse* of the marked depth, which is defined as the excess demand needed to push the price up by one unit (Bouchaud and Cont, 1998). The λ-variable has a clear connection to the work of Kyle (1985), who introduced "Kyle's λ", a common measure of market liquidity. Kyle (1985) also defined $\frac{1}{\lambda}$ as the market depth, but defined the market depth as *trade volume* needed to change the price with one unit, while here it is the excess demand needed to change the price with one unit. However, both formulations of λ follow the same intuition of being a quantification of the price elasticity. ϕ is the excess demand for the asset,

$$\phi_t = \phi_t^+ - \phi_t^-, \tag{2.2}$$

where ϕ_t^+ is the market demand and ϕ_t^- is the market supply. The idea behind the model is that a positive ϕ_t means a overvalued asset, which leads to an increased price, whereas a negative ϕ_t means it is undervalued, pushing the price down. The magnitude of the price change due to the market excess demand depends on the market depth, here quantified by parameter λ. The point is that a low market depth, i.e. a high value of λ, will make excess demand increase the price heavily, while a high depth means a small price change is sufficient to absorb the demand. Since this is merely an abstract model describing how one observable variable, P_t, is affected by two unobservable ones, ϕ_t and λ, the latter two have to be estimated to make the model useful in predicting market prices and or taking trading decisions. Therefore, Peng et al. (2003) uses a model proposed by Iino and Ozaki (2000), where P_t, ϕ_t and λ are described as continuous autoregressive processes

$$\mathrm{d}P_t = \lambda_t \phi_t + \lambda_t \, \mathrm{d}W_{1,t} \tag{2.3}$$

$$\mathrm{d}\phi_t = (\alpha_1 + \beta_1 \phi_{t-1}) + \gamma_1 \, \mathrm{d}W_{2,t} \tag{2.4}$$

$$\mathrm{dlog}\lambda_t = (\alpha_2 + \beta_2 \log \lambda_{t-1})\, \mathrm{d}t + \gamma_2\, \mathrm{d}W_{3,t} \tag{2.5}$$

where $W_{1,t}$, $W_{2,t}$ and $W_{3,t}$ are independent Wiener processes and α_i, β_i and γ_i for $i = \{1, 2\}$ are constant parameters. The goal of the model is to use filtering techniques to estimate the unobservable variables λ_t and ϕ_t. Note that the variable λ_t has also been added as a factor to the Wiener process in equation 2.3, to take the connection between market liquidity and price volatility into account. The DTMS model is then derived by Peng et al. (2003) by using Euler's discrete time approximation of equations 2.3-2.5 and can be described by

$$P_k = P_{k-1} + \lambda_{k-1}\phi_{k-1} + \gamma_3\lambda_{k-1}\xi_{1,k} \tag{2.6}$$

$$\phi_k = \alpha_1 + (1 + \beta_1)\, \phi_{k-1} + \gamma_1\xi_{2,k} \tag{2.7}$$

$$\log \lambda_k^2 = \alpha_2 + (1 + \beta_2) \log \lambda_{k-1}^2 + \gamma_2\xi_{3,k} \tag{2.8}$$

where $\xi_{i,k}$ for $i = \{1, 2, 3\}$ are independent standard gaussian white noise processes. Here, the parameter γ_3 has also been added to more closely model the relationship between market liquidity and asset price volatility. Note that this makes the price volatility become similar to that of the EGARCH model where the logarithmized volatility of the price follows a autoregressive process (Nelson, 1991). A simulated data set of the variables can be seen in figure 1. The variables of the DTMS model can be estimated using a number of techniques such as the Kalman filter (Peng et al., 2003) or the Bootstrap Particle filter (Peng et al., 2016). The theory behind these estimation methods are described in sections 2.2.1 and 2.2.2.

2.1.1 State space representation

To estimate the DTMS model variables using Kalman filtering as described in section 2.2.1, it has to be reformulated as a state space equation model with state-observability. For an introduction to state space-models, see appendix A. Peng et al. (2003) defines the hidden *state vector*

$$\boldsymbol{X}_k = \begin{bmatrix} P_k & \phi_k & \log \lambda_k^2 \end{bmatrix}^T \tag{2.9}$$

and the *observation vector* of outputs

$$\boldsymbol{Y}_k = \begin{bmatrix} P_k & \Delta P_k^* \end{bmatrix}^T \tag{2.10}$$

which follow processes

$$\boldsymbol{X}_{k+1} = \boldsymbol{A}\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right) \boldsymbol{X}_k + \boldsymbol{e}_{k+1} \tag{2.11}$$

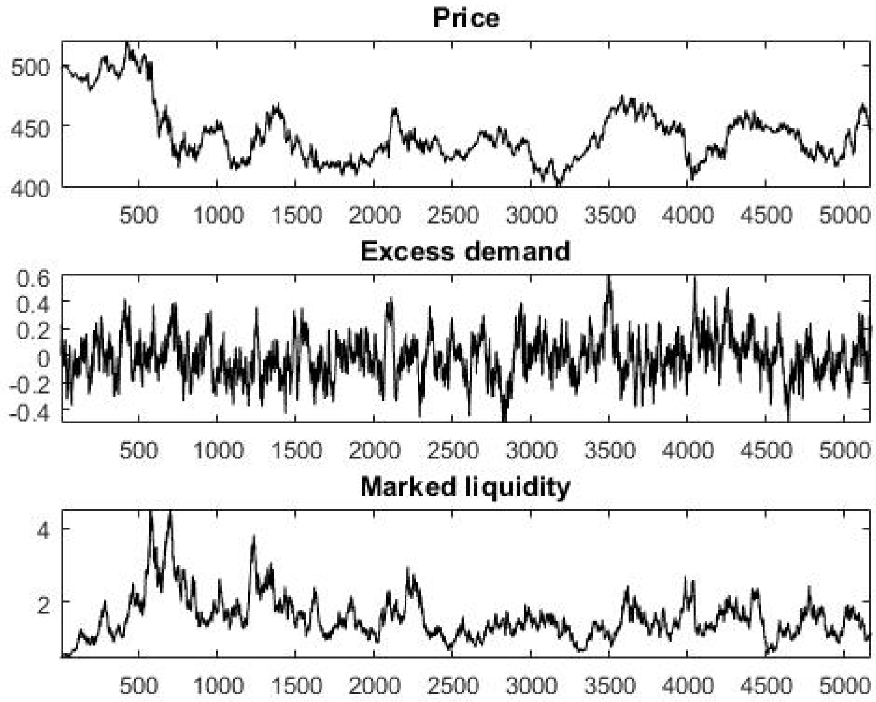

Figure 1: Simulated data from the DTMS model

and

$$\boldsymbol{Y}_k = \boldsymbol{C}\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right)\boldsymbol{X}_k + \boldsymbol{w}_k \tag{2.12}$$

where

$$\boldsymbol{e}_{k+1} \sim \mathcal{N}\left(0, \boldsymbol{S}_e\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right)\right) \tag{2.13}$$

$$\boldsymbol{w}_k \sim \mathcal{N}\left(0, \boldsymbol{S}_w\right) \tag{2.14}$$

$$\boldsymbol{A}\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right) = \begin{bmatrix} 1 & \lambda_k & 0 \\ \frac{\alpha_1}{P_k} & 1+\beta_1 & 0 \\ \frac{\alpha_2}{P_k} & 0 & 1+\beta_2 \end{bmatrix} \tag{2.15}$$

$$\boldsymbol{C}\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right) = \begin{bmatrix} 1 & 0 & 0 \\ \frac{\delta}{P_k} & 0 & 1 \end{bmatrix} \tag{2.16}$$

$$\boldsymbol{S}_e\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right) = \begin{bmatrix} \gamma_3^2\lambda_k^2 & 0 & 0 \\ 0 & \gamma_1^2 & 0 \\ 0 & 0 & \gamma_2^2 \end{bmatrix} \tag{2.17}$$

$$\boldsymbol{S}_w = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \tag{2.18}$$

and $\boldsymbol{\Theta}$ is a parameter vector consisting of $\{\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \gamma_3\}$. The variables σ_1 and σ_2 in matrix $\boldsymbol{S}_w$ are constants. The observable variable ΔP_k^* in equation 2.10 is added to make this model state-observable, i.e. making the state vector $\boldsymbol{X}_k$ estimatable (Jakobsson, 2013, p. 284) as suggested by Peng et al. (2003) and represents conditional variance. It is derived by first taking the square of both sides of equation 2.3 and ignoring higher order terms, yielding

$$(\mathrm{d}P_t)^2 = \lambda_t^2 \,\mathrm{d}t \tag{2.19}$$

which discretized and logarithmized becomes

$$\log (P_{k+1} - P_k)^2 \approx \log \lambda_k^2. \tag{2.20}$$

Since $(P_{k+1} - P_k)$ may be negative, the problem of logarithmizing zeros may arise. Fuller (1996) therefore suggests using the approximation

$$\log (P_{k+1} - P_k)^2 \approx \log \left((P_{k+1} - P_k)^2 + \eta\sigma_P^2\right) - \frac{\eta\sigma_P^2}{(P_{k+1} - P_k)^2 + \eta\sigma_P^2} = \Delta P_{k+1}^* \tag{2.21}$$

where η is a constant set to $\eta = 0.2$ and σ_P^2 is the sample variance of the asset price. The parameter δ has been added in matrix $\boldsymbol{C}(\boldsymbol{X}_k|\boldsymbol{\Theta})$ in equation 2.16 to adjust for the bias from the approximations in 2.20 and 2.21. The parameter vector $\boldsymbol{\Theta}$ can be estimated using, for example, maximum likelihood estimation (Peng et al., 2003).

2.1.2 Self-organizing State Space representation

The Self-organizing State Space model, henceforth denoted SOSS, was introduced by Kitagawa (1998), where "self-organizing" refers to the model's property to estimate its own parameters without the need of out-of-the-loop parameter estimation such as maximum likelihood. Peng et al. (2016) used a SOSS model reformulation of DTMS to more accurately estimate it in terms of predicting future market behaviour using current excess demand estimations, $\hat{\phi}_{k|k}$. First, the state space model described in equations 2.11 and 2.12 is rewritten in a generalized form

$$\boldsymbol{X}_{k+1} = \boldsymbol{f}(\boldsymbol{X}_k, \boldsymbol{e}_{k+1}, \boldsymbol{\Theta}) \tag{2.22}$$

$$\boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{w}_k, \boldsymbol{\Theta}). \tag{2.23}$$

A Bootstrap Particle filter, described in 2.2.2 could then be used to estimate the states, but the sampling error could make the maximum likelihood estimation of the parameters $\boldsymbol{\Theta}$ biased, if not a big amount particles are used, which would require a lot of computational power (Peng et al., 2016). To get around this problem, and to not use maximum likelihood estimation on a possibly biased estimate, the parameters are instead included in the state to make them estimateable *in* the filter rather than *outside* of it.

$$Z_k = \begin{bmatrix} \boldsymbol{X_k} \\ \boldsymbol{\Theta} \end{bmatrix} \tag{2.24}$$

is introduced by Peng et al. (2016), as well as transition functions

$$\boldsymbol{Z}_{k+1} = \begin{bmatrix} \boldsymbol{f}\left(\boldsymbol{X_k}, \boldsymbol{e}_{k+1}, \boldsymbol{\Theta}\right) \\ \boldsymbol{\Theta} \end{bmatrix} \tag{2.25}$$

$$\boldsymbol{Y}_k = \boldsymbol{H}\left(\boldsymbol{Z}_k, \boldsymbol{w}_k, \boldsymbol{\Theta}\right) = \boldsymbol{h}\left(\boldsymbol{X}_k, \boldsymbol{w}_k, \boldsymbol{\Theta}\right). \tag{2.26}$$

In this model formulation, the variables $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$ follow multivariate Gaussian distributions

$$\boldsymbol{X}_{k+1} \sim \mathcal{N}\left(\boldsymbol{A}\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right)\boldsymbol{X}_k, \boldsymbol{S}_e\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right)\right) \tag{2.27}$$

$$\boldsymbol{Y}_k \sim \mathcal{N}\left(\boldsymbol{C}\left(\boldsymbol{X}_k|\boldsymbol{\Theta}\right)\boldsymbol{X}_k, \boldsymbol{S}_w\right) \tag{2.28}$$

while $\boldsymbol{\Theta}$ is still a constant vector of parameters. Using this, the states and parameters can be estimated together recursively at every time step using probability densities derived from the DTMS model and a Bootstrap Particle filter as described in section 2.2.2.

2.2 Filtering techniques

2.2.1 The Kalman filter

The Kalman filter was introduced by Kalman (1960). It is a commonly used filter in signal processing and mathematical statistics. The following section does not aim to explain the derivation of the Kalman filter, but to shortly describe the estimation method using terminology and explanations from Jakobsson (2013). A linear state space representation of a discrete dynamic system with observable m-dimensional measurement vector $\boldsymbol{Y}_k$ and unobservable n-dimensional state vector $\boldsymbol{X}_k$ is described as

$$\boldsymbol{X}_k = \boldsymbol{A}\left(\boldsymbol{\Theta}\right)\boldsymbol{X}_{k-1} + \boldsymbol{e}_k \tag{2.29}$$

$$\boldsymbol{Y}_k = \boldsymbol{C}\left(\boldsymbol{\Theta}\right)\boldsymbol{X}_k + \boldsymbol{w}_k. \tag{2.30}$$

The matrices $\boldsymbol{A}\left(\boldsymbol{\Theta}\right)$ and $\boldsymbol{C}\left(\boldsymbol{\Theta}\right)$ are $(n \times n)$- and $(m \times m)$-matrices dependent on a set of known parameters $\boldsymbol{\Theta}$ and

$$\boldsymbol{e}_k \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{S}_e\right) \tag{2.31}$$

$$\boldsymbol{w}_k \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{S}_w\right) \tag{2.32}$$

(Jakobsson, 2013, p. 293). The system is assumed to be asymptotically stable. The optimal prediction of the state vector $\boldsymbol{X}_{k+l}$ in terms of mean square errors, l steps into the future is given by

$$\hat{\boldsymbol{X}}_{k+l} = E\left\{\boldsymbol{X}_{k+l}|\boldsymbol{Y}^k\right\} \tag{2.33}$$

where $\boldsymbol{Y}^k = \left[\begin{array}{cccc} \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \dots & \boldsymbol{Y}_k \end{array}\right]$ is a matrix containing all observations of $\boldsymbol{Y}$ up to and including point k (Jakobsson, 2013, p. 290). The optimal current state reconstruction is given by

$$\hat{\boldsymbol{X}}_{k|k} = E\left\{\boldsymbol{X}_k|\boldsymbol{Y}^k\right\} = \hat{\boldsymbol{X}}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\Gamma}_k \tag{2.34}$$

where the *Kalman gain* $\boldsymbol{K}$ is given by

$$\boldsymbol{K}_k = \boldsymbol{S}^{x,y}_{k|k-1}\left[\boldsymbol{S}^{y,y}_{k|k-1}\right]^{-1} \tag{2.35}$$

and the prediction error is given by

$$\boldsymbol{\Gamma}_k = \boldsymbol{Y}_k - \hat{\boldsymbol{Y}}_{k|k-1} \tag{2.36}$$

using the covariances

$$\boldsymbol{S}^{x,y}_{k|k-1} = C\left\{\boldsymbol{X}_k, \boldsymbol{Y}_k|\boldsymbol{Y}^{k-1}\right\} \tag{2.37}$$

$$\boldsymbol{S}^{y,y}_{k|k-1} = V\left\{\boldsymbol{Y}_k|\boldsymbol{Y}^{k-1}\right\} \tag{2.38}$$

(Jakobsson, 2013, p. 292). The point of the Kalman gain is to be a factor that is higher the less tolerant the system is to accepting prediction errors and lower the more tolerant it is. It is therefore formed by looking at the covariance between state and observation $\boldsymbol{S}^{x,y}_{k|k-1}$ as well as the inverse of autocovariance of the observation $\boldsymbol{S}^{y,y}$. The logic behind this is that the higher the covariance between the state and observation, the more accurate one would expect the prediction to be, while a high autocovariance in the observation could explain big prediction errors simply as stemming from high variance noise term realizations, making it plausible not to change the state estimate drastically. The covariance for the state reconstruction error, $\boldsymbol{S}^{x,x}_{k|k}$, can then be calculated as

$$\boldsymbol{S}^{x,x}_{k|k} = \boldsymbol{S}^{x,x}_{k|k-1} - \boldsymbol{K}_k\boldsymbol{S}^{y,y}_{k|k-1}\boldsymbol{K}_k^T. \tag{2.39}$$

Then, by using equations 2.29 and 2.30, the optimal linear reconstruction $\hat{\boldsymbol{X}}_{k|k}$ can be computed as

$$\hat{\boldsymbol{X}}_{k|k} = \hat{\boldsymbol{X}}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\Gamma}_k \tag{2.40}$$

using the observation prediction error

$$\boldsymbol{\Gamma}_k = \left(\boldsymbol{Y}_k - \boldsymbol{C}\left(\boldsymbol{\Theta}\right)\hat{\boldsymbol{X}}_{k|k-1}\right) \tag{2.41}$$

and the Kalman gain

$$\boldsymbol{K}_k = \boldsymbol{S}^{x,x}_{k|k-1}\boldsymbol{C}(\boldsymbol{\Theta})\left[\boldsymbol{S}^{y,y}_{k|k-1}\right]^{-1} \tag{2.42}$$

and the one step prediction $\hat{\boldsymbol{X}}_{k+1|k}$ can be computed as

$$\hat{\boldsymbol{X}}_{k+1|k} = \boldsymbol{A}(\boldsymbol{\Theta})\,\hat{\boldsymbol{X}}_{k|k} \tag{2.43}$$

(Jakobsson, 2013, p. 293). The covariance matrices can be computed as

$$\boldsymbol{S}^{x,x}_{k|k} = \boldsymbol{S}^{x,x}_{k|k-1} - \boldsymbol{K}_k\boldsymbol{S}^{y,y}_{k|k-1}\boldsymbol{K}_k^T = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}(\boldsymbol{\Theta}))\,\boldsymbol{S}^{x,x}_{k|k-1}, \tag{2.44}$$

$$\boldsymbol{S}^{x,x}_{k+1|k} = \boldsymbol{A}(\boldsymbol{\Theta})\,\boldsymbol{S}^{x,x}_{k|k}\boldsymbol{A}(\boldsymbol{\Theta})^T + \boldsymbol{S}_e, \tag{2.45}$$

and

$$\boldsymbol{S}^{y,y}_{k|k} = \boldsymbol{C}(\boldsymbol{\Theta})\,\boldsymbol{S}^{x,x}_{k+1|k}\boldsymbol{C}(\boldsymbol{\Theta})^T + \boldsymbol{S}_w. \tag{2.46}$$

As initial conditions $\hat{\boldsymbol{X}}_{0|0}$ and $\boldsymbol{S}^{x,x}_{1|0}$, some arbitrary but well suited values should be chosen, preferably as sample mean and covariances from previous realizations (Jakobsson, 2013, p. 294). A simple example of Kalman filter application on a linear problem, from Jakobsson (2013), can be seen in appendix A.2.

2.2.1.1 The Extended Kalman filter

If the state space representation of a system is *not* linear, equations 2.29 and 2.30 are not sufficient and can be replaced by equations

$$\boldsymbol{X}_{k+1} = \boldsymbol{F}(\boldsymbol{X}_k, \boldsymbol{\Theta}, \boldsymbol{e}_k) \tag{2.47}$$

$$\boldsymbol{Y}_k = \boldsymbol{H}(\boldsymbol{X}_k, \boldsymbol{\Theta}, \boldsymbol{w}_k) \tag{2.48}$$

where $\boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\Theta})$ and $\boldsymbol{H}(\boldsymbol{X}, \boldsymbol{\Theta})$ are possibly non-linear multivariate functions (Wan and van der Merwe, 2000). By using Taylor expansion of the functions $\boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\Theta})$ and $\boldsymbol{H}(\boldsymbol{X}, \boldsymbol{\Theta})$, the first order approximation of equations 2.47 and 2.48 can be made using equations 2.29 and 2.30 but with $\boldsymbol{A}$ and $\boldsymbol{C}$ replaced with the Jacobians

$$\boldsymbol{A} = \left.\frac{\partial\boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\Theta}, \boldsymbol{e}_k)}{\partial\boldsymbol{X}}\right|_{\boldsymbol{X}=\hat{\boldsymbol{X}}_{k|k}} \tag{2.49}$$

$$\boldsymbol{C} = \left.\frac{\partial\boldsymbol{H}(\boldsymbol{X}, \boldsymbol{\Theta}, \boldsymbol{w}_k)}{\partial\boldsymbol{X}}\right|_{\boldsymbol{X}=\hat{\boldsymbol{X}}_{k|k}} \tag{2.50}$$

The reconstruction $\hat{\boldsymbol{X}}_{k|k}$ and the one step prediction $\hat{\boldsymbol{X}}_{k+1|k}$ can then be calculated using the Kalman filtering algorithm described in section 2.2.1, but replacing equation

2.41 with

$$\boldsymbol{\Gamma}_k = \boldsymbol{Y}_k - \boldsymbol{H}\left(\hat{\boldsymbol{X}}_{k|k-1}, \boldsymbol{\Theta}, \boldsymbol{0}\right) \tag{2.51}$$

and equation 2.43 with

$$\hat{\boldsymbol{X}}_{k+1|k} = \boldsymbol{F}\left(\hat{\boldsymbol{X}}_{k|k}, \boldsymbol{\Theta}, \boldsymbol{0}\right) \tag{2.52}$$

respectively, and replacing the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ with their Jacobian counterparts from equations 2.49 and 2.50 respectively in the other algorithm equations (Wan and van der Merwe, 2000). Note that this is not necessarily the optimal reconstruction of the state variable $\boldsymbol{X}_k$, while the Kalman filter in the linear case is (Jakobsson, 2013). This is due to the fact that the Extended Kalman Filter estimates the nonlinear step-wise transition function as a linear function using the Jacobian as a transition matrix (Wan and van der Merwe, 2000). The point of this linearization is to capture the true conditional mean of the transition distribution, making it a "first order-linarization", which may lead to large errors in the true posterior mean and covariance of the state variable (Wan and van der Merwe, 2000). For the algorithm of estimating the states of the DTMS model using an Extended Kalman filter, see algorithm 1 on page 24.

2.2.1.2 The Unscented Kalman filter

To cope with the possible issues of the Extended Kalman filter linearization, the Unscented Kalman filter was introduced by Wan and van der Merwe (2000). By using a number of sample points, the goal of the filter is to capture the true mean *and* covariance of the system with accuracy up to the 2nd order and avoid computing Jacobians (Wan and van der Merwe, 2000). The unscented transform of the state variable $\boldsymbol{X}_{k|k}$ in the system described by equations 2.47 and 2.48 is done by creating a $(n \times (2n+1))$ matrix $\boldsymbol{\chi}_{k|k}$ with columns i

$$\boldsymbol{\chi}_{i,k|k} = \begin{cases} \hat{\boldsymbol{X}}_{k|k}, & i = 1 \\ \hat{\boldsymbol{X}}_{k|k} + \sqrt{(n+\eta)}\left(\sqrt{\boldsymbol{S}^{x,x}_{k|k}}\right)_i & i = 2, \dots, n+1 \\ \hat{\boldsymbol{X}}_{k|k} - \sqrt{(n+\eta)}\left(\sqrt{\boldsymbol{S}^{x,x}_{k|k}}\right)_i & i = n+2, \dots, 2n+1. \end{cases} \tag{2.53}$$

$\eta = \mu^2(n+\kappa) - n$ is a scaling parameter with μ set to a small value and κ is a secondary scaling parameter usually set to 0. $\left(\sqrt{\boldsymbol{S}^{x,x}}\right)_i$ denotes the i:th column of the matrix square root of $\boldsymbol{S}^{x,x}$ computed using lower triangular Cholesky factorization (Wan and van der Merwe, 2000). This gives a matrix $\boldsymbol{\chi}$ with every column representing its own estimate of $\hat{\boldsymbol{X}}_{k|k}$. The $\boldsymbol{\chi}$ matrix can then be transitioned to the next time step

$$\boldsymbol{\chi}_{k+1|k} = \boldsymbol{F}\left(\boldsymbol{\chi}_{k|k}, \boldsymbol{\Theta}, \boldsymbol{0}\right) \tag{2.54}$$

and used to predict the future state $\boldsymbol{X}_{k+1}$ as the weighted mean

$$\hat{\boldsymbol{X}}_{k+1|k} = \sum_{i=1}^{2n+1} W_i^m \boldsymbol{\chi}_{k+1|k}, \tag{2.55}$$

with weights

$$W_i^m = \begin{cases} \frac{\eta}{n+\eta} & i = 1 \\ \frac{1}{2(n+\eta)} & i = 2, \ldots, 2n+1. \end{cases} \tag{2.56}$$

(Wan and van der Merwe, 2000). Each column gets its own estimate of $\boldsymbol{Y}_{k+1|k}$, collected in matrix $\boldsymbol{\Upsilon}_{k+1|k}$ as

$$\boldsymbol{\Upsilon}_{k+1|k} = \boldsymbol{H}\left(\boldsymbol{\chi}_{k|k}, \boldsymbol{\Theta}, \boldsymbol{0}\right) \tag{2.57}$$

and $\boldsymbol{Y}_{k+1|k}$ can be estimated as a weighted mean

$$\hat{\boldsymbol{Y}}_{k+1|k} = \sum_{i=1}^{2n+1} W_i^m \boldsymbol{\Upsilon}_{i,k+1|k} \tag{2.58}$$

(Wan and van der Merwe, 2000). The covariance matrices are estimated as

$$\boldsymbol{S}_{k|k-1}^{x,x} = \sum_{i=1}^{2n+1} W_i^c \left(\boldsymbol{\chi}_{i,k+1|k} - \hat{\boldsymbol{X}}_{k+1|k}\right)\left(\boldsymbol{\chi}_{i,k+1|k} - \hat{\boldsymbol{X}}_{k+1|k}\right)^T + \boldsymbol{S}_e \tag{2.59}$$

$$\boldsymbol{S}_{k|k-1}^{y,y} = \sum_{i=1}^{2n+1} W_i^c \left(\boldsymbol{\Upsilon}_{i,k+1|k} - \hat{\boldsymbol{Y}}_{k+1|k}\right)\left(\boldsymbol{\Upsilon}_{i,k+1|k} - \hat{\boldsymbol{Y}}_{k+1|k}\right)^T + \boldsymbol{S}_w \tag{2.60}$$

$$\boldsymbol{S}_{k|k-1}^{x,y} = \sum_{i=1}^{2n+1} W_i^c \left(\boldsymbol{\chi}_{i,k+1|k} - \hat{\boldsymbol{X}}_{k+1|k}\right)\left(\boldsymbol{\Upsilon}_{i,k+1|k} - \hat{\boldsymbol{Y}}_{k+1|k}\right)^T \tag{2.61}$$

with weights

$$W_i^c = \begin{cases} \frac{\eta}{n+\eta} + (1 - \mu^2 + \nu) & i = 1 \\ \frac{1}{2(n+\eta)} & i = 2, \ldots, 2n+1 \end{cases} \tag{2.62}$$

where ν is used to incorporate prior knowledge of the distribution of $\boldsymbol{X}$. For Gaussian distributions, $\nu = 2$ is optimal (Wan and van der Merwe, 2000). The point here is to estimate the covariances without using the jacobians $\boldsymbol{A}$ and $\boldsymbol{C}$, and instead use the sample points that are based on prior covariance and state estimates. The Kalman gain is then calculated as

$$\boldsymbol{K}_k = \boldsymbol{S}_{k|k-1}^{x,y}\left[\boldsymbol{S}_{k|k-1}^{y,y}\right]^{-1}, \tag{2.63}$$

after which the current state variable

$$\hat{\boldsymbol{X}}_{k|k} = \hat{\boldsymbol{X}}_{k|k-1} + \boldsymbol{K}_k \boldsymbol{\Gamma}_k \tag{2.64}$$

and its covariance

$$\boldsymbol{S}_{k|k}^{x,x} = \boldsymbol{S}_{k|k-1}^{x,x} - \boldsymbol{K}_k \boldsymbol{S}_{k|k-1}^{y,y} \boldsymbol{K}_k^T \tag{2.65}$$

are updated (Wan and van der Merwe, 2000). The point of the Unscented Kalman filter is to estimate higher order statistics than the Extended Kalman filter by using the different values of the columns in $\boldsymbol{\chi}$. This is done to avoid the linearization of the Extended Kalman filter and based on the assumption that it is easier to estimate the probability density than to estimate the transition function Wan and van der Merwe (2000). The Unscented Kalman filter tries to do this without needing the computational power of a particle filter (Wan and van der Merwe, 2000). For the algorithm of estimating the states of the DTMS model using an Unscented Kalman filter, see algorithm 2 on page 25.

2.2.2 The Bootstrap Particle filter

The particle filter has grown in popularity in estimation of hidden state variables in the last few years and has shown a number of efficient applications in different fields of science (Peng et al., 2016). While the filter can seem quite simple, its derivation can be tedious and this section simply means to go through the basics of using a *bootstrap* particle filter—meaning a particle filter using sequential importance resampling as will be described soon—for hidden state estimation. For a more thorough derivation and explanation of particle filters and their application in finance, see Lindström et al. (2015). Now, consider the space state model given by equations 2.47 and 2.48. A way to rephrase this could be that variables $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$ belong to probability distributions

$$\boldsymbol{X}_{k+1} \sim \mathcal{N}\left(\boldsymbol{F}\left(\boldsymbol{X}_k, \boldsymbol{\Theta}, \boldsymbol{0}\right), \boldsymbol{S}_e\right) \tag{2.66}$$

$$\boldsymbol{Y}_k \sim \mathcal{N}\left(\boldsymbol{H}\left(\boldsymbol{X}_k, \boldsymbol{\Theta}, \boldsymbol{0}\right), \boldsymbol{S}_w\right). \tag{2.67}$$

First, M particles, or samples, are initiated as

$$\boldsymbol{x}_{0|0}^{(i)} = \boldsymbol{X}_{0|0} \tag{2.68}$$

for $i = 1, \ldots, N$ and some initial estimate $\boldsymbol{X}_{0|0}$ Doucet et al. (2001). Then, the particles are transitioned into the next time stem using the transition function $\boldsymbol{F}$ and the distribution of $\boldsymbol{X}_{k+1}$ given in equation 2.66, as *random samples* from the distribution

$$\boldsymbol{x}_{k|k-1}^{(i)} \sim \mathcal{N}\left(\boldsymbol{F}\left(\boldsymbol{x}_{k-1|k-1}^{(i)}, \boldsymbol{\Theta}, \boldsymbol{0}\right), \boldsymbol{S}_e\right). \tag{2.69}$$

using some computer randomizer. Each particle now represents an individual estimation of the state $\boldsymbol{X}_k$, using the expected value $\boldsymbol{F}\left(\boldsymbol{X}_{k-1|k-1}\right)$ and the covariance of the state $\boldsymbol{S}_e$. Then, each particle's strength in estimating the state can be evaluated using the distribution of $\boldsymbol{Y}_k$, giving them importance weights

$$\omega_k^{(i)} = p_N\left(\boldsymbol{Y}_k, \boldsymbol{H}\left(\boldsymbol{x}_{k|k-1}^{(i)}, \boldsymbol{\Theta}, \boldsymbol{0}\right), \boldsymbol{S}_e\right) \tag{2.70}$$

where $p_N\left(\boldsymbol{Y}, \bar{\boldsymbol{Y}}, \boldsymbol{S}_e\right)$ denotes the probability density function of $\boldsymbol{Y}$ for a Gaussian distribution with mean $\bar{\boldsymbol{Y}}$ and covariance matrix $\boldsymbol{S}_e$ Doucet et al. (2001). This will give higher weights to the particles that were better at predicting the current observation $\boldsymbol{Y}_k$ using their particle-specific state observations $\boldsymbol{x}_{k|k-1}^{(i)}$ and lower weights to the ones that gave inaccurate predictions. By normalizing the weights

$$\tilde{w}_k^{(i)} = \frac{\omega_k^{(i)}}{\sum_{i=1}^{M} \omega_k^{(i)}} \tag{2.71}$$

and resampling new particles from the present set, with weights representing probabilities, a new set of particles can be sampled from the old ones

$$\left\{\boldsymbol{x}_{k|k}^{(i)}\right\} \sim \left[\left\{\boldsymbol{x}_{k|k-1}^{(i)}, \tilde{w}_k^{(i)}\right\}\right] \tag{2.72}$$

using the normalized weights as discrete probabilities (Doucet et al., 2001). This means that particles that were given high weights in the prediction performance evaluation in equation 2.70 are likely to stay, and be duplicated, in the particle set while those who performed badly will be left out. The new particles can be used to form the current state estimate as the average of the resampled particles (Doucet et al., 2001)

$$\hat{\boldsymbol{X}}_{k|k} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{x}_{k|k}^{(i)}. \tag{2.73}$$

The new particles can then be transitioned into the next time step using equation 2.69 again, after which the weighting procedure can be repeated and a new current state variable estimated. The point of the particle filter is to try to construct a empirical distribution of the hidden state vector. For a simple example of applying a particle filter, see appendix A.3. For the algorithm of estimating the states of the DTMS model using a Bootstrap particle filter, see algorithm 3 on page 27.

2.3 Parameter estimation

In cases where parameters $\boldsymbol{\Theta}$ are unknown, such as in the DTMS model described in 2.1, they need to be estimated. (Peng et al., 2016) used two estimation methods, maximum likelihood and Bootstrap Particle using a Self-organizing State Space formulation of the DTMS model. Section 2.3.1 deals with maximum likelihood estimation as used by Peng et al. (2003) for the Extended Kalman filter, but the method is also applicable to the Unscented Kalman filter. Section 2.3.2 shortly describes the method of Peng et al. (2016), estimating the parameters as a part of the Bootstrap Particle filter loop.

2.3.1 Maximum likelihood parameter estimation

Consider the prediction error for $\boldsymbol{Y}_k$,

$$\boldsymbol{\Gamma}_k = \boldsymbol{Y}_k - \hat{\boldsymbol{Y}}_{k|k-1} \tag{2.74}$$

which is assumed to be a 2-dimensional Gaussian white noise vector and the covariance matrix $\boldsymbol{S}^{y,y}_{k|k-1}$ of observation prediction $\hat{\boldsymbol{Y}}_{k|k-1} = E\left\{Y_k|Y^{k-1}\right\}$ where $\boldsymbol{Y}^k$ denotes all observations up to and including $\boldsymbol{Y}_{k-1}$, $\boldsymbol{Y}^k = \{\boldsymbol{Y}_1, \boldsymbol{Y}_2, ..., \boldsymbol{Y}_{k-1}\}$. Both $\boldsymbol{\Gamma}_k$ and $\boldsymbol{S}^{y,y}_{k|k-1}$ can be retrieved from every time step of the Extended and Unscented Kalman filters described in 2.2.1.1 and 2.2.1.2. The joint conditional density function for $\boldsymbol{\Gamma}_k$ given all previous observations $\boldsymbol{Y}^{k-1}$ is

$$p\left(\boldsymbol{\Gamma}_k|\boldsymbol{Y}^{k-1}\right) = \frac{1}{2\pi\sqrt{\left|\boldsymbol{S}^{y,y}_{k|k-1}\right|}} e^{-\frac{1}{2}\boldsymbol{\Gamma}_k^T \boldsymbol{S}^{y,y}_{k|k-1}\boldsymbol{\Gamma}_k} \tag{2.75}$$

and the (-2) log-likelihood of equation 2.12 for K observations $\{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_K\}$ is then derived as

$$(-2)\log p\left(\boldsymbol{Y}^K|\boldsymbol{\Theta}\right) = \sum_{k=1}^{K}\left\{\log\left|\boldsymbol{S}^{y,y}_{k|k-1}\right| + \boldsymbol{\Gamma}_k^T\left[\boldsymbol{S}^{y,y}_{k|k-1}\right]^{-1}\boldsymbol{\Gamma}_k\right\} + 2N\log 2\pi \tag{2.76}$$

where $|\cdot|$ denotes matrix determinant (Peng et al., 2003). The maximum likelihood estimate of $\boldsymbol{\Theta}$, $\boldsymbol{\Theta}^*$ is then given by

$$\boldsymbol{\Theta}^* = \arg\min_{\boldsymbol{\Theta}}\sum_{k=1}^{K}\left\{\log\left|\boldsymbol{S}^{y,y}_{k|k-1}(\boldsymbol{\Theta})\right| + \boldsymbol{\Gamma}_k(\boldsymbol{\Theta})^T\left[\boldsymbol{S}^{y,y}_{k|k-1}(\boldsymbol{\Theta})\right]^{-1}\boldsymbol{\Gamma}_k(\boldsymbol{\Theta})\right\} + 2N\log 2\pi \tag{2.77}$$

where $\boldsymbol{\Gamma}_k$ and $\boldsymbol{S}^{y,y}_{k|k-1}$ are estimated through some filtering technique (Peng et al., 2003), some of which are described in section 2.2. Peng et al. (2003) suggests estimating the parameters with the Nelder-Mead method using MATLAB, running the filter loop numerous times to find the $\hat{\boldsymbol{\Theta}}$ with the highest likelihood.

2.3.2 Particle filter parameter estimation

In the Self-organizing State Space reformulation of the DTMS model, described in 2.1.2, the point is to avoid the maximum likelihood estimation performed in 2.3.1 since Peng et al. (2016) claims that the estimation might be biased due to the errors in the state estimates $\hat{\boldsymbol{X}}^K$. Instead, the parameter values of $\boldsymbol{\Theta}$ is added to the state vector, making their estimation a part of the filtering and estimating them in the same way as the state $\boldsymbol{X}_k$ using a particle filter. This requires some inital distribution for both $\boldsymbol{X}_k$ and $\boldsymbol{\Theta}$ to start the filter loop, which Peng et al. (2016) suggests achieving from using an Extended Kalman filter and maximum likelihood estimation prior to initiating the particle filter.

2.4 Portfolio strategy

Peng et al. (2003) used an Extended Kalman filter to estimate the DTMS models, and noted that the prediction error of the asset price was basically a white noise process, while the estimated current excess demand $\hat{\phi}_{k|k}$ had a higher autocorrelation, varied around zero and was much smoother. It was therefore argued that a trading strategy should be based on $\hat{\phi}_{k|k}$ rather than $\hat{P}_{k+1|k}$, due to the direct effect of previous periods' excess demand on price as seen in equation 2.6. Peng et al. (2003) suggests using a trading strategy that allocates different portions of the portfolio value in the asset depending on different threshold values $\boldsymbol{\tau}$ of $\hat{\phi}_{k|k}$. The strategy is described in algorithm 4 on page 28 where a is the portion of the portfolio invested in the asset and R_k is the asset return at point k. Another way to use $\hat{\phi}_{k|k}$ for portfolio could be to assume that a negative value of $\hat{\phi}_{k|k}$ implies declining prices and $\hat{\phi}_{k|k}$ implies rising prices, and then simply invest everything in the portfolio when $\hat{\phi}_{k|k} > 0$ and nothing when $\hat{\phi}_{k|k} < 0$. This strategy is described in algorithm 5 on page 5.

To find threshold values $\boldsymbol{\tau}$ for algorithm 4, Peng et al. (2003) defines a (negative) asset-valuation function

$$J(\boldsymbol{\tau}) = -A_K(\boldsymbol{\tau}) + \frac{\psi}{T}\sum_{k=1}^{N}\left|A_k(\boldsymbol{\tau}) - \left(A_0 + \frac{k}{T}\left(A_K(\boldsymbol{\tau}) - A_0\right)\right)\right|^2 \tag{2.78}$$

where A_k is the portfolio value at point k and ψ is a weighting factor. The first part of the function is simply the final portfolio value, while the second part is supposed to take portfolio value fluctuation into account and prefer smoother, less volatile portfolios (Peng et al., 2003). The weight ψ decides how much focus should be put on either one of these factors, with a high value preferring less volatile portfolios higher. The optimal threshold values are then given by

$$\boldsymbol{\tau}^* = \arg\min_{\boldsymbol{\tau}} J(\boldsymbol{\tau}) \tag{2.79}$$

which Peng et al. (2003) suggests estimating using MATLAB.

2.5 Portfolio performance evaluation

There are numerous known methods of evaluatiing the performance of portfolios. The most intuitive method would be to just look at the gross or expected return of the portfolio, claiming that the portfolio with the highest returns is the superior one. Many measures, such as the Sharpe ratio, also take *risk* into account, often measured using the sample variance of the returns. In portfolios that choose between investing and not investing in the market portfolio however, the returns for the strategic portfolios are between 0 and that of the passive portfolio, meaning sample variance of returns will almost always be lower than in a passive strategy, making that comparison less useful. Another disadvantage of the Sharpe ratio is that it loses its intuition in times of negative gross return, since negative Sharpe ratios *gain* from having high volatility. Even though the comparison of negative Sharpe ratios has been defended by for example McLeod and Van Vuuren (2004), it is not very intuitive to use in comparison between portfolios with negative expected returns. Other popular strategies are those including estimation of the portfolio β, such as the Jensen's alpha-measure Aragon and Ferson (2007). These measurements are made mostly for portfolios that have a constant covariance with the market portfolio, and since the strategies in this study clearly has not, these measurements are also left out. A more intuitive measure in these cases is the average daily return, shortly described in section 2.5.1.

2.5.1 Average daily return

The average daily return of the portfolios can simply be calculated as the sample mean

$$\bar{R}_p = \frac{1}{K}\sum_{k=1}^{K} R_{p,k} \tag{2.80}$$

for K time points, where $R_{p,k}$ is the portfolio's return at time point k. It is a very intuitive measure as it says how much this portfolio is expected to return every day. To test if a portfolio strategy can beat the *market portfolio* with returns $R_{m,k}$, a Welch's t-test (Welch, 1947) can be conducted as

$$t = \frac{\bar{R}_p - \bar{R}_m}{\sqrt{\frac{\hat{\sigma}_p^2}{K_p} + \frac{\hat{\sigma}_m^2}{K_m}}} \tag{2.81}$$

with $\hat{\sigma}^2$ being the sample variance. The test statistic t then follows a Student's t-distribution with

$$\nu = \frac{\left(\frac{\hat{\sigma}_p^2}{K_p} + \frac{\hat{\sigma}_m^2}{K_m}\right)}{\frac{\hat{\sigma}_p^4}{K_p^2(K_p-1)} + \frac{\hat{\sigma}_m^4}{K_m^2(K_m-1)}} \tag{2.82}$$

degrees of freedom. The null hypothesis of the portfolios having the same average return can then be tested using some significance level of choice.

2.5.2 Market timing

Another aspect that is of interest especially for actively managed portfolios is that of *market timing*, i.e. the ability to enter and exit the market at the right time points to avoid price declines and still benefit from price increases. One way to test this is the Henriksson-Merton non-parametric test.

2.5.2.1 Henriksson-Merton's non-parametric market timing test

Henriksson and Merton (1981) present an intuitive way to test market timing. Instead of estimating any regression models, the test examines the individual daily forecasts of an investor. Assuming that the investor at every day decides whether to invest in the market portfolio or keep the money in the bank by forecasting the next days return R_{k+1} and setting the portion of total assets to invest the next day as

$$a_{k+1} = \begin{cases} 1 & \hat{R}_{m,k+1|k} > 0 \\ 0 & \hat{R}_{m,k+1|k} \leq 0 \end{cases} \tag{2.83}$$

where $0 \leq a \leq 1$ and $\hat{R}_{m,k+1|k}$ is today's forecast of tomorrows market portfolio return. Henriksson and Merton (1981) then defines the conditional probability of a correct forecast given that $R_{m,k} \leq 0$ as

$$p_{1,k} = P\left(a_k = 0 | R_{m,k} \leq 0\right) \tag{2.84}$$

$$1 - p_{1,k} = P\left(a_k = 1 | R_{m,k} \leq 0\right) \tag{2.85}$$

and the conditional probability of a correct forecast given that $R_{m,k} > 0$ as

$$p_{2,k} = P\left(a_k = 1 | R_{m,k} > 0\right) \tag{2.86}$$

$$1 - p_{2,k} = P\left(a_k = 0 | R_{m,k} > 0\right). \tag{2.87}$$

According to Henriksson and Merton (1981), it is a sufficient condition that $p_{1,k} + p_{2,k} = 1$ for the investor's prediction to be of no use, while a optimal forecaster would have $p_{1,k} + p_{2,k} = 1 + 1 = 2$ and a forecaster that is always wrong would have $p_{1,k} + p_{2,k} = 0$. Since the probabilities $p_{1,k}$ and $p_{2,k}$ are generally not observable, they have to be estimated (Henriksson and Merton, 1981). By defining variable N_1 as the number of observations where $R_{m,k} > 0$, N_2 as the number of observations where $R_{m,k} \leq 0$, (i.e. $N_1 + N_2$ make ut the entire data set), n_1 as the number of successful predictions where $R_{m,k} \leq 0$ and n_2 as the number of *unsuccessful* predictions where $R_{m,k} > 0$ (i.e. $n_1 + n_2$ is the number of times the forecaster predicted negative market portfolio returns) (Henriksson and Merton, 1981). The estimated probabilities of correct forecasts are then

$$\hat{p_1} = \frac{n_1}{N_1} \tag{2.88}$$

and

$$\hat{p_2} = 1 - \frac{n_2}{N_2}. \tag{2.89}$$

Under the null hypothesis of no market timing, $p_1 = p_2 = p = 0.5$, meaning

$$\hat{p} = \frac{n_1 + n_2}{N_1 + N_2} \tag{2.90}$$

Since both $\frac{n_1}{N_1}$ and $\frac{n_2}{N_2}$ have expected value $p = 0.5$ under the null hypothesis and are drawn from independent subsamples, only one of them will have to be used as an estimate (Henriksson and Merton, 1981). In the rest of this derivation, n_1 is used solely. The probability of the binomially distributed variable n_1 getting value $n_1 = x$ from a subsample of N_1 drawings is then

$$P\left(n_1 = x | N_1, p\right) = \binom{N_1}{x} p^x \left(1 - p\right)^{N_1 - x}. \tag{2.91}$$

Henriksson and Merton (1981) then uses Baye's theorem to form the probability of $n_1 = x$ given N_1, N_2 and n, which is derived as

$$P\left(n_1 = x | N_1, N_2, n\right) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}, \tag{2.92}$$

meaning that n_1 follows a *hypergeometric* distribution (Henriksson and Merton, 1981). Using the probability mass function of the hypergeometric distribution, the null hypothesis of $p_1 + p_2 = 1$ versus the alternative hypothesis $p_1 + p_2 > 1$ can be tested (Henriksson and Merton, 1981).

3 Method

The algorithms and tests described below were all executed using MATLAB. The code in its entirety is available upon request.

3.1 Filtering algorithms

3.1.1 Extended Kalman filter

To estimate the DTMS model, described in section 2.1, an Extended Kalman filter, described in section 2.2.1.1, was used.

Using the method described in 2.2.1.1, algorithm 1 was used to calculate the one step prediction $\hat{\boldsymbol{X}}_{k|k-1} = E\left\{\boldsymbol{X}_k|\boldsymbol{Y}^{k-1}\right\}$ and the reconstruction $\hat{\boldsymbol{X}}_{k|k} = E\left\{\boldsymbol{X}_k|\boldsymbol{Y}^{k}\right\}$ with stepwise updated matrices $\boldsymbol{A}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)$, $\boldsymbol{C}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)$, and $\boldsymbol{S}_e\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)$. $\boldsymbol{S}^{x,x}_{k|k-1}$ and $\boldsymbol{S}^{x,x}_{k|k}$ are the conditional covariance matrices of $\hat{\boldsymbol{X}}_{k|k-1}$ and $\hat{\boldsymbol{X}}_{k|k}$ respectively. Note that algorithm 1 requires parameters $\boldsymbol{\Theta}$ to be estimated outside of the loop, which was done by minimizing function 2.77. Parameters δ, $\boldsymbol{S}_w$, $\hat{\boldsymbol{X}}_{0|0}$ and $\boldsymbol{S}^{x,x}_{0|0}$ were appended to the $\boldsymbol{\Theta}$-vector in the maximum likelihood estimation as suggested by Peng et al. (2003).

Algorithm 1 Extended Kalman filtering of the DTMS model

Initialize

$\hat{\boldsymbol{X}}_{0|0}$

$\boldsymbol{S}^{x,x}_{0|0}$

for $k = 1, \ldots, K$ **do**

Prediction

$\hat{\boldsymbol{X}}_{k|k-1} = \boldsymbol{A}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)\hat{\boldsymbol{X}}_{k-1|k-1}$

$\boldsymbol{S}^{x,x}_{k|k-1} = \boldsymbol{A}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)\boldsymbol{S}^{x,x}_{k-1|k-1}\boldsymbol{A}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)^T + \boldsymbol{S}_e\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)$

$\boldsymbol{S}^{y,y}_{k|k-1} = \boldsymbol{C}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)\boldsymbol{S}^{x,x}_{k|k-1}\boldsymbol{C}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)^T + \boldsymbol{S}_w$

Filtering

$\boldsymbol{K}_k = \boldsymbol{S}^{x,x}_{k|k-1}\boldsymbol{C}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)^T\left[\boldsymbol{S}^{y,y}_{k|k-1}\right]^T$

$\hat{\boldsymbol{X}}_{k|k} = \hat{\boldsymbol{X}}_{k|k-1} + \boldsymbol{K}_k\left(\boldsymbol{Y}_k - \boldsymbol{C}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)\hat{\boldsymbol{X}}_{k|k-1}\right)$

$\boldsymbol{S}^{x,x}_{k|k} = \left[\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{C}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)\right]\boldsymbol{S}^{y,y}_{k|k-1}$

end

3.1.2 Unscented Kalman filter

To use the *Unscented* Kalman filter, described in 2.2.1.2, rather than the Extended one for estimation of the DTMS model, algorithm 2 was used. This algorithm also needed parameters $\boldsymbol{\Theta}$ to be estimated outside of the loop. Once again, parameters δ, $\boldsymbol{S}_w$, $\hat{\boldsymbol{X}}_{0|0}$

and $\boldsymbol{S}^{x,x}_{0|0}$ were appended to the $\boldsymbol{\Theta}$-vector in the maximum likelihood estimation as suggested by Peng et al. (2003).

Algorithm 2 Unscented Kalman filtering of the DTMS model

Initialize

$\hat{\boldsymbol{X}}_{0|0}$

$\boldsymbol{S}^{x,x}_{0|0}$

$W^m_1 = \frac{\eta}{n+\eta}$

$W^c_1 = \frac{\eta}{n+\eta} + (1 - \mu^2 + \nu)$

for $i = 2, \ldots, 2n+1$ **do**

$\quad W^c_i = W^m_i = \frac{1}{2(n+\eta)}$

end

for $k = 1, \ldots, K$ **do**

$\quad$ *Prediction*

$\quad \boldsymbol{\chi}_{1,k-1|k-1} = \hat{\boldsymbol{X}}_{k-1|k-1}$

$\quad$ **for** $i = 2, \ldots, n+1$ **do**

$\quad\quad \boldsymbol{\chi}_{i,k-1|k-1} = \hat{\boldsymbol{X}}_{k-1|k-1} + \sqrt{(n+\eta)} \left(\sqrt{\boldsymbol{S}^{x,x}_{k-1|k-1}}\right)_i$

$\quad$ **end**

$\quad$ **for** $i = n+2, \ldots, 2n+1$ **do**

$\quad\quad \boldsymbol{\chi}_{i,k-1|k-1} = \hat{\boldsymbol{X}}_{k-1|k-1} - \sqrt{(n+\eta)} \left(\sqrt{\boldsymbol{S}^{x,x}_{k-1|k-1}}\right)_i$

$\quad$ **end**

$\quad \boldsymbol{\chi}_{k|k-1} = \boldsymbol{A}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)\boldsymbol{\chi}_{k-1|k-1}$

$\quad \boldsymbol{\Upsilon}_{k|k-1} = \boldsymbol{C}\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)\boldsymbol{\chi}_{k-1|k-1}$

$\quad \hat{\boldsymbol{X}}_{k|k-1} = \sum_{i=1}^{2n+1} W^m_i \boldsymbol{\chi}_{k|k-1}$

$\quad \hat{\boldsymbol{Y}}_{k|k-1} = \sum_{i=1}^{2n+1} W^m_i \boldsymbol{\Upsilon}_{k|k-1}$

$\quad \boldsymbol{S}^{x,x}_{k|k-1} = \sum_{i=1}^{2n+1} W^c_i \left(\boldsymbol{\chi}_{i,k|k-1} - \hat{\boldsymbol{X}}_{k|k-1}\right)\left(\boldsymbol{\chi}_{i,k|k-1} - \hat{\boldsymbol{X}}_{k|k-1}\right)^T + \boldsymbol{S}_e\left(\hat{\boldsymbol{X}}_{k-1|k-1}|\boldsymbol{\Theta}\right)$

$\quad \boldsymbol{S}^{y,y}_{k|k-1} = \sum_{i=1}^{2n+1} W^c_i \left(\boldsymbol{\Upsilon}_{i,k|k-1} - \hat{\boldsymbol{Y}}_{k|k-1}\right)\left(\boldsymbol{\Upsilon}_{i,k|k-1} - \hat{\boldsymbol{Y}}_{k|k-1}\right)^T + \boldsymbol{S}_w$

$\quad \boldsymbol{S}^{x,y}_{k|k-1} = \sum_{i=1}^{2n+1} W^c_i \left(\boldsymbol{\chi}_{i,k|k-1} - \hat{\boldsymbol{X}}_{k|k-1}\right)\left(\boldsymbol{\Upsilon}_{i,k|k-1} - \hat{\boldsymbol{Y}}_{k|k-1}\right)^T$

$\quad$ *Filtering*

$\quad \boldsymbol{K}_k = \boldsymbol{S}^{x,y}_{k|k-1}\left[\boldsymbol{S}^{y,y}_{k|k-1}\right]^{-1}$

$\quad \hat{\boldsymbol{X}}_{k|k} = \hat{\boldsymbol{X}}_{k|k-1} + \boldsymbol{K}_k\left(\boldsymbol{Y}_k - \hat{\boldsymbol{Y}}_{k|k-1}\right)$

$\quad \boldsymbol{S}^{x,x}_{k|k} = \boldsymbol{S}^{x,x}_{k|k-1} - \boldsymbol{K}_k \boldsymbol{S}^{y,y}_{k|k-1} \boldsymbol{K}_k^T$

end

3.1.3 Particle filter

The Bootstrap Particle filter described in section 2.2.2 was used to estimate either the DTMS model described in section 2.1 with externally estimated parameters $\boldsymbol{\Theta}$ or the SOSS model described in section 2.1.2. Peng et al. (2016) recommends using the SOSS model in favour of the DTMS model to avoid the possible error of biased maximum likelihood estimation. It is also argued that the main problem with using the Bootstrap Particle filter on the SOSS model is that of initial distributions for the particles. Peng et al. (2016) suggests using a Extended Kalman filtering technique and setting the initial conditions for the new state vector particles $\hat{\boldsymbol{z}}_{0|0}^{(i)} = \begin{bmatrix} \hat{\boldsymbol{x}}_{0|0}^{(i)} & \boldsymbol{\theta}_0^{(i)} \end{bmatrix}^T$ as $\boldsymbol{x}_{0|0}^{(i)} \sim \mathcal{N}\left(\hat{\boldsymbol{X}}_{0|0}, \boldsymbol{S}_{0|0}^{x,x}\right)$ and $\boldsymbol{\theta}_0^{(i)} \sim U\left(\hat{\boldsymbol{\Theta}}\right)$ where U denotes the uniform distribution in an interval close to the estimate $\hat{\boldsymbol{\Theta}}$. The algorithm used for the Bootstrap Particle filter can be seen in 3. Note that the particles $\boldsymbol{\theta}^{(i)}$ do not have a transition function and they will therefore not be updated particle-wise apart from in the resampling. This means there will be M different values for $\hat{\Theta}$ in the first time step, and then this number will reduce in every step.

3.1.3.1 Troubleshooting

There were two main issues with running the Bootstrap Particle filter in the way that Peng et al. (2016) suggests. One was the fact that the covariance matrix of observation noise, $\boldsymbol{S}_w$, was included in the $\boldsymbol{\Theta}$-vector. This gave all particles different values of $\boldsymbol{S}_w^{(i)}$, which was then used to evaluate their prediction performance and therefore importance weight using the probability density function. This led to particles with high values of $\boldsymbol{S}_w^{(i)}$ tending to get high weights even if their predictions were poor, due to the probability density function used to evaluate those particles performances being a high-variance one. The second issue was that the Extended Kalman filter estimates suggested very low values for $\boldsymbol{S}_w$, making MATLAB often truncate the probability density function value as 0 for a lot of the particles, and sometimes for *all* particles, breaking the filter down completely. Peng et al. (2016) does not mention this possible issue, but use a uniform interval for the first term in $\boldsymbol{S}_w^{(i)}$ from 0 to 500 times the Kalman filter estimate. To get around this problem in this thesis, the value of $\boldsymbol{S}_w$ was instead set as the same constat value for all particles. The value was chosen as the Kalman filter estimate times some arbitrary factor that completely eliminates the problem with all particles being given importance weight 0. While this tends to give the particles similar weights, it should at least give the highest weight to the most accurate prediction and lower weights for those less accurate. The factor was chosen as the smallest possible number that mitigated the issue of all zero-weights for every data set individually. This method, as described in algorithm 3, was used wtith initial distributions based on the Extended Kalman filter maximum likelihood estimates as uniformly distributed stochastic variables $\sim U\left[-min, max\right]$ as

$$\begin{array}{rc} \alpha_i \sim & U\left[-|\hat{\alpha}_i|, |\hat{\alpha}_i|\right] \\ \beta_i \sim & U\left[0, \hat{\beta}_i\right] \\ \gamma_i \sim & U\left[0, \hat{\gamma}_i\right] \\ \delta \sim & U\left[-|\hat{\delta}|, |\hat{\delta}|\right] \end{array}, \tag{3.1}$$

with 1000 particles.

Algorithm 3 Bootstrap particle filtering of the SOSS model

Initialize

for $i = 1, \ldots, M$ **do**

  Sample

  $\boldsymbol{x}_{0|0}^{(i)} \sim \mathcal{N}\left(\hat{\boldsymbol{X}}_{0|0}, \boldsymbol{S}_{0|0}^{x,x}\right)$

  $\boldsymbol{\theta}_0^{(i)} \sim U\left(\hat{\boldsymbol{\Theta}}\right)$

  $w_0^{(i)} = \frac{1}{N}$

  $\boldsymbol{z}_{0|0}^{(i)} = \left[\begin{array}{cc} \boldsymbol{x}_{0|0}^{(i)} & \boldsymbol{\theta}_0^{(i)} \end{array}\right]^T$

end

for $k = 1, \ldots, K$ **do**

  Prediction

  for $i = 1, \ldots, M$ **do**

    Sample

    $\boldsymbol{x}_{k|k-1}^{(i)} \sim \mathcal{N}\left(\boldsymbol{A}\left(\boldsymbol{x}_{k-1|k-1}^{(i)}|\boldsymbol{\theta}_{k-1}^{(i)}\right)\boldsymbol{x}_{k-1|k-1}^{(i)}, \boldsymbol{S}_e\left(\boldsymbol{x}_{k-1|k-1}^{(i)}|\boldsymbol{\theta}^{(i)}\right)\right)$

    Weight

    $\omega_k^{(i)} = p_N\left(\boldsymbol{Y}_k, \boldsymbol{C}\left(\boldsymbol{x}_{k|k-1}^{(i)}\right), \boldsymbol{S}_w\right)$

    Normalize

    for $i = 1, \ldots, M$ **do**

      $\tilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{i=0}^{M} w_k^{(i)}}$

    end

    Resample

    $\left\{\boldsymbol{z}_{k|k}^{(i)}\right\} \sim \left[\left\{\left[\begin{array}{cc} \boldsymbol{x}_{k|k-1}^{(i)} & \boldsymbol{\theta}_{k-1}^{(i)} \end{array}\right]^T, \tilde{w}_k^{(i)}\right\}\right]$

    Filtering

    $\boldsymbol{Z}_{k|k} = \frac{1}{M}\sum_{i=1}^{N} \boldsymbol{z}_{k|k}^{(i)}$

  end

end

3.2 Portfolio strategy

To compare the three different filtering techniques; Extended Kalman, Unscented Kalman and Bootstrap Particle, the two portfolio strategies described in section 2.4 were used based on the filter estimates of ϕ, yielding a total of 6 strategic portfolios. A passive portfolio that always invested all its money in the relevant asset, i.e. a proxy for tha market portfolio, was also used as a mean of comparison. The threshold values $\boldsymbol{\tau}$ for the threshold-based strategy were calculated with equation 2.79 with $\psi = 1$ to evenly focus on final portfolio value and smoothness. All tests were performed using a initial investment of 10 units of the local currency.

Algorithm 4 Threshold-based portfolio strategy τ

Initialize

A_0

$a_1 = 1$

Trading

for $k = 1, \ldots, K$ **do**
    $A_k = a_k R_k A_{k-1} + (1 - a_k) A_{k-1}$
    if $\hat{\phi}_{k|k} > \tau_1$ **then**
        $a_{k+1} = 1$
    else
        if $\tau_2 < \hat{\phi}_{k|k} \leq \tau_1$ **then**
            $a_{k+1} = 0.8$
        else
            if $-\tau_3 < \hat{\phi}_{k|k} \leq \tau_2$ **then**
                $a_{k+1} = 0.5$
            else
                if $-\tau_4 < \hat{\phi}_{k|k} \leq -\tau_3$ **then**
                    $a_{k+1} = 0.2$
                else
                    $a = 0$
                end
            end
        end
    end
end

	Index	Abbreviation	Country	Estimation	Testing
1.	All Ordinaries Index	AOI	Australia	2000-2005	2006-2011
2.	Deutscher Aktienindex	DAX	Germany	1999-2006	2007-2014
3.	Financial Times Stock Exchange 100 Index	FTSE	The U.K.	1987-1988	1989-1990
4.	Hang Seng Index	HSI	Hong Kong	1997-2000	2001-2004
5.	Índice Bolsa de Valores do Estado de São Paulo	IBOVESPA	Brazil	1998-2001	2002-2005
6.	Nihon Keizai Shimbun 225 Index	Nikkei	Japan	1992-1995	1996-1999
7.	OMX Stockholm 30	OMXS30	Sweden	1999-2003	2004-2008
8.	Russia Trading System Index	RTSI	Russia	1996-1999	2000-2003
9.	Standard & Poor's Bombay Stock Exchange Sensitive Index	S&P BSE	India	2009-2010	2011-2012
10.	Standard & Poor's 500	S&P500	U.S.A.	2001-2006	2007-2012
11.	Shanghai Stock Exchange Composite Index	SSE	China	2007-2011	2012-2016
12.	Shenzhen Stock Exchange Composite Index	SZSE	China	1997-2000	2001-2004

Table 1: Stock indices

Algorithm 5 Binary portfolio strategy B

Initialize

A_0

$a_1 = 1$

for $k = 1, \dots, K$ **do**

    $A_k = a_k R_k A_{k-1} + (1 - a_k) A_{k-1}$

    if $\hat{\phi}_{k|k} > 0$ **then**

        $a_{k+1} = 1$

    else

        $a_{k+1} = 0$

    end

end

3.3 Data

The data sets were chosen as stock indices on 12 big markets around the world and can be seen along with their estimation and testing periods in table 1. They were chosen as to serve as market portfolios in alignment with Peng et al. (2016), and the same transformation $P_k = 100\,(log\,(Z_k))$ where Z_k denotes the closing spot price of the stock index, was used.

All data sets were split into two equally long parts, one for estimating parameters and one for out-of-sample testing. The estimation periods were chosen as time periods containing at least one clear spike and one clear decline in the asset price as to not make the estimation biased for either, and to make the models better work with drastically changing markets. The time periods for the different stock indices are different, since big declines in index prices on one market are likely correlated with declines on others in the world economy, and therefore same period-tests would probably give similar results for all indices and not test the strategies properties in different kinds of markets.

The parameters $\boldsymbol{\Theta}$ for the Extended and Unscented Kalman filters were estimated on the estimation data, as were the initial distribution of $\boldsymbol{z}^{(i)}$ in the Bootstrap Particle filter. The test data was then be used to evaluate the performance of the portfolio strategies as described in 3.4.

3.4 Comparison

After running the 7 portfolio strategies on the 12 data sets, their results had to be compared. The most intuitive and simple way would be simply comparing the average daily return of the portfolio strategies over a test period and comparing it—simply claiming that the portfolio with the highest expected return has used the best strategy. Since this is very similar to the the method used in the work of Peng et al. (2003) and Peng et al. (2016), who compared final asset values, and also allows for the use of a simple t-test for equal average returns, this was be the main focus of performance evaluation in this thesis. The null and alternative hypotheses for the daily returns were

$$\begin{aligned} H_0 :& \quad \bar{R}_p = \bar{R}_m \\ H_1 :& \quad \bar{R}_p > \bar{R}_m \end{aligned} \tag{3.2}$$

Since the forecasts for the market portfolio returns in the binary strategies is basically used as

$$a_{k+1} = \begin{cases} 1, & \hat{R}_{m,k+1} > 0 \\ 0, & \hat{R}_{m,k+1} \leq 0 \end{cases} \tag{3.3}$$

Henriksson-Merton's non-parametric market time is applicable. Therefore, the three binary portfolios were tested using the procedure described in 2.5.2.1. Since the threshold-based strategies can not be tested using the Henriksson-Merton methodology, the binary strategies served as representatives for their respective filters in the market timing test. The estimates $\hat{p_1}$, $\hat{p_2}$ and $\hat{p} = \hat{p_1} + \hat{p_2}$ are presented and hypotheses

$$\begin{aligned} H_0 :& \quad p_1 + p_2 = 1 \\ H_1 :& \quad p_1 + p_2 > 1 \end{aligned} \tag{3.4}$$

are tested. The values $\hat{p}_1$ and $\hat{p}_2$ are also useful in the analysis of the average daily returns, since they can be examined to see a strategies specific accuracy in predicting positive or negative returns. A more qualitative examination was also conducted by viewing the portfolio value plots through time and looking for instances were the strategic portfolios successfully avoided significant declines in asset value.

Both tests were beforehand decided to require rejection of the null hypothesis on at least the 5% confidence level to be deemed statistically significant.

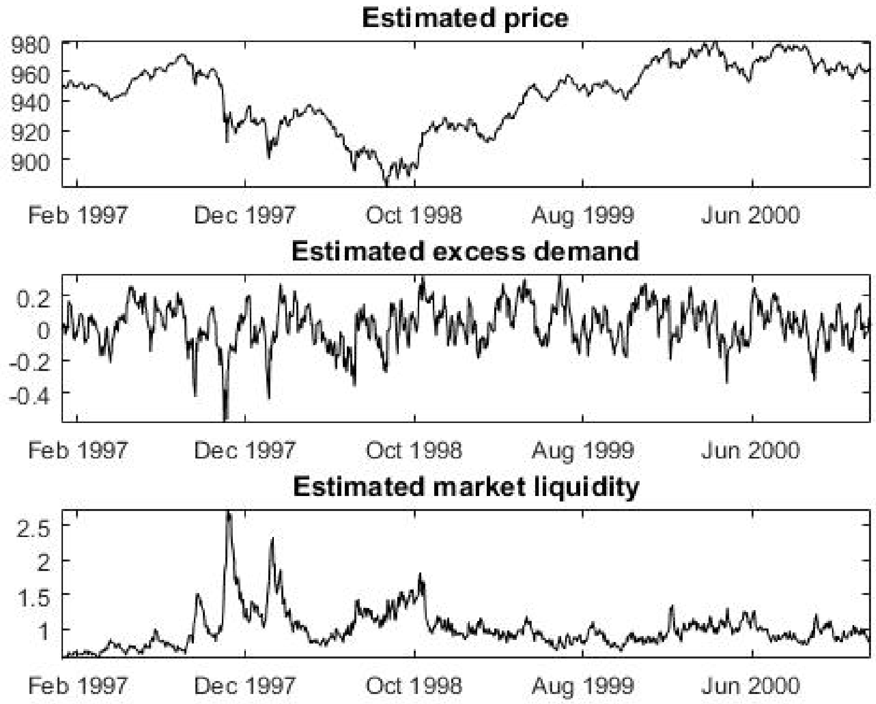

Figure 2: Estimated state variables using an Extended Kalman filter on the Hang Seng Index

4 Results

The average daily returns for the portfolios can be seen in table 2. The Extended Kalman filter is here shortened as EKF, the Unscented Kalman filter as UKF and the Bootstrap Particle filter as BPF. B denotes the binary strategies while τ denotes the threshold-based ones. All average daily returns that are less than those of the market portfolio are highlighted in grey. The results for the probability estimates $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}$ from Henriksson-Merton's market timing test can be seen in 3. All values of $\hat{p}_1$ or $\hat{p}_2$ that were less than 0.5, or values of $\hat{p}$ what were less than 1 are highlighted in grey. The best value for every data set is marked with a box. The subscripts *, ** and *** mean that the null hypothesis was rejected on the 5%, 1% or 0.1% significance levels respectively. The portfolio value plots can be found in figures 4-15 on pages 48-59.

An example of estimated state $\hat{\boldsymbol{X}}$ variables from the Extended Kalman filter on the Hang Seng Index can be seen in figure 2, with parameter estimates $\alpha_1 = 0.0029$, $\alpha_2 = -0.0003$, $\beta_1 = -0.12$, $\beta_2 = -0.0098$,$\gamma_1 = 0.0369$, $\gamma_2 = 0.0193$, $\gamma_3 = 4.2869$, $\delta = 0.7133$.

	Passive	EKF		UKF		BPF	
		τ	B	τ	B	τ	B
1. AOI	−0.17	0.02	0.06	−0.04	0.03	−0.19	−0.24
2. DAX	0.08	0.12	0.21	0.10	0.26	0.04	−0.02
3. FTSE	0.31	0.39	0.48	0.41	0.54	−0.01	0.03
4. HSI	−0.15	0.07	0.00	0.03	−0.04	−0.04	0.00
5. IBOVESPA	0.70	0.42	0.55	0.42	0.54	0.37	0.24
6. Nikkei	−0.19	−0.03	0.08	−0.05	0.11	−0.15	−0.08
7. OMXS30	−0.09	0.26	0.30	0.28	0.23	−0.32	−0.25
8. RTSI	0.84	0.56	0.42	0.58	0.56	0.12	0.25
9. S&P BSE	−0.17	−0.12	−0.17	−0.06	−0.14	−0.19	−0.10
10. S&P 500	−0.12	−0.02	0.00	−0.08	−0.01	−0.11	−0.26
11. SSE	0.19	0.35	0.37	0.38	0.37	0.37	0.32
12. SZSE	−0.53	−0.11	−0.09	−0.19	−0.02	−0.39	−0.56

Table 2: Average daily returns $R_{p,k}$ in ‰

	EKF			UKF			BPF		
	p_1	p_2	p	p_1	p_2	p	p_1	p_2	p
1. AOI	0.55	0.52	1.06^{**}	0.54	0.52	1.06^{**}	0.41	0.61	1.02^{*}
2. DAX	0.25	0.79	1.04^{**}	0.30	0.74	1.04^{**}	0.22	0.76	0.98
3. FTSE	0.49	0.57	1.05^{*}	0.49	0.59	1.07^{*}	0.50	0.56	1.06^{*}
4. HSI	0.44	0.56	1.00^{*}	0.45	0.55	1.00	0.52	0.48	1.00^{**}
5. IBOVESPA	0.43	0.59	1.02^{**}	0.43	0.59	1.03^{**}	0.18	0.81	0.99
6. Nikkei	0.72	0.27	1.00	0.75	0.25	1.00	0.64	0.37	1.01^{*}
7. OMXS30	0.43	0.65	1.08^{**}	0.36	0.74	1.10^{**}	0.60	0.36	0.96
8. RTSI	0.49	0.53	1.03^{**}	0.53	0.51	1.03^{**}	0.48	0.52	1.00^{**}
9. S&P BSE	0.52	0.44	0.96	0.51	0.45	0.96	0.44	0.55	0.99
10. S&P 500	0.56	0.43	0.99	0.56	0.43	0.99	0.45	0.53	0.98
11. SSE	0.51	0.51	1.02^{**}	0.50	0.53	1.02^{**}	0.48	0.55	1.03^{**}
12. SZSE	0.57	0.53	1.10^{***}	0.59	0.52	1.11^{***}	0.64	0.36	1.00^{***}

Table 3: Henriksson-Merton's non-parametric market timing test results

5 Discussion

The first impression from looking at the average daily returns in table 2 is that all four portfolios using Kalman filters yield higher returns than the passive strategy for all but two indices, even though the results are in no case statistically significant. This means that the comparative study failed to prove that any of the strategies yield statistically significantly higher returns than the passive strategy, i.e. the null hypothesis of the passive and strategic portfolios having the same mean daily return can not be rejected in any data set. Although this means nothing can, from a statistical perspective, be concluded, the positive results of these four strategies are still relevant from in an economical sense, since the strategic portfolios at least seem to produce higher returns.

Even though these four portfolios oftentimes yield negative returns, those returns are not as negative as the ones for the passive strategy. In a lot of cases, they are able to get positive average returns even in periods of financial decline—for example in the OMXS30 index where some of them all get big positive returns as compared to the $-0.09‰$ daily loss of the passive strategy. The IBOVESPA and RTSI indices both had test periods of very strong growth, with the passive portfolios having average daily returns 0.70‰ and 0.84‰ respectively and it seems that the strategic portfolios were not able to successfully follow the price upwards. Considering that the portfolios can never get a daily return of *more* than the passive portfolio, the DTMS-based portfolios strength contra the passive portfolio should be to avoid declines, while in times of financial posititvity they can—at best—tie with the passive portfolio. It therefore makes sense that they fail to outperform the passive portfolio in these two indices, since that would require them to very accurately pinpoint the few days of negative returns in the growth period. However, their actual average returns in these two periods are not very impressive as they are much lower than the passive one, making it reasonable to suspect that they consistently underestimate the excess demand in these two indices. Looking at the probabilities in table 3 can, however, not confirm this, since both binary portfolios based on Kalman filters have values of $\hat{p}_2 > 0.5$, meaning they were able to predict positive returns more often than they were unable. In the portfolio value plots in figures 8 and 11 it seems as if the excess return of the passive portfolio mostly come from short periods of very high positive returns that the strategic portfolios miss out on, explaining why they got lower average returns even though they had reasonable market timing. The Unscented Kalman filter using a binary asset allocation strategy actually has a higher portfolio value than the passive portfolio during the big financial peak of the series. When comparing the two, the Unscented Kalman filter yields the higher return in 6 of the data sets, compared to the Extended Kalman filter's 4. The binary strategies also seem to slightly outperform the threshold-based ones in terms of yielding high returns.

The particle filter performs worse in general. It is inefficient in most on the indices, rarely outperforming the passive portfolio, and never beating the Kalman filter-based portfolios. In the DAX data set, the binary particle filter portfolio has a negative return even though the passive portfolio has a positive return. It seems as if the particle filter has a tendency to severely underestimate the excess demand and stay out of potentially growing markets for long periods of time, see for example figures 6 and 10. This is also true for 7 out of 12 data sets in terms of market timing, where $\hat{p}_1$ is less than 0.5, and sometimes as low as around 0.2. It is not obvious where the source of this issue lies, but it might be connected to the high values of the observation covariance matrix chosen for the filter. The high values could have excessively smoothed out the particle weights, basically making the resampling not update the particles and thus simply keeping the same particles—that possibly underestimated the excess demand initially—for long time periods. This is something that could be examined more closely in future studies, and since this result contradicts that of Peng et al. (2016), it would be unfair to rule all formulations of particle filters out for trading applications.

As for the Henriksson-Merton test for market timing the results are similar to those from average returns but not identical. Here, a lot of the portfolios show statistical significance in correctly forecasting the next day market portfolio return, even when they were not able to achieve significantly higher returns. The null hypothesis of no market timing can be rejected for the Extended Kalman filter in 9 out of 12 data sets, for the Unscented Kalman filter in 8 of the data sets and for the Bootstrap Particle filter in 7 of the data sets. It is once again the two Kalman filters that prevail, but the particle filter is not as inefficient in market timing as it seems to be in yielding returns. The Extended Kalman filter shows significant market timing in the most data sets, but the Unscented Kalman filter gets the highest value of $\hat{p}$ in the most data sets. In a a statistical perspective, the Extended Kalman filter should be deemed the most efficient while in it is plausible to beliebe that the Unscented Kalman filter is a good, and sometimes better, option. The particle filter is able to get $\hat{p}$-values that are significantly higher than 1 in 7 data sets. This result is quite confusing as it doesn't fit well with the results in returns, but at least gives some hope to the filter. It should however be noted that it is the filter with the worst results in terms of market timing in general.

An interesting thing to note is that the values $\hat{p}_1$ are more often less than 0.5 than the values of $\hat{p}_2$ for all the portfolios. This implies that the filters are better at successfully predicting positive returns than negative ones. For example in the DAX data set, all the three portfolios have values of $\hat{p}_1$ around $0.2 - 0.3$. It is not very easy to spot this in figure 5, but it seems as if the portfolios make prediction errors on a lot of small declines while successfully avoiding the two big ones in 2008 and 2011—leading to a high return but a low $\hat{p}_1$. This also leads to the $\hat{p}_1$ giving a different implication than the average daily returns in terms of what strategic portfolio is the most efficient, since the

Henriksson-Merton test does not take return magnitude into account.

It is also worth noting that none of the portfolios using the Extended or Unscented Kalman filter manage to get $\hat{p} > 1$ in the S&P BSE and S&P 500 data sets even though they both yield higher returns than the passive portfolios. This seems to have to do with an inability to correctly forecast positive returns, something that can be seen in figures 12 and 13 where the market portfolio tends to get very volatile at times, with high negative returns following high positive ones, and the two strategic portfolios in question staying out of the market in these times. While this seems like a sensible strategy for avoiding too much risk and getting better returns overall, it also means missing out on days of positive returns.

When it comes to the more qualitative study of the plots the results are similar. The Unscented and Extended Kalman filters seem to be able to avoid price declines accurately in most cases, with the binary strategies being slightly more quick to adapt in big crisises. However, most of the portfolios take a few days to adjust to severe declines, meaning they follow the market down at first before they exit, as can be seen clearly in for example 5, 6 and 7. Oftentimes they also re-allocate the assets into the market portfolio too quickly and have to exit again, giving rise to a "stair-shaped" price decline as seen clearly in for example figures 7 or 8. In other times the portfolios instead take on a very careful strategy and do not re-invest in the market portfolio until after a very long time after a crisis, see for example on the Nikkei index in figure 9. This also gets the portfolios very low values of $\hat{p}_2$, since they miss a lot of positive returns, while $\hat{p}_1$ gets very high due to this over-pessimistic behaviour. It is unclear why any of these things happen, but one fault might lie in the parameter estimation. Since the testing periods sometimes differed a lot from the period where the parameters where estimated, the predictions might be badly adjusted for different times. While this problem is difficult to get past, since no one in practice would know anything about the testing period when doing the estimation, it should be possible to mitigate using longer estimation periods or recursively updated parameters. Though recursively updated parameters was one of the reasons to use the particle filters, it didn't live up to expectations in that sense. Instead, perhaps a Dual Kalman filter, i.e. two filters running in parallel with one estimating state and one estimating parameters, could be useful.

One suspicion that arose during the thesis process was that the DTMS model would overimply autocorrelation in the returns, conflicting too much with the random walk hypothesis. Due to the structure of the price dynamics, where returns depend on the autoregressive process of excess ϕ, the DTMS model might claim that the autocorrelation in the returns is significant while in the real data it is not. Even though this thesis is not meant to actually deduce whether or not the DTMS model is correct per se, it should be of concern for anyone seeking to to apply the model in portfolio strategies—no matter how good your filter is, it is improbable that any application will be success-

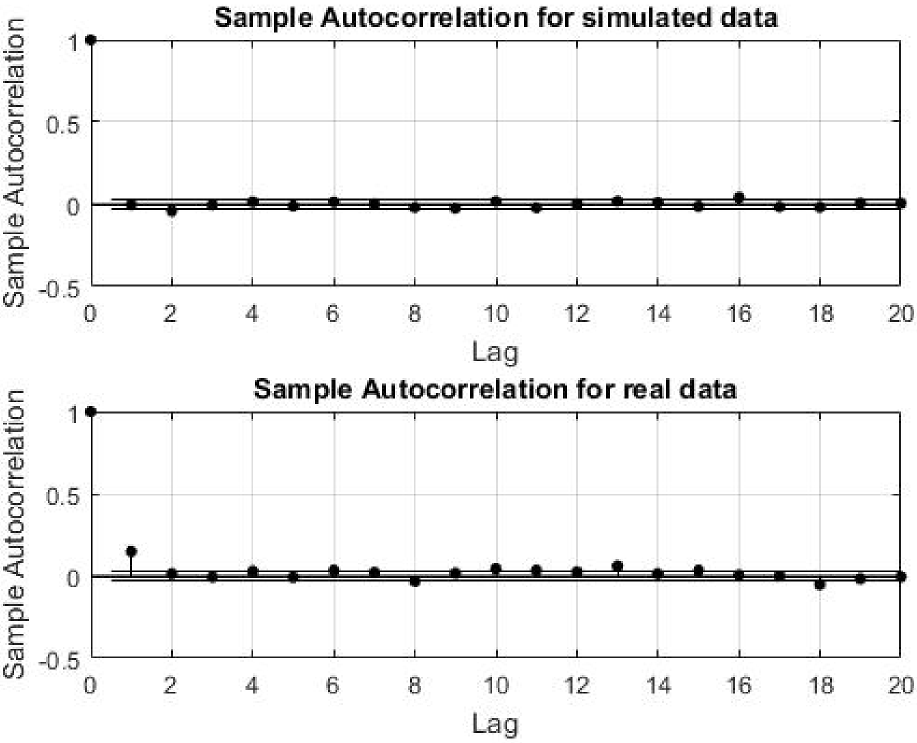

Figure 3: Autocorrelation of simulated and real data returns

ful if the underlying model specification is erroneous—and might have been a source to all the non-significant results in average daily returns. To just briefly address this problem, some price data was simulated as in figure 1, and the autocorrelation of its returns was compared to those of real data. This was examined in a lot of simulation runs and all the data sets, and the conclusion was, if anything, that the DTMS model rather *under*-implies the autocorrelation, since some data sets, especially those with big declines or peaks, showed significant autocorrelation in the returns for a few lags while the simulated data rarely did. An example of autocorrelations for simulated and real data can be seen in figure 3, where the simulated data is the same as the one in figure 1 and the real data is from the RTSI data set.

5.1 Conclusion

None of the strategies based on the DTMS model were able to produce returns that were significantly higher than those of the passive strategy, why the null hypothesis of the mean return being the same for the strategic and passive portfolios could not be rejected. However, using some filtering techniques, i.e. The Extended and Unscented formulations of the Kalman filter, the returns were in general higher than those of the passive strategy in most cases, excluding times of very high financial growth, though the lack of statistical significance makes it difficult to actually conclude this. The comparative study was unable to prove anything regarding the returns of these strategies, while the results imply, but do not prove, that they can produce higher returns than a passive strategy.

As for market timing, all filtering techniques Extended Kalman, Unscented Kalman and Bootstrap Particle were able to reject the null hypothesis of no market timing ability in 9, 8 and 7 of the data sets respectively. This shows great promise in the strategies and their ability to forecast returns, even though it was not sufficient to produce statistically significantly higher returns than a passive portfolio.

When it comes to choice of filtering technique, there are indications that the Unscented Kalman filter works better in terms of average daily returns. However, these results were not by any significant margin, why it is difficult to conclude which one of the filters is to prefer. In terms of market timing, the Extended Kalman filter rejected the null hypothesis of no market timing ability in one more data set that the Unscented Kalman filter, but got lower values for the market timing measure in most data sets where they were both significant. It is therefore difficult to conclude which of the filters is preferable, though once again they both outperform the Bootstrap particle filter. It should also be noted in the Extended Kalman filte's favour that it slightly less demanding in terms of computational power.

In terms of which strategy to use for asset allocation based on estimated excess demand, the binary strategy produces seemingly higher returns in most data sets. The difference is, however, not too apparent, why the threshold-based version might be more useful when taking transaction costs into account.

It is evident from the study that the Bootstrap Particle filter, as used in this thesis, is unreliable. It was rarely of use in portfolio strategy, no matter whether the threshold-based or binary strategy was used, even though it oftentimes had significant market timing. Since this contradicts the results of Peng et al. (2016), there is reason to believe that the alteration of the method, i.e. moving the measurement variance out of the sample particles and increasing it by a factor, was not successful. It is however clear that the application as used in this thesis is not to recommend for any investor.

In terms of in which kinds of markets the models worked and in which not, the strategic portfolios generally yielded higher returns than the passive portfolio in periods including at least one big decline, such as the 2008 financial crisis, while in times of financial positivity and generally rising markets, they performed worse. As for market timing, no specific type of market seemed to give different results, as the strategies managed to reject the null hypothesis in both rising and falling markets in some cases and not in others.

5.2 Suggested further studies

The Self-organizing State Space model, as used in this thesis, did not perform as well as one could have expected after reading Peng et al. (2016). It seems like this filter is sensitive to inital conditions and alterations in the algorithm, and more studies that closely examine the effect of these could be of use. Other ways to mitigate the issue of the particles having different covariance matrices for the importance weighting could also be of use.

Another interesting result is that the binary strategy often outperformed the threshold-based one in terms of high returns. This result could be tested further, especially when considering transaction costs, which are bound to be higher for the binary strategy.

The maximum likelihood parameter estimation, as proposed by Peng et al. (2003) and used in this thesis, was questioned by Peng et al. (2016), but in this thesis the strategic portfolios using maximum likelihood performed better than the ones using the Bootstrap Particle filter. There are other ways to carry out this estimation, such as dual Kalman filters, where two filters run parallel to each other. One filter estimates state and the other estimates parameters. This could also prove effective and be an interesting topic of study.

The threshold-based portfolio uses different values of $\hat{\phi}$ to decide how much should be allocated in the risky portfolio. Due to the dynamics of price as described in equation 2.6, it seems peculiar to base it solely on the $\hat{\phi}$-value, since the same excess demand can cause very different price changes depending on the value of λ. Therefore, these threshold-based strategies could possibly benefit from using the estimate $\hat{\lambda}$ too, something that could quite easily be tested.

The autocorrelation in returns seemed to be higher in the real data than in data simulated using the DTMS model formulation, though this was not tested using any significance test. Therefore, the DTMS model could benefit from adding parameters to the price estimate P_k that would increase the autocorrelation in returns—such as a two step lagged price P_{k-2}.

Acknowledgements

I would first and foremost like to thank my supervisor Birger Nilsson for providing help and inspiration as well as being supportive in the thesis project. I would also like to thank Johane Hedin for proof-reading and Johannes Solheim Karlsson for proof-reading and providing advice and feedback in the process of writing this paper.

Bibliography

Aragon, G. O. and Ferson, W. E. (2007). Portfolio performance evaluation. *Foundations and Trends in Finance*, 2(2):83–190.

Bouchaud, J.-P. and Cont, R. (1998). A Langevin approach to stock market fluctuations and crashes. *The European Physical Journal B - Condensed Matter and Complex Systems*, 6(4):543–550.

Brogan, W. L. (1991). *Modern Control Theory (3rd Ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Brooks, C. (2014). *Introductory Econometrics for Finance*. Cambridge University Press.

Doucet, A., De Freitas, N., and Gordon, N. (2001). An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer New York.

Drakakis, K. (2009). Application of signal processing to the analysis of financial data [in the spotlight]. *IEEE Signal Processing Magazine*, 26(5):160–158.

Fama, E. F. (1965). Random walks in stock-market prices. *Financial Analysts Journal*, 21:55–59.

Fuller, W. (1996). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. Wiley.

Henriksson, R. D. and Merton, R. (1981). On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills. *The Journal of Business*, 54(4):513–33.

Iino, M. and Ozaki, T. (2000). A nonlinear model for financial dynamics. In *Proceeding of the international symposium on frontiers of time series modeling*, pages 7–9.

Jakobsson, A. (2013). *An Introduction to Time Series Modeling*. Studentlitteratur AB.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

Kitagawa, G. (1998). A self-organizing state-space model. *Journal of the American Statistical Association*, 93(443):1203–1215.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335.

Lindström, E., Madsen, H., and Nygaard Nielsen, J. (2015). *Statistics for Finance*. Chapman and Hall.

Lo, A. and Hasanhodzic, J. (2011). *The Evolution of Technical Analysis: Financial Prediction from Babylonian Tablets to Bloomberg Terminals*. Bloomberg. Wiley.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1):59–82.

McLeod, W. and Van Vuuren, G. (2004). Interpreting the sharpe ratio when excess returns are negative. *Investment Analysts Journal*, 2004(59):15–20.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: a New Approach. *Econometrica*, 59:347–370.

Peng, H., Gan, M., Chen, L., and Zhang, C. Y. (2016). A self-organizing state space type microstructure model for financial asset allocation. *IEEE Access*, 4:8035–8043.

Peng, H., Kitagawa, G., Tamura, Y., Xi, Y., Qin, Y., and Chen, X. (2015). A modeling approach to financial time series based on market microstructure model with jumps. *Applied Soft Computing*, 29:40 – 51.

Peng, H., Ozaki, T., and Haggan-Ozaki, V. (2003). Modeling and asset allocation for financial markets based on a discrete time microstructure model. *The European Physical Journal B - Condensed Matter and Complex Systems*, 31(2):285–293.

Peng, H., Tamura, Y., Gui, W., and Ozaki, T. (2005). Modelling and asset allocation for financial markets based on a stochastic volatility microstructure model. *Intern. J. Syst. Sci.*, 36(6):315–327.

Strömberg, A. (2006). Optimal dynamic asset allocation using a non-linear discrete time liquidity driven microstructure market model and extended kalman filtering. Student Paper.

Varian, H. R. (2010). *Intermediate Microeconomics: A Modern Approach*. W.W. Norton Co., New York, eighth edition.

Wan, E. A. and van der Merwe, R. (2000). The Unscented Kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pages 153–158.

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35.

Appendices

A State space representation

A *state* is described by (Brogan, 1991, p. 72) as a complete summary of the status of a system. The state of a system can be described at any time point k as a vector of n variables $\boldsymbol{X}_k = [X_{1,k}X_{2,k}\ldots X_{n,k}]$, called the *state variables* (Brogan, 1991, p. 72). The state of the system then transitions step-wise through time using the *transition function* $\boldsymbol{F}$ as $\boldsymbol{X}_{k+1} = \boldsymbol{F}(\boldsymbol{X}_k, \boldsymbol{e}_{k+1})$, or in a simple linear case, $\boldsymbol{X}_{k+1} = \boldsymbol{A}\boldsymbol{X}_k + \boldsymbol{e}_{k+1}$ for a $(n \times n)$-matrix of parameters $\boldsymbol{A}$ (Brogan, 1991, p. 77) and noise term $\boldsymbol{e}_{k+1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_e)$ for some covariance matrix $\boldsymbol{S}_e$.

Then introduce an output, or *observation*, vector $\boldsymbol{Y}_k = [Y_{1,k}Y_{2,k}\ldots Y_{m,k}]$ of m variables that all depend on the current state of the system as $\boldsymbol{Y}_k = \boldsymbol{H}(\boldsymbol{X}_k, \boldsymbol{w}_k)$ for some vector valued function H or the simple linear system described by $\boldsymbol{Y}_k = \boldsymbol{C}\boldsymbol{X}_k + \boldsymbol{w}_k$ where $\boldsymbol{C}$ is a $(m \times n)$ matrix of parameters and noise term $\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_w)$ for some covariance matrix $\boldsymbol{S}_w$ (Brogan, 1991, p. 78). The system can then be described in its entirety as

$$\boldsymbol{X}_{k+1} = \boldsymbol{A}\boldsymbol{X}_k + \boldsymbol{e}_{k+1} \tag{A.1}$$

$$\boldsymbol{Y}_k = \boldsymbol{C}\boldsymbol{X}_k + \boldsymbol{w}_{k+1} \tag{A.2}$$

$$\boldsymbol{e}_{k+1} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_e) \tag{A.3}$$

$$\boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{S}_e) \tag{A.4}$$

(Brogan, 1991, p. 79). For more complex systems, it might be impossible to write the system equations in this linear form, while in other systems it is possible by making the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ depend on time and or the current state or measurement vectors $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$.

State observability is a property of such a system as the one described above where $\boldsymbol{X}_k$ *can* be estimated, and is closely connected to the number of state- and observation variables (Jakobsson, 2013, p. 284). For example, a system with one observed variable and 10 state variables is in general not observable.

Since $\boldsymbol{X}_k$ is generally not observeable, some filtering technique has to be used to estimate the current and future states of the system, and through that predict future output $\boldsymbol{Y}_{k+1}$

A.1 Space state representation example

Consider the *growth* of two different plants 1 and 2, denoted G_1 and G_2 respectively. The growth of these specific plants is dependent on three things—the current outside temperature T, the level of rain R and the amount of healthy minerals in the ground m. The growth at time point k can be calculated using the functions

$$G_{1,k} = c_{1,1} + c_{1,2}T_k + c_{1,3}R_k + c_{1,4}m_k + w_{1,k} \tag{A.5}$$

$$G_{2,k} = c_{2,1} + c_{2,2}T_k + c_{2,3}R_k + c_{2,4}m_k + w_{1,k} \tag{A.6}$$

for noise terms $w_{1,k} \sim \mathcal{N}\left(0, \sigma^2_{w,1}\right)$ and $w_{2,k} \sim \mathcal{N}\left(0, \sigma^2_{w,2}\right)$. The variables T, R and m follow processes

$$T_{k+1} = W_k + e_{1,k+1} \tag{A.7}$$

$$R_{k+1} = a_{1,1} + a_{1,2}T_k + e_{2,k+1} \tag{A.8}$$

$$h_{k+1} = a_{3,1} + a_{3,2}T_k + a_{3,4}h_k + e_{3,k+1} \tag{A.9}$$

with noise terms $e_{1,k} \sim \mathcal{N}\left(0, \sigma^2_{e,1}\right)$, $e_{2,k} \sim \mathcal{N}\left(0, \sigma^2_{e,2}\right)$ and $e_{3,k} \sim \mathcal{N}\left(0, \sigma^2_{e,3}\right)$. By introducing matrices

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ a_{3,1} & a_{3,1} & 0 & a_{3,4} \end{bmatrix} \tag{A.10}$$

and

$$\boldsymbol{C} = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{3,4} \end{bmatrix} \tag{A.11}$$

the system can then be formulated using its *state space representation*

$$\boldsymbol{X}_{k+1} = \boldsymbol{A}\boldsymbol{X}_k + \boldsymbol{e}_k \tag{A.12}$$

$$\boldsymbol{Y}_k = \boldsymbol{C}\boldsymbol{X}_k + \boldsymbol{w}_k \tag{A.13}$$

for state vector

$$\boldsymbol{X}_k = \begin{bmatrix} 1 & T_k & R_k & m_k \end{bmatrix}^T, \tag{A.14}$$

observation vector

$$\boldsymbol{Y}_k = \begin{bmatrix} G_{1,k} & G_{2,k} \end{bmatrix}^T, \tag{A.15}$$

and noise vectors $\boldsymbol{e}_k \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{S}_e\right)$ and $\boldsymbol{w}_k \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{S}_w\right)$ with covariance matrices

$$\boldsymbol{S}_e = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{e,1} & 0 & 0 \\ 0 & 0 & \sigma^2_{e,2} & 0 \\ 0 & 0 & 0 & \sigma^2_{e,3} \end{bmatrix} \tag{A.16}$$

$$\boldsymbol{S}_w = \begin{bmatrix} \sigma^2_{w,1} & 0 \\ 0 & \sigma^2_{w,2} \end{bmatrix}. \tag{A.17}$$

The state of the system can then be estimated using the observations and filtering techniques such as the Kalman filter.

A.2 Kalman filter example

Consider the following example of Kalman filter usage from Jakobsson (2013). The unknown constant

$$x_k = x_{k-1} \tag{A.18}$$

has measurements corrupted as

$$y_k = x_k + w_k \tag{A.19}$$

with observation noise term $w_k \sim \mathcal{N}(0, \sigma_w^2)$. The space state representation of this can, for clarification, be written as

$$\boldsymbol{X}_k = \boldsymbol{A}\boldsymbol{X}_{k-1} \tag{A.20}$$

$$\boldsymbol{Y}_k = \boldsymbol{C}\boldsymbol{X}_k + \boldsymbol{w}_k \tag{A.21}$$

where $\boldsymbol{X}_k = x_k$, $\boldsymbol{Y}_k = y_k$, $\boldsymbol{A} = \boldsymbol{C} = 1$ and $\boldsymbol{w}_k = w_k$. The optimal linear reconstruction of x_k is then

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left(y_k - \hat{x}_{k|k-1}\right) \tag{A.22}$$

and since the covariance of x here is constant,

$$\boldsymbol{S}^{x,x}_{k+1|k} = \boldsymbol{S}^{x,x}_{k|k} = \sigma^2_{x,k} \tag{A.23}$$

and the conditional covariance of y is

$$\boldsymbol{S}^{y,y}_{k+1|k} = \boldsymbol{S}^{x,x}_{k+1|k} + \sigma_w^2 = \sigma^2_{x,k} + \sigma^2. \tag{A.24}$$

Given initial estimates x_0 and $\boldsymbol{S}^{x,x}_{1|0} = \sigma^2_{x,0}$, then

$$\begin{aligned}
\sigma^2_{x,1} &= \frac{\sigma^2_{x,0}}{\left(1+\frac{\sigma^2_{x,0}}{\sigma^2_w}\right)} \\
\sigma^2_{x,2} &= \frac{\sigma^2_{x,1}}{\left(1+\frac{\sigma^2_{x,1}}{\sigma^2_w}\right)} = \frac{\sigma^2_{x,0}}{\left(1+2\frac{\sigma^2_{x,0}}{\sigma^2_w}\right)} \\
&\vdots \\
\sigma^2_{x,k} &= \frac{\sigma^2_{x,0}}{\left(1+k\frac{\sigma^2_{x,0}}{\sigma^2_w}\right)}
\end{aligned} \tag{A.25}$$

leading to

$$K_k = \frac{1}{\left(k + \frac{\sigma^2_w}{\sigma^2_{x,0}}\right)} \tag{A.26}$$

which gives the state estimate

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \frac{1}{\left(k + \frac{\sigma_w^2}{\sigma_{x,0}^2}\right)} \left(y_k - \hat{x}_{k|k-1}\right). \tag{A.27}$$

Then, consider there is no prior knowledge about the value of x, why $\sigma_{x,0}^2$ is set as high as possible, leading to

$$\lim_{\sigma_{x,0}^2 \to \infty} K_k = \frac{1}{k} \tag{A.28}$$

making the state estimate

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \frac{1}{k} \left(y_k - \hat{x}_{k|k-1}\right) \tag{A.29}$$

which is the same as a recursive formulation of the *sample mean*

$$\hat{x}_{k|k} = \frac{1}{k} \sum_{k=1}^{k} y_k \tag{A.30}$$

which is the estimator most commonly used for such an estimation where the observation is corrupted by a zero-mean white noise process.

A.3 Bootstrap particle filter example

Consider a random walk process

$$x_{k+1} = x_k + e_k \tag{A.31}$$

which is observed with a measurement noise as

$$y_k = x_k + w_k \tag{A.32}$$

where $e_k \sim \mathcal{N}(0, \sigma_e^2)$ and $w_k \sim \mathcal{N}(0, \sigma_w^2)$ for known variances σ^2. Now, make a first estimation guess $\boldsymbol{x}_{k|k}$ and sample 100 particles

$$x_{k+1|k}^{(i)} \sim \mathcal{N}\left(\boldsymbol{x}_{k|k}, \sigma_e^2\right). \tag{A.33}$$

Some of these particles probably go in the right direction, and get values fairly close to the actual value of x_{k+1}, while others end up inaccurate. To tackle this, evaluate their accuracy using the probability density function for y and give them weights

$$\omega_{k+1}^{(i)} = p_N\left(y_1, x_{k+1|k}^{(i)}, \sigma_w^2\right) \tag{A.34}$$

where $p_N(y, \bar{y}, \sigma^2)$ means the probability density function or y given mean $\bar{y}$ and variance σ^2. Now, particles who ended up close to the true state x_k should have gotten higher weights than those far away from it. Now, normalize the weights

$$\tilde{w}_{k+1}^{(i)} = \frac{\omega_{k+1}^{(i)}}{\sum_{i=1}^{M} \omega_{k+1}^{(i)}} \tag{A.35}$$

and sample a new set of particles from the old ones, with the normalized weights as discrete probabilities of every specific particle being chosen

$$\left\{x_{k+1|k+1}^{(i)}\right\} \sim \left[\left\{x_{k+1|k}^{(i)}, \tilde{w}_{k+1}^{(i)}\right]\right\} \tag{A.36}$$

yielding a set of particle hopefully all pretty close to the current state x_{k+1}. Form the current estimate as the sample mean

$$\hat{x}_{k+1|k+1} = \frac{1}{M} \sum_{i=1}^{M} x_{k+1|k+1}^{(i)}. \tag{A.37}$$

Repeat the process for every time step to follow the state using the particles estimate, and reevaluate the weights for every observation y_k.

B Portfolio value plots

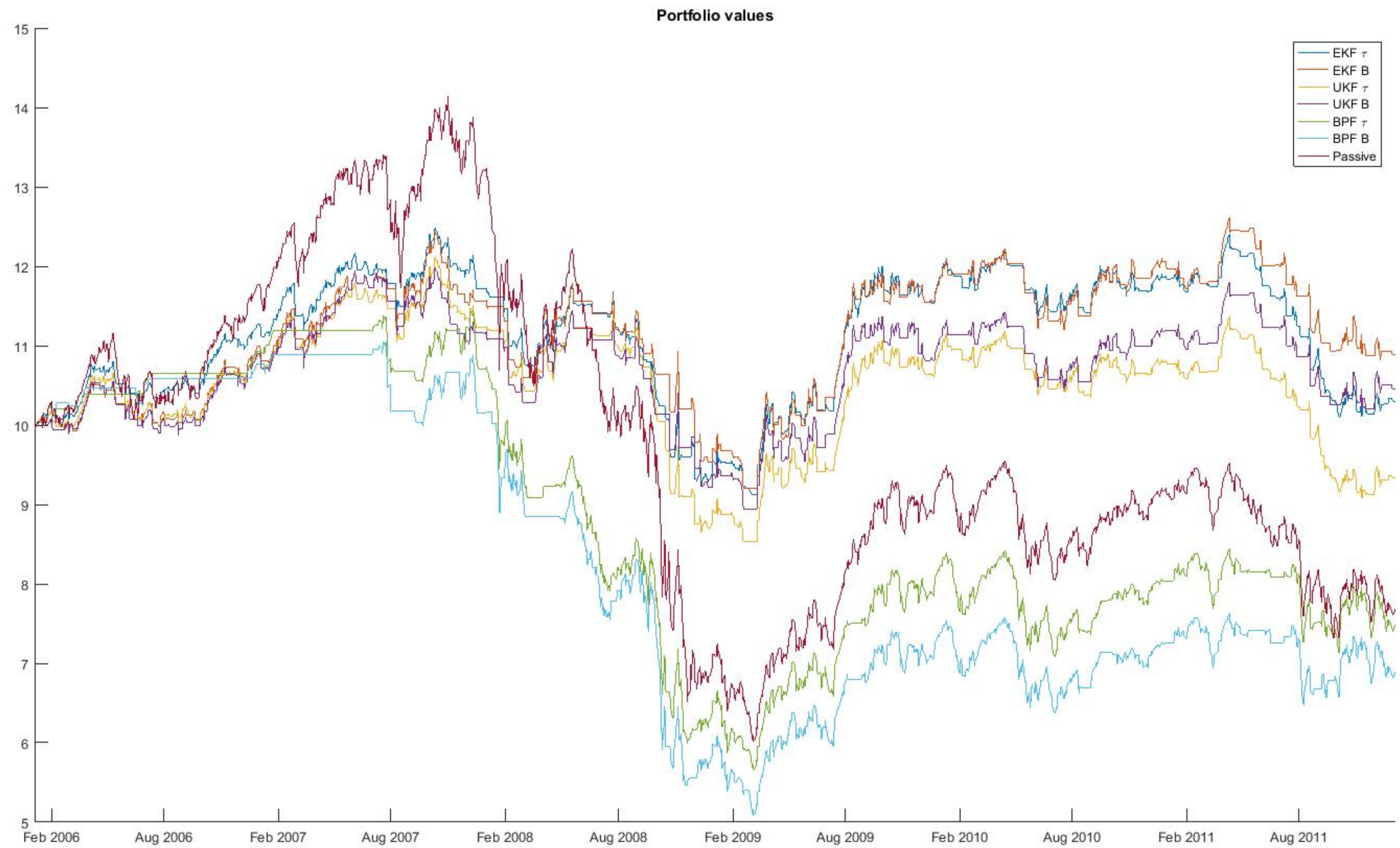

Figure 4: All Ordinaries

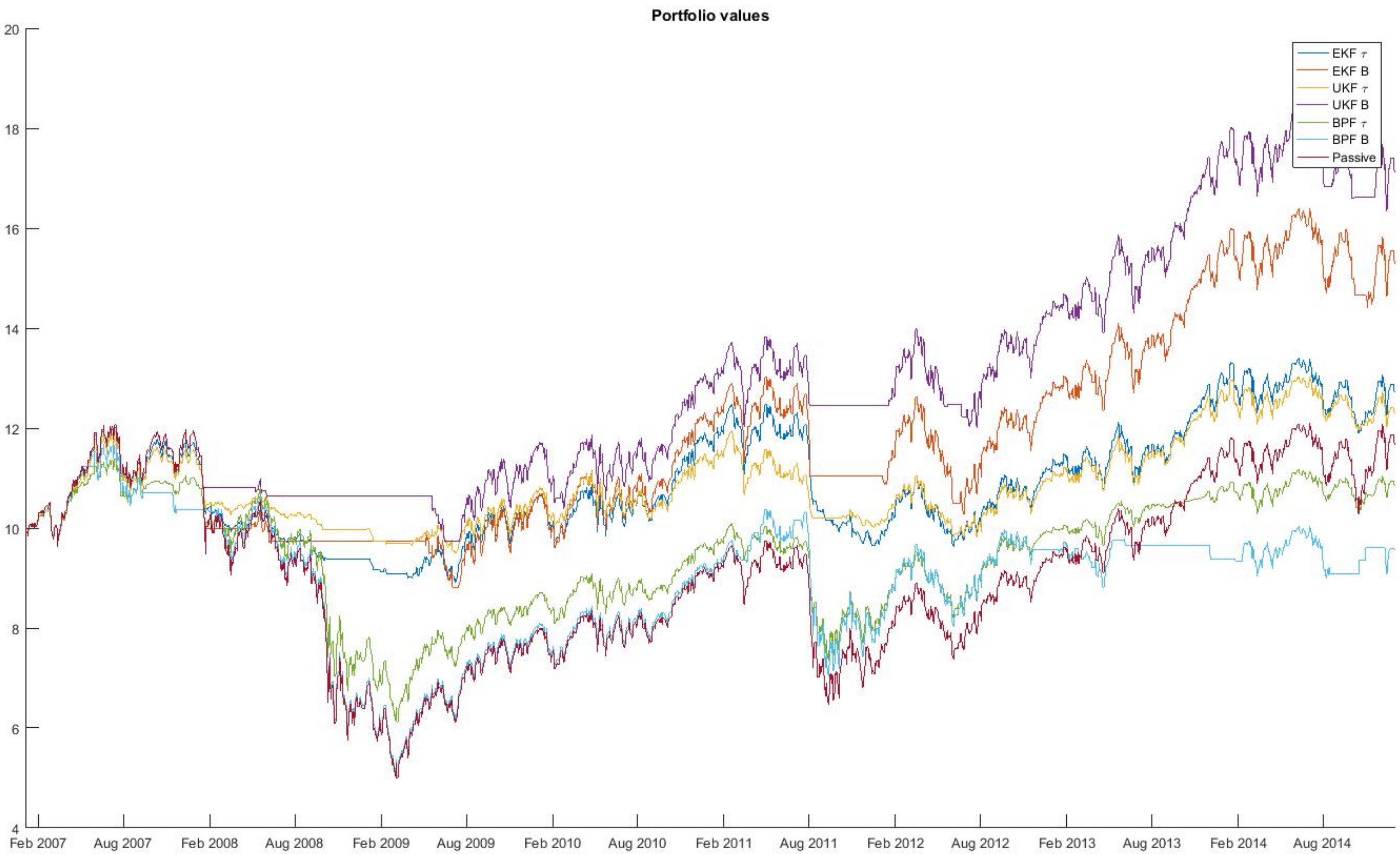

Figure 5: Deutscher Aktienindex

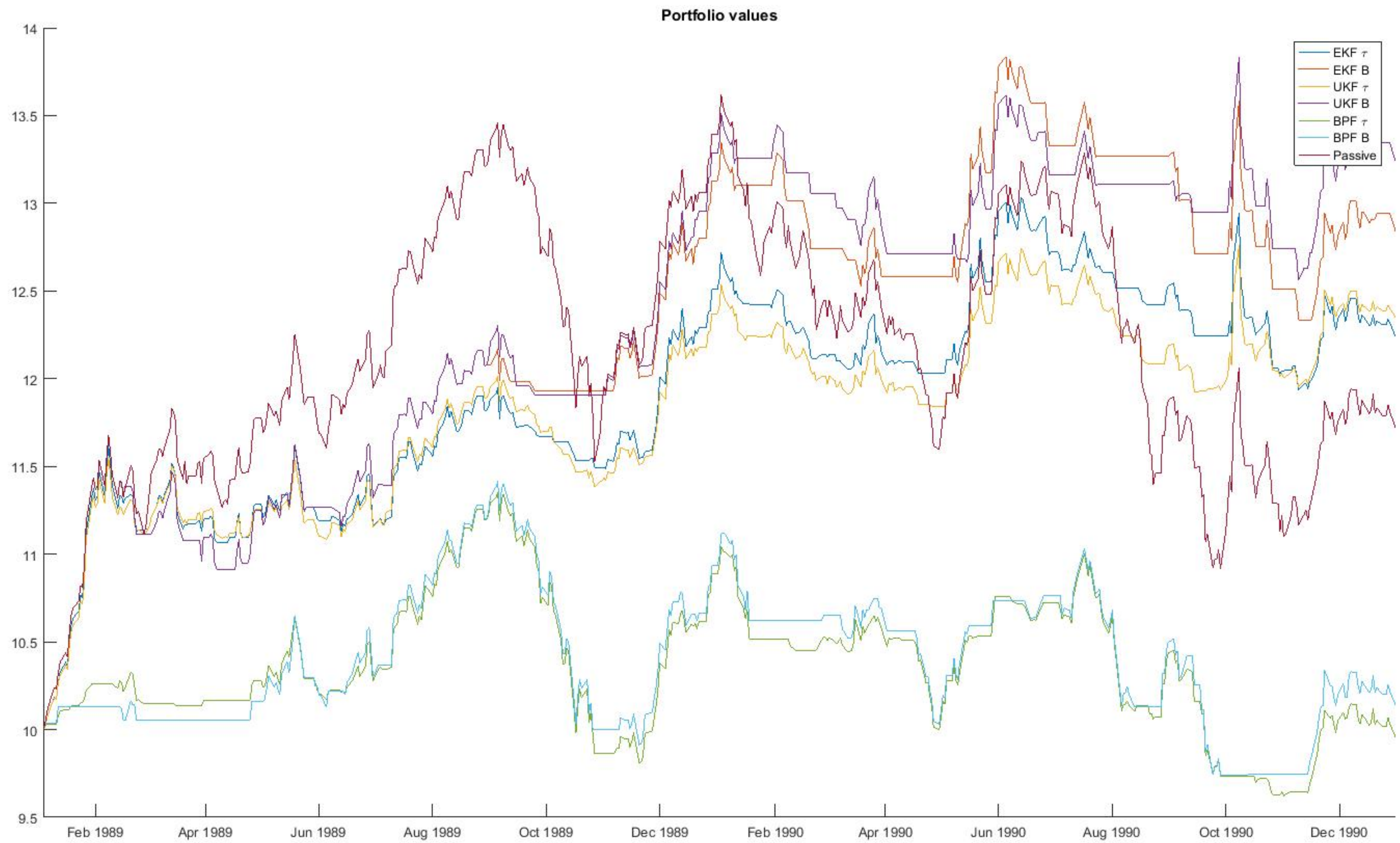

Figure 6: Financial Times Stock Exchange 100

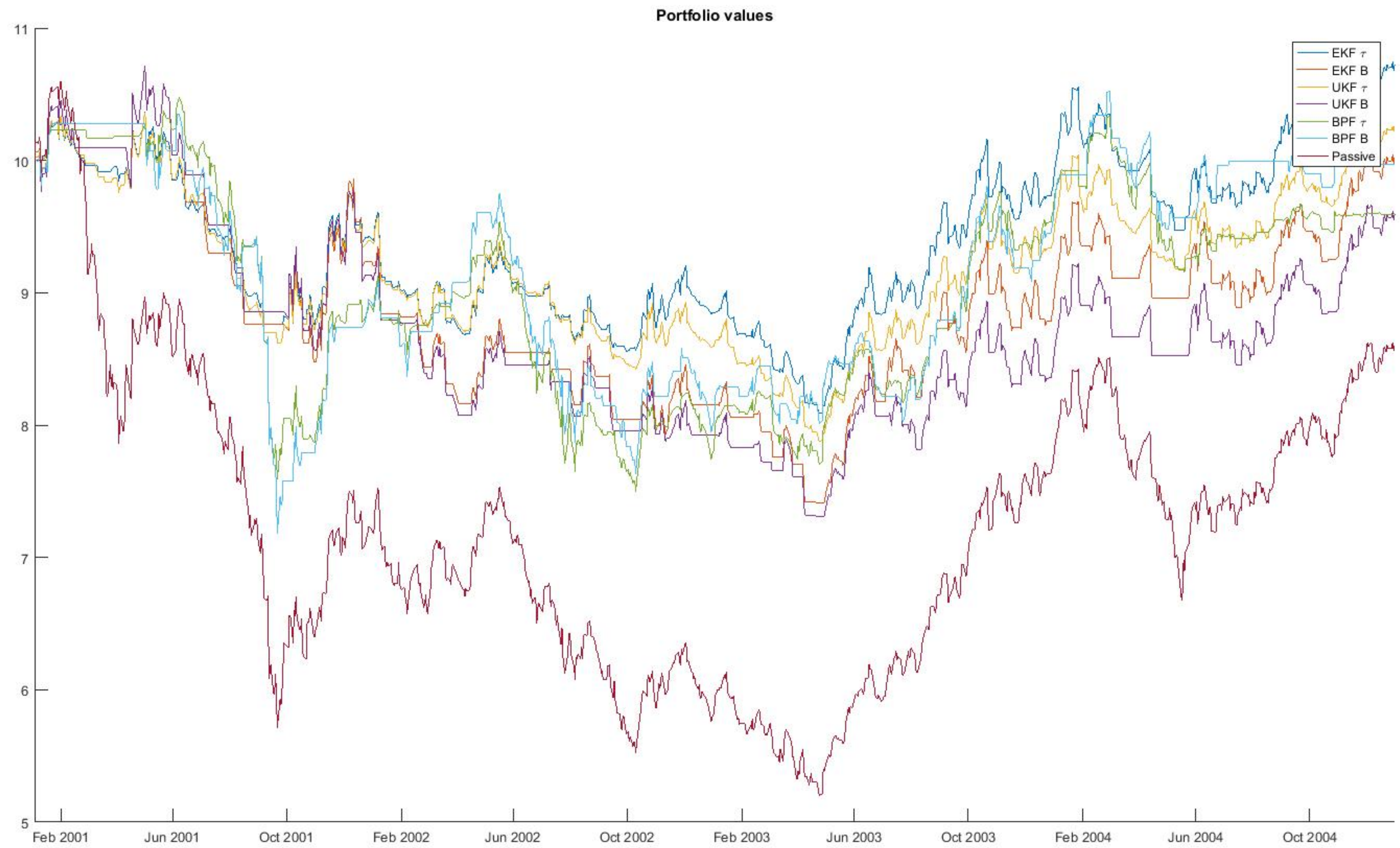

Figure 7: Hang Seng Index

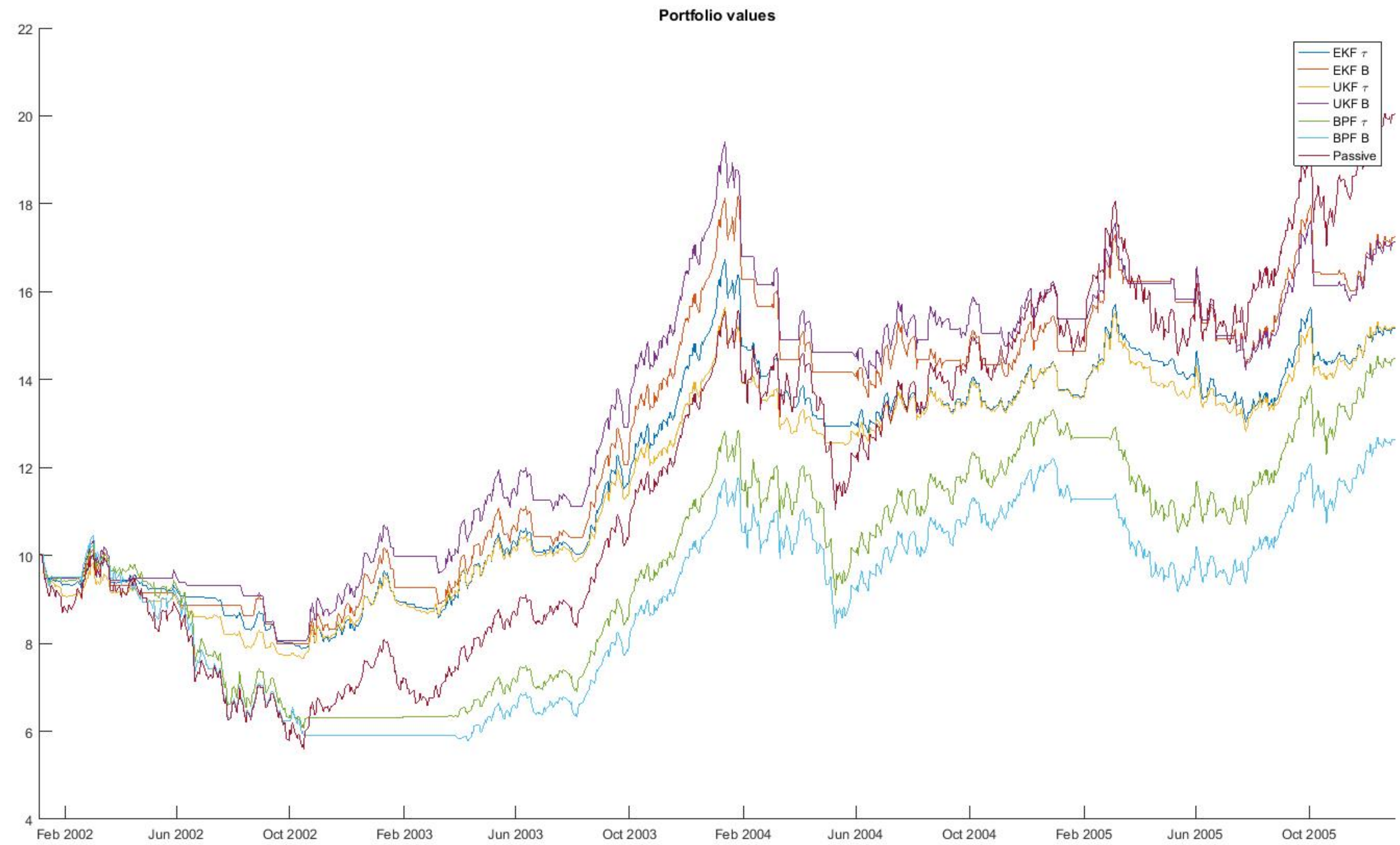

Figure 8: IBOVESPA

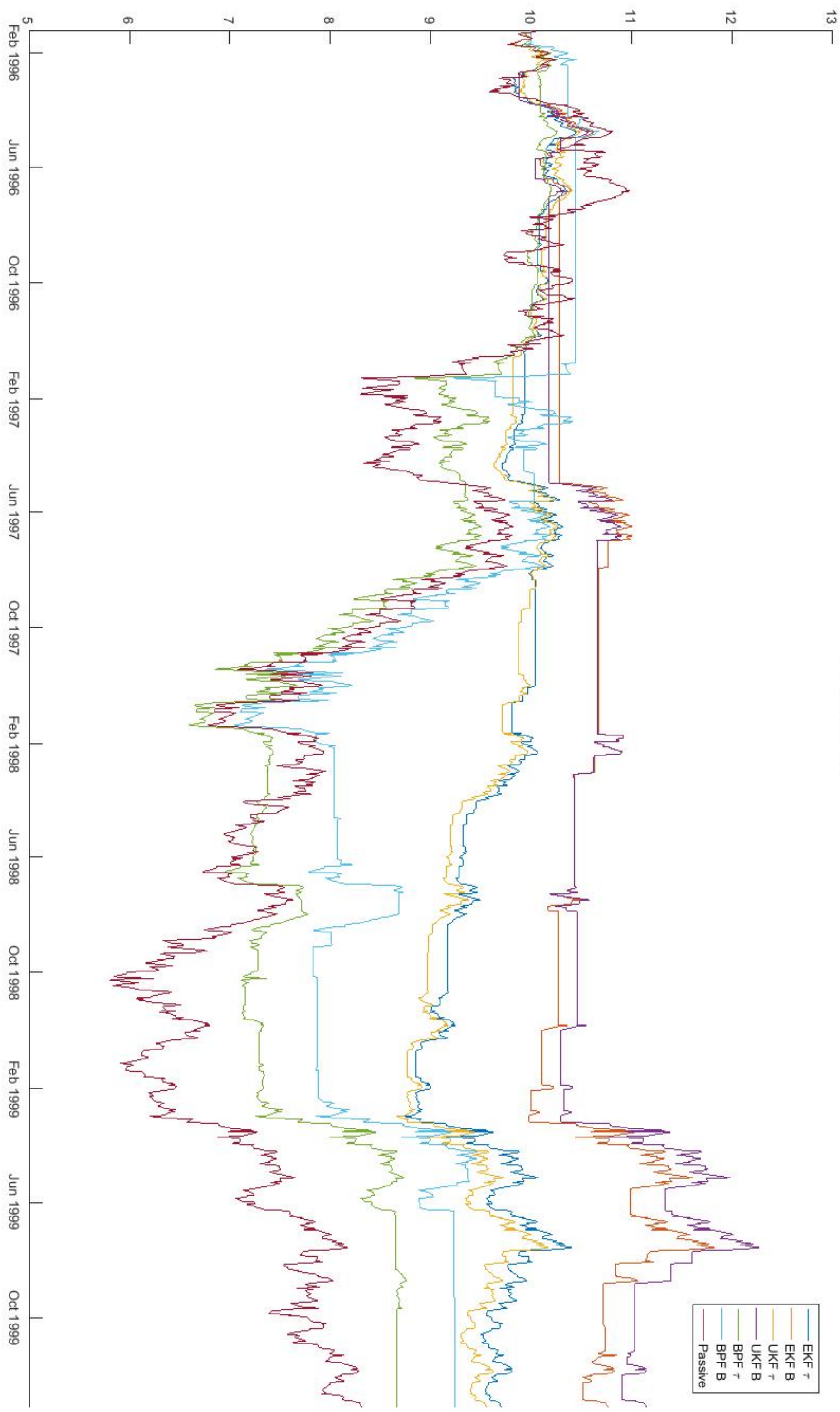

Figure 9: Nikkei 225

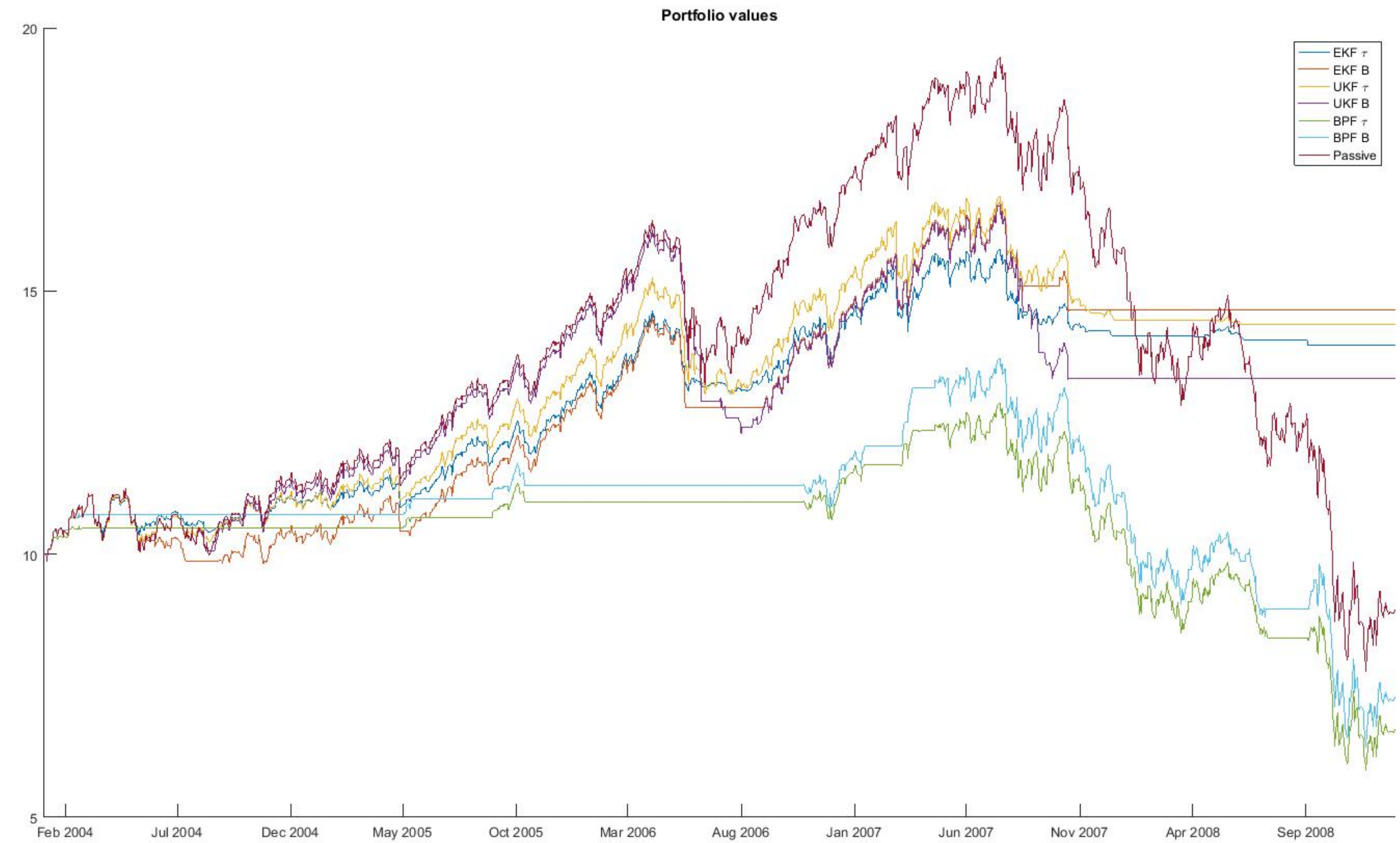

Figure 10: OMX Stockholm 30

Figure 11: Russian Trading System Index

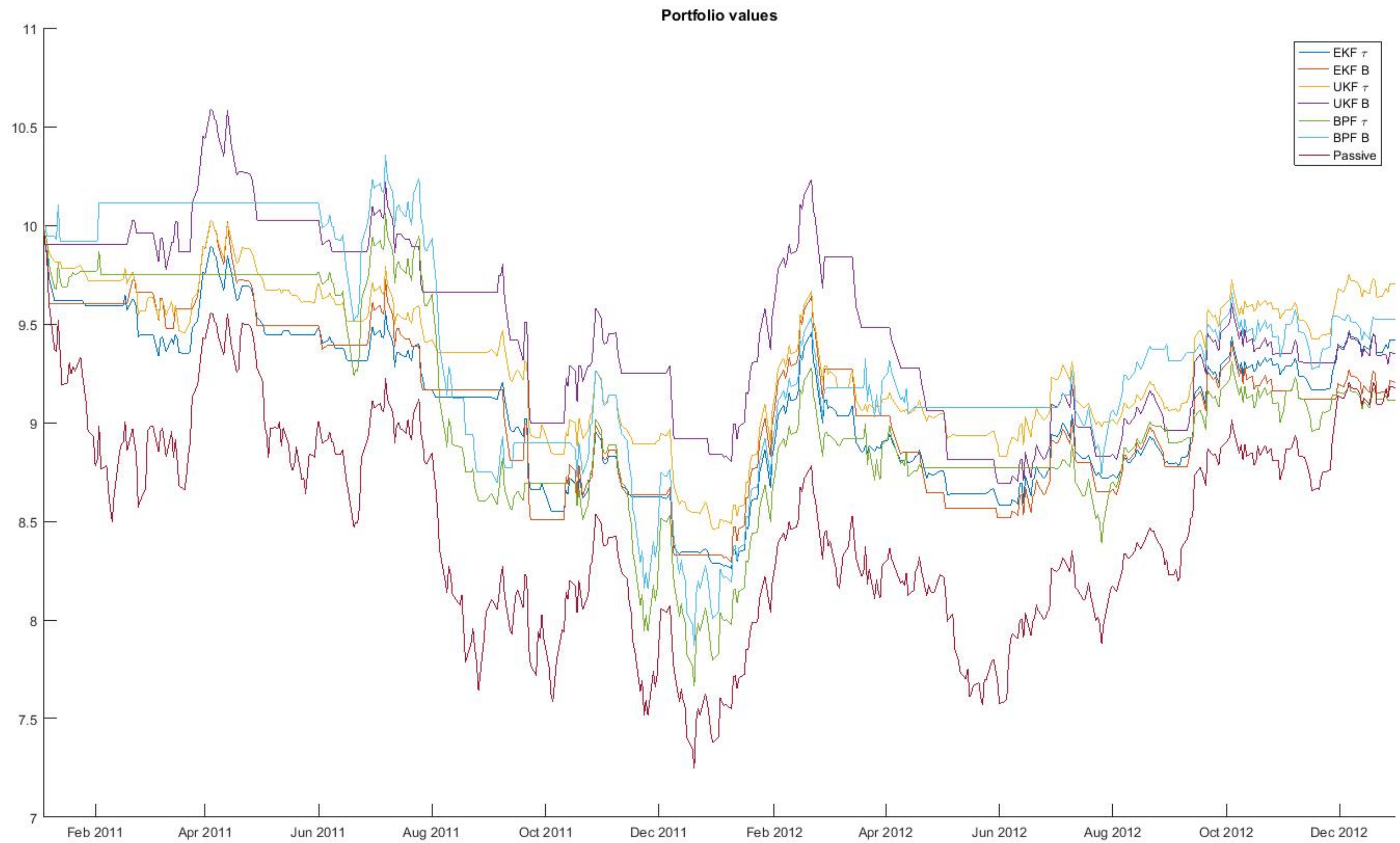

Figure 12: S&P Bombay Stock Exchange

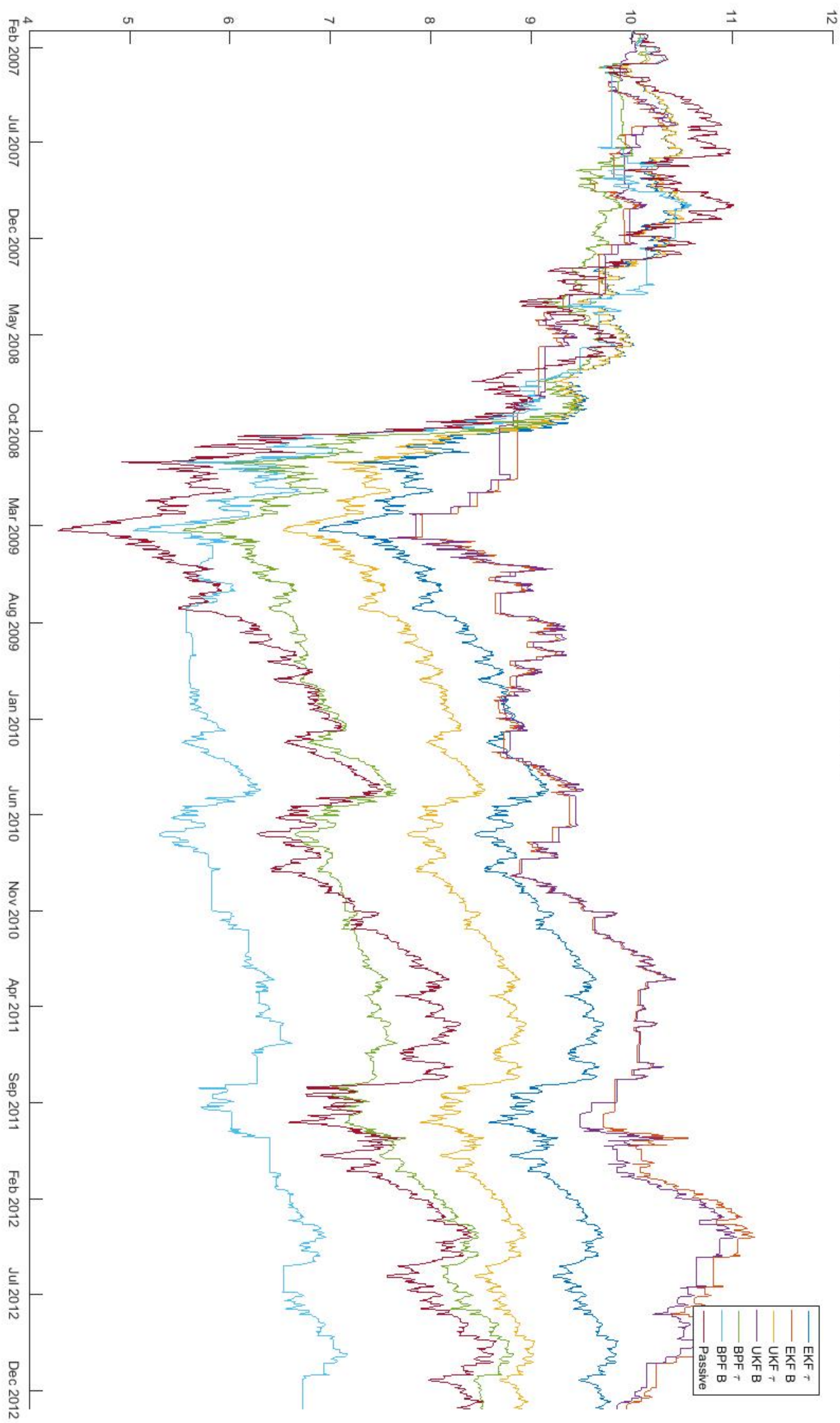

Figure 13: S&P 500

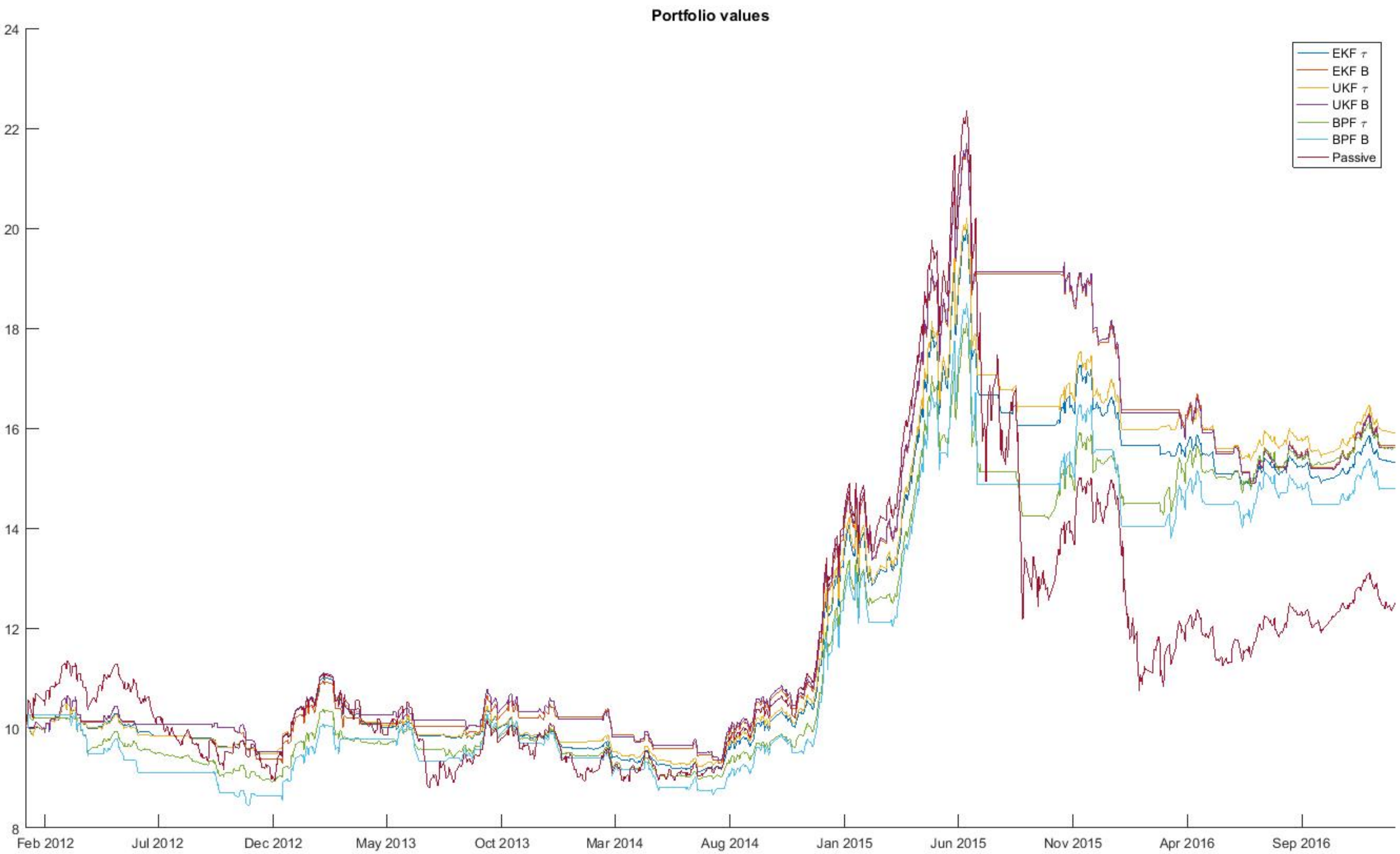

Figure 14: Shanghai Stock Exchange

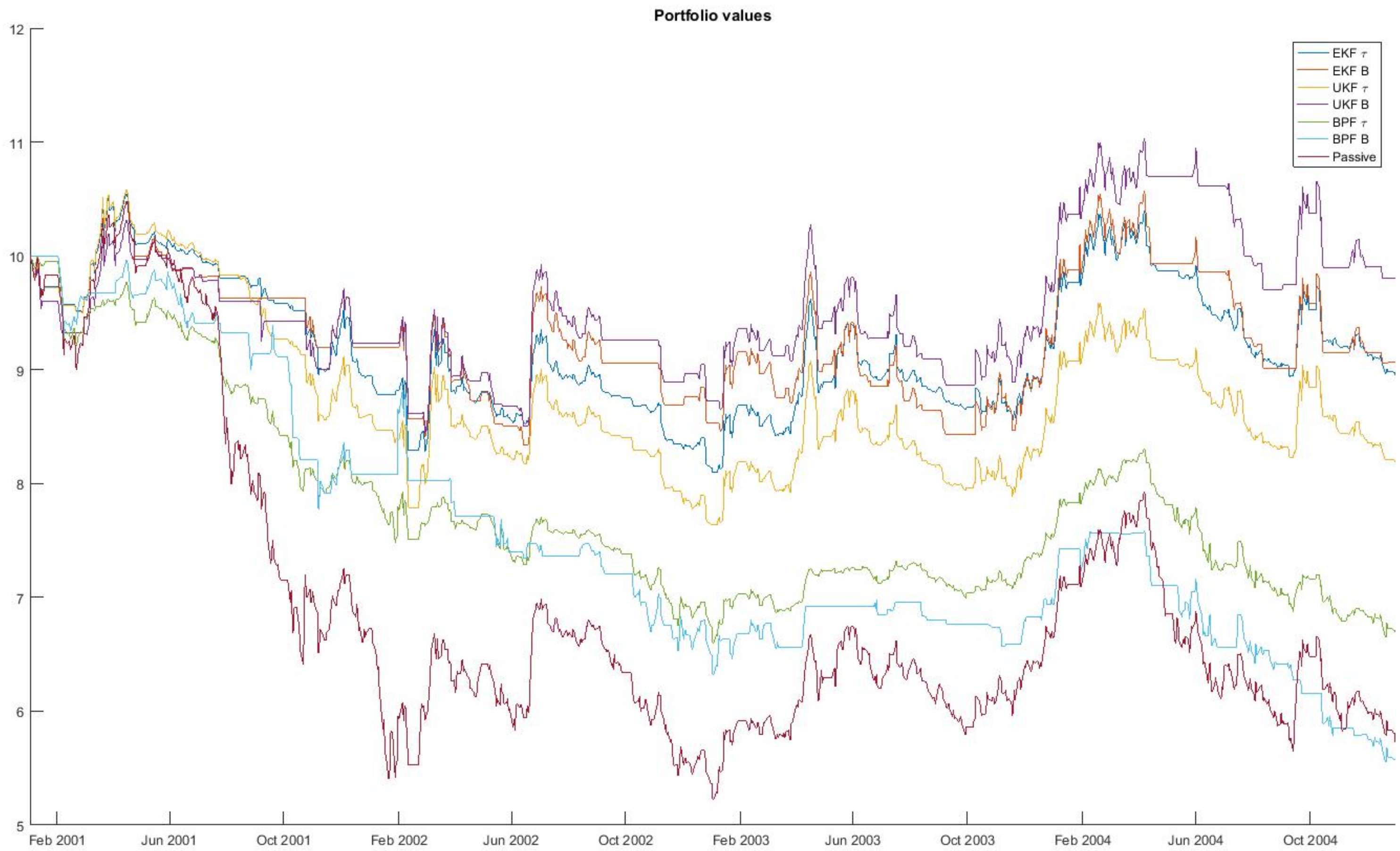

Figure 15: Shenzen Stock Exchange