Matematisk statistik
Stockholms universitet

Modelling Time Dynamic Alphas and Betas for Mutual Fund Returns

Veronica Wahlberg

Examensarbete 2007:8

ISSN 0282-9169

Postadress:
Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:
http://www.matematik.su.se/matstat

Matematisk statistik
Stockholms universitet
Examensarbete **2007:8**,
http://www.matematik.su.se/matstat

Modelling Time Dynamic Alphas and Betas for Mutual Fund Returns

Veronica Wahlberg*

May 2007

Abstract

This thesis implements a model for estimating the time dynamics of mutual funds' alphas and betas. Funds are often managed according to an active trading strategy and shifted across securities, and that gives rise to time variation in their beta. By estimating beta using non-overlapping-window OLS one detects autocorrelation in the beta series for most funds. Capturing this time variation in the beta will result in more accurate estimate of a fund's alpha, which serves as a measure of a portfolio manager's "managerial talent". Previous attempts to estimate funds' alphas and betas have used OLS on a version of CAPM, which does not allow for time dynamics in alpha and beta. This project implements and tests an alternative model, where alpha and beta are estimated via an Extended Kalman filter. The Kalman model is based on an assumption that assets are reallocated according to an unobservable factor and it allows for changes in the parameters over time. It consists of a system of equations where the unobservable factor follows an AR(1) process, and by that yields the aspect of time dynamics in the parameters. Testing the two models on a fund universe of approximately one hundred funds shows that the Kalman model can be more successful in capturing the time variation than the OLS model, but that the OLS model on a large scale seems to possess better alpha prediction ability despite its static nature. This is due to the complexity of estimating the EKF model. However, when we examine funds individually and compare the two models when the EKF model has converged, it is shown that the EKF model can be superior in capturing the time dynamics of alpha and beta.

*E-post: veronica.wahlberg@gmail.com. Handledare: Thomas Höglund.

Acknowledgements

This project constitutes a 20-credit thesis in mathematical statistics at Stockholm University and has been performed with Kaupthing Bank Sverige AB. I would like to express my sincere gratitude and appreciation to my supervisor at Kaupthing Bank, PhD Anders Blomqvist, for sharing his skills and providing valuable guidance and support in all phases of the project. I also want to thank the team at Tactical Asset Allocation and Quantitative Research at Kaupthing Bank for giving an insight in the application of this project.

I also want to acknowledge my supervisor at the department of mathematical statistics at Stockholm University, Thomas Höglund, for his advice and comments.

Contents

Chapter 1

Introduction

In the past few decades the mutual fund industry has rapidly grown and become object of quite extensive research in the academic and financial community. One issue of prime interest is how to measure portfolio managers' performance, which is determined relative to the over-all market. It has been heavily debated whether or not abnormal risk adjusted returns, α, exists and previous research have used different models in order to detect alpha. However, those models have one feature in common; they all model funds' alpha and beta as constant parameters. That is a great limitation in the analysis since funds' betas seem to exhibit autocorrelation. Therefore, estimates of alpha and beta from a static model will be biased. The gain of obtaining accurate estimates of beta is that it results in a more accurate estimate of alpha; miss estimated beta will induce miss estimated alphas, see [9]. This thesis studies and implements an alternative model, where we let alpha and beta vary with time, and our aim is to obtain more accurate parameter estimates.

1.1 Origin of time dynamics

Since funds are managed according to an active trading strategy and shifted across securities, funds' betas will generally not be constant. For example, a fund heavily invested in stocks at the beginning of a year might have a beta of one. If however money is moved to e.g. bonds later in the year, the fund's beta will drop.

This time dependence can be captured by estimating beta using Ordinary Least Squares (OLS) with non-overlapping windows. Examining the obtained series of beta estimates reveals that some mutual fund returns have indications of β autocorrelation. The experiment is illustrated in Figure 1.1. By taking the LPC estimate of each fund the autocorrelation is heavily re-

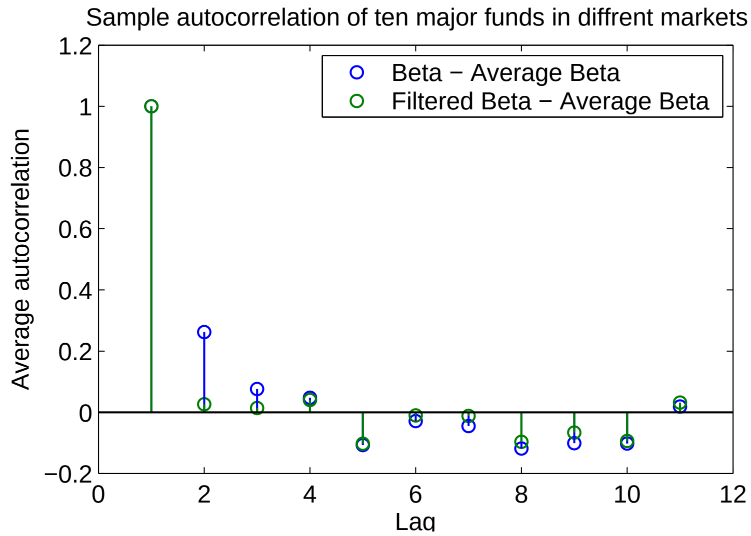


Figure 1.1: Average autocorrelation of ten funds' beta before and after filtering

duced. One way to capture this time dependence is to use an AR(1) model, and we want to take advantage of that when examining fund manager's performance. Since most models that evaluate portfolio returns do not allow for dynamic parameters we need to find a model that does take into account the time dependence in the parameters.

1.2 Two models

Time Dynamic model

The model, used to estimate time varying alphas and betas for mutual fund returns, is based on an assumption that portfolio managers must possess private information in order to produce alpha. Portfolios are then shifted across securities due to this information, which is modelled as an unobservable factor. That factor is assumed to follow an autoregressive process of order one, since the value of the information will decline with time. Both a fund's alpha and beta are influenced by this unobservable private information, and it leads to a model consisting of a system of equations. The system is of so-called state space form and is estimated via an Extended Kalman Filter, EKF.

Attempts to detect alpha by using this model have been made by Mamaysky, Spiegel and Zhang in *Estimating the Dynamics of Mutual Fund Alphas and Betas, 2004* and *Improved Forecasting of Mutual Fund Alphas and*

Betas, 2005. This thesis will implement the EKF model used by the authors of [9] and it will be tested and compared to a static model.

Static model

The static model is a version of CAPM, that allows for the parameter α:

$$r_t - r_f = \alpha + \beta(r_t^M - r_f) + \epsilon_t. \tag{1.1}$$

The model in (1.1) is estimated via rolling window regression, which allows for some time variation in its parameter. However, the trade-off between parameter accuracy and time dependence is not obvious. However, we will in this thesis use a window length of four years.

1.3 Objective

By detecting autocorrelation in funds' beta we have reason to believe that the dynamic model will produce more accurate estimates of alpha and beta than a model based on OLS. This thesis aims at examining and illustrating the differences in alpha and beta estimates between the two models. We also want to be able to say something regarding the two models' performance in predicting alpha and beta.

In chapter two we will derive the time dynamic model used in this thesis, the EKF model. Chapter three contains a description of the Kalman filter and gives a theoretical background of how to estimate the Kalman model. In chapter four we describe the system parameter identification problem associated with the model calibration, and examine the two models' alpha and beta prediction ability on one fund. Chapter five treats an out-of-sample test, where each model's ability to predict alpha and beta is tested on a large sample of funds. Conclusions are found in chapter six.

Chapter 2

Model description

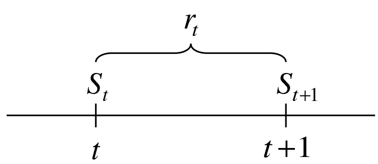
As discussed earlier, a model that captures time variation in funds' betas is highly preferable since poorly estimated betas result in a systematic bias in the alpha estimates. In this chapter we will derive a model, which is used in [9] and [10] that does allow for time dynamics in a fund's alpha and beta.

2.1 Background

A portfolio manager's performance needs to be compared to the overall market returns. According to CAPM an asset's rate of return can be modelled as

$$r_t - r_f = \beta(r_t^M - r_f) \tag{2.1}$$

where $r_t = \log(\frac{S_{t+1}}{S_t})$ is the asset's return and S_t the its price at time t,



r_t^M is the market return and β the funds beta:

$$\beta = \frac{\text{Cov}\left(r, r^M\right)}{\sigma_M^2} = \rho_{r,r^M}\frac{\sigma}{\sigma_M} \tag{2.2}$$

where σ and σ_M is the fund and the market volatility respectively.

A fund's beta measures to what degree the fund covariates with the over all market. For example, if an asset has a beta of 1.2 it is theoretically 20% more volatile than the market, whereas a beta of less than 1 means that the asset will be less volatile than the market. A beta value of 1 means that the asset will move with the market. CAPM describes the relationship between risk and expected return and is used in the pricing of risky securities. It states that the expected rate of return of an asset equals its beta times the rate of return of the overall market. This holds under a number of assumptions, of which one is that all investors have the same expectations about security returns, also referred to as assumption of perfect information [8]. Thus, in order to generate any positive risk-adjusted returns, portfolio managers must possess some private information that he or she uses to forecast returns. The financial measure of such abnormal returns is α, and a model that would better describe asset i's return under the assumption of existing alpha would be

$$r_{i,t} - r_f = \alpha_i + \beta_i(r_t^M - r_f) + \epsilon_{i,t} \tag{2.3}$$

The returns of a portfolio consisting of n assets $I = \{i_1, i_2, \ldots, i_n\}$ can be written as:

$$r_{P,t} = \sum_{i \in I} w_i r_{i,t} \tag{2.4}$$

where $w_{i,t}$ is the fraction of the portfolio invested in asset i and

$$\sum_{i \in I} = 1. \tag{2.5}$$

We now assume that (2.3) accurately describes the return of an individual stock [9]. Then a portfolio's time t returns are described by:

$$\begin{aligned} r_{P,t} - r_{f,t} &= \sum_{i \in I} w_{i,t}\Big(\alpha_{i,t} + \beta_{i,t}(r_t^M - r_{f,t}) + \epsilon_{i,t}\Big) + k_t \\ &= \alpha_{P,t} + \beta_{P,t}(r_t^M - r_{f,t}) + \epsilon_{P,t}, \end{aligned} \tag{2.6}$$

where

$$\alpha_{P,t} = \sum_{i \in I} w_{i,t}\alpha_{i,t} + k_t, \tag{2.7}$$

$$\beta_{P,t} = \sum_{i \in I} w_{i,t}\beta_{i,t}, \tag{2.8}$$

$$\epsilon_{P,t} = \sum_{i \in I} w_{i,t}\epsilon_{i,t}, \tag{2.9}$$

and where k_t corresponds to the transaction costs and is assumed to be proportional to the funds under management. Note that if the CAPM holds for each investment period, then $\alpha_{P,t} = 0$ due to the assumption of perfect information.

One assumption under (2.3) is that all single assets' stocks have constant betas, and we see from Equation 2.6 that portfolio's returns are modelled with a linear model. However, the time $t+1$ portfolio weights will have changed from the time t weights, so even if the portfolio returns are linear in r_t, due to the weight changes, the model will be governed by time varying coefficients. Thus, a linear model with constant coefficients will be miss specified. Another limitation of the above model is that in order to estimate the parameters of (2.6) one needs information regarding the portfolio's underlying assets and their composition. In most cases when evaluating fund managers' performance, one has no access to such information. We need a model that takes into account the aspect of time dependence and that does not require inaccessible information.

2.2 Deriving time dynamic model

According to the assumptions in the previous section, a fund manager will only produce positive abnormal risk adjusted returns, α, as a result of having some private information which is used in the investment decisions. This amount of information is assumed only to occur occasionally and the value of it to decrease with time. The authors of [9] model this amount of information as a variable x, which is assumed to follow an autoregressive process of order one:

$$x_t = ax_{t-1} + \eta_t, \tag{2.10}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$. The parameter a is a measure of how the value of the information lasts over the next investment period, and in order to obtain causality in the autoregressive process we must let $a \in [0, 1)$, see [2].

For non-zero x, it is assumed to influence the fund's present holding and future expected stock return according to the following equations:

$$w_{i,t} = \bar{w}_i + l_i x_t, \tag{2.11}$$

$$\alpha_{i,t} = \bar{\alpha}_i x_t. \tag{2.12}$$

Here $\bar{w}_i$ represents a steady state fraction of the strategy invested in a given security and l_i is a stock's loading of x that allows for the weight to deviate from its steady state value. An asset's alpha is assumed to be influenced by x via $\bar{\alpha}_i$, which measures to what degree a stock's expected return is predictable by x. Note that there is no constant term in Equation 2.12, which is a result of the fact that α does not exist without superior information.

Using (2.7) and (2.8) with the formulation of the signal's impact on single assets' weights and alphas in (2.11) and (2.12), some algebra yields:

$$\begin{aligned} r_{P,t} - r_f &= \sum_{i \in I} (\bar{w}_i + l_i x_t)\Big(\bar{\alpha}_i x_t + \beta_i (r^M_{i,t} - r_f) + \eta_{i,t}\Big) + k_t \\ &= k_t + c x_t + e x_t^2 + (d + x_t)(r^M_{i,t} - r_{f,t}) + \epsilon_{P,t} \end{aligned} \quad (2.13)$$

where

$$c = \sum_{i \in I} \bar{w}_i \bar{\alpha}_i \quad (2.14)$$

$$e = \sum_{i \in I} \bar{\alpha}_i l_i \quad (2.15)$$

$$\epsilon_{P,t} = \sum_{i \in I} (\bar{w}_i + l_i x_t)\epsilon_{i,t} \quad (2.16)$$

This model clearly shows how a fund's alpha is allowed to systematically depend on a fund's trading strategy x, and the dependence comes via parameter e and c. A more detailed description of each parameter's economic interpretation is found in [9]. The authors state that parameter c measures to what degree a fund's strategy is related to the instantaneous alphas of individual stocks, and that since x takes both positive and negative values, a non-zero value of c does not indicate either over- or underperformance. The quadratic term e however, does indicate just that; to what degree a fund systematically goes long (short) in positive (negative) alpha stocks, and it can be thought of as the covariance between a fund's security weights and the underlying alphas.

2.3 Time Dynamic model

The final model forms a system of equations and it will be used in this thesis for predicting alphas and betas of mutual funds:

$$\begin{aligned} x_t &= a x_{t-1} + \eta_t \\ r_t - r_f &= k + c x_t + e x_t^2 + (d + x_t)(r^M_t - r_f) + \epsilon_t \end{aligned} \quad (2.17)$$

where $\eta_t \sim N(0, \sigma^2_\eta)$ and $\epsilon_t \sim N(0, \sigma^2_\epsilon)$ and where c, e and ϵ_t are as in (2.14) - (2.16).

The time variation of the model comes via the variable x in the way it follows an auto regressive process. Note that nowhere in this model do we need any information about a fund's composition of securities. The only two sources of information that this model is based on are historical returns of the fund and a relevant index. Still it allows for changes in alpha and beta due to time.

We see from Equation 2.17 that given x_{t-1} we can make predictions of the time t alpha and beta:

$$\alpha_{P,t} \equiv k_t + c\tilde{x}_t + e\tilde{x}_t^2 \tag{2.18}$$

$$\beta_{P,t} \equiv d + \tilde{x}_t \tag{2.19}$$

and where

$$\tilde{x}_t = ax_{t-1}. \tag{2.20}$$

In order to make the predictions, we need to estimate the variable x up to time $t-1$. The next chapter treats the most common algorithm for estimating x, the Kalman filter.

Chapter 3

Theoretical background

In the previous chapter we derived a model that takes into account the fact that alphas and betas of mutual funds vary with time. The model relies on the assumption that the time dependence originates from the unobservable variable x according to the following model:

$$\begin{aligned} x_t &= ax_{t-1} + \eta_t, \\ r_{P,t} - r_f &= k + cx_t + ex_t^2 + (d + x_t)(r_t^M - r_f) + \epsilon_t, \end{aligned} \tag{3.1}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$ and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.

As discussed in the previous chapter, in order to make predictions of α_t and β_t at time $t-1$, x_{t-1} is required, and a method for estimating x is needed. We will here present such a method and discuss how it can be applied to our model.

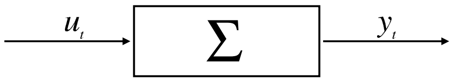
3.1 State space model and Kalman filter

The model in (3.1) consists of a system of equations of so-called state space form that in recent years has become a powerful tool in handling time series. Once a model has been put in state space form, a *Kalman filter* can be applied to obtain an estimate of the state, x. Techniques of state space representations and the associated Kalman recursions were originally developed in connection with the control of linear systems [3]. The next sections give a theoretical background to the state space representation and treat the algorithm for state estimation.

3.1.1 Linear Control Systems and State Space Form

In mathematical systems and control theory one refers a linear control system Σ to a "black box" that produces noisy output y as a result of an input signal

u. The internal shape of the box, also referred to as the *state* of the system,



is unobservable and that is what we want to obtain a proper description of [6].

The state space form provides a description of the control system Σ via two equations; the *state*- and *measurement* equation

$$
\begin{aligned}
x_t &= A_{t-1}x_{t-1} + B_{t-1}u_{t-1} + C_{t-1}v_{t-1}, \\
y_t &= D_t x_t + E_t u_t + F_t w_t.
\end{aligned} \tag{3.2}
$$

In (3.2), x constitutes the state of the system, y the measurements, u_t deterministic input signal and v_t and w_t uncorrelated white noise processes. The problem one faces is how to make optimal estimates of the state, x, given the input and output signals, u and y. We want to make *least squares* optimal estimates of the state x in the sense that it should be chosen to minimize the prediction error. As will be shown in the next section, that is just what the Kalman filter provides.

A state space model can be set up in several ways. The system 3.2 is a discrete time variant model, since the parameters A_{t-1},B_{t-1},C_{t-1},D_t,E_t and F_t are all dependent of time. The system can also be modelled in continuous time and as a time invariant system type. The model derived in the previous chapter, and the one that will be used in this thesis, is assumed to be a discrete time invariant system. That means that the matrices A,B,C,D,E and F in (3.2) are independent of time; they have no time subscripts.

3.1.2 The Kalman filter

State Space models can be estimated by a number of algorithms, but the most common one is the Kalman filter. It has been widely used within different fields of engineering where the state of a system is of prime interest, e.g., tracking problems and positioning systems.

The true state x, that corresponds to the unobservable internal shape of the box, is assumed to be a Markov process, and the measurements y are the observed states of the hidden Markov model. The Kalman filter is a *recursive* estimator, in the sense that it only needs the state estimate from previous step, x_{t-1}, and the current measurement y_t in order to estimate the current state x_t. For further reading on the probabilistic origins of the filter see [11] and [4].

The Kalman filter gives an optimal estimate of the state in the sense that it for each step minimizes variance of the prediction error:

$$E[x_t - \hat{x}_t]^2. \tag{3.3}$$

The filter consists of two phases; a prediction and an update phase. The prediction phase provides an *a priori* estimate of the state from previous step's state estimate via the Projection Theorem, see [6]. The *a priori* estimate is then updated with information about the current measurement via the Kalman gain in the update phase. The refinement of the estimate will hopefully deliver a more accurate current state estimate, which can be thought of as an *a posteriori* estimate of the state.

Filter dynamics

We define $\tilde{x}_t$ to be the *a priori* estimate of the state at time t, given the knowledge of the time $t-1$ process, and define $\hat{x}_{t|t}$ be the *a posteriori* state estimate at time t given the measurement y_t. We can then write expressions for the *a priori*- and *a posteriori* estimate errors as

$$\begin{aligned} \tilde{e}_t &\equiv x_t - \tilde{x}_t \\ e_t &\equiv x_t - \hat{x}_{t|t} \end{aligned}$$

We then let the *a priori* estimate error covariance be

$$\tilde{P}_t = \mathrm{E}\left[\tilde{e}_t \tilde{e}_t^T\right] \tag{3.4}$$

and the *a posteriori* estimate error covariance

$$P_t = \mathrm{E}\left[e_t e_t^T\right] \tag{3.5}$$

In the derivation of the Kalman equations, the goal is to write the *a posteriori* state estimate $\hat{x}_{t|t}$ as a linear combination of the *a priori* state estimate $\tilde{x}_t$ and a weighted difference between the measurement y_t and measurement prediction $\tilde{y} = D\tilde{x}_t$:

$$\hat{x}_{t|t} = \tilde{x}_t + K_t(y_t - \tilde{y}_t) \tag{3.6}$$

Here $y_t - \tilde{y}_t$ is called the measurement innovation. K is the so-called Kalman gain, which is chosen to minimize P_t. One form of the resulting K that minimizes P_t is given by

$$\begin{aligned} K_t &= \tilde{P}_t D^T (D\tilde{P}_t D^T + R)^{-1} \\ &= \frac{\tilde{P}_t D^T}{D\tilde{P}_t D^T + R}, \end{aligned} \quad (3.7)$$

where R is the measurement error covariance, $R = FF'$. Looking at the expression of the Kalman gain in (3.7) we see that as R approaches zero, the Kalman gain weighs the innovations in (3.6) more heavily. One can think of the weighing by K as the following: when measurement error covariance R approaches zero, the actual measurement y_t is trusted more and more, while the predicted measurement $\tilde{y}_t$ is trusted less and less. On the other hand, as the *a priori* estimate error covariance $\tilde{P}_t$ approaches zero, the Kalman gain weighs the innovations less heavily implying that the actual measurement y_t is trusted less and less while the predicted measurement $\tilde{y}_t$ is trusted more and more.

A complete derivation of the Kalman filter and the properties of its components is found in [6] where the following theorem with proof is stated:

Theorem 1 *Given a linear stochastic system 3.2 with non-deterministic output process y, the linear least squares estimate $\hat{x}_t$ of the state x_t, given the observations $y_0, y_1, \ldots, y_{t-1}$is generated by the Kalman filter.*

3.2 The Extended Kalman filter

The theorem above states that the Kalman filter produces least-squares-optimal estimates of the states, given a *linear* system. However, the model in Equation 3.1 is governed by a non-linear measurement equation and the standard Kalman filter cannot be applied. We need a method that manages to estimate the state despite the non-linearity in the observation equation. The solution to the problem of how to estimate the state of a non-linear system is the Extended Kalman filter, also referred to as the EKF.

3.2.1 Non-linear state space model

The EKF estimate gives an approximation of the optimal estimate and the non-linearities of the systems dynamics are approximated in a linearization around the last estimate. The standard Kalman filter is thereafter applied to obtain an estimate of the state.

The state space form now becomes

$$\begin{aligned} x_t &= f(x_{t-1}, u_{t-1}, v_{t-1}) \\ y_t &= g(x_t, u_t, w_t) \end{aligned}$$

In our case, with our dynamic model 3.1, the system will look like

$$x_t = f(x_{t-1}, u_{t-1}, v_{t-1}) = x_{t-1} + v_t \tag{3.8}$$

$$y_t = g(x_t, u_t, w_t) = cx_t + ex_t^2 + (d + x_t)u_t + w_t \tag{3.9}$$

The idea behind the extended Kalman filter is to linearize the system around the current state estimate via a first order Taylor expansion and then apply the standard Kalman filter to the resulting linear system. However, that will not result in a least-squares optimal state estimate. Also, as a result of the non-linearity convergence of the EKF is not guaranteed. However, among all recursive identification algorithms, the extended Kalman filter is no doubt the best-known and most widely used example [7]. A more detailed description and background of the EKF as well as the algorithm can be found in Appendix A. In short one can say that, like the Kalman filter, the EKF consists of a prediction and a measurement update phase, but the algorithm is based on the following consecutive steps:

1. Consider the last filtered state estimate $\hat{x}_{t-1|t-1}$,
2. Linearise $x_t = f(x_{t-1}, u_{t-1}, v_{t-1})$ around $\hat{x}_{t-1|t-1}$,
3. Apply the prediction step of the Kalman filter to the linearized system dynamics obtained. This yields $\tilde{x}_t$ and $\tilde{P}_t$,
4. Linearize $y_t = h(x_t, u_t, w_t)$ around $\tilde{x}_t$,
5. Apply the update cycle of the Kalman filter to the linearized observation dynamics. That yields $\hat{x}_{t|t}$ and P_t.

For further details see [12].

3.2.2 EKF on dynamic model

The EKF is applied to our dynamic model in Equation 3.1 in order to obtain an estimate of the state x. This is done using returns of the fund and its benchmark up to time $t-1$. Having the estimate of x_{t-1} we can make predictions of time t alpha and beta:

$$\begin{aligned} \hat{\alpha}_t &= k + c\tilde{x}_t + e\tilde{x}_t^2 \\ \hat{\beta}_t &= d + \tilde{x}_t \end{aligned}$$

where $\tilde{x}_t = ax_{t-1}$.

We complete this chapter by stating that the EKF seems promising to use for state estimation given our model in Equation 3.1. In order to use the EKF for our purpose we need to know the system parameters $\theta = \{a, \sigma_\eta, c, d, e, \sigma_\epsilon, k\}$. The next chapter will treat the problem of identifying θ.

Chapter 4

Model calibration

In the previous chapters we derived our dynamic model in state space representation, and described how an extended Kalman filter can be applied in order to estimate the state of the system. This chapter will describe how to implement the model and how it performs in predicting funds' alpha and beta.

4.1 System parameter estimation

As we saw in the previous chapter, given a set of model parameters $\theta = \{a, \sigma_\eta, c, d, e, \sigma_\epsilon, k\}$ the extended Kalman filter, also called EKF, will produce least-squares estimates of x. However, θ is in this case not known and needs to be estimated. We do that via a prediction error optimisation, and thereafter apply the EKF on the resulting model to obtain a state estimate.

To determine a set of parameters that produces as good estimates of the state as possible, we aim at minimizing the prediction error of the model w.r.t θ. The prediction error is formulated to constitute the discrepancy between step t return of the fund and the Kalman predicted step t return of the fund:

$$PE_t(\theta) = r_t - \tilde{r}_t = r_t - (k + c\tilde{x}_t + e\tilde{x}_t^2 + (d + \tilde{x}_t)r_t^M) \tag{4.1}$$

Remember from previous chapter, the projection theorem gives

$$\tilde{x}_t = ax_{t-1} \tag{4.2}$$

Before starting the implementation of the optimisation routine we need to study our model and system's behaviour. We need to learn what we can expect to detect from return series and to obtain information regarding system parameters. This is done via simulations and the information obtained is used in the optimisation procedure in order to succeed with the parameter estimation.

4.2 Simulation

One way to learn about the behaviour of our system is to simulate returns series from different systems, and then observe how the Kalman filter performs when applied to them. Whether or not the filter manages to adapt its output to different input signals will reveal information that can be used in the optimisation.

The reason for simulating returns is that we want to be able to rule out unrealistic parameter values, and learn about the influence of different parameter settings in the state estimate. We also want to learn how and if the system parameters are mutually related and how that influences the filtering. This is examined via two simulations; one where we let the real x to be a deterministic process that follows a sinus curve, and one where the real state is governed by an $AR(1)$-process.

4.2.1 Deterministic state process

Earlier we discussed the existence of autocorrelation in a funds beta, and saw that it exists for some funds. According to the EKF model, that is a result of autocorrelation in the state x. One can think of the autocorrelation in the state as inertia in the information process which is captured in a model where x is assumed to follow a sinus curve. By modelling x this way we let it equal zero on average, and it illustrates how the value of the information decreases with time. All this is in line with our expectation; alpha-generating information is assumed to only occur occasionally and to decrease with time. Another aspect of modelling real state as a deterministic curve is that it gives an opinion of the size of parameters a and b, which serves as a measure of the inertia in the state process and the state noise.

The return series are simulated out of the following system:

$$x_t = 0.2 \sin\left(\frac{t}{17}\right) \tag{4.3}$$

$$r_t = k + cx_t + ex_t^2 + (d + x_t)r_t^M + \epsilon_t \tag{4.4}$$

Modelling x this way allows for changes in the market beta by letting it deviate from its long run beta with 0.2 in absolute value.

The EKF is applied to the simulated return series and a state estimate is obtained. The estimated state is then compared to the real state in order to examine what parameter settings that make the estimated x capture the sinus behaviour.

Parameter influence in state estimate

Examining our system using the sinus model gives us some understanding of what set of parameters to use in the optimisation. Parameter d which corresponds to the steady state beta is set to calculated market beta:

$$\beta \quad = \quad \frac{\text{cov}(r, r^M)}{\sigma_M^2}$$

Thereafter we test how the filter captures a sinus behaviour for different relationships between the two noise variances, σ_η^2 and σ_ϵ^2.

One way to get an opinion of the size of σ_η is to examine what amount of noise a sinus shaped state corresponds to. If we fix the autocorrelation parameter a to e.g. 0.85 we can measure the amount of noise by taking the standard deviation of the series:

$$\eta_t = \{x_s - 0.85x_{s-1}, 1 \leq s \leq t\} \tag{4.5}$$

. Repeating this for different values of a yields the following figure: We

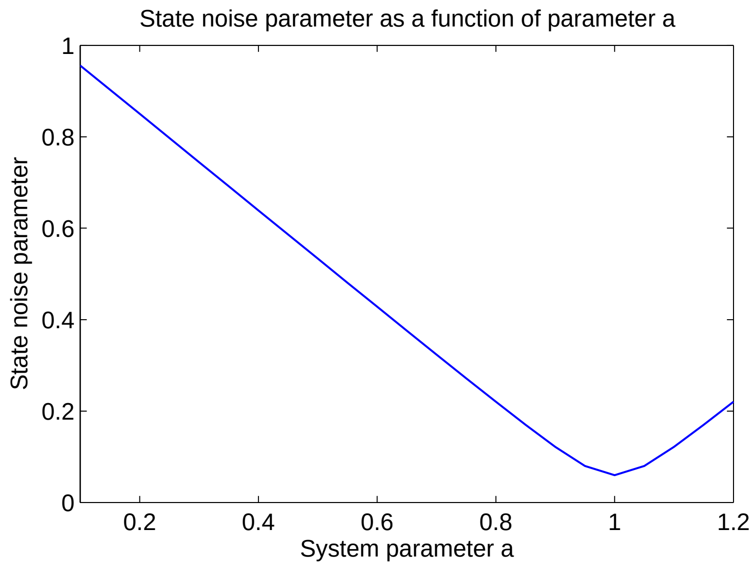


Figure 4.1: State noise as a function of system parameter a.

see that when $a = 0.85$ the sinus shaped state generates an amount of noise, corresponding to σ_η, of 0.17 on yearly basis, and that the noise is a decreasing function of parameter a, when $a \in [0.1, 1)$. This gives us an idea of what size of σ_η to expect. Also, when testing the filter's ability to capture the sinus behaviour of the estimated state, we conclude that the EKF performs best when σ_η is considerably larger than σ_ϵ. The fact that the ratio $\frac{\sigma_\eta}{\sigma_\epsilon} > 1$ reveals that we should trust our observations more than the predictions. Remember from Chapter 3, the so-called Kalman Gain gives a measure of how to weigh the innovations $y_k - A\tilde{x}_t$ in the updating of the a priori estimate in the measurement update phase. If the observation noise is reduced, the Kalman gain increases implying that the measurements should be trusted more than the predictions [6].

Measuring the autocorrelation in state

To get an idea of the autocorrelation in the state, the prediction error is minimized w.r.t. a. By that we obtain an idea of its value, if we believed in the sinus model. The optimisation routine `fminbnd` in MATLAB'S Optimisation toolbox estimates $a \in (0.87, 0.97)$. Also, when testing other values of parameter a, we observe that as it approaches 1 the system becomes unstable and produces abnormally large state in absolute value. That would be a result of the resulting non-causality in the autoregressive process.

c and e influence

We also want to examine how different values of c and e influence the sinus behaviour of state estimate. When we simulate returns with small values of e and c, the EKF fails in detecting the sinus shaped state. It is not until we let $|c| \approx 0.3$ we seem to detect state. For values less than 0.3, the EKF produces a state estimate around zero with only some noise. It turns out that in order to obtain the sinus shaped pattern in estimated x we need quite large values of e and c in absolute value. However, it is not likely that they will be infinitely large; a phenomena that occurs when letting c and/or e become too large in absolute value is that the estimated state becomes the reflection of the real state, see Figure 4.3. It shows that the filter fails picking up positive/negative values of the state. According to the sinus model, ideal values would be $|c| < 2$ and $|e| < 3$. Figure 4.2 shows the resulting state estimates for a specific θ.

Allowing for a constant k

Simulations shows that having rather large values of e and allowing for constant parameter k diminishes the reflections of the real state described above. The correlation between parameter e and k seems to be negative. A possible explanation to that is that since e always contributes to the system with positive x via the square term, the system needs to be compensated. That is what the constant k does. The Figure 4.3 above illustrates the state estimate

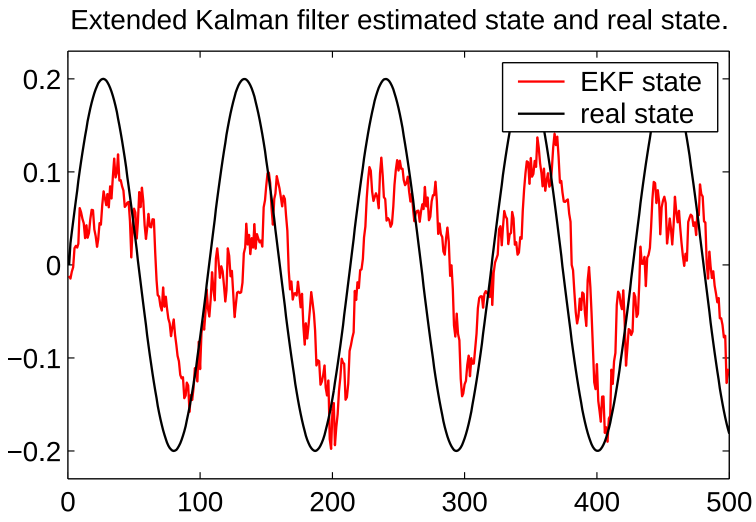


Figure 4.2: $\theta = \{0.96, 0.29, -1.88, 0.85, 2.24, 0.09, -0.0479\}$

when we let $k = -0.01$ and $e = 20$, and we see the reflections clearly. In Figure 4.4 however we have set $k = -0.3$ which compensates for the under estimation of the state due to the large positive value of e.

4.2.2 AR(1) state process

The same analysis is made by simulating data out of the real system, where x constitutes the autoregressive process:

$$\begin{aligned} x_t &= ax_{t-1} + \eta_t \\ r_t &= k + cx_t + ex_t^2 + (d + x_t)r_t^M + \epsilon_t \end{aligned} \tag{4.6}$$

where $\eta_t \sim N(0, \sigma_\eta^2$ and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$.

We get similar results as in the sinus case, and it is illustrated in Figure 4.5. The simulations also reveal that rather large values of c and e seem to capture fluctuations in the state even if the system from time to time struggles with estimating the right sign of the real state. The constant k seem to play an important role here as well.

4.3 Non-linear-least-square optimisation

We implement the parameter estimation procedure by optimising the prediction error with respect to θ. The optimisation toolbox in MATLAB contains

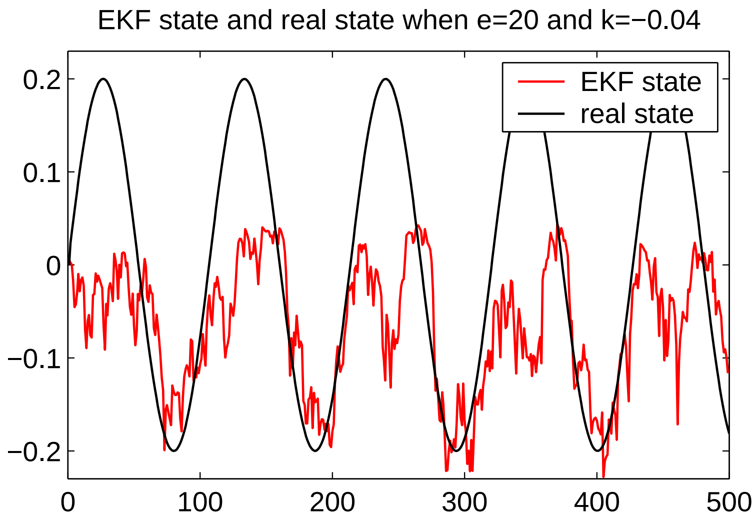


Figure 4.3: $\theta = \{0.96, 0.29, -1.88, 0.85, 20, 0.09, -0.0479\}$

the function `lsqnonlin`, which uses the Gauss Newton algorithm or the Levenberg Marquardt method if specified in the search for local minimum. The function to be minimized is the sum of squares of the prediction error.

$$\min_{\theta} f(\theta) = \min_{\theta} \frac{1}{2}||F_s(\theta)||^2$$

In our case

$$F_s(\theta) = PE_s(\theta), \; 0 \leq s \leq t \tag{4.7}$$

where $PE_s(\theta)$ is as in Equation 4.1

The knowledge of the system obtained via simulations is used in the optimisation set up when choosing starting values and parameter boundaries. When that is done we want to assure that starting values converge to the same and right optimum. To determine which is right and to get an opinion of global and local optima we examine the prediction error surfaces for different sets of parameters. Figures 4.6 - 4.8 shows the prediction error surface for Fidelity Funds - Australia Fund A Inc.

The optimisation is very time-consuming since the Kalman filtering is executed recursively. Also, since we use weekly data, an alternative formulation of the function of the prediction error is used in the *lsqnonlin* function. It is composed by the prediction errors obtained by sampling from all weekdays of a week.

$$PE_s(\theta) = \sum_i PE_{i,s}(\theta), 0 \leq s \leq t$$

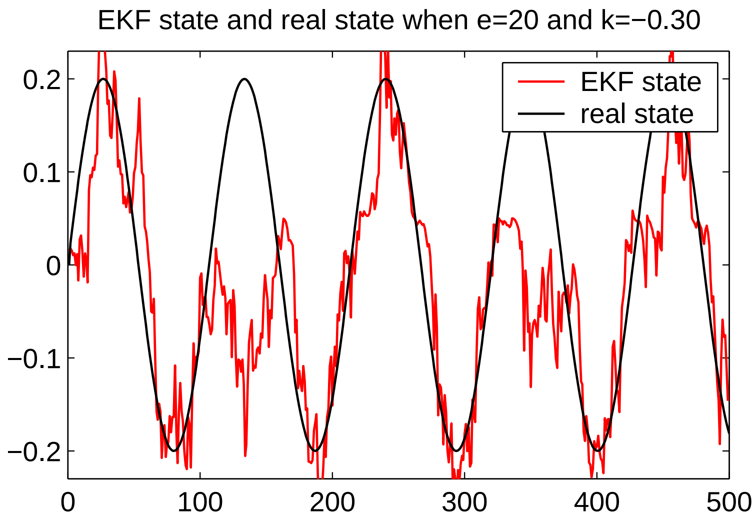


Figure 4.4: $\theta = \{0.96, 0.29, -1.88, 0.85, 20, 0.09, -0.30\}$

for all weekdays i, Monday to Friday. This is used in (4.7).

The prediction vector thereby become five times as large and thus further slows down the computations. One way to gain efficiency in the procedure is to narrow down the dimension of the parameter space over which we optimise. For example, fixing σ_ϵ to 0.04, which seems reasonable, still produces rather stable optima. Also, according to the sinus model, the autocorrelation in state would correspond to a value of $a \in [0.87, 0.97]$, and fixing parameter a to some value is also an option. We could exclude parameter k for a faster optimisation procedure. However, with the analysis of the sinus model in mind we decide to include it, since it appears to be an important component for stabilizing our state estimate.

4.4 Result on a single fund

Testing for different parameter settings, start values and boundaries resulted in a final parameter estimation of a fund universe consisting of approximately one hundred funds from the Morningstar database. The EKF model was estimated during a four-year training period starting in January 2000 and ending in December 2003. We here display the result of a single fund and examine its state estimate and alpha prediction ability.

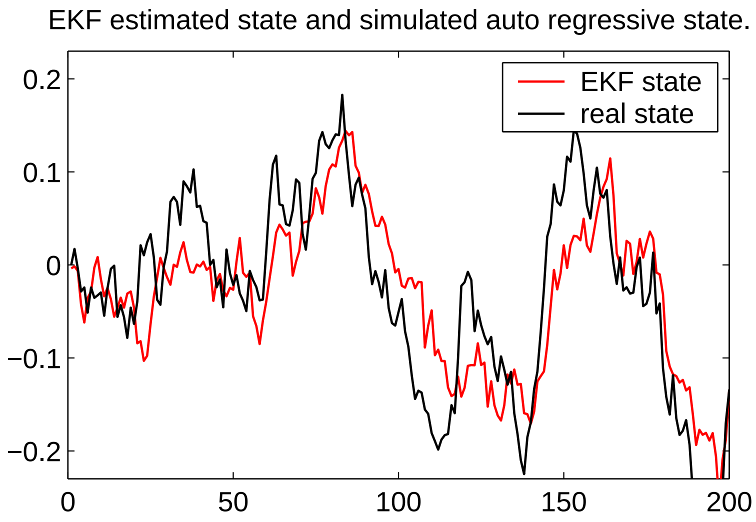


Figure 4.5: EKF state and real state

4.4.1 EKF model

Data used to test the fund's alpha and beta prediction ability is the weekly returns of *Länsförsäkringar's* fund Globalfonden and the S&P Global 1200 Total Composite Return Index over the period January 2004 to April 2007. This period is referred to as the out-of-sample period, since the predictions of alpha and beta are independent of data that was used for system parameter identification in the training period.

The estimated model for this particular fund is the following:

$$\begin{aligned} x_t &= 0.93x_{t-1} + \eta_t \\ r_t - r_f &= -0.048 - 1.88x_t + 2.24x_t^2 + (0.85 + x_t)(r_t^M - r_f) + \epsilon_t \quad (4.8) \\ \sigma_\eta &= 0.29 \text{ and } \sigma_\epsilon = 0.09 \end{aligned}$$

Applying the extended Kalman filter to this model produces a state estimate at each point in time which is shown in Figure 4.10. We see from Figure 4.10 that on average state equals zero, but fluctuates with amplitude around 0.1. The resulting time dynamic beta, here referred to as the EKF beta, is shown in Figure 4.11 and illustrates how it varies around its long run beta.

4.4.2 Rolling window regression

We would like to compare the time dynamic beta to a beta produced by a model that does not allow for any time dynamics in its parameters. We do that by estimating the fund's alpha and beta via rolling window regression,

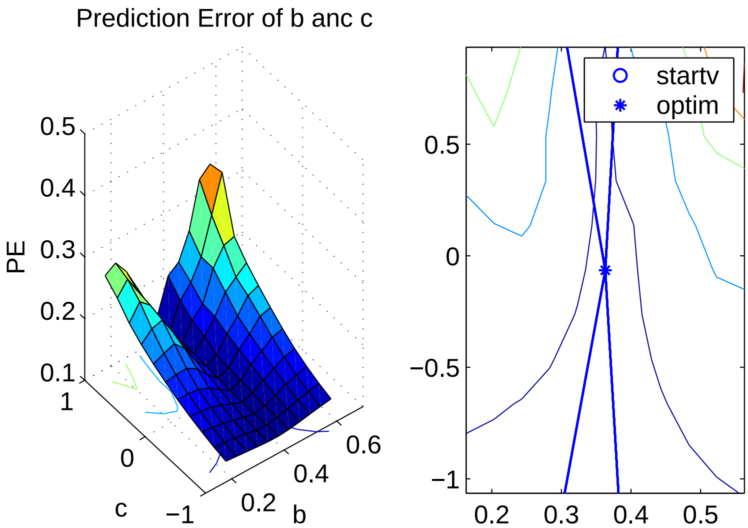


Figure 4.6: min PE w.r.t b and c

here referred to as the OLS model. We thereafter compare its results to the EKF model.

Using a rolling window regression instead of regular linear regression allows for some time dynamics in the parameters; a shorter window length is preferred in order to capture the time dependence. However, to obtain accurate parameter estimates the window should not be too short and the trade-off between window length and accuracy is not at all obvious. For our purpose we will use a window length of four years, even if the authors of [9] use five years.

The OLS time t predictions of alpha and beta are obtained from a regression, based on $t-208$ to $t-1$ fund and index returns, on the following model.

$$r_t - r_f = \alpha + \beta(r_t^M - r_f) + \epsilon_t \tag{4.9}$$

Repeating the regression at each point in time in the out-of-sample period will result in a series of OLS predicted alpha and beta. Figure 4.11 displays the beta estimate obtained from a rolling window regression with a window length of four years. We call this beta estimate OLS-beta, and we see how this model fails in picking up any time dynamics in the fund's beta. What remains to do is to determine which model, the EKF or the OLS, is preferable to the other in predicting alpha.

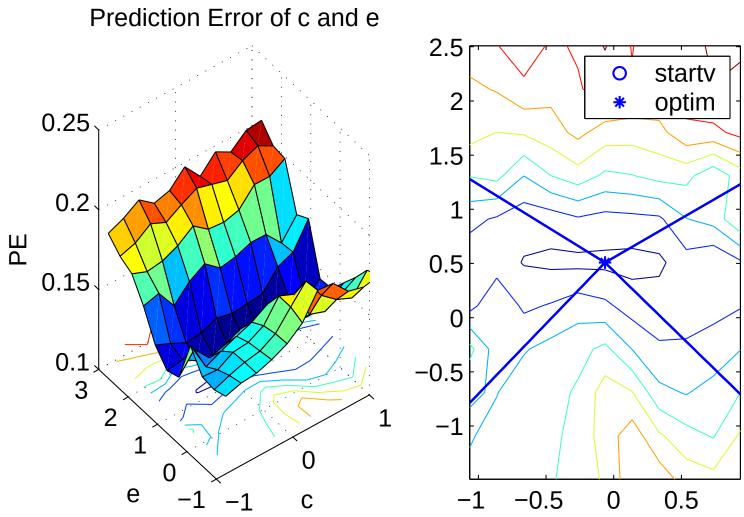


Figure 4.7: min PE w.r.t.c and e

4.4.3 Dynamic vs. static model

We se a distinct difference between the EKF estimated beta and the OLS beta. Both estimates have an average around 0.85, but the OLS estimated beta does not deviate from its average value nearly as much as the EKF beta.

As Figure 4.11 shows, the period starts with an EKF beta around 0.93, e.g. the expected return of the fund at this point, according to the EKF model, would be 0.93 of the index return. The OLS model however estimates it to be 0.844. The OLS model predicts a constant beta over the whole out-of-sample period, whereas the EKF beta decreases down to 0.78 during the first 20 week period and goes up to 0.96 in the following 30 week period, indicating how the fund covariates with the index.

In this particular example the OLS beta is more or less constant through out the whole out-of-sample period. However, this is not the case for all funds. Some funds do actually show some time dynamics in their OLS beta, and alpha as well, but in comparison to the EKF beta it picks up these changes much later than the EKF beta. The EKF model is quicker to adapt itself to changing conditions.

As discussed earlier, the gain of accurately estimating funds' betas is that it results in better estimated alphas. What remains to examine is which model's predicted alpha best agrees with the actual realized alpha.

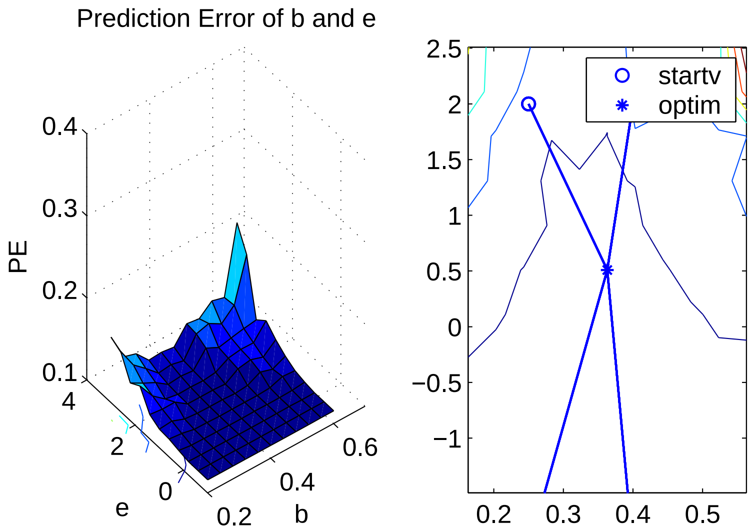


Figure 4.8: min PE w.r.t. b and e

Predicting alpha

The prediction of a fund's beta over the next investment period will also generate a signal of a funds upcoming alpha. We here illustrate the differences between the OLS and EKF models' alpha signal by comparing it to the realized alpha. That would reveal which of the OLS and EKF estimated beta that better describes a funds beta. Figure 4.12 illustrates the EKF and OLS predicted alpha compared to the realized alpha. The realized returns are composed by the average of time $t+1$ and time t realized alpha.

$$\alpha_{2,t} = \frac{\alpha_{t+1} + \alpha_t}{2} \tag{4.10}$$

$$\hat{\alpha}_t^{EKF} = -0.04 - 1.88\tilde{x}_t + 2.2412\tilde{x}_t^2 \tag{4.11}$$

where $\tilde{x}_t = 0.93x_{t-1}$

We see from Figure 4.12 that the two models predict alpha very differently. Acting on the signal from the EKF model results in more accurately estimated alpha than the corresponding for the OLS model in the way it follows the variations in realized alpha. In periods where the fund on average produces positive alpha, the EKF alpha manages to indicate just that. Even if the EKF predicted alpha is somewhat delayed, it picks up the variations in the realized alpha to a much greater extent than the does OLS model, which instead predicts alpha to constantly be below zero. This serves as evidence that the time dynamic model can be preferable to the static one, and that

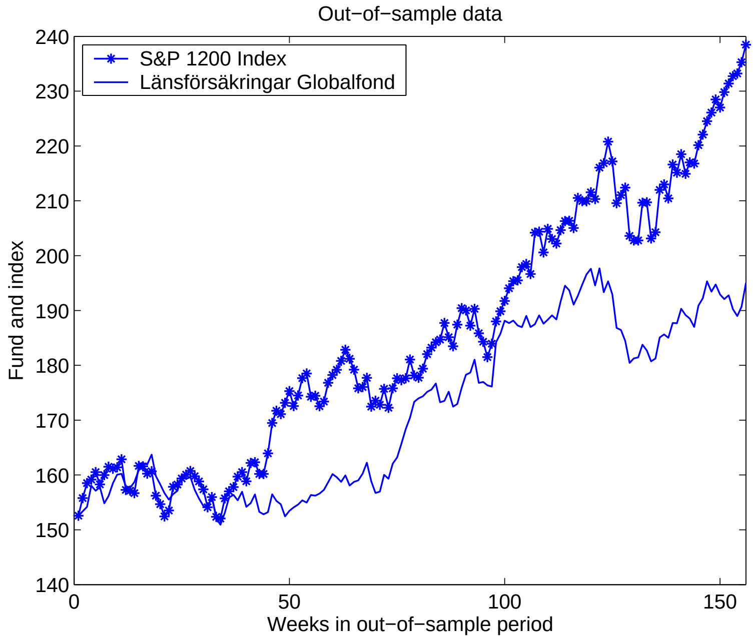


Figure 4.9: Länsförsäkringar's Globalfonden & S&P Global 1200 Jan 2004 - April 2007

capturing the time dynamics of beta yields more accurate alpha predictions.

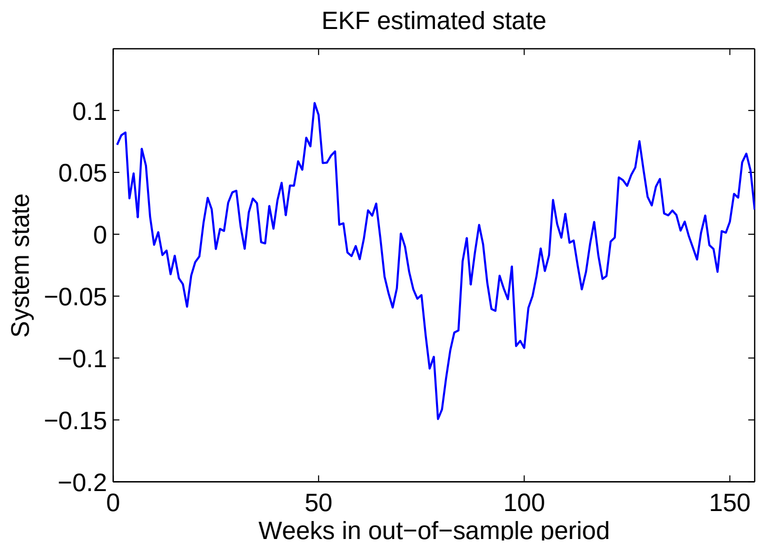


Figure 4.10: EKF estimated state

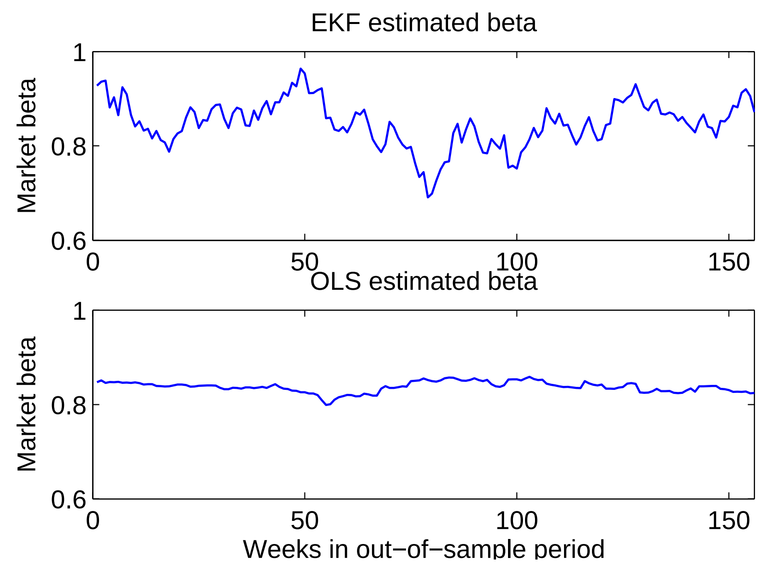


Figure 4.11: EKF and OLS estimated beta

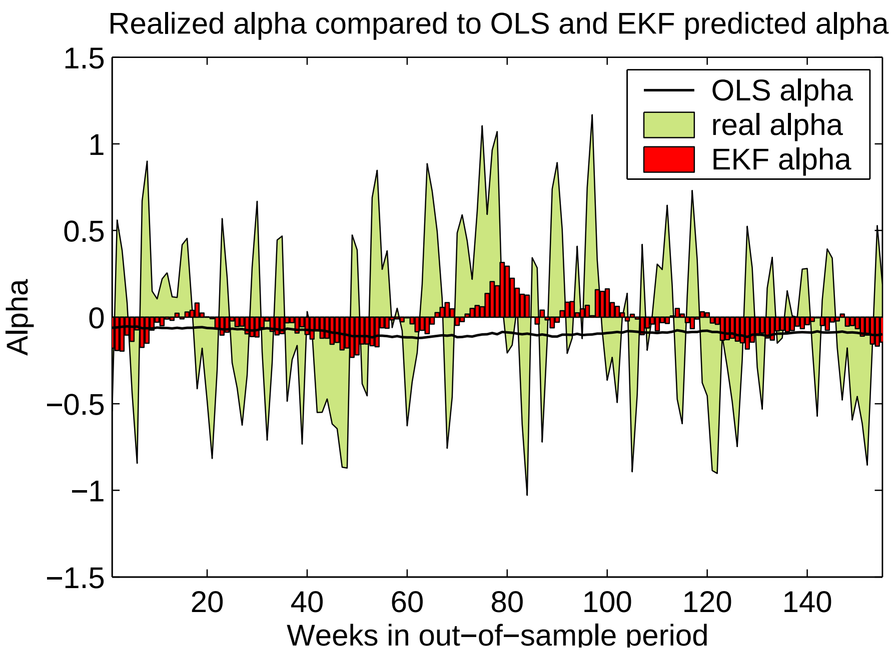


Figure 4.12: Comparing real and predicted Alpha

Chapter 5

Out-of-sample test

In the previous chapter we examined both the EKF and OLS model applied to a single fund and concluded that the EKF model can perform better in predicting alpha than the OLS model in the way it captures the time dynamics of beta. In this chapter we will test and compare the two models on a larger sample of funds. This is done with an out-of-sample test in order to try to illustrate the differences alpha and beta prediction ability on a large scale between the two models.

5.1 Out-of-sample period and fund universe

We will perform the out-of-sample test on a fund universe consisting of approximately 100 equity funds from Morningstar's database. The whole back testing period uses 7 years of data, beginning in January 2000 and ending in April 2007. The EKF model is identified for each fund using the first four years of data, and each model's alpha and beta prediction ability is then tested on the remaining three-year out-of-sample period, January 2004 - April 2007.

This test will handle investment periods of length one week. One could be interested in the two or three step prediction ability or perform the test on monthly basis. However, the out-of-sample test will be outlined in a way that illustrates how the alpha signal lasts over time.

5.2 Fund ranking

Having the approximately 100 funds at hand we determine which model that best describes the time dynamics of alpha and beta via a ranking procedure. Using data up to and including time $t-1$, we calculate the time t predictions of alpha, $\hat{\alpha}_t$, and rank the funds into deciles thereafter. The tenth decile represents the funds that at time $t-1$ had the 10 per cent highest predicted alphas, etc. We thereafter compare the deciles' predicted alpha to their realized alpha.

The realized alpha is composed in different ways in order to capture different aspects of the two models' accuracies. Firstly we want to capture how the alpha signal lasts over time and we do that by comparing predicted alpha to a moving average of the future realized alphas of order one to three. For example, when setting moving average order $q = 2$, we will compare the time t predicted alpha to the average of time t and time $t+1$ realized alpha. This gives us some understanding of how the alpha signal lasts over more than one investment period

$$\alpha_{q,t} = \frac{\alpha_{t+q-1} + \alpha_{t+q-2} + \cdots + \alpha_t}{q}. \tag{5.1}$$

Secondly, we use three different definitions of realized alpha to compare the predictions with. One is independent of both the OLS- and EKF beta estimates, the other two use the EKF and the OLS beta respectively.

$$\begin{aligned} \alpha_t^1 &:= r_t - r_t^M, \\ \alpha_t^{\beta_{EKF}} &:= r_t - \beta_{EKF} r_t^M, \\ \alpha_t^{\beta_{OLS}} &:= r_t - \beta_{OLS} r_t^M. \end{aligned} \tag{5.2}$$

Carrying out the ranking procedure described above in each time-step throughout the whole three-year out-of-sample period will result in a time series of realized alpha for each decile i:

$$\begin{aligned} \boldsymbol{\alpha}^i &= \left\{\alpha_s^i,\ 0 \le s \le t\right\}, \\ \text{where } \alpha_s^i &= \{\alpha_j\},\ j \in I_s^i, \end{aligned}$$

where $I_s^i = \{\text{all funds that at time } s \text{ are ranked to decile } i\}$.

To illustrate how the differently ranked funds have performed on average, the geometric mean of each decile's alpha series $\boldsymbol{\alpha}^i$ is calculated. Naturally one expects the relation to be positive, that the low-alpha ranked funds will also produce low alpha and the corresponding for the highly ranked funds.

5.3 Results of out-of-sample test

Figures 5.1 and 5.2 display the results from the out-of-sample test. The bars represent the geometric mean of the realized alpha, $\boldsymbol{\alpha}^i$, $i = \{1, 2, \ldots, 10\}$

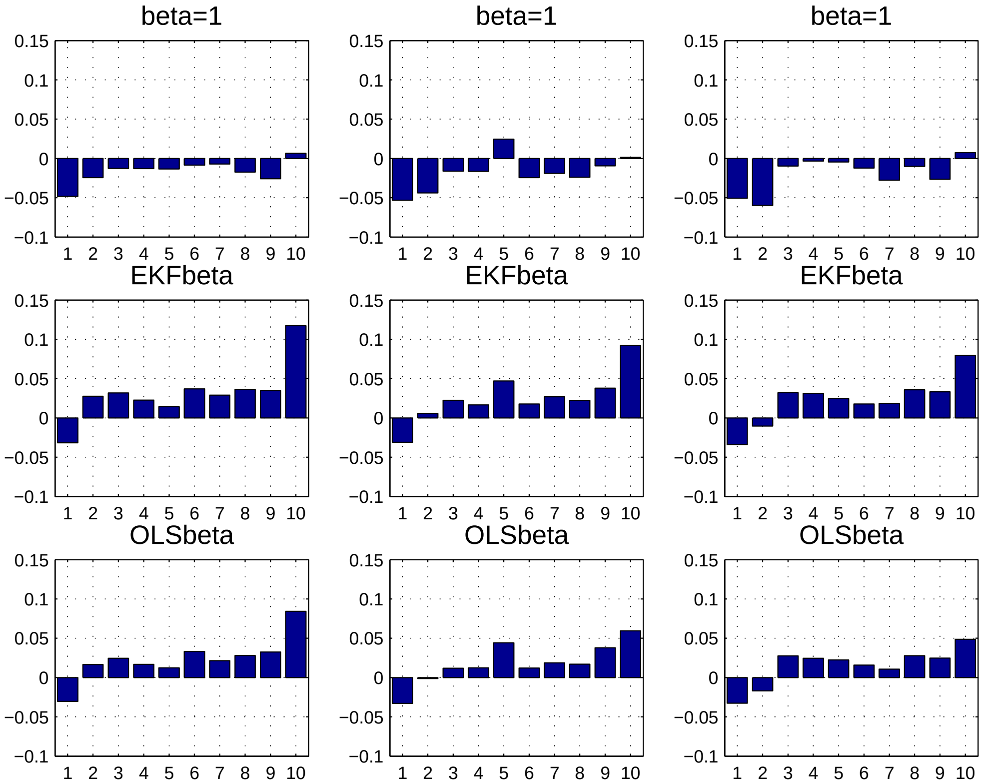


Figure 5.1: EKF average alpha, decile 1 to 10

over the three-year out-of-sample period. A complete table of the geometric mean for the differently defined realized alpha and moving average orders q are found in appendix B, and the results for Figure 5.1 and 5.2 are found in Tables1 5.3 and 5.4. The alpha predictions for each model are found in Figure 5.3. This shows that the EKF model predicts higher alpha than the OLS for the top and ninth decile ranked funds.

Figure 5.1 displays the results from the EKF ranking and Figure 5.2 shows the corresponding for funds sorted on OLS alpha predictions. Row one to three in the figure compares the predicted alpha to the realized alpha in (5.2) and the columns one to three is the comparison to the moving average of realized alpha, $q = 1, \ldots, 3$.

5.3.1 Differences in alpha prediction ability generally

The Figures 5.1 and 5.2 reveal that levels of average realized alpha calculated with EKF beta are higher for the top and the lower EKF deciles than the corresponding for the OLS model. They also show that $\boldsymbol{\alpha}^{10}$ of both models

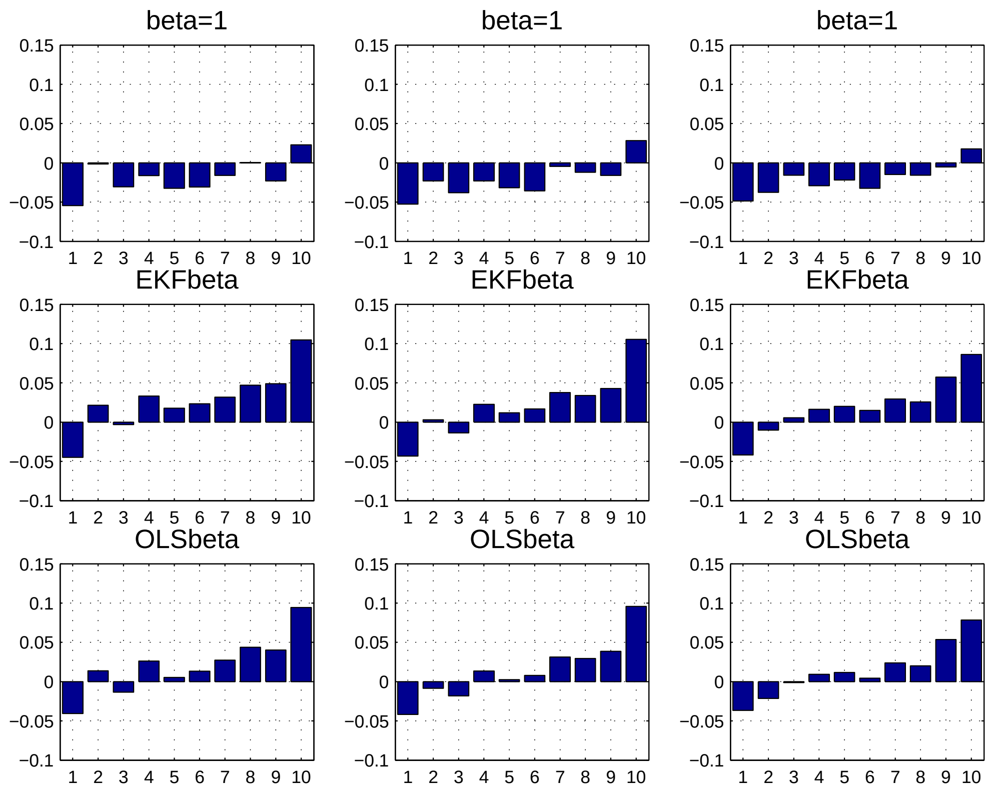


Figure 5.2: OLS average alpha, decile 1 to 10

decreases with time, which is in line with our expectation; alpha generating information will decline with time, which makes the alpha signal decay. However, the figures show that the EKF alpha signal seems to decrease faster with time than the corresponding for OLS.

Another interesting feature Figures 5.1 and 5.2 show is that the levels of average alphas are higher for realized alpha calculated using the EKF beta. Looking at the top row of Figures 5.1 and 5.2, which shows the average alpha calculated with $\beta = 1$ that is independent of both the EKF and OLS estimates, we see that the levels of average alpha are actually higher for the OLS model than the EKF. It seems that, when applied on a large sample of funds, the OLS model exhibits better prediction ability than the EKF model.

The variation around the geometric mean is very large, so even if the level of average alpha is higher for the tenth EKF decile than the OLS, we cannot say that is a statistically significant difference. The standard deviations of the deciles' can be found in Table 5.1. Note that the extreme deciles, $\boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}^{10}$ have the greatest variation around its geometric mean. This could indicate that top ranked funds' predictions are somewhat unreliable, especially for

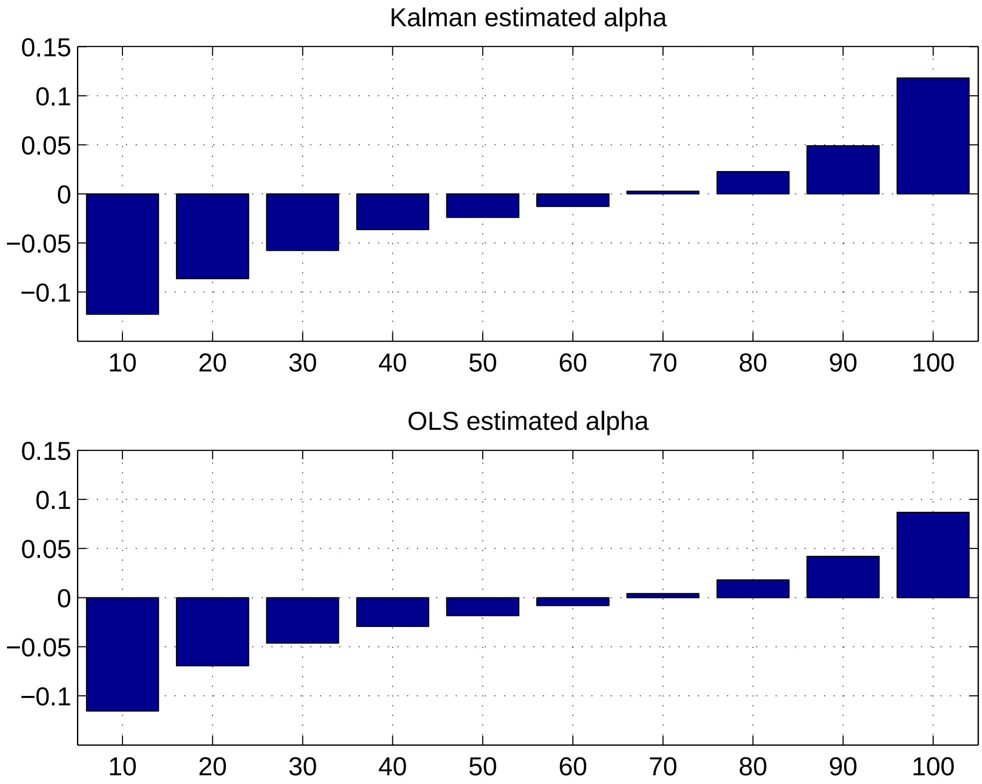


Figure 5.3: OLS and EKF predicted alpha

the EKF model.

Table 5.1: Standar deviation of $\boldsymbol{\alpha}^i$, $i = 1, 2, \ldots, 10$.

Model	1	2	3	4	5	6	7	8	9	10
EKF	13.89	11.48	10.63	8.78	8.32	7.94	7.58	7.09	7.53	12.46
OLS	13.14	11.68	10.03	9.55	9.30	8.66	7.61	7.14	8.07	9.85

5.4 Improved ranking

We noticed in the previous section that the geometric mean of $\boldsymbol{\alpha}^{10}_{EKF}$ decreases faster than the corresponding OLS as we look at investment periods of more than one step ahead. We also saw, by examining the top row of the figures, that the OLS model seem to perform better in predicting alpha on average than the EKF model. There might be several reasons for that. One

Table 5.2: **Geometric mean and standard deviation of whole-sample-alpha** *All funds' differently defined realized alphas in the three-year out-of-sample period*

Differently defined alphas			
	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$
$\bar{\alpha}$	-0.022	0.0296	0.0218
σ	0.753	0.711	0.686

Table 5.3: **Geometric mean of EKF $\boldsymbol{\alpha}^i$.** *Average alpha in per cent generated by EKF alpha signals in the out-of-sample period. Columns represent decile i=1,...,10, and rows use realized alphas as defined in (5.2). Moving average order* $q = 1$.

α	1	2	3	4	5	6	7	8	9	10
α^1	-5.05	-2.45	-1.27	-1.29	-1.34	-0.84	-0.07	-1.74	-2.58	0.64
$\alpha^{\beta_{EKF}}$	-3.45	2.77	3.19	2.27	1.42	3.69	2.89	3.63	3.47	11.74
$\alpha^{\beta_{OLS}}$	-3.17	1.65	2.46	1.67	1.23	3.33	2.15	2.80	3.25	8.43

explanation could be that the system parameters θ obtained in the optimisation are miss specified for some funds. For non-linear models, convergence of the parameter search algorithm is not guaranteed, and that can cause miss specified system parameters, which in turn produce unreliable predictions of alpha. This is one drawback of the EKF model. Also, the EKF model is much more computationally demanding and because of the occasionally non-convergence, some funds need to be treated individually. The OLS model, on the other hand, is much more rewarding to directly apply on large sample of funds in the aspect of convergence.

The authors of [10] take note of these facts and present an improvement of the ranking in order to reduce the impact of miss specified system param-

Table 5.4: **Geometric mean of OLS $\boldsymbol{\alpha}^i$** *Average alpha in per cent generated by OLS alpha signals in the out-of-sample period. Columns represent decile* $i = 1, \ldots, 10$, *and rows use realized alphas as defined in (5.2). Moving average order* $q = 1$.

α	1	2	3	4	5	6	7	8	9	10
α^1	-5.44	-0.14	-3.06	-1.63	-3.23	-3.07	1.60	0.04	-2.31	2.28
$\alpha^{\beta_{EKF}}$	-4.47	2.15	-0.31	3.33	1.78	2.34	3.19	4.70	4.89	10.46
$\alpha^{\beta_{OLS}}$	-4.07	1.38	-1.34	2.61	0.53	1.33	2.73	4.37	4.02	9.43

eters. The formulation of the restriction takes into account a fund's previous prediction ability and rules unrealistic estimates according to the following scheme:

- If time $t-1$ Kalman predicted alpha and time $t-1$ realized alpha had the same sign, the fund is included in the ranking at time t, otherwise it is ignored.
- If Kalman time $t-1$ predicted alpha and beta lie within specific bounds:

$$-0.4 \leq \hat{\alpha} \leq 0.4, \tag{5.3}$$

$$0 \leq \hat{\beta} \leq 2, \tag{5.4}$$

they will be including in the ranking, otherwise ignored.

In that way one can avoid ranking a fund highly as a result of highly predicted but miss specified alpha. Checking for previous prediction ability and prediction bounds in the EKF ranking results that can be found in Figure 5.4 and Table 5.5.

Table 5.5: **Geometric mean of EKF $\boldsymbol{\alpha}^i$ in per cent with restriction.** *Average alphas in Figure 5.4 for decile i=1,...,10 and q=2*

α	1	2	3	4	5	6	7	8	9	10
α^1	-5.87	-0.35	-0.94	-2.45	2.38	-1.60	4.93	-0.61	0.13	6.03
$\alpha^{\beta_{EKF}}$	-5.97	2.70	-0.32	-1.21	2.09	1.04	7.86	4.28	4.86	11.19
$\alpha^{\beta_{OLS}}$	-5.07	1.75	-0.20	-1.61	1.95	0.24	6.73	3.46	4.07	8.77

Table 5.6: **Geometric mean of OLS $\boldsymbol{\alpha}^i$ in per cent with restriction.** *Average alphas in Figure 5.4 for decile i=1,...,10 and q=2*

α	1	2	3	4	5	6	7	8	9	10
α^1	-5.87	-3.67	-2.26	3.80	-2.01	-3.05	-0.78	-1.75	1.29	8.00
$\alpha^{\beta_{EKF}}$	-4.13	-2.18	4.33	0.10	0.25	0.32	3.03	2.23	5.71	13.66
$\alpha^{\beta_{OLS}}$	-3.80	-2.44	3.59	-0.74	-0.32	-0.14	2.71	1.26	5.44	13.45

The results show that the top Kalman decile's average alpha does not decrease as rapidly as without the restriction, and the levels of average alpha are still higher for realized alpha calculated with β_{EKF}. Putting the same restriction on the ranking of OLS sorted funds does too improve its alpha prediction ability, see Figure 5.5.

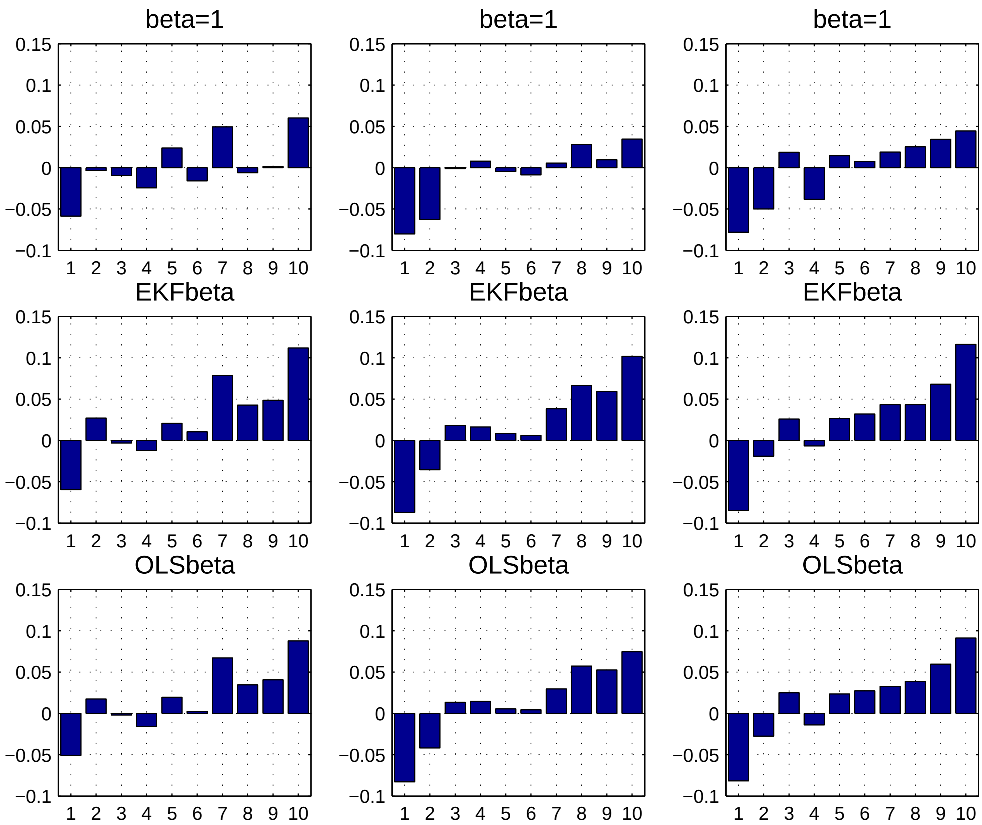


Figure 5.4: EKF average alpha with restriction. Decile $i = 1, \ldots, 10$.

The restriction rules out several funds at each step in time and the $\boldsymbol{\alpha}^i$, $i = 1, \ldots, 10$ now become half as long. Without the restriction, the alpha series has approximately 1,100 observations, whereas using the restriction shortens the alpha series to approximately 600 observations.

We conclude from the out-of-sample test that the OLS model exhibits better alpha prediction ability than the EKF model, when applied on a large sample of funds. Due to the non-linearity of the EKF model, the search algorithm might diverge and cause miss specified system parameters. That will in turn produce abnormal and unreliable alpha predictions for some funds. However, as we saw in the single fund case in previous chapter, the EKF model can still perform much better than the OLS. Therefore, when examining funds performance individually at a deeper level, the EKF is to prefer to the OLS, since we can fine-tune the model calibration procedure by hand.

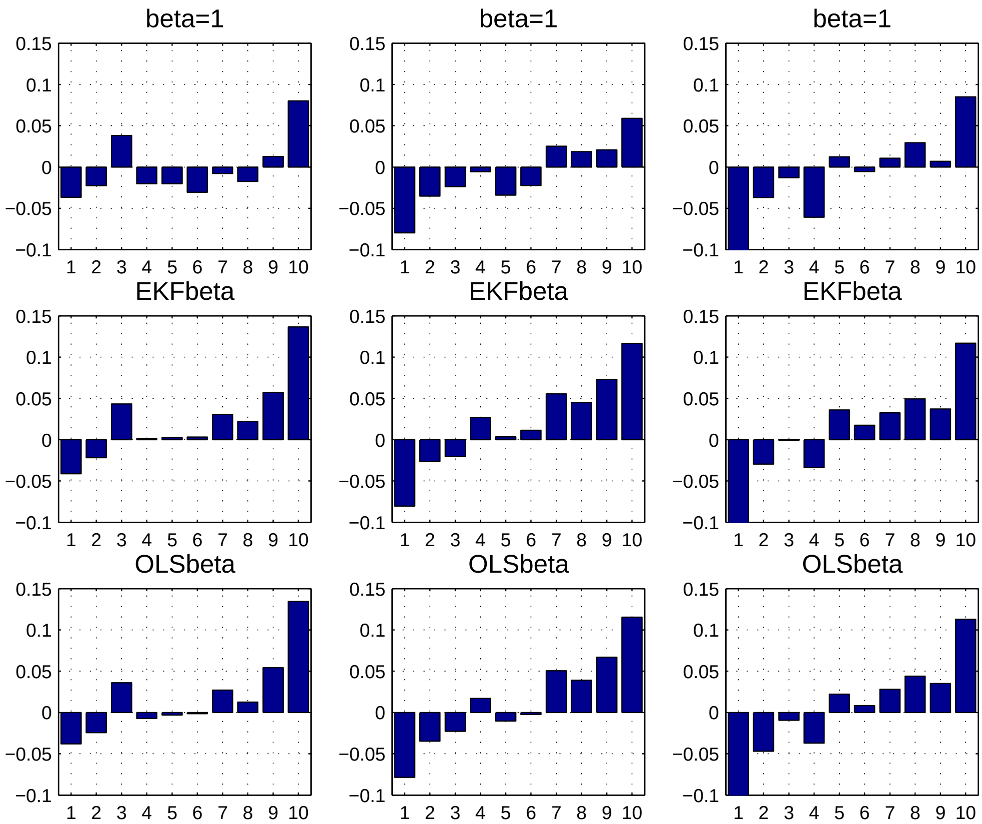


Figure 5.5: OLS average alpha with restriction. Decile $i = 1, \ldots, 10$.

Chapter 6

Conclusions

The discovery of autocorrelation in funds' betas arises the question of whether a static model, as the CAPM, really is suitable for estimating funds' beta. It is reasonable to expect that beta estimates from an OLS regression will be biased, and miss estimated betas will induce miss estimated alphas. The authors of [9] and [10] actually find that sorting on funds' α and β estimates from such a model in fact sorts the funds on estimation error. Thus, a model that allows for time changes in its parameter is required in order to obtain accurate estimates of funds' alphas and betas.

The time dynamic model that we use in this thesis for the purpose of estimating time varying alpha and betas is the so-called EKF model. It consists of a system of non-linear equations and is estimated via an Extended Kalman filter, EKF. The properties of the time dynamic model gives us reason to believe that the EKF model will produce more accurately estimated alphas and betas for funds.

We see from the single fund case in chapter four that the EKF estimated alphas pick up the time variation in the realized alphas to a much greater extent than the model based on OLS.

An out-of-sample test on a large sample of funds shows that the EKF model can generate higher alphas on average for tenth decile ranked funds over one investment period. However, the alpha signal generated by the EKF model seem to decay faster than the corresponding for the OLS. The OLS model, when applied to a large set of funds, ranks funds well and in some cases even better than the EKF model, despite its disability to adapt to changing conditions. The reasons for that have to do with the non-linearity of the system of which the EKF model is based. The non-linear system is first locally linearized via a first order Taylor expansion around the current estimate, and a standard Kalman filter is then applied to the resulting linear system in the prediction and the update phase of the algorithm. Due to these approximations the EKF will not be an optimal filter in least squares, and because of the non-linearity, convergence in the parameter search algorithm is not guaranteed. This can cause misidentified system parameters in the system identification routine, which could result in unreliable estimates of alpha and beta. By that, the EKF model loses some of its predictive power for some funds. In the single fund cases however, when the algorithm seem to have converged, we can see that the EKF is preferable to the OLS model. Putting a restriction on the ranking, designed to filter out misidentified system parameters and unreliable predictions, improves the results in the out-of-sample test for both models.

There is some room left for improvements of the EKF model. For example, we could consider a time variant system of equations in our state space representation. This means that for each point in time we want to make a prediction, we re-identify our system. We can then take advantage of previously identified parameters and use them as start values in the upcoming optimisation. Also, we can formulate the function of the prediction error in different ways in the optimisation routine.

Since the OLS model performed well in the out-of-sample period, it can be meaningful to improve the OLS model too. We could try different window lengths, and use a weighted OLS that weighs observations in the past less heavily than recent observations.

We conclude by stating that both models perform well in predicting alphas and betas, but in different ways. On a large scale, the OLS model is preferable to the EKF since convergence of the parameter search algorithm can not be assured. However, for single funds, when the system identification has succeeded, the EKF model performs better than the OLS in predicting time dynamic alphas and betas.

Bibliography

[1] Welch, G. & Bishop, G. (2006): *An Introduction to the Kalman Filter*. Technical report. Department of Computer Science, University of North Carolina, Chapel Hill.

[2] Brockwell, P.J & Davis, R.A. (2002): *Introduction to Time Series and Forecasting*. Springer.

[3] Brockwell, P.J & Davis, R.A. (1991): *Time Series: Theory and Methods, Second Edition*. Springer.

[4] Brown, R.G & Hwang, P.Y.C. (1992): *Introduction to Random Signals and Applied Kalman Filtering, Second Edition*. John Wiley & Sons, Inc.

[5] Harvey, A.C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University press.

[6] Lindquist, A. & Sand, J.: *An Introduction To Mathematical Systems Theory*. Lecture notes. Optimization and Systems Theory, Royal Institute of Technology, Stockholm, Sweden.

[7] Ljung, L. & Söderström, T. (1983): *Theory and Practise of Recursive Identification*. The Massachusetts Institute of Technology.

[8] Luenberger, D.G. (1998): *Investment Science*. Oxford University Press, Inc.

[9] Mamaysky, H., Spiegel, M. & Zhang, H. (2004): *Estimating the Dynamics of Mutual Fund Alphas and Betas*. Technical report. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id= 389740.

[10] Mamaysky, H., Spiegel, M. & Zhang, H.(2005): *Improved Forcasting of Mutual Fund Alphas and Betas*. Technical report. Yale International Center for Finance.

[11] Maybeck, P.S. (1979): *Stochastic Models, Estimation, and Control, Volume 1.* Academic Press, Inc.

[12] Ribeiro M.I. (2004): *Kalman and Extended Kalman Filters: Concept, Derivation and Properties.* Technical report. Institute for Systems and Robotics, Instituto Superior Tcnico, Lisboa, Portugal.

Appendices

Appendix A

The EKF equations

We here give a description of the derivation of the EKF equations. Some of the steps are based on the derivation of the standard Kalman filter, which can be found in [6].

Consider our process having state vector $x \in \mathbb{R}^n$ that is governed by the non-linear stochastic difference equation:

$$x_t = f(x_{t-1}, v_t) \tag{A.1}$$

The measurement $y \in \mathbb{R}^r$ is of the form

$$y_t = h(x_t, u_t, w_t) \tag{A.2}$$

We have that

$$f : \mathbb{R}^n \to \mathbb{R}^n \tag{A.3}$$

$$h : \mathbb{R}^n \to \mathbb{R}^r \tag{A.4}$$

where $v_t \in \mathbb{R}^n$ and $w_t \in \mathbb{R}^r$ and $v_t \sim N(0, Q)$ and $w_t \sim N(0, R)$ represent the state and measurement noise processes. Also we have $\mathrm{E}\left[v_t w_t\right]^2 = 0$.

Due to the non-linearity in the state and measurement equation we can approximate x_t and y_t with the following:

$$\tilde{x}_t = f(\hat{x}_{t-1|t-1}, 0) \tag{A.5}$$

$$\tilde{y}_t = h(\tilde{x}_t, u_t, 0) \tag{A.6}$$

where $\hat{x}_{t-1|t-1}$ is the *a posteriori* estimate from time step $t-1$.

Now, consider the last filtered estimate of the state $\hat{x}_{t-1|t-1}$ We want to linearize x_t around this last estimate $\hat{x}_{t-1|t-1}$ as follows:

$$x_t \cong f(\hat{x}_{t-1|t-1}, 0) + A(x_{t-1} - \hat{x}_{t-1|t-1}) + V v_t \tag{A.7}$$

$$y_t \cong h(\tilde{x}_t, u_t, 0) + H(x_t - \tilde{x}_t) + W w_t \tag{A.8}$$

where

$$
\begin{aligned}
A &= \frac{\partial f}{\partial x}(\hat{x}_{t-1|t-1}, 0) && \text{(A.9)} \\
V &= \frac{\partial f}{\partial v}(\hat{x}_{t-1|t-1}, 0) && \text{(A.10)} \\
H &= \frac{\partial h}{\partial x}(\tilde{x}_t, u_t, 0) && \text{(A.11)} \\
W &= \frac{\partial h}{\partial w}(\tilde{x}_t, u_t, 0) && \text{(A.12)}
\end{aligned}
$$

Now we define a notation for the prediction error:

$$\tilde{e}_{x_t} \equiv x_t - \tilde{x}_t \quad \text{(A.13)}$$

and the measurement residual

$$\tilde{e}_{y_t} \equiv y_t - \tilde{y}_t \quad \text{(A.14)}$$

Remember, one does not have access to x_t in Equation A.13 since it is the actual state which we want to estimate. However, we do have access to y_t which we use to estimate x_t. We can now write equations for *error processes* as

$$
\begin{aligned}
\tilde{e}_{x_t} &\cong \frac{\partial f}{\partial x}(\hat{x}_{t-1|t-1}, 0)(x_{t-1} - \hat{x}_{t-1|t-1}) + \eta_t && \text{(A.15)} \\
\tilde{e}_{y_t} &\cong \frac{\partial h}{\partial x}(\tilde{x}_t, u_t, 0)\tilde{e}_{x_t} + \epsilon_t && \text{(A.16)}
\end{aligned}
$$

where ϵ_t and η_t represent new independent random variables with zero mean and covariance matrices VQV^T and WRW^T, where V and W are as in Equation A.10 and A.12.

Note that Equations A.15 and A.16 are linear and they closely resemble the difference and measurement equation in a standard discrete Kalman filter. For a detailed description of the standard Kalman filter see [6]. We can thereby use the actual measurement residual $\tilde{e}_{y_t}$ in Equation A.14 to estimate the prediction error $\tilde{e}_{x_t}$ given by Equation A.15.

Let the prediction error obtained from the standard Kalman filter be $\hat{e}_t$. If we use it in Equation A.13 we get an *a posteriori* state estimate $\hat{x}_{t|t}$:

$$\hat{x}_{t|t} = \tilde{x}_t + \hat{e}_t \quad \text{(A.17)}$$

The random variable A.15 is approximately (see footnote in [1]):

$$\tilde{e}_{x_t} \sim N(0, \mathrm{E}\left[\tilde{e}_{x_t}\tilde{e}_{x_t}^T\right] \quad \text{(A.18)}$$

By letting the predicted value of $\hat{e}_t$ be zero, the Kalman filter equation used to estimate $\hat{e}_t$ is

$$\hat{e}_t = K_t \tilde{e}_{y_t} \tag{A.19}$$

By subsituting Equation A.19 back into Equation A.17 and using Equation A.14 we get an expression of the a posteriori estimate of the time t state:

$$\begin{aligned} \hat{x}_{t|t} &= \tilde{x}_t + K_t \tilde{e}_{y_t} && \text{(A.20)} \\ &= \tilde{x}_t + K_t(y_t - \tilde{y}_t) && \text{(A.21)} \end{aligned}$$

Equation A.21 is now used for the measurement update in the extended Kalman filter, with $\tilde{x}_t$ and $\tilde{y}_t$ coming from Equations A.5 and A.6, and K_t is the Kalman gain coming from the standard Kalman filter.

The Algorithm

The extended Kalman filter consist of one prediction phase and one measurement update phase according to the following equations:

Time update equation - Prediction

$$\begin{aligned} \tilde{x}_t &= f(\hat{x}_{t-1|t-1}, 0) \\ \tilde{P}_t &= A\hat{P}_{t-1}A^T + VQV^T \end{aligned}$$

The prediction equations project the state and covariance estimates from the previous time step $t-1$ to the current step t.

Measurement update equations

$$\begin{aligned} K_t &= \tilde{P}_t H_t^T (H_t \tilde{P}_t H_t^T + WRW^T)^{-1} \\ \hat{x}_{t|t} &= \tilde{x}_t + K_t(y_t - h(\tilde{x}_t, u_t, 0) \\ P_t &= (I - K_t H_t)\tilde{P}_t \end{aligned}$$

where

$$\begin{aligned} A &= \frac{\partial f}{\partial x}(\hat{x}_{t-1|t-1}, 0) \\ V &= \frac{\partial f}{\partial v}(\hat{x}_{t-1|t-1}, 0) \\ H &= \frac{\partial h}{\partial x}(\tilde{x}_t, u_t, 0) \\ W &= \frac{\partial h}{\partial w}(\tilde{x}_t, u_t, 0) \end{aligned}$$

Appendix B

Out-of-sample test

We here display the results from the out-of-sample test.

Table B.1: **Geometric mean of EKF and OLS α_i, q=1.** *Average alpha generated by EKF and OLS alpha signals in the out-of-sample period. Rows represent decile i=1,...,10, and columns use realized alphas as defined in Equations 5.2.*

Decile	**EKF model**			**OLS model**		
i	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$
1	-0.0483	-0.0317	-0.0303	-0.0544	-0.0447	-0.0407
2	-0.0245	0.0277	0.0165	-0.0014	0.0215	0.0138
3	-0.0127	0.0319	0.0246	-0.0306	-0.0031	-0.0134
4	-0.0129	0.0227	0.0167	-0.0163	0.0333	0.0261
5	-0.0134	0.0142	0.0123	-0.0323	0.0178	0.0053
6	-0.0084	0.0369	0.0333	-0.0307	0.0234	0.0133
7	-0.0070	0.0289	0.0215	-0.0160	0.0319	0.0273
8	-0.0174	0.0363	0.0280	0.0004	0.0470	0.0437
9	-0.0258	0.0347	0.0325	-0.0231	0.0489	0.0402
10	0.0064	0.1174	0.0843	0.0228	0.1046	0.0943

Table B.2: **Geometric mean of EKF and OLS $\boldsymbol{\alpha_i}$, q=2.** *Average alpha generated by EKF and OLS alpha signals in the out-of-sample period. Rows represent decile i=1,...,10, and columns use realized alphas as defined in Equations 5.2.*

Decile	**EKF model**			**OLS model**		
i	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$
1	-0.0533	-0.0310	-0.0329	-0.0524	-0.0432	-0.0419
2	-0.0438	0.0056	-0.0013	-0.0229	0.0030	-0.0085
3	-0.0162	0.0224	0.0118	-0.0380	-0.0136	-0.0181
4	-0.0164	0.0166	0.0124	-0.0230	0.0228	0.0136
5	0.0246	0.0469	0.0443	-0.0316	0.0120	0.0026
6	-0.0244	0.0177	0.0121	-0.0357	0.0169	0.0079
7	-0.0190	0.0270	0.0187	-0.0045	0.0377	0.0313
8	-0.0239	0.0223	0.0171	-0.0121	0.0340	0.0294
9	-0.0095	0.0379	0.0379	-0.0161	0.0428	0.0385
10	0.0012	0.0921	0.0596	0.0283	0.1055	0.0958

Table B.3: **Geometric mean of EKF and OLS $\boldsymbol{\alpha_i}$, $\boldsymbol{q = 3}$.** *Average alpha generated by EKF and OLS alpha signals in the out-of-sample period. Rows represent decile i=1,...,10, and columns use realized alphas as defined in Equations 5.2.*

Decile	**EKF model**			**OLS model**		
i	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$
1	-0.0508	-0.0340	-0.0325	-0.0486	-0.0417	-0.0367
2	-0.0597	-0.0105	-0.0169	-0.0375	-0.0102	-0.0214
3	-0.0096	0.0320	0.0275	-0.0157	0.0056	-0.0012
4	-0.0034	0.0310	0.0246	-0.0291	0.0164	0.0093
5	-0.0046	0.0245	0.0224	-0.0218	0.0202	0.0115
6	-0.0124	0.0177	0.0159	-0.0323	0.0149	0.0045
7	-0.0278	0.0181	0.0107	-0.0149	0.0295	0.0239
8	-0.0102	0.0358	0.0278	-0.0157	0.0257	0.0202
9	-0.0266	0.0333	0.0248	-0.0051	0.0574	0.0536
10	0.0073	0.0795	0.0487	0.0177	0.0862	0.0783

Table B.4: **Geometric mean of EKF $\boldsymbol{\alpha_i}$, ranking restricted and q=2.** *Average alpha generated by EKF alpha signals in the out-of-sample period, when checking previous prediction ability. Rows represent decile i=1,...,10, and columns use realized alphas as defined in Equations 5.2.*

Decile	**EKF model**			**OLS model**		
i	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$
1	-0.1109	-0.0874	-0.0860	-0.0524	-0.0432	-0.0419
2	-0.0459	-0.0024	-0.0085	-0.0229	0.0030	-0.0085
3	-0.0332	-0.0005	-0.0091	-0.0380	-0.0136	-0.0181
4	-0.0415	-0.0117	-0.0130	-0.0230	0.0228	0.0136
5	0.0279	0.0328	0.0322	-0.0316	0.0120	0.0026
6	-0.0438	-0.0061	-0.0132	-0.0357	0.0169	0.0079
7	0.0231	0.0753	0.0674	-0.0045	0.0377	0.0313
8	-0.0264	0.0464	0.0387	-0.0121	0.0340	0.0294
9	-0.0154	0.0534	0.0507	-0.0161	0.0428	0.0385
10	0.0176	0.1108	0.0772	0.0283	0.1055	0.0958

Table B.5: **Geometric mean of EKF $\boldsymbol{\alpha_i}$, ranking restricted and $\boldsymbol{q}$=3.** *Average alpha generated by EKF alpha signals in the out-of-sample period, when checking previous prediction ability. Rows represent decile i=1,...,10, and columns use realized alphas as defined in Equations 5.2.*

Decile	**EKF model**			**OLS model**		
i	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$
1	-0.0944	-0.0806	-0.0762	-0.0486	-0.0417	-0.0367
2	-0.0705	-0.0338	-0.0377	-0.0375	-0.0102	-0.0214
3	-0.0076	0.0321	0.0278	-0.0157	0.0056	-0.0012
4	0.0000	0.0322	0.0248	-0.0291	0.0164	0.0093
5	-0.0196	0.0116	0.0083	-0.0218	0.0202	0.0115
6	-0.0115	0.0164	0.0148	-0.0323	0.0149	0.0045
7	-0.0200	0.0349	0.0274	-0.0149	0.0295	0.0239
8	-0.0017	0.0550	0.0482	-0.0157	0.0257	0.0202
9	-0.0263	0.0443	0.0359	-0.0051	0.0574	0.0536
10	0.0093	0.0984	0.0697	0.0177	0.0862	0.0783

Table B.6: **Geometric mean of EKF $\boldsymbol{\alpha_i}$, ranking restricted and $\boldsymbol{q}$=4.** *Average alpha generated by EKF alpha signals in the out-of-sample period, when checking previous prediction ability. Rows represent decile i=1,. . . ,10, and columns use realized alphas as defined in Equations 5.2.*

Decile	**EKF model**			**OLS model**		
i	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$	α^1	$\alpha^{\beta_{EKF}}$	$\alpha^{\beta_{OLS}}$
1	-0.0892	-0.0824	-0.0786	-0.0424	-0.0349	-0.0320
2	-0.0851	-0.0321	-0.0411	-0.0283	-0.0053	-0.0160
3	-0.0170	0.0113	0.0086	-0.0082	0.0187	0.0099
4	-0.0674	-0.0153	-0.0201	-0.0250	0.0098	0.0085
5	0.0101	0.0366	0.0342	-0.0186	0.0217	0.0078
6	-0.0215	0.0143	0.0116	-0.0308	0.0128	0.0059
7	0.0113	0.0500	0.0455	0.0052	0.0394	0.0358
8	0.0100	0.0281	0.0244	-0.0066	0.0336	0.0270
9	0.0272	0.0659	0.0602	-0.0297	0.0269	0.0242
10	0.0123	0.1020	0.0814	0.0295	0.0888	0.0819