1-1-2011

Time-Varying Coefficient Models And The Kalman Filter : Applications To Hedge Funds

Ana G.S. Punales
Ryerson University

Time-Varying Coefficient Models and the Kalman Filter. Applications to Hedge Funds

by

Ana Gloria Suarez Punales

Bachelor of Science, Havana University, 2002

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Science

in the Program of

Applied Mathematics

Toronto, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Time-Varying Coefficient Models and the Kalman Filter. Applications to Hedge Funds

Master of Science 2011

Ana Gloria Suarez Punales

Applied Mathematics

Ryerson University

There are various studies concerned with the estimation of stochastically varying coefficients for the hedge fund series but just few are available in the literature that study the model with time-varying coefficients and non-linear factors, or make a comparison of the series before and during the financial crisis. This work studies a model with linear and non-linear factors with stochastically varying coefficients to obtain better estimation of the exposure of the hedge fund and accuracy in the results. Better exposure estimates implies better hedging against negative changes in the market hence a reduction in the risk taken by the hedge fund manager. Besides, different techniques have been studied, implemented and applied in this thesis to estimate and analyze time varying exposures of different HFRX Index (an index that describes the hedge fund industry performance).

The study shows that option-like models with time-varying coefficients perform the best for most of the HFRX indexes analyzed. It also shows that the Kalman Filter technique combined with the Maximum Likelihood Estimator is the best approach to estimate time-varying coefficients. In addition, we provide evidence that Kalman Filter is in a better position to capture changes in the exposure to the market conditions.

Keywords: Time-varying coefficients, Kalman filter, hedge funds, financial crisis

Acknowledgements

First and foremost, I would like to take the opportunity to express my humble gratitude to all people who support me and without whom this thesis would not have been possible.

My deepest gratitude to my supervisors, Dr. Pablo Olivares and Dr. Marcos Escobar, who spend their time and shared their knowledge for helping me to complete my thesis. They supported me throughout my thesis with their patience, encourage and advise, while allowing me the space to do the work in my own way. I could not wish for better or friendlier supervisors.

My most sincere gratitude also goes with the Department of Mathematics at Ryerson University, which assisted me for the last couple of years. Particularly to those who were my professors for sharing with me their expertise and experience. I would like to thank all of them for their trust, support and guidance that helped me late with my thesis research and studies.

I also thank my fellow classmate for their interesting discussions, for all the working together before deadlines and for all of the enjoyable and fun time we have had during the last two years. Especially to Azi, that more than a classmate becomes one of my best friend, supporting and helping me in tough times. Thank you for your patience, good company and for being my friend.

Lastly but most importantly, I would like to thank my husband Carlos for his support, patience, encouragement, friendship and love that were the foundation upon which the past twelve years of my life have been built. I thank my family and my husband's family, without they help and support I would not be here and in this stage of my life now.

I would not finish without saying how grateful I am with my parents, Teresa and Raymundo for their understanding and faith in me. They have been a constant source of support, emotional, moral and of course financial during my entire life. They have always supported and encouraged me to do my best in all matters of life. To them I dedicate this thesis.

> *"In the end we will conserve only what we love; we will love only what we understand; and we will understand only what we have been taught."*
>
> — Baba Dioum

Dedication

To the pillars of my life: my parents and my husband. Especially to my dad that is no longer with us but who inspired my love for math.

"The happiest moments of my life have been the few which I have passed at home in the bosom of my family."

—Thomas Jefferson

Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

Beta is a systematic risk measure, also called a market risk measure. It represents one of the most widely used concepts in finance. In a sense, beta is an incomplete explanation of risk and returns because a low beta does not necessarily mean low risk, it simply means low exposure to the market, or more simply, low market risk [LHabitant, 2001]. Beta is used by financial economists and practitioners to estimate asset's sensitivity to the overall market, to apply a variety of valuation models, to evaluate the performance of the asset managers, among others.

Our objective in this thesis is to model beta varying on time, such as a random walk or as a periodic function. Then, we can estimate beta by using different techniques in a variety of models.

In the context of the Capital Asset Pricing Model (CAPM), [LHabitant, 2001], beta is assumed to be constant over time and it is estimated using an Ordinary Least Squares (OLS) approach. However, due to the fact that beta depends on economic's factors some studies have found evidence of time dependent betas. [Fabozzi and Francis, 1978] suggest that beta is random for a significant number of securities. [Sunder, 1980] introduces a model where some variations of beta are not independent in time. His results indicate a higher level of non-stationarity beta in some period of time but not in others. [Bos and Newbold, 1984] study a market model in which they allow a stochastic beta to obey a first order autoregressive process. In their empirical study they found strong evidence indicating stochastic systematic risk. Support studies for the hypothesis of time dependent exposure are provided by [Daniel W. Collins and Rayburn, 1987] and [Kim, 1993] among others.

The most intuitive and straightforward explanation for the exposure of hedge fund be time dependent states in their nature. For example, the portfolio exposure can change due to time variant market premiums or beta underlying assets, while portfolio weights are kept constant. But also, the hedge fund managers are allowed to change the exposures in order to adapt their portfolios to changing market environments, thus the portfolio weights are dynamically changed over time.

Several alternatives have been used in the financial literature to model time-varying returns and risk. One idea to make beta time dependent is that there is a linear relationship between the coefficients at time t and the coefficients at time $t-1$. The idea is that at time $t-1$ the hedge fund manager observes the market and adjusts his portfolio accordingly. [Yao and Gao, 2006] use a dynamic model

1

and recursive filtering on the Australian industry portfolios to confirm that betas on stock portfolios vary in time.

In the fund industry there exist various papers studying the estimation and the use of regression models with stochastic coefficients but just few of them explain the properties of the estimator and the parameters, and it is what [Pagan, 1980] does. Under some conditions he obtains that the estimator is consistent and asymptotically normal. He also provides sufficient conditions for the estimation of regression models with stationary stochastically varying coefficients [Pagan, 1980]. As there exist different techniques that allow for the modeling and estimation of time varying betas we decided to focus on three of them: the rolling windows, the Kalman filter approach and seasonal periodic betas. Theses techniques provide the estimation of beta series through time allowing to examine and analyze the time varying behavior of betas.

Regression analysis is one of the most widely used and misused techniques in economics and finance as well as in several other fields. However, the standard assumptions of these models are most of the time violated principally when time series data are used. One of the main assumptions of the standard regression analysis is that the variance of the errors is constant. [Brealey and Kaplanis, 2001] studied whether the exposure in hedge fund returns is constant. They found that the null hypothesis of constant coefficients is strongly rejected when using a rolling window regression, which indicates that the factor exposures of hedge fund are time varying.

It is well known that the role of regression analysis is to describe exactly the specific relationship that exists between the two variables under the assumption that the coefficients of the regression model are constant over time. But there is an ample group of practical applications where this is not a realistic assumption, such as in financial or economical studies. For example, in financial investment, it is generally agreed that stock markets' volatility is rarely constant over time [Wei, 2006], therefore an option-like factor in the hedge fund models indicate that the exposure of it is not constant but changing over time. It is precisely the base of the OLS-rolling window methodology that it is a regression analysis applied in each window assuming a fixed style allocation over time. It is hardly sustainable, that at the end beta will be time dependent.

By using rolling windows, the fund's investment style is constantly monitored, making classification and style exposures readily identifiable. Such a rolling style benchmark tracks the manager's actual returns much more closely than static benchmarks. It determines how the fund's styles may have changed over time. It also allows faster reactions to changes in management style, and so it provides an early warning of potential changes in whether funds adhere to their stated investments styles over time.

From a theoretical point of view, there is no argument to justify the use of rolling windows in model analysis. The rolling window technique even creates a contradiction between the estimation model assuming constant exposures and the final output that shows time-varying exposures. If it is true that the exposure can change over time, their variations should be explicitly modeled rather than estimated by rolling window assuming them to be constant first in some intervals of time. The Kalman Filter is a technique that allows precisely this sort of modeling.

Named after Rudolph E. Kalman (1940), the Kalman filter approach is a computational algorithm that makes optimal use of imprecise data in a linear or near linearly system with Gaussian errors to contin-

uously update the best estimate of the past of the data, current and even future states. In a sense, the Kalman filter is similar to least squares fitting, but allows for real time updates to fit.

The filter only uses past information, but it reacts very quickly to a change in its environment. This confirms its usefulness and applicability to the changing nature of hedge fund portfolios and the importance of it to detect crisis or some big changes in the market. Indeed, the Kalman filter is suited to take into consideration the multiple investment style variations of actively managed hedge funds [Swinkels and Sluis, 2006], but its use has been rather limited in the literature due to the limited size of hedge fund databases. In addition, due to the complexity of understanding and implementing the Kalman filter, it has not been widely used over the traditional regression analysis in most of the statistical inference problems.

There are some studies where the researchers use Kalman filter to estimate different time-varying coefficients models in order to see which model suits better the data. For example, [Wells, 1994] estimates a dynamic version of the market model (CAPM) by using the Kalman filter technique. They allow the parameters to be dependent on their past values. There were no gain model but somehow the random walk model was preferred by using the Akaike information criterion to compare the out-of-sample forecasting ability of the models. The researches not only compare models but also most of the studies are based in the comparison of different estimator's techniques including Kalman filter approach. In the paper of [Robert D. Brooks and McKenzie, 1999] the estimation of the conditional time varying betas for Australian portfolios using monthly data was made. In this paper it can be found evidence supporting the Kalman filter approach, based on in-sample and out-of-sample forecast errors. Also, [Robert W. Faff and Hillier, 2000] estimate the time dependent exposure of 32 different UK industry sectors by some models, including the regression model with time varying beta where the Kalman filter was used for estimation. Their study concludes overwhelmingly that market model betas are unstable and betas estimated using Kalman filter approach are consistently more efficient than other methods used in the paper. [Ebner and Neumann, 2004] evaluate the estimation of beta using a rolling window regression, a random walk using Kalman filter and a flexible least square model for individual German stocks. They found evidence that the rolling window regression is even worse than the constant beta estimated using OLS, in spite of being widely used in the financial field.

On the other hand it is known that the analysis of the time varying beta of hedge fund is an important part of the performance analysis of it, however there are not too many studies related to it in the field. [Racicot and Theoret, 2009] test the assumption that the conditional alpha and beta follow a random walk by using different models with a data of hedge fund. They propose the use of the Kalman filter approach with some supposed dynamic risk factors and measure the error of the estimation process.

While some of the models assume that the relation between the market and the hedge fund depends on time, we also analyze whether there is a close relation with nonlinear option-like exposures of the hedge fund to standard asset classes. There is a suggestion in the literature, that hedge fund returns exhibit option-like features. [Henriksson and Merton, 1981] introduce one option in the index portfolio trying to separate the portfolio manager ability to pick an option and to determine the market timing.

[Fung and Hsieh, 1997] show some results indicating that there exist some strategies of hedge funds highly dynamic. They find five dominant investment styles in hedge funds that combined with the asset

class factor model, can provide an integrated framework for style analysis of both buy-and-hold and dynamic trading strategies [Fung and Hsieh, 1997]. Then, [Agarwal and Naik, 2004] show that some of the equity hedge fund strategies present a non-linear pay-off structure. They found that the option-like strategies explain better the variation in hedge fund returns.

My thesis research in this area has been concentrated on analyzing the existence of time dependent beta in the hedge fund series and to estimate them by using a Kalman filter approach. Using linear regression methods as benchmark, since it is the most used technique by the practitioners nowadays, we work with models and methods that allow us to obtain the estimation of the time dependent parameters for a hedge fund series. Therefore, we implement a Kalman filter in order to use it with different models and hedge fund strategies.

We expect that not only the models with time dependent exposures perform better but also the model with option like and time dependent parameters will have the best performance for most of the styles of hedge funds. Since our interest is to make a comparative analysis of the behavior of the time-varying betas estimated by each technique in each model and to compare the techniques by their accuracy to forecast the parameters, we use in-sample performance criteria. Some studies on the modeling hedge fund series with stochastic coefficients and the estimation of them have emerged in recent years, but only few of them are available in the literature on analysis of the series before and after a financial crisis. Therefore, we make an analysis if this issue, before and after a financial crisis basis in order to examine the effect of different data frequency on the result of the study as well as the timing to locate the crisis by each technique of parameter estimation.

The rest of the thesis is organized as follows: an overview of models and methods with constant and time dependent parameters used for testing is presented in Chapter 2. Most of the estimation of the parameters for these models requires a Kalman filter technique, thus Chapter 3 establishes the assumptions of the models that can be used by Kalman filter as well as an explanation of the Kalman filter approach, some of its characteristics and its implementation. Chapter 4 describes and analyzes the data, and also shows the test and results. At the end the conclusions and further future work are described to complete the thesis.

Chapter 2

Models and estimation methods

Hedge fund returns can be expressed as the sum of the exposure to the market, which is measured by beta, and the abnormal returns, defined by alpha. Alpha characterizes the managers selection skill. Now, a portfolio may exhibit time-varying alpha and beta for three reasons: Firstly, the underlying assets in the portfolio may have changing alphas and betas. Secondly, due to active trading strategies and thus dynamic change of portfolio weights, alpha and beta of the portfolio may be time-varying. Thirdly, both of these characteristics may be inherent in the portfolio. While the first behavior is not connected to the fund manager, the second is an indication of market timing skills of the manager.

A study of some hedge fund return model regression are presented below. In this chapter we describe first the factor models and then the cases of one, two or three factors models when the parameters are constants or time dependent. Then, we explain some of the methods used to estimate the parameters, their ideas and the assumptions needed to apply them.

2.1   Factor Models

Factor models are useful for asset pricing, portfolio management, risk measurement and, more generally for any discipline that needs information about the co-movements of different assets. Technically, a factor model does no more than condensing the dynamics of a large series of fund returns into smaller series of explanatory factors, whose influence is common to all funds. In a sense, the small set of factors and factor exposures provide a parsimonious representation of the large set of funds. That is, it explains most of the variance and covariance of the funds considered, see [LHabitant, 2001]

In practice, the use of factor analysis is supported by the observation that hedge fund returns tend to react together to some extent. This confirms the intuition that fund returns are likely to be affected by the same factors at the same time. Consequently, it is meaningful to attempt to capture the common behavior of a series of funds by one or several factors.

Multi-factor modeling is a general form of factor model, and it is the most popular model for the return

generating process. The return y_t is represented as,

$$y_t = \alpha_t + \sum_{i=1}^{k} \beta_{it} F_{it} + \varepsilon_t \tag{2.1}$$

where:

y_t = Return of the hedge fund at time t.

F_{it} = Factors affecting the returns at time t.

α_t = The "zero" factor which is the value added by the manager at time t

β_{it} = Exposure of the hedge fund return to the factor i at time t.

ε_t = random disturbances of the return at time t.

k: The number of factors and it is a positive integer larger than zero.

The statistical noise ε_t, which corresponds to the residual return is a zero mean random variable. It is generally assumed that the covariance between ε_t and factor F_{it} is zero.

Some domains of applications for hedge fund multi-factor models are the following:

- The identification of the relevant drivers of performance of a portfolio of funds or a hedge fund index.

- The understanding of the factors that together explain the total risk of a portfolio of funds or a hedge fund index.

- The creation of a benchmark.

- The construction of a portfolio of funds tilted toward specific risk factors of choice.

- The construction of index trackers or enhanced index portfolio.

- The determination of alpha which describes the degree to which a product outperforms a comparable product or index.

Furthermore, factor models provide an interesting and natural explanation for the change in correlations observed between hedge funds.

In this section we consider one, two and three factor models. We focus on the parameters' characteristic of these models, i.e. if they are constant or changing in time, therefore, in the equation of the models and the properties of their parameters.

2.1.1 Case of one factor models: $k = 1$

In this case the model represented in 2.1 can be written as,

$$y_t = \alpha_t + \beta_t X_t + \varepsilon_t \tag{2.2}$$

where X_t is the market return indicators at time t.

Linear Regression with constant coefficients

The classical regression model is given by,

$$y_t = \alpha + \beta X_t + \varepsilon_t \tag{2.3}$$

where ε_t is a white noise. It assumes that the relationship between the explanatory and explained variables remains constant through the estimation period.

Linear Regression with Time-Varying Coefficients

In this model the parameters alpha and beta can be taken in many different ways, but here we assume that they evolve according to a random walk which is confirmed by a unit root test in Chapter 4. This means that the current exposure to the market is a normally distributed random variable taking as mean the exposure of the last period. The corresponding noises are normally distributed and not correlated. Therefore, the system can be written as:

$$
\begin{aligned}
y_t &= \alpha_t + \beta_t X_t + \varepsilon_t \\
\alpha_t &= \alpha_{t-1} + \epsilon_t \\
\beta_t &= \beta_{t-1} + \xi_t
\end{aligned}
\tag{2.4}
$$

where:
$\epsilon_t = $ Alpha disturbances at time t.
$\xi_t = $ Exposure disturbances at time t.
and,

$$
\begin{aligned}
\varepsilon_t &\sim \aleph(0, \sigma^2_{\varepsilon_t}) \\
\epsilon_t &\sim \aleph(0, \sigma^2_{\epsilon_t}) \\
\xi_t &\sim \aleph(0, \sigma^2_{\xi_t})
\end{aligned}
\tag{2.5}
$$

Also there are some time varying characterizations of the exposure that can be considered dependent on the style of the hedge fund return used,

$$
\begin{aligned}
\beta_t &= \phi(\beta_{t-1} - \bar{\beta}) + \bar{\beta} + \epsilon_t, &&\text{AR(1)} \\
\beta_t &= \bar{\beta} + \xi_t, &&\text{Random Coefficient} \\
\beta_t &= \beta_{t-1} + \xi_t, &&\textbf{Random walk used in this work} \\
\beta_t &= a(b + \sin(wt)), &&\textbf{Periodic Beta used in this work}
\end{aligned}
$$

We just use two of them and in section 4.2.3 in Chapter 4 the empirical testing of the hypothesis for using these models are explained.

2.1.2   Case of two factors models: $k = 2$ or option-like models

The literature has identified several problems with the above linear models when used for the task of performance evaluation. For example, these models restrict the relationship between risk factors and returns to be linear, and thus do not properly evaluate the assets with non-linear payoffs.

[Henriksson and Merton, 1981] were the first to suggest using options to explain the performance of managed portfolios. Their model is basically a simplified two-state framework for performance evaluation in which a fund manager attempts to forecast whether the market return will be higher or lower than the risk free rate [LHabitant, 2001].

In this context hedge fund returns may be viewed as a linear combination of the market's returns and a call option or a put option on the underlying market. A call option is an agreement that gives an investor the right (but not the obligation) to buy a stock, bond, commodity, or other instrument at a specified price within a specific time period. The put option is the opposite of a call option, it is an option contract giving the owner the right, but not the obligation, to sell a specified amount of an underlying security at a specified price within a specified time.

The exercise price of this call option is taken as the risk free rate, hence the option becomes more valuable only if the market return is lower than the risk free rate. Consequently, the following equation of regression provides consistent estimates for timing and selectivity:

$$y_t = \alpha_t + \beta_{1t} X_t + \beta_{2t} \max\{X_t - K_t, 0\} + \varepsilon_t \tag{2.6}$$

in the case of a call option. In the case of the put option the model can be written as:

$$y_t = \alpha_t + \beta_{1t} X_t + \beta_{2t} \max\{K_t - X_t, 0\} + \varepsilon_t \tag{2.7}$$

where K_t is the risk free rate.

[Henriksson and Merton, 1981] show that $\beta_{2t} > 0$ if and only if the fund manager has a superior market timing ability of the fund while $\alpha_t > 0$ still indicates selection ability. In fact, a positive parameter β_{2t} can be seen as the number of no cost options on the market portfolio provided by the market timing strategy. A $\beta_{2t} < 0$ and $\alpha_P = 0$ [1] are equivalent to being short a number of options on the market without receiving any cash.

Option-like model with constant coefficients

Here, we assume the parameters to be constant through time. It means that at time t and $t - 1$ the parameters must be:

$$\begin{aligned} \alpha_t &= \alpha \\ \beta_{1t} &= \beta_1 \\ \beta_{2t} &= \beta_2 \end{aligned} \tag{2.8}$$

---

[1]P is a portfolio

where α, β_1, β_2 are constants.

Option-like model with Time-Varying Coefficients

In this case we assume that the parameters have a linear relationship with the previous values in the series. Moreover, we assume that the parameters at time t are random walks given by,

$$
\begin{aligned}
\alpha_t &= \alpha_{t-1} + \epsilon_t \\
\beta_{1t} &= \beta_{1,t-1} + \xi_{1t} \\
\beta_{2t} &= \beta_{2,t-1} + \xi_{2t}
\end{aligned}
\tag{2.9}
$$

but as we said before betas can be characterized by other time dependent models.

2.1.3   Case of three factors models: $k = 3$

In this case the model represented in 2.1 can be written as,

$$
y_t = \alpha_t + \beta_{1t}X_{1t} + \beta_{2t}X_{2t} + \beta_{3t}X_{3t} + \varepsilon_t
\tag{2.10}
$$

with the parameters constant or changing on time as follows:

$$
\begin{aligned}
\alpha_t &= \alpha_{t-1} + \epsilon_t \\
\beta_{1t} &= \beta_{1,t-1} + \xi_{1t} \\
\beta_{2t} &= \beta_{2,t-1} + \xi_{2t} \\
\beta_{3t} &= \beta_{3,t-1} + \xi_{3t}
\end{aligned}
\tag{2.11}
$$

where X_{1t}, X_{2t}, X_{3t} can be taken as different market return indicators at time t or as a combination of a call-option and put-option.

2.2   Estimation methods

After having established a set of models that are well suited to describe the returns of a particular fund we can attempt to estimate the fund's exposure, with respect to the factor and all of the parameters of the model by using different approaches. Since there is an ample variety of techniques to estimate the beta, it is critical that we adopt the most appropriate modeling and estimation technique. In this section we describe few methods used by researchers to estimate the exposure and the parameters mentioned above.

2.2.1   Ordinary Least Squares

A popular technique used to estimate the unknown parameters, such as exposures, is Ordinary Least Squares (OLS) [LHabitant, 2001]. The OLS is a statistical technique that uses sample data to estimate

the relationship between two variables. This technique is used frequently because its implementation is straightforward and it is easy to understand.

The basis of the OLS method is to minimize the sum of the squared difference between the observed responses in the data set, y, and the responses predicted by the linear approximation model 2.3, $\hat{y}$. That is,

$$\min_{\theta} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \tag{2.12}$$

where θ is the set of parameters to estimate and T is the total number of observations.

One virtue of the OLS approach is that it is very easy to implement computationally. In particular this approach has three most desirable properties:

- Gives simple closed-form solutions for the parameters estimation.

- Penalizes large individual errors and ensure that all errors remain small on average.

- Has good statistical properties under plausible assumptions about the error term. One can show:

  1. the regression line always passes through the points of means $(\overline{X}, \overline{y})$.

  2. the errors have zero covariance with the sample of X values and also with the predicted $\widehat{y}$ values.

  3. the estimation of the parameters, $\widehat{\alpha}$ and $\widehat{\beta}$, are the best linear unbiased estimators of the values α and β.

These properties make the OLS technique the most widely used in practice when performing regression analysis [LHabitant, 2001]. However, a critical assumption necessary to obtain robust exposures from OLS estimation is that they remain constant over the estimation period. Consequently, it will be difficult to capture the diverse and dynamic behavior of the hedge funds using a model based solely on regression estimation. Indeed, it would not be real to assume that hedge fund managers do not change their factor exposures over the life of their fund. A lack of rigid investment restrictions provides hedge fund managers with the flexibility to make rapid and significant changes in their style, sector or market bets according to their future expectations. As a result, hedge fund managers can be much more dynamic in their investment approach than traditional managers. As said before, the technique is used when the model has constant coefficients, for example it can be used in models 2.3, 2.8

2.2.2 Rolling Window Regression

A method used frequently within the regression framework to account for the dynamic behavior of hedge funds is the Rolling Window (RW) regression. The method involves using a shorter and more recent data window to estimate the regression parameters by using the Ordinary Least Square method at each window. Discarding past data in this manner will allow the model to capture recent changes in the exposures more rapidly.

The RW method comes at a cost of statistical accuracy since the estimation is performed using a

smaller data sample. In addition, the method will not capture exposure changes over short windows [Robert W. Faff and Hillier, 2000]. The estimation window can be chosen in an ad-hoc manner and it may not be optimal for every hedge fund. Choosing a size of the window is not easy since this may result in the exclusion of critical data from the estimation process of the parameters for some windows. Therefore, it will strongly depend on the individual dynamic of each fund. Consequently, the use of a shorter data window may produce noisy estimates and inferior forecasts. The RW method is used when the model has time varying coefficients, for example it can be used in models 2.4, 2.9. The use of RW comes with the easy way to implement and explain it combined with the assumptions that the coefficients are time dependent.

2.2.3 Kalman Filter approach

As we mentioned before, rolling window regression is unable to capture immediate changes in the exposures as the estimated parameters are depending upon the length of the estimation window. Therefore, when the exposures of the hedge funds vary over an estimation window, then the use of a more general method can improve the estimation of the parameters. Here we present the Kalman Filter (KF) State-Space Model, since Kalman filter overcomes the issues observed in rolling window regression. The Kalman Filter estimation technique is presented with more detail in Chapter 3.

Kalman Filter (KF)

The Kalman filter method, originally developed by Kalman (1960) within the context of linear systems, is a recursive algorithm procedure for computing the optimal estimator of the state vector at time t, based on the information available at time $t - 1$ ([Harvey, 1989]) and for forecasting variances in time series models.

It was invented to solve a problem in spacecraft navigation, but the technique nowadays is relevant for helping to solve many problems where incomplete observations must be combined with a state of a system. As a mathematical point of view, Kalman filter does not solve any problem by itself, it is only a mathematical tool that help us to make the problem easier to understand. As statistical point of view of an estimation problem, Kalman filter is more than an estimator because it propagates through the prediction step of the filter the entire probability distribution of the variables are asked to estimate. This is a complete statistical characterization of the current state of knowledge of the dynamic system, including the influence of all past measurements.

A Kalman filter combines all available measurement data, plus prior knowledge about the system and measuring devices to produce an estimation of the desired variables such that any quadratic function *(estimation error)* is statistically minimized. Hence it is going to minimize, in this research, the square of the errors of the regression model.

The Kalman filter provides a linear estimation method for any time series model that can be represented in a state space form. The origin of state space models can be traced to dynamical system in engineering branches including automatic control, communications, robotics, and aerospace systems such as spacecraft altitude control [Kedem and Fokianos, 2002].

If Y(t) represent the output, A(t) the state vector, and U(t), the input or error, then the general state-space equations that creates the relationship between the variables are the nonlinear equations,

$$Y(t) = G(A(t), U(t), t)$$
$$\frac{dA(t)}{dt} = F(A(t), U(t), t) \qquad (2.13)$$

The discrete-time linear case is,

$$Y(t) = A(t)X(t) + U(t)$$
$$A(t+1) = M(t)A(t) + V(t) \qquad (2.14)$$

The statistical adaptation of equations 2.13 and 2.14 is widely used in discrete time regression-like models made of two interconnected equations, the observation equation and the system equation, which may assume various linear and nonlinear forms and commonly referred to as state space models and their application in prediction, filtering and smoothing or interpolation is crucial [Kedem and Fokianos, 2002]. The state space model provides a flexible approach to time series analysis, especially for simplifying maximum likelihood estimation and handling missing values [Tsay, 2010].

Then the application of a Kalman filter is just a matter of problem's appearance. Once the problem is formulated in terms of state-space equations, the standard Kalman filter algorithm can be applied in a straightforward manner.

State Space Model

The linear state space has been demonstrated to be an extremely powerful tool in handling all linear and many classes of nonlinear time series models [Harvey, 1989]. The state-space form provides a general framework for representing a wide range of time series models. It consists in the measurement equation (measurement model) that indicates the relationships among unobserved state variables and observed variables. And the transition equation (dynamic model) that describes the function as we described in the previous section.

A filtering problem can be put into the state space form by defining the state vector represented by certain parameters. The equation representing the state vector is known as the transition equation, and it is not observed directly. Instead, the state of the system is conveyed by an observed variable called the signal equation, which is subject to contamination by disturbance or measurement error. Now we are going to be more specific in the definition of the State-Space Model (SSM) since we need to set some assumptions over the error of the model for our cases.

Gaussian State Space Models

The state space model is similar to a regression model, but does not assume that the exposures are constant over the estimation window. Instead, it introduces stochastic elements which allows the

exposures to vary over the estimation period. Therefore, writing the model 2.1 in the state space representation is going to be useful for the estimation. The new representation of the model using this framework is given by:

$$Measurement\ equation: y_t = \alpha_t + \sum_{i=1}^{k} \beta_{it} X_{it} + u_t, \qquad u_t \sim \aleph(0, U_t) \qquad (2.15)$$

$$State\ equations: \beta_{it} = M_{1i} \beta_{i,t-1} + v_{it}, \qquad v_{it} \sim \aleph(0, V_{it}) \qquad (2.16)$$

$$\alpha_t = M_2 \alpha_{t-1} + w_t, \qquad w_t \sim \aleph(0, W_t) \qquad (2.17)$$

where $t = 1, \ldots, T$, $\{u_t\}$, $(\{v_{1t}\}, \ldots, \{v_{kt}\})$ and $\{w_t\}$ are independent and mutually independent random variables. M_{jt} are the transition matrices, $j = 1, 2$ and the order of the matrix varies depending on the order of the autoregressive process[2] taken to describe the coefficients of the model, see Chapter 3.

The measurement equation describes the relation between the data and the unobserved variables. While the transition equation describes the dynamics of the state variables based on the information from the past such that the future behavior of the system can be completely described by the knowledge of the present state and the future input. The most important difference between the conventional linear models and the state space models is that in the former the state of nature is not assumed to be constant but can change over time.

In model 2.16 - 2.17, calculation of the exposure estimates for every point in time involves the use of the Kalman filter approach. The noises u_t (*that is the amount by which X_t has changed since the prior period, time $t - 1$*) and $(v_{1t}, \ldots, v_{kt}), w_t$ are time-independent Gaussian noise processes. The exposures $(\beta_{1t}, \ldots, \beta_{kt})$, the variance of u_t, and the variance of $(v_{1t}, \ldots, v_{kt}), w_t$ are usually estimated using a maximum likelihood technique, see Chapter 3 for more details. In addition, the state-space framework and Kalman filter allow for the computation of contemporaneous, predicted and smoothed values of the exposures.

In principle, this model may be better due to its ability to capture the dynamic of exposures of the time series analyzed. However, this additional flexibility sometimes has a drawback. When the model specification is inaccurate, or when there are too few return data points available, this approach inadvertently will fit the excess noise. As was the case with the regression approach, the state space model also can suffer from outdated data if there is a significant shift in a hedge funds risk profile. However, the dynamic quality of the state space model allows it to be more adaptive, and therefore more robust than the constant exposure modeling assumption used in the regression analysis [Tsay, 2010].

The automated exposure estimation approach can vary by hedge fund. This flexibility is necessary in order to capture the dynamic behavior of a funds exposures with respect to the common factor returns. As a result, the exposure estimation methodology can vary over time and across hedge funds, or even for a particular hedge fund. For example, OLS multivariate regression might be the best choice for a hedge fund whose exposures vary little over time. Conversely, a Kalman Filter approach might be selected as the superior estimation approach for a hedge fund during a period when its common factor exposures

---

[2]Autoregressive process of order p, AR(p), can be written as: $B_t = c + \sum_{i=1}^{p} \varphi_i B_{t-i} + \varepsilon_t$ where ε_t is white noise, φ_i are the parameters of the model and c is a constant.

have been very volatile, as we can see in the last chapter of the thesis.

In the state space equation 2.16-2.17 the goal of the analysis is to infer the properties of the state β_{it}, $i = 1, \ldots, k$, from the data y_t and the chosen model. There are three important types of inference where N is the length of the data vector [Tsay, 2010]:

- Filtering for $t = N$, to recover the state variable given the information available at time t to remove the measurement errors from the data.

- Prediction for $t > N$, to forecast $\beta_{i,t+m}$ or y_{t+m} for $m > 0$ given the information available at time t, where t is the forecast origin.

- Smoothing for $t < N$, to estimate β_{it} given the information available at time T, where $T > t$

A simple analogy of the three types of inference is reading a handwritten note. *Filtering* is figuring out the word you are reading based on knowledge accumulated from the beginning of the note, *predicting* is to guess the next word, and *smoothing* is deciphering a particular word once you have read through the note.

One of the main advantages of the Kalman filter is that it can be applied in real time. That is, for any value observed of the time series, the forecast for the next observation can be computed. This makes the method very practical and important in the financial field. Therefore in the next chapter a more detailed explanation of the Kalman filter as well as its algorithm is given.

Chapter 3

Kalman Filter

The model of a stochastic system driven by a noise disturbance requires the use of a filter to estimate it. One optimal state space based estimation tool that is widely used in statistics and engineering is the Kalman filter, as it is described in Chapter 2. The filter is known to be able to support estimations for past, present, and future states even when the precise nature of the modeled system is unknown.

In this chapter we define the equations of the Kalman Filter technique and its algorithm. At the end of the chapter we use the Kalman filter approach with simulated data in order to see how it works in the environment where we have the perfect model. We use this study to choose the initial values and a right sample size we need in order to get good estimators. Good estimation in the context where the Kalman filter estimator display the minimum mean square of the innovations.

3.1   Betas varying on time

When dealing with time series data, researchers usually write the regression model as,

$$y_t = \alpha + \sum_{i=1}^{k} \beta_i X_{it} + u_t$$

for $t = 1, ..., T$ and i the number of factors, where y_t is the dependent variable, X_{it} are the vectors of exploratory variables, β_i are the unknown coefficients to be estimated, and u_t is the error with mean zero and variance σ^2

As we described in Chapter 2, the Ordinary Least Square (OLS) is one of the methods that is used to estimate the equation mentioned above. In this method it is assumed that the parameters to be estimated are constant over time. However, there are ample evidence that the parameters used may have some variations over time. Therefore, it is going to be useful to consider a model such that the parameters are a function of time, which is usually called the time-varying parameter model. There are few examples of time varying models commons in finance and economics fields. They are mostly estimated by a combination of Kalman filter technique and the log-likelihood function.

The Kalman filter technique is used to estimate the state of alpha and beta based on recursive observations of the unknown parameters for the following general state-space model with $t = 1, 2, \ldots, T$,

$$y_t = \alpha_t + \sum_{i=1}^{k} \beta_{it} X_{it} + \varepsilon_t$$

$$\beta_{it} = M_{1i} \beta_{i,t-1} + \epsilon_{it} \qquad (3.1)$$

$$\alpha_t = M_2 \alpha_{t-1} + \xi_t$$

where $\varepsilon_t \sim \aleph(0, \sigma_{\varepsilon_t}^2)$, $\epsilon_{it} \sim \aleph(0, \sigma_{\epsilon_{it}}^2)$, $\xi_t \sim \aleph(0, \sigma_{\xi_t}^2)$, and they are independent.
Assuming that the parameters alpha and beta are autoregressive of order p the matrices M_{1i} and M_2 represent the autoregressive structure matrix of the time varying coefficients β_t and α_t, and can be written as:

$$M_{1i} = \begin{bmatrix} \phi_{1i}^{(1)} & \phi_{2i}^{(1)} & \cdots & \phi_{pi}^{(1)} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \qquad M_2 = \begin{bmatrix} \phi_1^{(2)} & \phi_2^{(2)} & \cdots & \phi_p^{(2)} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

where $\{\phi_l^{(m)}\}_{l=1,\ldots,p}^{m=1,2}$ are the parameters of the autoregressive process and i are the number of factors in the model. For our study we are going to focus in a more particular case that it is when the parameters can be represented by a first order autoregressive, and in this case the transition matrices are constants. Therefore, $M_{1i} = \phi_{1i}$ and $M_2 = \phi_2$ for our specific case and the model can be written in the following matrix form,

$$y_t = Z_t \Gamma_t + \varepsilon_t$$

$$\Gamma_t = \Phi \Gamma_{t-1} + \Psi_t \qquad (3.2)$$

this is the notation being used in this chapter assuming $i = 1, \ldots, k$ is the number of factors in the model, where $\Gamma_t = (\alpha_t \quad \beta_{1t} \quad \beta_{2t} \quad \ldots \quad \beta_{kt})'$, $Z_t = (1 \quad X_{1t} \quad X_{2t} \quad \ldots \quad X_{kt})$, $\Psi_t = (\xi_t \quad \epsilon_{1t} \quad \epsilon_{2t} \quad \ldots \quad \epsilon_{kt})'$ and $\Phi_t = (\phi_2 \quad \phi_{11} \quad \phi_{12} \quad \ldots \quad \phi_{1k})'$ [1]
The error is assumed to be distributed with conditional expectation zero and covariance matrix H_t, $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = H_t$, [Harvey, 1989]. In our case we assume H_t constant over time, $H_t = H = \sigma_\varepsilon^2$. In addition, $E(\Psi_t) = (0 \quad 0)'$ and $Var(\Psi_t) = \begin{bmatrix} R_t & 0 \\ 0 & Q_t \end{bmatrix}$ where Q_t and R_t are diagonal matrices with the variance of ξ_t and $(\epsilon_{1t} \quad \epsilon_{2t} \quad \ldots \quad \epsilon_{kt})'$ on it respectively, as

---

[1]$(\bullet)'$ is the transpose matrix.

$$R_t = \begin{bmatrix} \sigma_{\xi_1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\xi_t}^2 \end{bmatrix} \qquad Q_t = \begin{bmatrix} \sigma_{\epsilon_{1t}}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\epsilon_{kt}}^2 \end{bmatrix}$$

In our study we take $R_t = R = \sigma_{\xi}^2$ as constant for any t and $Q_t = Q$ is a vector depending on $i = 1, \ldots, k$ the number of factors in the model. Vector Q contains the constant variance for each factor in our model, i.e. $Q = (\sigma_{\epsilon_1}^2, \ldots, \sigma_{\epsilon_k}^2)$.

The problem now is how to estimate a model given by 3.2. As can be seen there are two sets of unknowns: the parameters of the model and the elements of the state vectors Γ_t. However, once a model is cast into his state-space representation, the addition of certain assumptions allow the model to be estimated using maximum likelihood and the Kalman filter algorithm as we have seen in the previous chapter. That is, estimation of the state variables and system parameters usually involves two well-developed inter-related algorithms: the Kalman filter and the Gaussian Maximum Likelihood, more specifically the maximum of the Gaussian log-likelihood.

3.2 Kalman filter equations

In this section we work with a general model 3.2 with the assumptions mentioned before. Usually the parameters of the model are unknown hence by using the Kalman filter approach the problem of how to estimate the parameters of the state equation of the model, Γ_t, is solved. Therefore, finding the "best" linear estimates of the state, in the sense of minimum mean square error can solve the problem. [Pagan, 1980], [Harvey, 1989], [Grewal and Andrews, 1993], [Robert W. Faff and Hillier, 2000], [Racicot and Theoret, 2009] The difference between the best estimation of the state given the information up to $t-1$ and the result obtained at time t,

$$\begin{aligned} \varepsilon_t &= y_t - \widehat{y_t} \\ &= y_t - Z_t \widehat{\Gamma}_{t|t-1} \end{aligned} \qquad (3.3)$$

where $\hat{\Gamma}_{t|t-1}$ is the estimator at time t given the information up to time $t-1$, is called the prediction error.

Consider a_t as the optimal estimator of Γ_t based on all of the information at time t. Then, the estimator could be written as $a_t = E_t(\Gamma_t)$, i.e. the conditional expectation of the state variables up to time t. The covariance of the estimators, denoted as P_t, is defined by $P_t = E_t[(a_t - \Gamma_t)(a_t - \Gamma_t)']$. Therefore, the optimal estimator of Γ_t based on all the observations at time $t-1$ could be denoted by $a_{t-1} = E_{t-1}(\Gamma_t)$ and consequently the covariance of this estimator is defined by $P_{t|t}$.

The Kalman filtering consists in the following recursive set of equations:

$$
\begin{cases}
a_{t|t-1} = \Phi a_{t-1} & \text{(state prediction)} \\
P_{t|t-1} = \Phi P_{t-1}\Phi' + (R \quad Q)' & \text{(prediction dispersion)} \\
y_{t|t-1} = z_t a_{t|t-1} & \\
\eta_t = y_t - y_{t|t-1} & \text{(prediction error)} \\
F_t = z_t P_{t|t-1} z_t' + H & \text{(error dispersion)} \\
G_t = P_{t|t-1} z_t' F_t^{-1} & \text{(Kalman gain)} \\
a_t = a_{t|t-1} + P_{t|t-1} z_t' F_t^{-1} \eta_t & \text{(state estimate)} \\
P_t = (I_2 - K_t z_t) P_{t|t-1} & \text{(estimate dispersion)}
\end{cases}
$$

where $P_{t|t-1}$ is the covariance matrix of the error of $a_{t|t-1}$, as we defined before, hence the one-step predictor. And η_t is a one-period prediction error for y_t, called innovation, F_t^{-1} is the inverse of the covariance matrix of the innovation at time t while G_t is called the gain of the Kalman filter. The above equations are derived in [Grewal and Andrews, 1993].

The Kalman gain vector depends on Φ, $(R \quad Q)$, $H = \sigma_\varepsilon^2$ and the past data vector z_t. The Kalman gain function plays an important role in updating the estimates because it determines how heavily the innovations are weighted. When the system is linear and the normality assumptions are valid, this specific form of the Kalman gain function optimally weighs the innovations, which makes $a_t = \widehat{\Gamma}_t$ (the expectation of the conditional distribution of Γ_t given the information y_t). This is the Minimum Mean Square Estimator (MMSE) of Γ_t based on the information up to t as it is given in the theorem below. If the assumptions mentioned above are violated then the Kalman filter estimator is no longer the MMSE. However, it is still the Minimum Mean Square Linear Estimator, which means it is the optimal among all the estimators that are composed of linear combinations of the observations [Harvey, 1989].

In other words, it is well known that the Kalman filter method gives unbiased and efficient estimators of the state vector $E\left(\Gamma_{t|t}\right) = E\left(\Gamma_{t|t-1}\right) = E\left(\Gamma_t\right) = a_t$, when the initial conditions a_0 and P_0 and the matrices Φ, $(R \quad Q)$, H are known. When $(R \quad Q)$, H are unknown will be better to use the log-likelihood of the innovations combine with the Kalman filter approach to obtain consistent and asymptotically efficient estimators of Φ, $(R \quad Q)$, H; otherwise the filter cannot be optimal.

Kalman Filter estimates properties

In theorem 1 we give a result showing a property of the estimation of the Kalman filter technique and the proof can be found in [Grewal and Andrews, 1993].

Theorem 1. *Assume that the white noise ε_t and Ψ_t in model 3.2 are Gaussian and uncorrelated, i.e*

$$
\varepsilon_t \sim \aleph(0, H_t) \qquad and \qquad \Psi_t \sim \aleph(0, \Sigma_t)
$$

where $\Sigma_t = (R_t \quad Q_t)$ Then the Kalman filter gives the minimum-variance estimate of β_t. That is, the

covariances $P_{t|t}$ and $P_{t|t-1}$ are the smallest possible. Also, the estimates are the conditional expectations

$$a_{t|t} = E(\Gamma_t|y_t)$$
$$a_{t|t-1} = E(\Gamma_t|y_t - 1)$$

$\square$

Therefore, we can conclude that the Kalman filter is the optimal filter for a linear model subject to Gaussian noise and it can be derive by using conditional expectations.

The equations of the recursive least square and the Kalman filter equations described above have the same characteristic of estimation as expected, it means that both try to find the estimator that minimize the mean square error. The equations are relevant for the estimation of time-varying coefficient models. Models that can be written into state-space form so that the likelihood function can be calculated easily by the Kalman filter algorithm as it was showed in [Pagan, 1980]. The equation 3.3 is usually used as the prediction error decomposition of the likelihood function.

3.3 Likelihood estimators

In the estimation stage, the maximum likelihood estimator is used. In this way the Kalman filter can be used with exact maximum likelihood in the time domain to estimate the parameters of the model and the state space vector Γ_t via a prediction error decomposition function. As for all numerical procedures, attention has to be paid to starting values in order to avoid local minima. [Manly, 1994], [Racicot and Theoret, 2009]

The classic likelihood function for independently and identically distributed data is not applicable for time series data, since the observations are time dependent. Thus, the classic likelihood function in the context of time series analysis is usually used in the sense of Gaussian likelihood, which means that the likelihood is computed under the assumptions that the series is Gaussian, fact that could be false for some time series as we can see in Chapter 3. Nevertheless, estimators of the ARMA[2] coefficients computed by maximization of the Gaussian likelihood have good large-sample properties, i.e. they are unbiased when T is large even if they are not Gaussian.

However, it is known that the data y_t, $t = 1, \ldots, T$, conditional on all observation up to $t-1$ (denoted as Y_{t-1}) are independent. Therefore, the likelihood function of the observations is the joint conditional probability density function with respect to Θ defined for all $y = (y_0, \ldots, y_T)$ where $\Theta = (\Phi, H, R, Q)$

---

[2]Forecasting model or process in which both auto-regression analysis and moving average methods are applied to a well-behaved time series data. ARMA assumes that the time series is stationary-fluctuates more or less uniformly around a time-invariant mean. An ARMA(p,q) model is given by, $B_1 = A + \sum_{i=1}^{p} \varphi_i B_{t-i} + \sum_{j=1}^{q} \psi_j \varepsilon_{t-j} + \varepsilon_t$ where ε_t are the error terms and are assume to be independent identically-distributed random variable sampled from a normal distribution with zero mean and variance σ^2 and φ_i, ψ_j are the parameters of the model while A is a constant.

the vector of the parameters to be estimate. The function is given by:

$$L_{\Theta|y} = f(y_T|Y_{T-1})f(y_{T-1}|Y_{T-2})\ldots f(y_2|Y_1)f(y_1)$$

$$= \prod_{t=1}^{T} f(y_t|Y_{t-1}, \Theta) \tag{3.4}$$

Assuming that $f(y_1)$ does not depend on Θ, it is irrelevant. Most of the practitioners assume the state one is known.

Based on the assumptions that the initial values and innovation are multivariate normally distributed, and $y_t|Y_{t-1}$ is multivariate normal with $E(y_t|Y_{t-1}) = Z_t a_{t|t-1}$ and $cov(y_t|Y_{t-1}) = F_t = Z_t P_{t|t-1} Z_t' + H_t$. Hence, as define before

$$y_t - E(y_t|Y_{t-1}) = y_t - Z_t a_{t|t-1} = \eta_t$$

and F_t is the covariance matrix of η_t, then the likelihood function can be written as a function of innovations by:

$$L_{\Theta|y} = \prod_{t=1}^{T} (2\pi)^{-\frac{p}{2}} |F_t|^{-\frac{1}{2}} e^{-\frac{1}{2}(y_t - y_{t|t-1})_t' F_t^{-1}(y_t - y_{t|t-1})}$$

$$= \prod_{t=1}^{T} (2\pi)^{-\frac{p}{2}} |F_t|^{-\frac{1}{2}} e^{-\frac{1}{2}\eta_t' F_t^{-1} \eta_t} \tag{3.5}$$

where y is the vector $(y_1, y_2, \ldots, y_T)$ and $|F_t|$ is the determinant of the matrix F_t

Let us now define,

$$l_{\Theta|y} = \log(L_{\Theta|y}) \tag{3.6}$$

as the log-likelihood function. Then, by the properties of logarithm and 3.5

$$l_{\Theta|y} = -\frac{pT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log(|F_t|) - \frac{1}{2} \sum_{t=1}^{T} \eta_t' F_t^{-1} \eta_t \tag{3.7}$$

and it is known as the *prediction error decomposition* form of the likelihood [Harvey, 1989].

By the prediction error decomposition formula, the log-likelihood function of the model 3.2 is essentially decomposed into segments of different instances of time. The log-likelihood at each time can be represented by the innovations η_t and their variance F_t, both of which are functions of unknown parameters $\Theta = (\Phi, H, R, Q)$. Since η_t and F_t involve recursive terms, the Kalman Filter technique is used to estimate the fixed parameters of the model, and obtain them from the prediction of future values. In other words, the maximum likelihood estimators obtained by maximizing the summation of the log-likelihood over the time with respect to Θ are based on information of the entire time series.

Having estimated the vector of the variances of the model $(\sigma_\varepsilon^2, \sigma_\xi^2, \sigma_{\epsilon_1}^2, \ldots, \sigma_{\epsilon_k}^2)$, the prediction based on the fitted state space model can be made and the mean square error can be minimized.

3.4   Algorithm

The algorithm used to estimate the parameters of the model 3.2 was based in the recursive Kalman filter equations for the state space equations of the regression model with time-varying coefficients. The Kalman and the Gaussian maximum likelihood are related to each other very closely in this algorithm. In order to apply the Kalman filter, a system of matrices must be fixed and their maximum likelihood estimated. In the other hand, the log-likelihood value is computed from the products of the values obtained by the Kalman filter algorithm. The procedure begins providing initial values for the innovation and their covariance matrix, and also the starting values of the parameters, for the Kalman filter, that are going to be estimated.

The Kalman filter provides innovations and innovation covariance matrices that are needed to calculate the log-likelihood. Then, the initial values of the parameters, Θ, are updated based on some conventional recipe of numerical optimization. Once the parameters are updated, the Kalman filter is carried out again to calculate the new log-likelihood. The procedure is repeated until a convergence criteria is reached. The state and parameters estimates corresponding to the maximization of the log-likelihood function are regarded as the final estimates.

Steps of the estimation method:

1. Initialized the variables, Φ^0, $(R_0 \quad Q_0)$, H^0, a_0, P_0 where P_0 is the covariance matrix of the vector a_0 as we saw before.

2. Apply the Kalman filter to the model 3.2 to generate η_t, the prediction errors, and F_t, the innovation covariance to define the log-likelihood function:

$$l_{\Theta|y} = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{N}\log\left(|F_t|\right) - \frac{1}{2}\sum_{t=1}^{N}\eta_t^T F_t^{-1}\eta_t$$

3. Using numerical optimization, maximize the likelihood function with respect to the unknown parameters: the volatility of the residuals, η_t, and the volatility of the stochastic components, F_t. In this step we need to find,

$$l_{\widehat{\Theta}|y} = \max_{\Theta}\ l_{\Theta|y}$$

where Θ is the vector of the variances to estimate that represents the variance of each equation of the model. Thus, the region where the variances must be estimated are all the real positive values.

In brief, the algorithmic loop can be summarized in the following diagram,
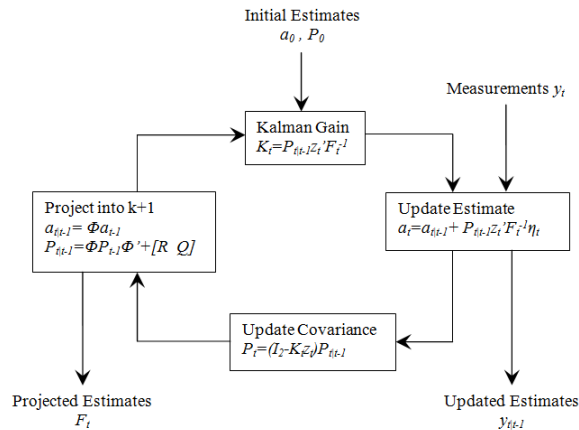
Figure 3.1: Kalman Filter recursive algorithm

Note that the roundoff is a problem in Kalman filter implementations when we are working in MATLAB environments, there are some examples in [Grewal and Andrews, 1993] that show how a well-conditioned problem can be made ill-conditioned by the filter implementation. Therefore the next step is, to use decomposition methods for the covariance matrices and other matrices in the implementation that could be singular, to avoid roundoff errors. We can use for example, Cholesky decomposition algorithms.

3.5   Application to simulated data

Before running the models with real data of Hedge Funds, a simulation can show the performance of the Kalman filter approach for the regression model with one or both parameters varying in time using a one factor model describe in Chapter 1. Both models are under artificial conditions.

A simulation of the data are computed using the equations 2.4 taking in account that the time-varying coefficients alpha and beta follow a random walk. First, we worked with different data size to see how it could affect the results. Then, we use the same size of data but different values for the parameters we want to estimate.

In the next table, table 3.1, we are going to show how the algorithm works with different initial values in order to see if it depends on them or if the result are independent of them. As we can see the results do not depend on the initial values in general. This table is only an example with simulation data of $N = 1500$.

Simulated	$\sigma^2_{\varepsilon_t} = 0.09$	$\sigma^2_{\epsilon_t} = 0.04$	$\sigma^2_{\xi_t} = 0.01$
Initial Values $[\sigma^2_{\varepsilon_t} \quad \sigma^2_{\epsilon_t} \quad \sigma^2_{\xi_t}]$	$\lvert\sigma^2_{\varepsilon_t} - \widehat{\sigma}^2_{\varepsilon_t}\rvert$	$\lvert\sigma^2_{\epsilon_t} - \widehat{\sigma}^2_{\epsilon_t}\rvert$	$\lvert\sigma^2_{\xi_t} - \widehat{\sigma}^2_{\xi_t}\rvert$
$[0.5, 0.2, 0.1]$	0.001288	0.013923	0.005091
$[0.1, 0.4, 0.2]$	0.001290	0.013917	0.005091
$[0.4, 0.03, 0.1]$	0.001289	0.013905	0.005091
$[0.3, 0.5, 0.008]$	0.001291	0.013906	0.005091
$[0.1, 0.03, 0.008]$	0.001289	0.013913	0.005091

Table 3.1: Table of estimation error for each parameter using different vectors of initial values

The table shows different starting points and the error estimating the parameters in each case. We can see that the starting point is not important for this model since the estimated parameters are very close to the real values of the parameters taking to simulate the data.

Next, we show how fast the estimators converge to the parameters. In the graph we can see that a big sample size is not needed to have good estimation using Kalman filter algorithm with the regression model in the case the parameters are varying on time.
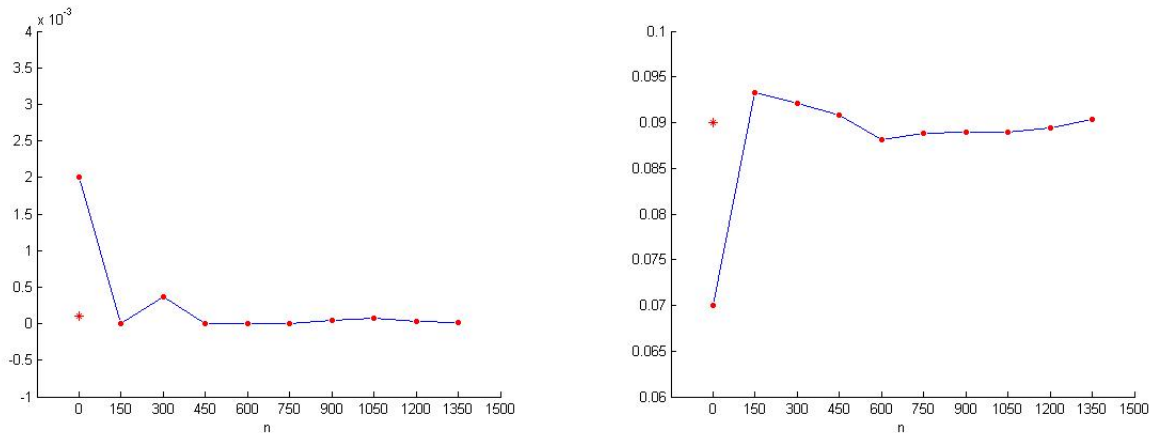
Figure 3.2: Estimation of the variance of the model equation and beta equation when alpha is taken constant over time.

As we show in table 3.1 the initial point does not make a big difference but also the sample size is not a critical issue. As we can see in the graphs above from $N = 450$ the estimated parameters are very close to the real parameters used to simulate the data. The outlier point in both figures represent the real value uses to simulate the data.

In order to know the adequate sample size to use when working with time dependent parameters we are going to simulate a data from a linear regression. The tables 3.2 and 3.3 show the values of the parameters used to simulated the data and the error when estimating the parameters using Kalman filter approach.

Beta varying on time and alpha constant over time.

Based on an increasing constantly beta and alpha kept constant. Table 3.2 show the error estimating the variance of beta and the regression equation for different sample sizes using Kalman filter technique, as well as the parameters used to simulate the data. Initial values are taken as: $\sigma^2_{\varepsilon_t} = 0.7$, $\sigma^2_{\epsilon_t} = 0.3$

Simulated	$\sigma^2_{\varepsilon_t} = 0.16$	$\sigma^2_{\epsilon_t} = 0.04$
	$\|\sigma^2_{\varepsilon_t} - \widehat{\sigma}^2_{\varepsilon_t}\|$	$\|\sigma^2_{\epsilon_t} - \widehat{\sigma}^2_{\epsilon_t}\|$
N=500	0.001640	0.012573
N=1500	0.000829	0.012234
N=2500	0.000908	0.003559

Table 3.2: Error of estimation using different sample size and using the model with alpha constant and beta time dependent

Table 3.2 shows the estimator of the parameters with different sample sizes. Once again we can see that we do not need to have a big sample size since after $N = 2000$ (approximately) the estimation values are very similar.

Both parameters, alpha and beta, changing over time.

Here we can also show that even for the model with both parameters time dependant, the estimation stabilizes after certain value of the sample size, which will help us with the application since there is very little data available in the hedge fund industry. The simulated data was done as we described at the beginning of the section.

Simulated	$\sigma^2_{\varepsilon_t} = 0.36$	$\sigma^2_{\epsilon_t} = 0.09$	$\sigma^2_{\xi_t} = 0.04$
	$\|\sigma^2_{\varepsilon_t} - \widehat{\sigma}^2_{\varepsilon_t}\|$	$\|\sigma^2_{\epsilon_t} - \widehat{\sigma}^2_{\epsilon_t}\|$	$\|\sigma^2_{\xi_t} - \widehat{\sigma}^2_{\xi_t}\|$
N=500	0.018007	0.009808	0.003687
N=1500	0.005023	0.013130	0.001567
N=2500	0.008399	0.009210	0.009583

Table 3.3: Error of estimation using different sample size and using the model with both parameters time varying.

Initial values are taken as: $\sigma^2_{\varepsilon_t} = 0.7$, $\sigma^2_{\epsilon_t} = 0.3$, $\sigma^2_{\xi_t} = 0.2$

As we can see also when the model has both parameters varying on time the sample size is not so important after $N = 1500$ (approximately).
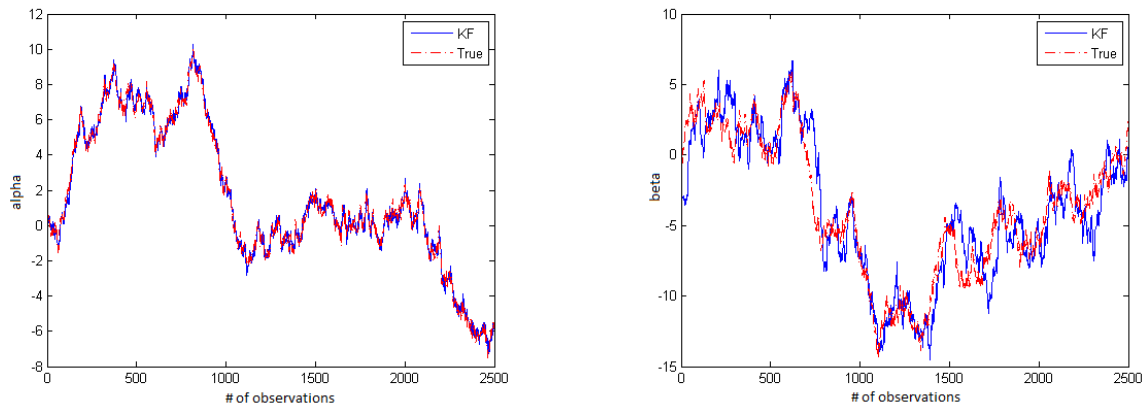
Figure 3.3: Vector of alpha(left) and beta(right) estimated by Kalman Filter

Figure 3.3 shows the Kalman filter estimation of the parameters, α and β, versus the true values of the simulation vectors. As we can see the estimation of the Kalman filter is very close to the real values for the given values of the variance. Looking at the graph of alpha we almost cannot distinguish the estimation and real curves, just in few points small differences appear. In the case of the estimation of beta the discrepancies between the real and the estimation curves are more visible. Therefore, in the next chapter we are going to apply this approach to real values of a Hedge Fund series.

Chapter 4

Applications to Hedge Funds

Hedge funds, best known for their unique strategies, high returns, and capital inflows, have also attracted considerable requests for stronger regulation. The substantial out-performance prior to the financial crisis suggests that in particular hedge funds offer investors significant portfolio benefits by enhancing the risk-return trade-off of their portfolios. In fact, hedge fund strategies generate high and steady returns that appear to be rather uncorrelated with returns of conventional asset classes such as stocks and bonds. During the recent financial crisis, however, the investment performance of hedge funds has deteriorated substantially as they suffered significant losses on their portfolio holdings. Moreover, recent studies indicate that diversification benefits of hedge funds have continuously declined due to a slow but persistent upward trend in the co-movement of hedge fund returns with conventional asset classes [Bressler and Holler, 2010].

In this chapter we first analyze the statistical properties of one hedge fund index and a market index we are going to work with. Then, we test the forecasting in sample of each model using the index of hedge fund and the market index previously analyzed. We do some comparison not only between models but also between the methods used to estimate the parameters in each model in order to show the relative importance of using a beta varying on time when modeling hedge fund indexes. It is well known that there is difference between the performance of the hedge funds prior and during the crisis, hence we examine both series separately and the analysis is done for each model and method to see the behavior of each of them.

As we mentioned before there is and indication that hedge funds may include derivatives, therefore we study the models with put option and call option using beta varying on time. Finally, we work with some strategies of hedge funds. For these strategies, a statistical analysis and a research of some of the styles that work in general with options is performed. The styles taken in this chapter are styles that work with options as they are defined in [LHabitant, 2001].

4.1 Hedge Funds data

"Hegde funds are unregistered private investment partnerships, funds or pools that may invest and trade in many different markets, strategies and instruments (lets say securities, non-securities and derivatives), and are not subject to the same regulatory requirements as mutual funds. It is not surprising that hedge funds are a diverse lot, given that even the majority of existing definitions of them reveal little about their process. Some researchers describe a hedge fund as an "actively managed, pooled investment vehicle that is open to only limited group of investors and whose performance is measured in absolute units of return". The term "hedge" suggests that these funds reduce their volatility by taking positions that offset their exposures to various risk factors" [LHabitant, 2001]. The degree of hedging that actually occurs varies widely among these funds.

In this thesis we work with a hedge fund index: the hedge funds Absolute Return Index (HFRX-Abs). We take the daily closure prices from July of 2004 to June of 2010 and we chose the Standard and Poor with 500 companies ($S\&P500$) as the index describing the market in NYSE.

The HFRX Absolute Return Index is designed to be a representative compilation of all available Hedge Funds that pursue various strategies. Some of these strategies are: convertible arbitrage, distressed securities, equity hedge, equity market neutral, event-driven, macro, merger arbitrage, and relative value. As a component of the optimization process, the Index selects those funds that exhibit lower volatilities and a closer correlation to traditional markets for example $S\&P500$.

HFRX indices are a series of benchmarks of hedge funds industry performance in order to reach a representative performance of larger universe of hedge fund strategies. Hedge Fund Research (HFR) utilizes the HFRX methodology to construct the HFRX Hedge Fund Indices. The methodology is based on defined and predetermined rules and objective criteria to select and re-balance components to maximize representation of the Hedge Fund Universe. This methodology includes robust classification, cluster analysis, correlation analysis, Monte Carlo simulations and advanced optimization techniques to ensure that each index is a pure representation of its corresponding investment focus. In the construction of the HFRX methodology a model output is used to select funds that, when added and weighted, have the highest statistical likelihood of producing a return series that is most representative of the reference universe of strategies. As a result, a sub-set of strategies which are representative of a larger universe of hedge fund strategies, geographic constituencies or groupings of funds maintaining certain specific characteristics are selected.

Figure 4.1 shows the series of HFRX and the series of the market index $S\&P500$ corresponding to the same period and their series of return. The return of the index is the relative change in the index over a certain holding period of time, and it is calculate using a single period as,

$$R_t = \frac{r_t - r_{t-1}}{r_{t-1}}$$

where r_t is the index value at time t.
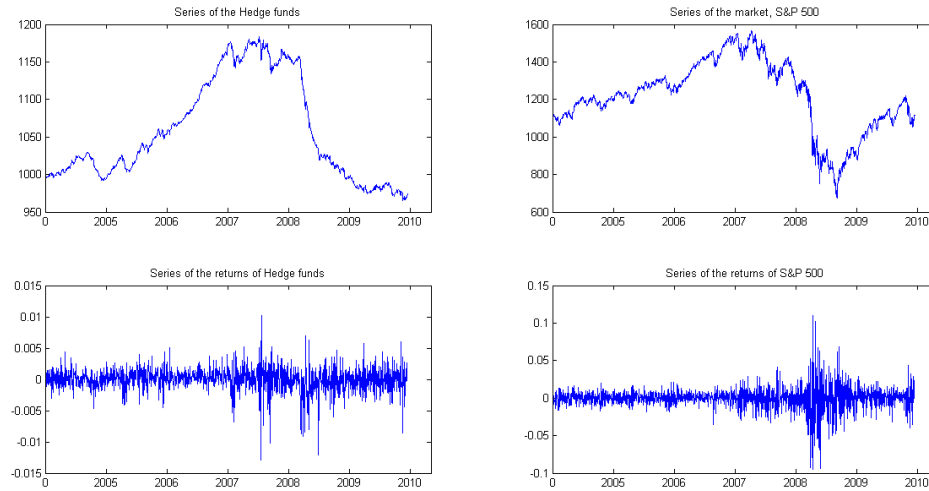
Figure 4.1: Series and Return series of the HFRX-Abs and S&P500 for a period of 6 years, from July 2004 to June 2010. The returns are daily and include the financial crisis of 2007

In Figure 4.1, in the series of returns, we can see the change in the volatility around mid-2007, when the financial crisis started. The graph of the series of HFRX-Abs shows the increment from 2006 to 2007 and how it rapidly decays around 2008 leaving a lag in between when the financial crisis really started and when it affects the HFRX.

Table 4.1 presents a summary of descriptive statistics including mean, standard deviation, Sharpe Ratio, Skewness, Kurtosis, minimum, maximum, and the proportion of positive months for the series of returns of HFRX-Abs and S&P500.

	Hedge Fund	S&P 500
Mean	-1.740049×10^{-5}	$-0.7656518 \times 10^{-5}$
Standard Deviation	0.001798	0.014468
Minimum	-0.012921	-0.094695
Maximum	0.010219	0.109572
Skewness	-0.909301	-0.251452
Kurtosis	9.342557	13.505853
Number of positive months	787	826
Sharpe	-0.009680	-0.000529

Table 4.1: Summary of Descriptive Statistics of the return series

Let us first explain some of these statistics showed in table 4.1 that may be used to quantify the effect of large deviations from normality in the data:

Skewness is the third central moment. It measures the symmetry of the probability distribution around its mean. Zero skewness indicates a symmetrical distribution. A positive skewness is the outcome of

rather small negative changes but large positive changes, so the distribution has a long right-tail which is desirable. While, a negative skewed distribution is the outcome of many small gains but larger losses are less frequent, so it has a long tail on the left-hand side of the distribution. [LHabitant, 2001]

The skewness is define as:

$$Sk = \frac{T}{(T-1)(T-2)} \sum_{t=1}^{T} \left(\frac{R_t - \overline{R}}{S} \right)^3$$

where T is the number of observations, $R_t, \overline{R}$ are the return at time t and the mean of the return respectively, and S is its standard deviation.

Kurtosis is the fourth moment of a probability distribution. It measures the degree of "peakedness" and heaviness of the tail. Distributions where a relative large proportion of the observed values lie towards the extreme display a positive kurtosis. Therefore, they have heavy tails compared with a normal distributions whose kurtosis is equal to three. On the other hand, negative kurtosis display a flat top near the mean.

The kurtosis is computed as:

$$Kurt = \frac{T(T+1)}{(T-1)(T-2)(T-3)} \sum_{t=1}^{T} \left(\frac{R_t - \overline{R}}{S} \right)^4 - \frac{3(T-1)^2}{(T-2)(T-3)}$$

where T is the number of observations.

Table 4.1 represents the descriptive statistics of the daily data of the hedge funds and market returns from 2004 to 2010 in order to provide a general understanding of the nature of the return series. The data sets show that the market return series is greater than the mean of the Hedges Funds index. Both have negative average of the returns implying the fact that the overall market returns have decreased over the sample period. The volatility is smaller in the HFRX-Abs than in the S&P500 but the hedge fund index has higher average losses in the period in general.

As usual features in any financial time series, high kurtosis or heavy tails and excess are features appearing in the return series. As we can observe the kurtosis for both series is higher than three, therefore, the tails are heavier than the normal distribution for both of them in the period analyzed.

We also found negative skewness in both series indicating that the tail on the left side of the probability density function is longer than the right side. It means, that there is now a high probability of having big negative returns.

The Sharpe Ratio is a measure of the mean return per unit of risk in an investment portfolio or trading strategy. More particularly, it is the average of excess returns divided by the volatility of excess returns taking, for example, the risk free rate of return as the benchmark. The Sharpe Ratio is calculate by:

$$\text{Sharpe Ratio} = \frac{\overline{R} - R_f}{S}$$

where R_f is the risk-free interest rate taken as reference (LIBOR), $\overline{R}$ and S are the average and the standard deviation of returns on portfolio respectively.

LIBOR is the interest rate that banks charge each other for one-month, three-month, six-month and

one-year loans and it is an acronym for London Inter-Bank Offer Rate. LIBOR is the most widely used benchmark for short term interest rates in the world, primarily because most of the world's largest borrowers borrow money on the London market. It is also important because it is the rate at which the world's most preferred borrowers are able to borrow money. LIBOR is compiled daily by the British Bankers Association (BBA), and derived from a filtered average of the world's most creditworthy banks' interbank deposit rates for larger loans [1]. It is officially fixed once a day by small group of large London banks, but the rate changes throughout the day.

The interpretation of the Sharpe Ratio is straightforward: the higher the ratio the better. A high Sharpe Ratio means that the fund is question delivered a higher return for its level of volatility, see [LHabitant, 2001]. In general, the Sharpe Ratio is used to characterize how well the asset returns compensate the investor for the risk taken.

The period we are analyzing has included the profound financial crisis started in 2007. We can divide the series between before and after the financial crisis in order to analyze the effects separately.

	July 2004 - July 2007	August 2007 - June 2010
Mean	2.040404×10^{-4}	-2.537871×10^{-4}
Standard Deviation	0.001262	0.002209
Minimum	-0.004683	-0.012921
Maximum	0.005011	0.010219
Skewness	-0.200983	-0.758414
Kurtosis	4.432529	7.587643
Sharpe	0.161745	-0.114961

Table 4.2: Descriptive statistical analysis to the HFRX-Abs before and during the crisis.

In table 4.2 we can see the difference between the series before and during the financial crisis and we can compare them with the complete series that include both periods, see table 4.1. In the series that include both periods the mean is negative but not in the series before the crisis. We can also see the decreasing Kurtosis for the series before and during the crisis with respect to the complete series, and between them. More interesting is the change on the Sharpe Ratio which is negative in the complete series and in the series during the crisis but in the series before the crisis is positive together with the higher volatility after the crisis showing higher risk after August 2007.

4.2 One-factor models

The simplest factor model is one-factor model, i.e. $k = 1$. It expresses the return on each hedge fund as a linear function of a factor F. One-factor models with market index as the factor variable are called market models. However, factor models do not restrict the factor to be the market index. Researchers use different approaches in factor models, see [Nai-Fu Chen and Ross, 1986] and [Manly, 1994]. The

---

[1]Countries that rely on the LIBOR for a reference rate include the United States, Canada, Switzerland, and the United Kingdom.

first one assumes that the sensitivities to factors are known, and the factors are estimated from the security returns. The second approach assumes that some known fundamental factors are the factors that influence the security and β's are evaluated accordingly. In this section we are going to use HFRX-Abs index as a dependent variable while the market index, S&P500, is the factor of the models.

One-factor models provide a simple but effective framework for understanding and predicting returns. As beta is the factor's exposure indicator, then a higher beta means more reaction to factor movements, while a lower beta means less reaction. Also, a positive beta means that hedge fund returns generally follow the market's returns, in the sense that they both tend to be above or below their respective averages together, while a negative beta means that the asset's returns generally move opposite the market's returns. The knowledge of the future value of F could be used to predict assets returns, albeit not perfectly, given the presence of a random error term. One-factor models also provide a very simple framework for understanding and predicting risk.

4.2.1 Deterministic Coefficients Model

In model 2.3 where we assume the parameters are constant over time, we can estimate them by using the OLS method as we already explained in Chapter 1. Now, we are going to assume that $y_t = R_t$ is the return of the hedge fund at time t and $X_t = R_{Mt}$ is the market index return, in this case S&P500, but could be Russell3000, MSCI World, among others.

The OLS is the classic method used for estimation of the regression coefficients. Several technical measures are produced as part of a regression output and usually serve as indicators of confidence in the results of the regression. For instance, R^2 is the Coefficient of Determination defined as:

$$R^2 = 1 - \frac{SSE}{SStot}$$

where,

$$SSE = \sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \qquad \text{is the sum of the square of the errors}$$

$$SStot = \sum_{t=1}^{T}(y_t - \overline{y})^2 \qquad \text{is the total sum of the squares}$$

The Coefficient of Determination indicates the total variation observed in the dependent variable R_t that can be explained using the linear model prediction $\widehat{R_t}$ compared with just using the mean $\bar{R}_t$. R^2 can not be used as selection criteria for accuracy of forecast but it gives an indicator of the in-sample fit. Another measure we are going to look at is the F-statistic that is used as statistical test to measure if there is a linear relationship between the dependent and independent variables.

Based on the analysis, using MATLAB functions we get $R^2 = 0.0637\%$ and $F - stat = 0.955$ with $p - value = 0.329$. Also the OLS estimation of the parameters are: $\alpha = -0.0000174$ and $\beta = 0.00314$. Looking at the values of R^2 we can see that the percentage of the explanation using a classic linear

regression model is very low, giving evidence that the chosen model does not fit well the data. The F value shows that there is no reason to assume a linear relationship between the variables since the $p-value$ is bigger than 0.05. Suggesting that the use of a Regression Model with time-varying coefficients might be better to the hedge fund's returns series analyzed.

4.2.2 Periodical Coefficient

A particular dynamic formulation is an oscillator model for beta with sinusoidal terms. The motivation of using this model was that the model might capture high frequency dynamics of data variation.

In this case, we implemented a code that allows us to find the parameters that minimize the sum of the squared errors. To do that we need to create a grid to help the algorithm find the best estimator since it depends on the initial values. The grid is taken depending on the number of years and the frequency is taken small to capture as many market's downwards and upwards as possible. The following values are taken to create the mentioned grid,

$$
\begin{aligned}
w &= [0.010, 0.013], &&\text{with the step size of } 0.001 \\
a &= [1.5, 3], &&\text{step size equal to } 0.5 \\
b &= [0.5, 2], &&\text{step size equal to } 0.5
\end{aligned}
$$

The code implemented is very effective in the case of simulations. The estimations obtained are close to the real values taken to simulate the data. In the case of real values, working with HFRX and S&P500 we get the following graph of beta and the estimation of the model ($\hat{Y}$),
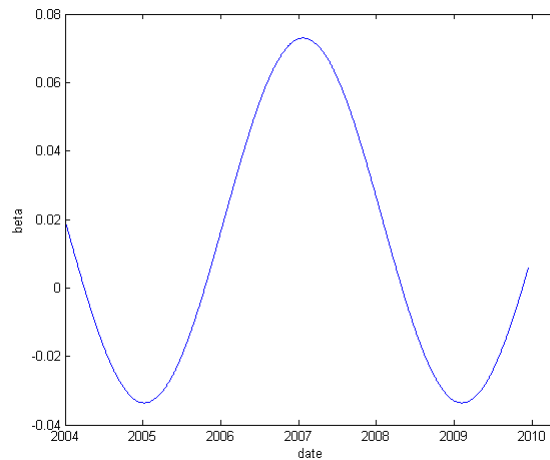

Figure 4.2: Beta estimation using periodical model. The beta's parameters are estimated using the code describe above.
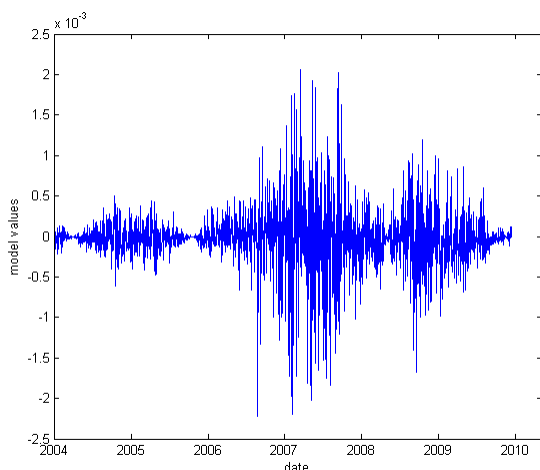
Figure 4.3: Estimation of Y from the periodical model using the above approach to estimate all the parameters of the model. The entire hedge fund series is used.

Figure 4.2 shows the estimation of the parameter beta, and we can see that there is also a peak around 2007 where the financial crisis started. While figure 4.3 shows large changes around 2007 when the financial crisis began. Though, there are some changes around 2009, the largest changes are concentrated around the crisis and here we can see also some evidence that beta should be stochastic.

The values using to measure the performance of the model can be found in the following subsection. There, a comparison between models and also between methods used to estimate the parameters are given.

4.2.3 Stochastic Coefficients

Over the past years, hedge fund researches have been studying the stability of beta. When estimating the capital market model it is common practice to assume betas to be invariant over time. However, this stability assumption has been questioned and a considerable amount of empirical evidence reports the importance of variation over time (see among others [Fabozzi and Francis, 1978], [Ledolter, 1979], [Bos and Newbold, 1984], [Daniel W. Collins and Rayburn, 1987], [Kim, 1993]). Therefore, the common admitted opinion that hedge funds offer positive absolute returns independent of market conditions can not be maintained. Moreover, it seems that the low exposure of hedge funds to market risk, measured with traditional regression methods is just an average value of a dynamic, time-varying exposure of the portfolio.

There are plenty of reasons why linear regression models with constant coefficients do not have a good performance in describing hedge fund returns; an example is the inability of these models to measure time-varying exposures. Some papers investigate whether the factor exposure of the hedge funds are constant or not, see [Sunder, 1980], [Racicot and Theoret, 2009], [Schwert and Seguin, 1990].

Using a moving window regression with OLS method to estimate the parameters for each window

taken with same size (same amount of observation) and having no intersection, we found that the null hypothesis of stable coefficients is strongly rejected, as it is in Figure 4.4, which indicates that factor exposures of hedge funds are indeed changing over time. In the figure we can see that in each window beta estimation is different, suggesting that beta might be not constant over a period of time.

In order to study the time variation of the hedge funds, we divide the sample data into evenly-spaced subperiods. Figure 4.4 represents one of this subperiods when a window of 12 months is taken to fit the parameters and analyze the time variation.

As showed in Figure 4.4 in each window the beta and alpha estimated by OLS appear to be different.
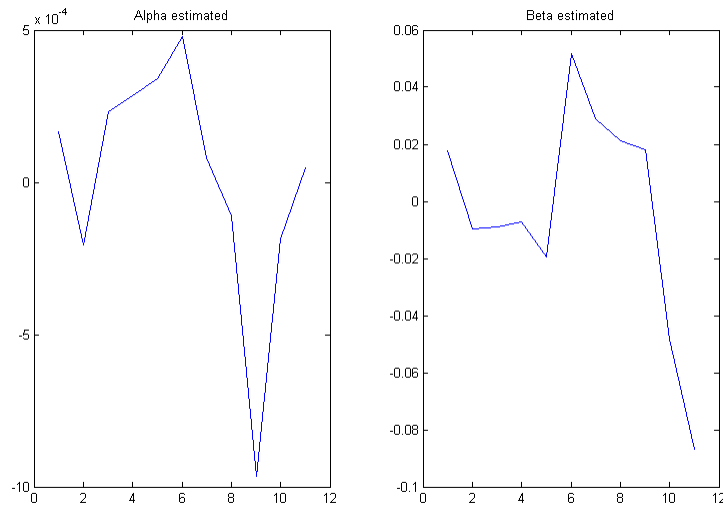
Figure 4.4: Alpha (left) and beta (right) estimation in different disjoint periods of time. The x-axis is the subperiod, yearly, in which we divided the sample data while the y-axis represents the estimation values of the parameters.

In this graph the $x-axis$ represent the subperiods taken to estimate the parameters and the $y-axis$ are the values of the parameters in each window.

In the following graph, Figure 4.5, we can see the confidence intervals for the OLS estimator obtained from disjoint windows of the return series. The size of the window taken was 500 observations, therefore with the sample size of 1500 we had 3 disjoint windows as it is shown in the following figures.
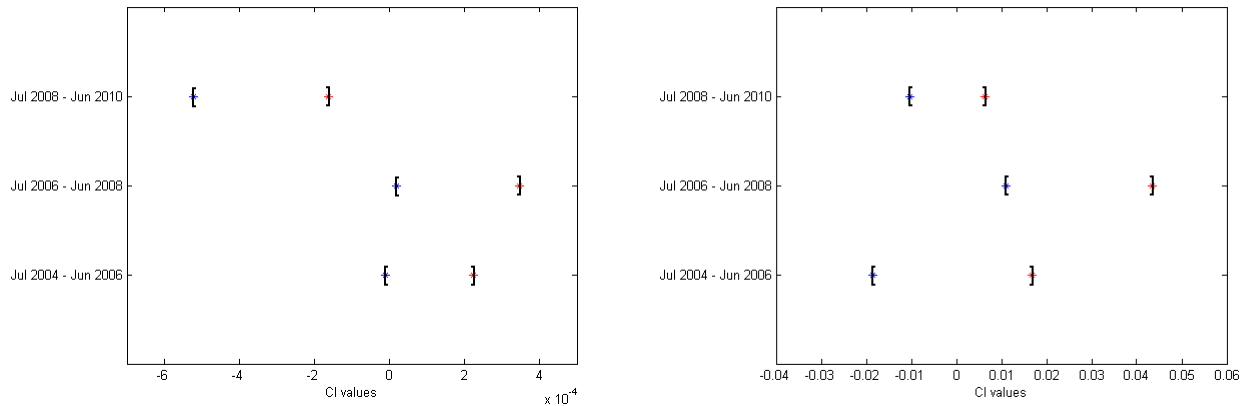
Figure 4.5: Confidence interval for OLS alpha (left) and beta (right) estimators using disjoint windows. The confidence level used was 95%

Figure 4.5 suggests that for the real data used the beta exposure of the hedge fund index (HFRX) and the market index (S&P500) varies in time, since there are at least two intervals do not overlapping.

Testing a random walk (the unit root tests)

In order to test whether the series of hedge funds utilized in this study can be modeled by a Random Walk. We consider the equations

$$
\begin{aligned}
y_t &= \alpha_t + \beta_t X_t + \varepsilon_t \\
\beta_t &= M_1 \beta_{t-1} + \xi_t \\
\alpha_t &= M_2 \alpha_{t-1} + \epsilon_t
\end{aligned}
\tag{4.1}
$$

and test for $M_1 = 1$ and $M_2 = 1$ applying a unit root test. It translate into the following hypothesis test for M_1, an it is the same for M_2,

$$
\begin{aligned}
H_0 &: M_1 = 1, \quad \text{(unit root)} \\
H_a &: M_1 < 1, \quad \text{(no unit root)}
\end{aligned}
$$

this is a well-known unit root testing problem, [Schwert, 1989].
A unit root process is a data-generating process whose first difference is stationary. It attempts to determine whether a given time series is consistent with a unit root process. There are many tests to prove unit root process; in this thesis we are going to apply Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests that are two of the most commonly used.
A convenient test statistics is the $t - ratio$ of the least square estimate of the M_1. The $t - ratio$ of $\widehat{M_1}$

is given by,

$$ADF - test = \frac{\widehat{M_1} - 1}{se}$$

where $\widehat{M_1}$ is the least square estimate of M_1, and se is the standard error in the alternative model.
In the case of KPSS test the statistic is:

$$KPSS - test = \frac{1}{s_{nw}^2 T^2} \sum_{t=1}^{T} (s_t)^2$$

where $s(t)$ is the sum of the residuals from the regression, T is the the sample size of the time series, in our case HFRX, and s_{nw}^2 is the Newey-West estimator of the long-run variance.
To test whether M_1 and M_2 are equal to one for the specific data analyzed in this study, we assume that both (M_1 and M_2) are different of one and estimate them using the Kalman filter approach. Then, using as data the estimation obtained we can applied the test mentioned above.
The hypothesis of the KPSS test is the opposite of the ADF test. Meaning that the null hypothesis is that an observable time series is trend stationary, it means stationary[2] around a deterministic trend, against the alternative that it is a non-stationary process.

	Alpha	**Beta**
ADF test		
Statistic	-1.0576	-0.6561
p-value	0.2645	0.4116
KPSS test		
Statistic	1.7454	16.5289
p-value	0.01	0.01

Table 4.3: Unit root test statistics of the return series using a significance level of 0.05

The results given in table 4.3 suggest that the series of the beta is a random walk when we use the ADF test and using the KPSS test with 0.05 as the confidence level, then we can assume that both series are random walks. The results are obtained from the test equations including both intercept and trend terms.
For the remaining thesis, we are going to assume that the beta and alpha exposure change over time. Furthermore, they follow a random walk process.

4.2.4   Models Comparison

Now we make a comparison of the parameters obtained as a combination of Kalman filter and maximum likelihood, and by OLS-rolling windows (with 1 year as window size) approaches. The series are showed in the following figure. Figures 4.6 and 4.7 show the vector of beta and alpha estimated by both techniques.

---

[2]A stationary process is a stochastic process whose joint probability distribution does not change when shifted in time or space.

Here we can see that in the case of alpha estimated by KF approach is most of the time higher than using OLS-RW method except from 2007 to late 2008, and this is exactly a period of financial crisis.
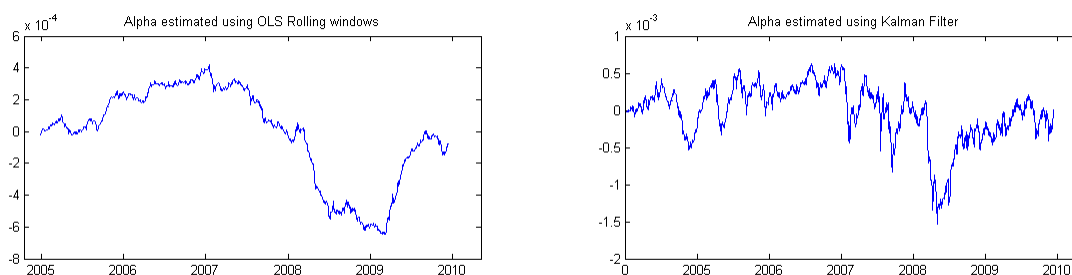
Figure 4.6: Alpha estimated using RW-OLS with a window size of 12 months (left) and KF (right) techniques
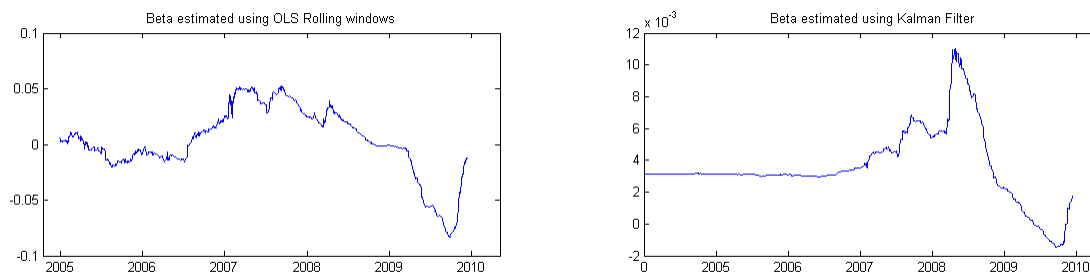
Figure 4.7: Beta estimated using RW-OLS with a window size of 12 months(left) and KF (right) techniques

Figure 4.7 represents the estimation of the beta exposure by RW-OLS and KF approaches. In both graphs we can see a high peak but more representative in the KF. The peak occurs approximately one year after the crisis started and then after that the values of the exposure decrease until late 2009. This might be caused due to hedge funds managers trying to balance the returns. But more interesting is that we can see that by late 2007, when the crisis started, the volatility increases, it is clearer in the graph of beta estimated by KF technique suggesting that KF is better to locate changes and it will show some high volatility before some crash in the market.

The following figure 4.8 shows the fitted residuals when using the one factor models from Chapter 1.
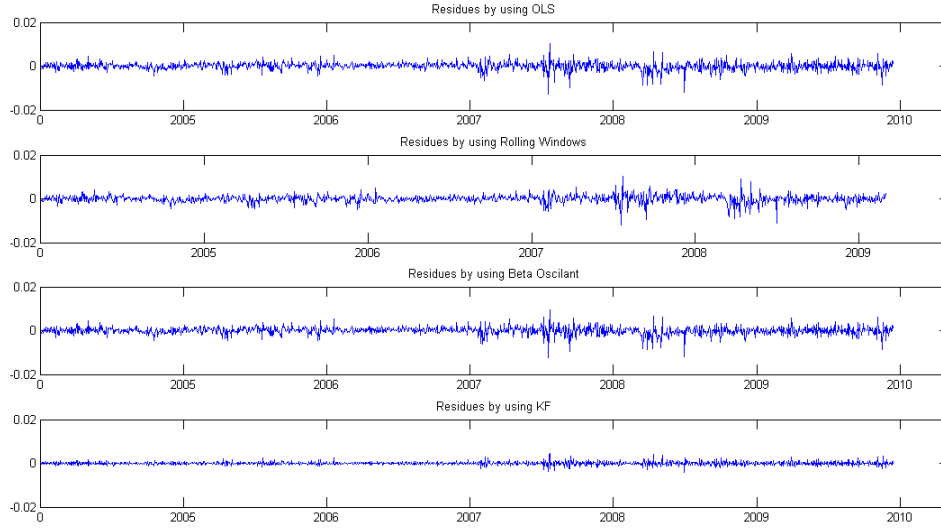
Figure 4.8: Residues of the estimation by each method used in this study

The fitted residuals are the differences between the observed values $y_1, ..., y_n$ of the vector Y (the HFRX) and the values $\widehat{y_1}, ..., \widehat{y_n}$ fitted by the model,

$$\widehat{y_i} = \widehat{\alpha} + \widehat{\beta} * X_i$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ denote the estimators of the model by four different approaches and X is the market index ($S\&P500$). That is, the fitted residuals are given by,

$$\widehat{\epsilon_i} = y_i - \widehat{y_i}$$

The residuals give an indication of the errors the regression would make in a forecasting application. Looking to the graph of the residues it shows that the error in the estimation of the parameters by a linear regression is larger than that using the time varying regression. Smallest errors are showed by the estimation of the time-varying coefficients using Kalman filter techniques, as we can see in Figure 4.8.

Residuals analysis.

The adequacy of the model and its underlying assumptions were examined further by analyzing the statistical properties of the residuals.

39
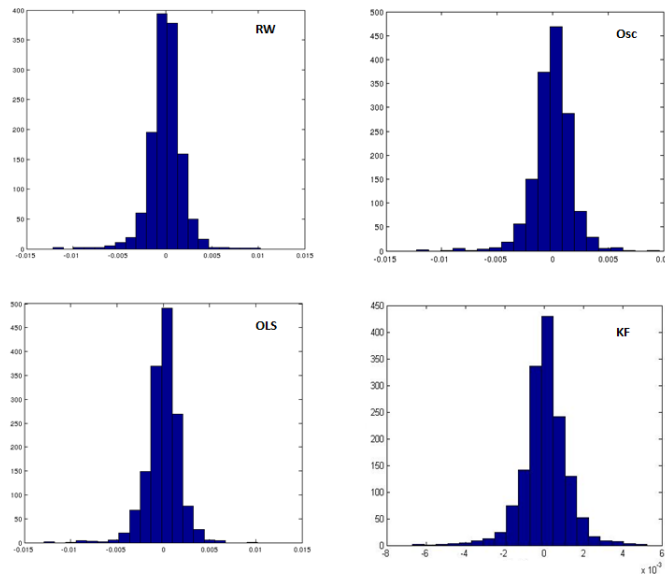
Figure 4.9: Histogram of the residues of the estimation by each estimation method used in the thesis

Residuals from the full sample estimates were in general well behaved. In all cases they were distributed with approximate symmetry and with means near zero as we can see in histograms in figure 4.9. Most of the values are concentrated in the middle, and there is no heavy tail. Looking at the graphs, they suggest that there is a normal distribution in the residuals but for further analysis we are going to do some test to confirm or not the assumptions.

In order to see if the provided assumptions regarding the random error term in the models are satisfied we can analyze the residues of each method by running a diagnostic tests on residuals. We obtained the following table 4.4 in which the values of the Ljung-Box-Q (LBQ) test and the Kolmogorov-Smirnov (KS) test are shown.

The Ljung-Box-Q test is used to test the the hypothesis that all of the autocorrelations are zero; that is, that the series is a white noise. Under the null hypothesis the statistics used, Q, is distributed as Chi-square. In general the Ljung-Box-Q test can be defined as:

$$H_0 : \quad \text{The data is independently distributed.}$$
$$H_a : \quad \text{The data is not independently distributed.}$$

and the Q-statistic is given by the following equation:

$$Q = T(T+2) \sum_{k=1}^{L} \frac{r_k^2}{T-k}$$

where T is the sample size, r_k^2 is the squared sample autocorrelation at lag k, and L is the number of lags being tested. The Kolmogorov-Smirnov test is a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution. The test can be modified to serve as a goodness of fit test. In the special case of testing for normality of the distribution, samples are standardized and compared with a standard normal distribution. Kolmogorov-Simirnov test is defined by:

H_0 :   The data (x) follows a standard normal distribution.
H_a :   The data (x) does not follow that distribution.

and the test statistic is given by:

$$D = \max(|F(x) - G(x)|)$$

where $F(x)$ is the empirical cumulative distribution function and $G(x)$ is the standard normal cumulative distribution function.

Tests		OLS	RW-OLS	KF	bOsc
LBQ	*stats*	82.3083	57.3156	76.7338	186.5346
	p-value	$< 10^{-3}$	$< 10^{-3}$	0	$< 10^{-3}$
KS	*stats*	0.4967	0.4959	0.4984	0.4967
	p-value	$< 10^{-3}$	$< 10^{-3}$	0	$< 10^{-3}$

Table 4.4: Values of the statistics and p-value of the Ljung-Box-Q test and Kolmogorov-Smirnov test with 0.05 as the significance level to test independence and normality respectively.

Table 4.4 shows that the residues for each method are not normal and independent distributed because for each method we have to reject the null hypothesis at 0.05 of significance level. Therefore, the models fit are not adequate. Consequently, we decided to work with a GARCH model to overcome these two issues, [Terasvirta, 2009]. But working with GARCH model we face another problem, the parameters estimation.

The GARCH model we use is GARCH(1,1) and it can be written by the following equations,

$$y_t = \alpha_t + \beta_t X_t + \varepsilon_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$
$$\alpha_t = \alpha_{t-1} + \vartheta_t$$

where ζ_t and ϑ_t are white noise, and $\varepsilon_t \sim \aleph(0, \sigma_{\varepsilon_t}^2)$, with

$$\sigma_{\varepsilon_t}^2 = C_0 + C_1 \varepsilon_{t-1}^2 + D_1 \sigma_{\varepsilon_{t-1}}^2$$

where GARCH(1,1) coefficients have the following constrains $C_0 > 0$, $C_1 \geq 0$, $D_1 \geq 0$ and $C_1 + D_1 < 1$. For more information see Appendix A.

Studying the GARCH model we realize that the estimation obtained by simulation were not good when compared to the real values used to simulate the data. Therefore, we decide not to continue working with this model for the thesis. It remains an open problem.

In the case of in-sample forecast we are going to measure the accuracy of OLS, rolling windows (RW-OLS), beta periodic (bOSC) and KF approaches by using a measure of forecast error which compares the forecasts to the actual observed data in-sample. One option is the mean absolute forecasting error (MAE), defined as:

$$MAE = \frac{1}{T} \sum_{j=1}^{T} |\widehat{y_j} - y_j| \tag{4.2}$$

An alternative approach is the root of the mean square forecasting error (RMSE) approach,

$$RMSE = \sqrt{\frac{1}{T} \sum_{j=1}^{T} (\widehat{y_j} - y_j)^2} \tag{4.3}$$

The use of the square in the equation places a heavier penalty on outliers than the MAE measure. Therefore, we are going to compare the performance of each estimation technique by using these two measures of forecasting error.

Methods/	**S&P500**		**Russell3000**		**MSCI**	
Models	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*
OLS	1.2741	1.7970	1.2736	1.7972	1.2691	1.7775
RW-OLS	1.2022	1.7109	1.2018	1.7100	1.1820	1.6432
KF	0.8321	1.0758	0.9955	1.3939	0.7493	1.0133
bOsc	1.2687	1.7886	1.2655	1.7882	1.2558	1.7643

Table 4.5: MAE and RMSE values for different models and market indices. Russell 3000 index is comprised of the 3000 largest and most liquid stocks based and traded in the U.S. and the MSCI World index is a composite of 24 developed and 27 emerging market indices.

From table 4.5 we can see that the best estimation is given by the model with time varying beta where a KF-likelihood approach is used. For the rest of the models presented in the table there is no significant difference between their values. Therefore, the best goodness of fit is given by the KF techniques followed by RW-OLS. Also a comparison between different indexes can be done. The Russell 3000 index is composed of the 3,000 largest U.S. companies as measured by market capitalization, and represents about 98% of the U.S. stock market. It can be subdivided into two segments: the Russell 1000 (consisting of the 1000 largest market-cap companies) and Russell 2000 (consisting of 2000 small-cap companies). As we can notice, the index is market-cap weighted, hence, the largest firms have the biggest impact on the index's value. On the other hand, the MSCI World index is a leading provider of equity, fixed-income and hedge fund indexes. Its acronym stands for *Morgan Stanley Capital*

International. MSCI global equity benchmarks have become the most widely used international indexes by institutional investors worldwide. This consistent approach makes it possible to aggregate individual indexes to create meaningful composite, regional, sector and industry benchmarks. As it is shown in the table, the best index could be the MSCI. It might be due to the fact that MSCI is a broad and investable global equity benchmark.

4.2.5 Before and after the financial crisis

From the graph of beta filtered, Figure 4.7, we have decided to paid special attention separately to the series before and during the crisis, in order to compare the performance of hedge fund strategies in good and bad times. Figure 4.7 showed that the KF approach will locate the crisis better, as we can see in the peak that the plot has is when the financial crisis started in late 2007.
We want to find which model fit better before and during a financial crisis, therefore in the following table we get some performance values for each model.

	Before		After	
	MAE	*RMSE*	*MAE*	*RMSE*
OLS	0.9775	1.2889	1.5592	2.2233
RW-OLS	0.9777	1.2792	1.5437	2.1544
KF	0.9728	1.2739	0.8363	1.1492
bOsc	0.9756	1.2783	1.5314	2.1909

Table 4.6: Performance values before and after the financial crisis started in late 2007. The values are re-scaled to 10^3.

Table 4.6 shows the performance of the series before and during the financial crisis respectively. In the first table we can see that the behavior of all the approaches are similar with just a little bit of variation that indicates KF provides a better fitting. However, in the performance values for the series after the financial crisis the difference between the values of each model is more relevant, given the KF as the best estimation since it can adapt to changes in the data.

4.3 Multi-factors model

Another factor, as for example options, in the regression model may explain the data better. In our case, we are going to work with a two factor model as described in Chapter 1, equation 2.7, as well as a three factor models when we combine both equations. The equation can be written as:

$$R_t = \alpha_t + \sum_{i=1}^{3} \beta_{it} R_{it} + \varepsilon_t$$

where R_{it} can be taken as the market return or as the option based factor for different values of i.

4.3.1 Models Comparison

In order to see if a KF approach with two or more factors models improves the fit, we are going to make a comparison with the OLS method. The models are given by:

$$R_t = \alpha_t + \beta_{1t}R_{Mt} + \beta_{2t}\max\{R_{Mt} - K_t, 0\} + \varepsilon_t, \quad \text{a call option added to the model}$$
$$R_t = \alpha_t + \beta_{1t}R_{Mt} + \beta_{2t}\max\{K_t - R_{Mt}, 0\} + \varepsilon_t, \quad \text{a put option added to the model}$$

or the case where both, put and call options are added to the model and the equation can be written as follow:

$$R_t = \alpha_t + \beta_{1t}R_{Mt} + \beta_{2t}\max\{R_{Mt} - K_t, 0\} + \beta_{3t}\max\{K_t - R_{Mt}, 0\} + \varepsilon_t$$

where K_t is the daily return in the LIBOR series. The parameters are either constant over time or follow a random walks in the case of time dependent parameters.

Methods/	**MAE**			**RMSE**		
Models	*call-option*	*put-option*	*both*	*call-option*	*put option*	*both*
OLS	1.2534	1.2562	1.2494	1.7774	1.7794	1.7743
KF	0.5732	0.6218	0.0473	0.7941	0.8704	0.0649

Table 4.7: Performance of the model with call options, put options, and the model with three factor that include call and put options. The values are re-scaled to 10^3

In table 4.7, as before, we can conclude that the KF estimation is better than OLS. The estimation is much better using two factor models than one factor model as we suggested before. But also using a call option the performance is better than put option, the following graph represents this better since the call option should have positives values when the market performance is good contrary to the put option.

Figure 4.10 shows both graphs, in the first (call option) we can see that in general most of the values are positive except during the crisis where the market performance were bad, hence the exposure to the option should take opposite position with respect to the market. The put option has the same behavior concluding that in this case the model with put option do not have better performance than the call option.
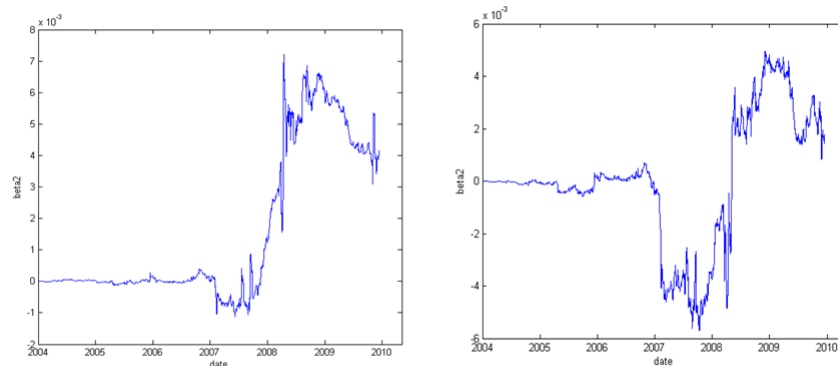
Figure 4.10: Exposure to the market for the call-option (left) and put-option (right) in a two factors models

Ideally, an equity fund should participate in the upside performance of the market and limits its negative values in the downside. The fund will typically increase its beta (market exposure) when the market is expected to perform well, and reduce its market exposure when market is expected to perform badly. In figure 4.10 we can see that before the financial crisis most exposures values are positive, inversely after the financial crisis in late 2007 the market exposure is reduced in order to minimize the losses. The figure, also shows the timing of the manager changing the investments, middle of 2008,as soon as they notice the change due to the crisis.
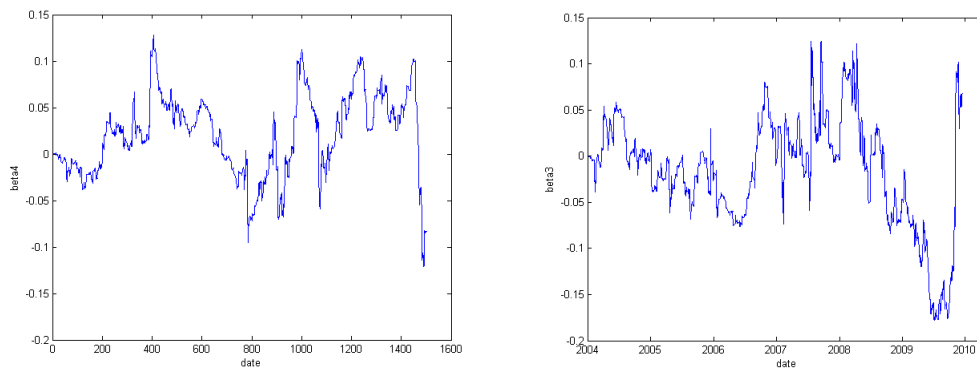

Figure 4.11: Exposure to the market for the call-option and put-option when working with a three factors model that takes in account the market, the call options and the put options.

Figure 4.11 represents the beta estimated for the factor call and put when working with a three factors model. In the graph we can observe the difference between the put and call option. In the case

of put option we can see that before the financial crisis most of the values are negatives, meaning that it has a negative correlation with the negative values of the market. The opposite happens during and after the financial crisis. The same analysis can be done for the graph of call options, but we have to take into account that here the values are positive, therefore the correlation is positive with positive values of the market.

Another three factor model interesting to study, is the model where the factor are different market indexes. In this case we chose S&P500, Russell3000 and MSCI indexes, hence the model can be written as:

$$R_t = \alpha_t + \beta_{1t}R_{SP500} + \beta_{2t}R_{Rusell3000} + \beta_{3t}R_{MSCI} + \varepsilon_t$$

where the parameters can be either constant or time dependent. With this model we get the following graphs,
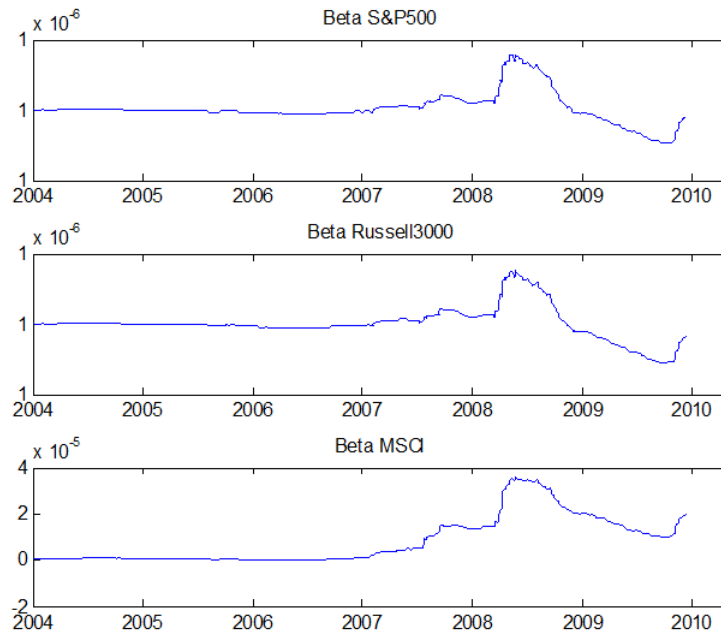
Figure 4.12: Exposure to each market return used in the model describe above. Market returns used: S&P 500, Russell 3000 and MSCI

Figure 4.12 we can see that the pattern of each exposure to the market return is very similar to the exposure when we use only S&P500 as the market return. We can see the difference in values between before and after the crisis, but the values of the exposure for MSCI are slightly bigger that the others two. It might be caused due to the fact that the MSCI is a global index, hence it includes international markets but not S&P500 and Russell3000 that are based fundamentally in the US market.

The performance value in this case are: $MAE = 0.5514$ and $RMSE = 0.7790$ both lesser values than

the previous models, suggesting that the model fit better for the data analyzed.

4.3.2   Before and after the financial crisis with two factors

Now, working with the option-like model and the data before and after the crisis we get the following values to measure the performance of each series.

	Call Option		Put Option	
	MAE	*RMSE*	*MAE*	*RMSE*
OLS	0.9474	1.2548	0.9495	1.2567
KF	0.9358	1.2359	0.8571	1.1356

Table 4.8: Before the Financial Crisis in late 2007. The series are divided as following: before the crisis are the returns from July 2004 to July 2007 and after the crisis are the returns from August 2007 to June 2010. The values are re-scaled to 10^3

	Call Option		Put Option	
	MAE	*RMSE*	*MAE*	*RMSE*
OLS	1.5266	2.1951	1.5663	2.1950
KF	0.4784	0.7902	0.8302	1.1501

Table 4.9: After the Financial Crisis in late 2007. The series are divided as following: before the crisis are the returns from July 2004 to July 2007 and after the crisis are the returns from August 2007 to June 2010. The values are re-scaled to 10^3

In tables 4.8 and 4.9 the conclusion is similar to the tables in the previous section. Therefore, the behavior of the model is similar but the estimation is better using two factor models. Also, before the crisis the values of MAE and RMSE of the KF and OLS methods are closer but not the same after the crisis, since the values of KF are much more smaller than OLS which implies that the KF is the better approach to use since in normal time it behaves like the OLS but in time with large volatility the estimation is much better. Also, as we said before, the KF approach is very useful in the location of the crisis time.

4.4   Hedge Fund Styles

The term "hedge funds" is often used generically, in reality hedge funds are not all alike. In fact, there are several investment styles with different approaches and objectives, where the returns, volatilities and risk vary not only according to the fund manager, but also to the target market and the investment strategies. As it is critical to have a basic understanding of the underlying hedge fund strategies and their difference. In order to develop a coherent plan to exploit the opportunity offered by hedge funds, consultants, investors and managers often classify the hedge fund market into a range of investment styles.

We examined eight hedge fund indexes corresponding to different strategies. We use 8 strategies of hedge fund: Convertible Arbitrage, Distressed Securities, Event Drive, Global, Equity Market Neutral, Merger Arbitrage, Macro, Relative Value Arbitrage. See Appendix A for more details on the definitions of these strategies. The data consist of daily returns from June 2003 through to July 2010, a period that covers market up and downturns.

Since the crisis cause large volatility variation and high kurtosis in the returns data, it is interesting to analyze the behavior of different strategies of hedge funds in good and bad times. As each style has different investment strategies, therefore they will have a diverse way to handle the crisis.

Hedge Fund Style	Mean	SD	Sharpe	Skew	Kurtosis	Min	Max
Convertible Arbitrage	5.39×10^-5	0.0021	0.025	-0.417	5.128	-0.0095	0.0090
Distressed Securities	$0.3210 \times^- 5$	0.0015	0.216	0.397	6.062	-0.0074	0.0077
Event Drive	0.362×10^-5	0.0027	0.137	-0.829	7.928	-0.019	0.011
Global	0.242×10^-5	0.0022	0.109	-1.213	8.420	-0.013	0.0081
Equity Market Neutral	4.13×10^-5	0.0025	0.017	-0.241	49.386	-0.031	-0.031
Merger Arbitrage	0.23×10^-5	0.0022	0.108	-1.137	15.675	-0.018	0.013
Macro	0.16×10^-5	0.0045	0.035	-1.764	14.554	-0.037	0.020
Relative Value Arbitrage	0.19×10^-5	0.0019	0.098	0.404	12.532	-0.0095	0.016

Table 4.10: Descriptive analysis of the return series of different strategies of hedge fund before the crisis

Table 4.10 presents a summary of descriptive statistics of daily returns for eight HFR indexes from July 2003 through to July 2007. The summary statistics include mean, standard deviation (SD), sharpe ratio (Sharpe), skewness (Skew), kurtosis, minimum (Min) and maximum (Max), minimum and maximum values of each of the statistic among the indexes.

Hedge Fund Style	Mean	SD	Sharpe	Skew	Kurtosis	Min	Max
Convertible Arbitrage	-0.744×10^-5	0.0073	-0.103	-3.253	23.682	-0.066	0.032
Distressed Securities	-0.56×10^-5	0.0037	-0.151	-2.972	32.448	-0.043	0.017
Event Drive	0.16×10^-5	0.0041	-0.039	-1.060	13.641	-0.032	0.025
Global	-0.18×10^-5	0.0032	-0.058	-1.073	10.652	-0.020	0.019
Equity Market Neutral	-4.93×10^-5	0.0072	-0.015	-0.171	4.672	-0.014	0.013
Merger Arbitrage	0.17×10^-5	0.0044	0.039	2.091	48.575	-0.024	0.056
Macro	0.35×10^-5	0.0050	0.007	-0.180	3.801	-0.019	0.015
Relative Value Arbitrage	-0.15×10^-5	0.0046	-0.033	-1.315	19.966	-0.038	0.032

Table 4.11: Descriptive analysis to the return series of different styles of hedge fund during the crisis

Table 4.11 presents a summary of descriptive statistics of daily returns for eight HFR indexes from August 2007 through to July 2010. The summary statistics includes mean, standard deviation (SD), sharpe ratio (Sharpe), skewness (Skew), kurtosis, minimum (Min) and maximum (Max), minimum and maximum values of each of the statistic among the indexes.

When comparing the performance of eight hedge fund strategies before and after the crisis we observe that the average returns for all of the chosen strategies are positives for the series before crisis, but as we expected during the crisis period the behavior is worse, since most of them are negative. When we look at the volatility we find, similarly to the average return, that for the series before the crisis is lower

than the series after the crisis period. We need to point out that the worst behavior from before and after the crisis is for the Equity Market Neutral with respect to the average return and volatility.

We observe that the kurtosis for both series is higher than three for all of the strategies of hedge fund we are analyzing. Therefore, the probability distribution of the returns have heavier tails than the normal distribution for all of them in each period. We find positive skewness for the following strategies: Distress Securities and Relative Value Arbitrage in the series before the crisis, and Merger Arbitrage for the series after the crisis that indicates the fat left tail for these strategies. The rest of the strategies have a negative skewness for both series (before and after the crisis).

In terms of Sharpe ratio as well, the strategies for the series before the crisis exhibit better risk-return tradeoffs compared to the series after the crisis. Thus, overall before the crisis strategies seem to have delivered better risk-return tradeoff compared to the after the crisis strategies across a wide range of risk-return metrics during the period measuring in each case. Furthermore, we can conclude that the series of the strategies before the crisis perform better that those after the crisis period. We need to emphasize that the Merger Arbitrage strategy performs well in both periods, before and after crisis.

4.4.1 Strategies Analysis

The suggested option-like model should be more useful for the type of hedge fund styles that use option strategies. Therefore, we are going to work with three of these styles: Equity Hedge Index, Global Hedge Fund Index and Macro Index, see [LHabitant, 2001].

	MAE		RMSE	
	O-L	No O-L	O-L	No O-L
Equity	1.1931	3.0867	1.7712	4.3474
Global	0.5684	1.8066	0.8138	2.6350
Macro	2.6841	2.8622	3.7498	4.0015

Table 4.12: Performance of the option-like model with time-varying parameters (O-L) and the regression model with time-varying parameters (No O-L) for different styles of hedge funds. The values are re-scaled to 10^3

In table 4.12 we can note the higher difference in the Equity and Macro styles since they are the styles that use more option strategies. The comparison between the styles shows that the styles Equity and Global have better behavior when using option-like models. This might be due to the tendency of these styles to have long exposure that might be suggesting a significant correlation with the market. The performance of the Global style is much better since the manager's investments could be worldwide.

Conclusions

Over the past decades the financial literature has abounded with constant coefficient model just to make it easy to analyze the stochastic dynamics of asset prices. However, this assumption has been criticized by many empirical analysts. Consequently, at the beginning of the thesis we stated that dynamics asset allocations of hedge fund managers could not be captured using constant coefficient regression methods. We realized that as consequence of this, time varying exposure of hedge funds as beta would be inaccurately computed with these methods.

Given the regression model with time varying coefficients as our state space, we applied a Kalman filter based state estimation technique to our model. Anchored to the recursive Kalman filtering mechanism, we performed a numerical analysis with real data of Hedge Funds. The same was done for a regression model with two factors and three factors. Taking into account the relationship between some hedge funds styles and options, we used a model that includes a call or put option and performance also a numerical analysis. The numerical analysis in both cases include the comparison between models and methods, since we worked not only with stochastic coefficient models but also with deterministic coefficients models.

We found that the models with stochastic coefficients are superior to the model with constant coefficient in most of the cases. Moreover, the estimation of the parameters of this model with Kalman filter approach perform better than the OLS-Rolling Windows method. Through the thesis a study of the models with beta varying on time have been done and this study suggests that the time varying beta models behave better when working with forecasting in-sample. In general, concerning the evolution of beta over time, results indicates that models with time varying are better able to estimate the current value of the parameters than OLS.

When using option-based models, the advantages of models with time varying parameters against OLS decline while the forecast error becomes smaller than in the univariate case for both, classic linear regression and linear regression with time-variant parameters.

It is precisely the base of the OLS-rolling window In order to analyze the impact of the financial crisis in the exposure of the hedge funds industry; we made use of the series of return of some styles of hedge funds that includes the financial crisis. An analysis of the series before and during the financial crisis was completed for every model studied in this thesis. When working with the series before the financial crisis we concluded that the use of Kalman filter approach or OLS technique might be a choice of the practitioners since the performance of both methods for this period was similar. In the case of the series

during the financial crisis the Kalman filter approach improved the OLS method suggesting that the Kalman filter can locate better the changes in the market showing high volatilities even before the crash. For further works a forecasting out-of-sample may have been done to conclude if time-varying exposure considered in the estimation models fit better when modeling hedge funds or some financial series that involves managing. It will also be interesting to use Kalman filter approach with GARCH models when we do not have the assumptions needed in the regression equation.

Appendix A

GARCH Model

Financial economists are concern with modeling volatility in assets returns. This is important in the financial field because volatility is considered a measure of risk, and investor want a premium for investing in risky assets. Modeling and forecasting volatility is therefore important. The fact that volatility in returns vary on time has been known for a long time. Observations in return series of financial assets observed at weekly and higher frequencies are in fact not independent. While observations in these series are uncorrelated or nearly uncorrelated, the series contain higher order dependence. Models of Autoregressive Conditional Heteroskedasticity (ARCH) form the most popular way of parameterizing this dependence. This ARCH model is the first model of conditional heteroskedasticity. Let ε_t be a random variable that has mean and variance conditionally on the information known at time $t-1$. The ARCH model follows the properties related to conditional mean and variance in t given the information at $t-1$. It means that the conditional mean and conditional variance of $\varepsilon_{t|t-1}$ has the following properties: first, $E(\varepsilon_{t|t-1}) = 0$ and second, $Var(\varepsilon_{t|t-1}) = \sigma_{\varepsilon_t}^2$) is a nontrivial positive-valued parametric information function. The sequence ε_t in our case is the error sequence of the regression model. Lets consider the general regression model with conditional heterocedasticity ARCH of order q structure for the error variance and parameters α and β varying in time,

$$
\begin{aligned}
y_t &= \alpha + \beta_t X_t + \varepsilon_t \\
\beta_t &= \beta_{t-1} + \zeta_t \\
\alpha_t &= \alpha_{t-1} + \vartheta_t
\end{aligned}
\tag{A.1}
$$

where ζ_t and ϑ_t are white noise, and $\varepsilon_t \sim \aleph(0, \sigma_{\varepsilon_t}^2)$, with

$$
\sigma_{\varepsilon_t}^2 = C_0 + \sum_{i=1}^{q} C_i \varepsilon_{t-i}^2
\tag{A.2}
$$

where $C_0 > 0$, $C_i \geq 0$ $i = 1, \ldots, q-1$ and $C_q > 0$. The parameter restrictions in A.2 form a necessary and sufficient condition for positivity of the conditional variance.

In many applications, the ARCH model has been replaced by the Generalized ARCH (GARCH) model. In this model proposed by [**?**], the conditional variance is also a linear function of its own lags and can be written of the following form for a GARCH(p,q) model,

$$\sigma^2_{\varepsilon_t} = C_0 + \sum_{i=1}^{q} C_i \varepsilon^2_{t-i} + \sum_{j=1}^{p} D_j \sigma^2_{\varepsilon_{t-j}} \tag{A.3}$$

where the GARCH(p,q) coefficients has the following constrains $C_0 > 0$, $C_i \geq 0$ $i = 1, \ldots, q$, $D_j \geq 0$ $j = 1, \ldots, p$ and $\sum_{i=1}^{q} C_i + \sum_{j=1}^{p} D_j < 1$. The parameter restrictions in A.2 form a necessary and sufficient condition for positivity of the conditional variance. The conditional variance define above in A.3 has the property that the unconditional autocorrelation function of ε_t^2, if exists, can decay slowly although still exponential. For the ARCH family, the decay rate is too rapid compare to what is typically observed in financial time series, unless the maximum lag q in A.2 is long. As A.3 is more parsimonious model of the conditional variance than a high-order ARCH model [Terasvirta, 2009]

Appendix B

Hedge Fund Styles

To further investors understanding of hedge funds, data providers have classified them into more homo-geneous categories or styles. The funds within each category serve as a peer group for the style. Hedge fund classifications differ in both definition and granularity. Therefore, it is important to understand the differences between the various hedge fund strategies because all hedge funds are not the same – investment returns, volatility, and risk vary enormously among the different hedge fund strategies. Some strategies which are not correlated to equity markets are able to deliver consistent returns with extremely low risk of loss, while others may be as or more volatile than mutual funds. A successful fund of funds recognizes these differences and blends various strategies and asset classes together to create more stable long-term investment returns than any of the individual funds.

We now describe 8 Hedge Fund investments styles with which we are going to do some statistical com-parison,

- **Convertible Arbitrage**
  Convertible arbitrage is a market-neutral investment strategy often employed by hedge funds. This style is identified by hedge investing in the convertible securities of a company. While some hedge funds simply invest in convertible bonds, a hedge fund using convertible arbitrage is actually taking positions in both the convertible bonds and the stocks of a particular company. A convertible bond can be converted into a certain number of shares. A typical investment is to be long the convertible bond and short the common stock of the same company. Positions are designed to generate profits from the fixed income security as well as the short sale of stock, while protecting principal from market moves. The approach typically involves a medium-term holding period and results in low volatility.

- **Distresses Securities**
  Distressed Securities is the investing by the fund managers invest in the debt, equity or trade claims of companies in financial distress and generally bankruptcy. Also, short-selling the stocks of those corporations. Such companies are generally in bankruptcy reorganization or are emerging from bankruptcy or appear likely to declare bankruptcy in the near future. Because of their

distressed situations, the manager can buy such companies' securities at deeply discounted prices. The manager stands to make money on such a position should the company successfully reorganize and return to profitability. Also, the manager could realize a profit if the company is liquidated, provided that the manager had bought senior debt in the company for less than its liquidation value. Results generally not dependent on the direction of the markets. Expected Volatility: Low - Moderate.

- **Event drive**
  This strategy is defined as equity-oriented investing designed to capture price movement generated by an anticipated corporate event. Some managers who employ Event-Driven trading strategies may shift the majority weighting between Merger Arbitrage and Distressed Securities, while others may take a broader scope. Typical trades and instruments used may include long and short common and preferred stocks, debt securities, options and credit default swaps. Leverage may be employed by some managers. Such funds, which include risk-arbitrage vehicles and entities that buy distressed securities, typically employ medium-term holding periods and experience moderate volatility.

- **Global**
  The Global strategy is an all-round category for funds that invest is assets beyond those based in their home market. Other than that, no more specific technique is associated with this. An example would be a Hedge Fund investing in an emerging market such as India.

- **Equity Market Neutral** Equity market neutral is a hedge fund strategy that seeks to exploit investment opportunities unique to some specific group of stocks while maintaining a neutral exposure to broad groups of stocks defined, for example, by sector, industry, market capitalization, country, or region. Due to the portfolio's low net market exposure, performance is insulated from market volatility. The Market-neutral funds typically employ long-term holding periods and experience moderate volatility.

- **Merger Arbitrage**
  Merger Arbitrage Fund is a fund strategy in which the stocks of two merging companies are simultaneously bought and sold to create a risk-less profit. A merger arbitrageur looks at the risk that the merger deal will not close on time, or at all. Because of this slight uncertainty, the target company's stock will typically sell at a discount to the price that the combined company will have when the merger is closed. This discrepancy is the arbitrageur's profit. In other words, merger arbitrage is an investment strategy that simultaneously buys and sells the stocks of two merging companies.

- **Macro**
  Global macro are the strategies that have the highest risk/return profiles of any hedge fund strategy. Global macro funds is an approach in which a fund manager seeks to anticipate broad trends in the worldwide economy due to economic, political, or government related events. Based on

those forecasts, the manager chooses investments from a wide variety of market, e.g. stocks, bonds, currencies, commodities, options, futures, forwards and other forms of derivative securities. They tend to place directional bets on the prices of underlying assets and they are usually highly leveraged. Most of these funds have a global perspective and, because of the diversity of investments and the size of the markets in which they invest, they can grow to be quite large before being challenged by capacity issues. Many of the largest hedge fund that "blow-ups" were global macros. The approach typically involves a medium-term holding period and produces high volatility. Many of the largest hedge funds follow global-macro strategies. They are sometimes called "macro" or "global directional-investment" funds. Expected Volatility: Very High.

- **Relative Value Arbitrage**
  Relative Value Arbitrage is an investment strategy that seeks to take advantage of price differentials between related financial instruments, such as stocks and bonds, by simultaneously buying and selling the different securitiesthereby allowing investors to potentially profit from the "relative value" of the two securities. The underlying concept is that a hedge fund manager is purchasing a security that is expected to appreciate, while simultaneously selling short a related security that is expected to depreciate. Related securities can be the stock and bond of a specific company; the stocks of two different companies in the same sector; or two bonds issued by the same company with different maturity dates and/or coupons. In each case, there is an equilibrium value that is easy to calculate since the securities are related but differ in some of their components.

Appendix C

MATLAB Code

```
% Working with option−like model y_t = a_t + b1_t * X_t + b2_t * X1_t + ey_t
% where a_t = a_t−1 + ea_t; b1_t = b1_t−1 + eb1_t; b2_t = b2_t−1 + eb2_t;
% X_t is the vector of the market returns; y_t is the vector of the hedge
% fund returns and X1_t = max(X_t−K_t,0) with K_t the LIBOR series.

clear all
format long
global XT Y m bKF nt
load Y; load X; load IR;
%%%%%%
K = IR;
X1 = max(X−K,0);
XT = cat(2,ones(length(X),1),X,X1);
% Initial Values
par0 = [0.07 0.02 0.03 0.01];   %[ey,ealpha,ebeta,elambda]
% Option for the optimization function
options = optimset('Display','iter−detailed','TolX',10^(−10),...
      'TolCon',10^(−10),'TolFun',10^(−10));
lbv = 10^(−9);        % lower bound
ubv = 200;            % upper bound
lb = [lbv lbv lbv lbv];
ub = [ubv ubv ubv ubv];
% Calling the optimization function implemented for Matlab
[par,fval,eflag] = fmincon(@loglik1,par0,[],[],[],[],lb,ub,[],options);
% Error of estimation by Kalman Filter approach
res = Y−(bKF(1,:)'+bKF(2,:)'.*X+bKF(3,:)'.*X1);
```

```
% Estimation performance
perf1 = mae(res);
perf2 = mse(res); perf2 = sqrt(perf2);
% Plot the vector of the coefficient estimated by KF
subplot(3,1,1); plot(bKF(1,:)); h1=gca;
title('Alpha')
subplot(3,1,2); plot(bKF(2,:)); h2=gca;
title('Beta 1')
subplot(3,1,3); plot(bKF(3,:)); h3=gca;
title('Beta 2')
hh = [h1 h2 h3];
set(hh,'XTick',1:252:length(Y)+252) % setting the axes of the figures
set(hh,'XTickLabel',{'2004','2005','2006','2007','2008','2009','2010'})
```

Log-likelihood Function

```
% V=(v1,d1,d2,v2,v3)
% v1 -- the variance of e1, z(t)=(bt)*z(t-1)+e1
% d1 -- coefficient of the tansition matrix M (alpha)
% d2 -- coefficient of the tansition matrix M (beta)
% v2 -- the variance of e2, s(t)=M1*s(t-1)+e2 (alpha)
% v3 -- the variance of e3, z(t)=M2*z(t-1)+e3 (beta)


function L = loglik(V)
global XT Y m bKF nt h
% Initialization
m = length(Y);     % size of returns
Xt = XT;
Yt = Y;
Bu = [0.000001 0.000001 0.000001]';
Pu = zeros(m,1);
H = V(1);
M = 1;                % we are working with a random walk
h = length(V);
vtemp = V(2:h);
Q = diag(vtemp);     % matrix of variance of the error
% Calling the Kalman Function
[Ft,nt, lt, bKF] = kalman(Bu,M,Q,Pu,Xt,Yt,H);
L = -sum(lt);          % the negative of the log-likelihood function since
                       % we need to find its maximum and the optimization
                       % function implemented in Matlab finds the minimum.
```

Kalman Function

```matlab
% The input of the function are the initial values needed for the Kalman
% equations since the equations work with the past data. The output are
% the vector values of the log-likelihood function in each step, and the
% matrix that each column contains the estimation vector of each coefficient
% of the model.

function [kalm, error, LL, betaKF] = kalman(Bu,M,Q,Pu,Xt,Yt,H)
global m h
for j=1:m
    if j==1
        Btu = M'.*Bu;          % Bt|t-1 -- state prediction
        Ptu = Q;               % prediction dispersion
    else
        Btu = M'.*Bu;          % Bt|t-1 -- state prediction
        Ptu = M*Pu*M'+Q;       % prediction dispersion
    end
    Zt = Xt(j,:);
    Ytu(j) = Zt*Btu;
    nt(j) = Yt(j)-Ytu(j);      % prediction error
    Ft(j) = Zt*Ptu*Zt'+H;      % error dispersion
    InvFt(j) = Ft(j)\eye(1);
    Kt = (Ptu*Zt').*InvFt(j);      % Kalman gain
    Bu = Btu+Kt.*nt(j);        % state estimate
    for k=1:h-1
        BuT(k,j) = Bu(k);
    end
    Pu =Ptu-Kt*Ft(j)*Kt';      % estimate dispersion
    kalm1(j) = Ft(j);
    error(:,j) = nt(j);
    LL(j) = -((h/2)*log(2*pi)+(1/2)*log(abs(Ft(j)))+(1/2)*nt(j)*InvFt(j)*nt(j)');
end
kalm = kalm1;
betaKF = BuT;
```

Bibliography

[Agarwal and Naik, 2004] Agarwal, V. and Naik, N. Y. (2004). Risks and portfolios decisions involving hedge funds. *The Review of Financial Studies*, 17:63–98.

[Bos and Newbold, 1984] Bos, T. and Newbold, P. (1984). An empirical investigation of the possibility of stochastic systematic risk in the market model. *Journal of Business*, (1):35–41.

[Brealey and Kaplanis, 2001] Brealey, R. A. and Kaplanis, E. (2001). Changes in the factor exposures of the hedge funds. *London Business School*.

[Bressler and Holler, 2010] Bressler, W. and Holler, J. (2010). Hedge funds and asset allocation: Investor confidence, diversification benefits, and a change in investment style composition. In Fink, A., Lausen, B., Seidel, W., and Ultsch, A., editors, *Advances in Data Analysis, Data Handling and Business Intelligence*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 441–450. Springer Berlin Heidelberg.

[Daniel W. Collins and Rayburn, 1987] Daniel W. Collins, J. L. and Rayburn, J. (1987). Some further evidence on the stochastic properties of systematic risk. *Journal of Business*, 60:425–448.

[Ebner and Neumann, 2004] Ebner, M. and Neumann, T. (2004). Time-varying betas of german stock returns. *Financial Market and Portfolio Management*, 19:29–43.

[Fabozzi and Francis, 1978] Fabozzi, F. J. and Francis, J. C. (1978). Beta as a random coefficient. *Journal of Financial and Quantitative Analysis*, (13):101–115.

[Fung and Hsieh, 1997] Fung, W. and Hsieh, D. A. (1997). Empirical characteristics of dynamic trading strategies. the case of hedge funds. *The Review of Financial Studies*, 10:63–98.

[Grewal and Andrews, 1993] Grewal, M. S. and Andrews, A. P. (1993). *Kalman Filtering: Theory and Practice*. Prentice-Hall.

[Harvey, 1989] Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press.

[Henriksson and Merton, 1981] Henriksson, R. D. and Merton, R. C. (1981). On market timing and investment performance. ii. statistical procedures for evaluating forecasting skills. *Journal of Business*, 54(4):513–533.

[Kedem and Fokianos, 2002] Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. Wiley-Interscience.

[Kim, 1993] Kim, D. (1993). The extent of non-stationarity of beta. *Review of Quantitative Finance and Accounting*, (3):241–254.

[Ledolter, 1979] Ledolter, J. (1979). A recursive approach parameter estimation in regression and time series models. *Comm. Statistics*, A8:1227–1245.

[LHabitant, 2001] LHabitant, F.-S. (2001). *Market Models: A Guide to Financial Data Analysis*. John Wiley & Sons.

[Manly, 1994] Manly, B. (1994). *Multivariate statistical methods: A primer*. Chapman and Hall.

[Nai-Fu Chen and Ross, 1986] Nai-Fu Chen, R. R. and Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, 59:383–403.

[Pagan, 1980] Pagan, A. (1980). Some identification and estimation results for regression models with stochastically varying coefficientes. *Journal of Econometrics*, 13:341–363.

[Racicot and Theoret, 2009] Racicot, F.-E. and Theoret, R. (08-2009). The modeling hedge fund returns using the kalman filter: an errors-in-variables perpective. Research paper, Department of Administrative Sciences, University of Quebec.

[Robert D. Brooks and McKenzie, 1999] Robert D. Brooks, R. W. F. and McKenzie, M. D. (1999). Time-varying beta risk of australian industry portfolios: A comparison of modelling techniques. *MIT Press*.

[Robert W. Faff and Hillier, 2000] Robert W. Faff, D. H. and Hillier, J. (2000). Time varying beta risk: An analysis of alternative modelling techniques. *Journal of Business, Finance & Accounting*, 27(6):523–553.

[Schwert, 1989] Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business and Economic Statistics*, 7:147159.

[Schwert and Seguin, 1990] Schwert, G. W. and Seguin, P. J. (1990). Heterscedasticity in stock returns. *Journal of Finance*, (45):1129–1155.

[Sunder, 1980] Sunder, S. (1980). Stationarity of market risk: Random coefficients tests for individual stocks. *Journal of Finance*, (4):883–896.

[Swinkels and Sluis, 2006] Swinkels, L. and Sluis, P. J. V. D. (2006). Return-based style analysis with time-varying exposures. *European Journal of Finance*, 12:529–552.

[Terasvirta, 2009] Terasvirta, T. (2009). An introduction to univariate garch models. *In: Andersen, T.G., Davi,s R. A., Mikosch, T. (Eds.): Handbook of Financial Time Series*, pages 17–42.

[Tsay, 2010] Tsay, R. S. (2010). *Analysis of Financial time series*. John Wiley & Sons.

[Wei, 2006] Wei, W. S. (2006). *Time series analysis: univariate and multivariate methods*. Pearson Addison Wesley.

[Wells, 1994] Wells, C. (1994). Variable betas on the stockholm exchange 1971-1989. *Applied Financial Economics*, (4):75–92.

[Yao and Gao, 2006] Yao, J. and Gao, J. (2006). Computer-intensive time-varying model approach to the systematic risk of australian industrial stock returns. *Australian Journal of Management*, 29.