Bayesian State Space Modelling Of Factor Investing

A Quantitative Equity Strategy Based On Kalman Filter

Master's Thesis

Author: Giovanni Venturato

Student Number: 107244

MSc. Economics and Business Administration - Finance and Investments (FIN)

Supervisor: Rasmus Tangsgaard Varneskov

Department of Finance, Copenhagen Business School

Number of characters (No space): 101,933

Number of characters (With space): 121,475

Number of pages: 70

Date of submission: 16 September 2018

Abstract

This document focuses on the implementation and development of a momentum investment strategy. Instead of considering the past movements of stock prices as signals for investments, the positions are based on forecast estimations of asset prices. Predictions are based on a dynamic linear representation of the Fama French three factors model. Kalman recursion allows to filter out noise free estimates of the regression coefficients and to predict asset returns. The simulated investment strategies are able to overperform the market portfolio with statistical significance in a control environment, but performances start to suffer once market frictions are introduced.

Acknowledgement

The first person I would like to thank is my supervisor, professor Rasmus Tangsgaard Varneskov. His valuable guidance and help with computational issues had a major impact on my thesis. Also, his friendly and genuine support helped me during the whole process.

I need to thank the Department of Statistics of the University of Padua for the opportunity to participate to a PhD seminar with professor Siem Jan Koopman. I also need to thank professor Koopman for his advice of the model developed in this document.

I need to thank to all my co-workers at Derivative Partners for their support and advice while writing all the codes. Using the office machines helped me to reduce the computational time of the recursions and save a huge amount of time.

I also need to thank all the special people I met in Denmark. These two years have been an important period of my life for personal development, and this would not have been possible without them.

Lastly, I want to thank my family for their unconditional love and support.

Zürich, 16 September 2018

Giovanni Venturato

Contents

1 Introduction

In economics, researchers are used to analyze relevant variables by analyzing historical data and try to model their relationships. They can be interested in assessing the relationship between the wage and education level among individuals, or measuring the group effect of separating males and females in order to catch the gender effect on wealth redistribution. For these basic questions, economic literature provides powerful tools for estimating relationships and try to model causality among variables. For example, a least square regression model can be applied to model the correlation between dependent and independent variables. Assessing if these relationships will continue or variate in the future is a different question. Information about the past can be used for forecasting, but uncertainty will always be a major problem for future estimations. Every economic model requires and imposes assumptions in order to control uncertainty, and relaxing these assumptions may to a dramatic rise of model complexity.

The forecasting of financial variables has always been a prominent area of finance and economics literature. Every investor seeks to wisely predict future movements of stocks, bonds and other financial instruments to trade based on this information. One of the most famous model for stock return analysis is the Fama French three factor model. In their model, the two authors describe the returns of a portfolio or stock with the returns of three factors. It allows to describe the returns of a specific asset or portfolio by analysing its sensitivities to the returns of three factors. The historical co-movements with the three factors signal how the analysed stock react to the factors movements. This information can be used to predict future developement of the stock price. Future predictions are usually based on modelling techniques able to analyze past trend and incorporate current information for future estimations. These estimates are created from employing time-series, cross-sectional or longitudinal data in statistical models. These quantitative models are used to forecast data by taking historical information as input. A powerful tool which started to be employed in economic forecasting during the last century is the Kalman filter. This is a linear quadratic estimator able to model time series observations and filter out statistical noise to produce estimates of unknown variables. Proper tool of engineering, the Kalman filter has been widely accepted as a valuable tool for econometricians for its powerful flexibility and robust results.

The aim of this document is to create an investment strategy where the signals for

portfolio rebalancing are backed and controlled by a forecasting recursion in order to develop a quantitative equity momentum strategy. The goal is to combine the flexibility of the Kalman filter and the predictive power of the Fama French three factors model to create an investment strategy and test it on real data. These strategies fall under the category of quantitative equity investments, proper of active portfolio manager.

The outline of the document is the following. First, the theoretical foundations required to understand the models are presented. This accounts for general notions of investment strategies, asset pricing techniques, Bayesian statistics and State Space modelling. Then, the documents presents the model for two different amount of financial assets. Once data are presented and analysed, the results of the filtering recursion and the discussion of the performances conclude the document. Proposal for further studies are written after the document conclusion.

2 Active and Passive Investment

People invest in financial assets for different reasons but with the same goal of increasing the value of the initial invested capital. Investors can be generally separated into two categories: active and passive investors. Active investors believe that financial markets are not completely efficient, and try to extrapolate returns from those inefficiencies. By studying market anomalies, hedge fund managers undertake an active position in the market and trade against those anomalies. For example, if a company is believed to present over-stated earnings or flawed business plan which does not seem suitable to create profits, an informed hedge fund manager can undertake a short position on the underlying. The investment position is based on the idea that the real financial situation of the company is not reflected in the price, and the hedge fund manager believes the price will burst soon. While active investors focus on seeking price anomalies, passive investors firmly believe in the Efficient Market Hypothesis developed by Fama(1970). These investors believe that prices carry all the available market information and that they will never be able to extrapolate positive reward from actively seeking market inefficiencies. In this sense, passive investors rely on the basic argument of portfolio diversification. By creating portfolios of well diversified stocks, passive investors will plug their savings in mean-variance optimal portfolios and risk free assets.

Passive investment strategies are proper of investors placing their resources in well-diversified portfolios of stocks. Passive investors believe financial markets to be efficient. Market prices fully reflect the available information so that it is not possible to trade against mispriced stocks. Stock returns develop as stationary stochastic processes for which no investor can develop profitable predictions of future price developments. A passive investor creates a mean-variance optimal portfolio and gets rewarded by the equity risk premium over the risk-free asset. A mean-variance optimal portfolio refers to the Modern Portfolio theory introduced by Markowitz(1952). Investors are assumed to be risk-adverse and invest in different stocks by maximizing the expected return for a given level of risk. It formalizes the concept of diversification by showing how the portfolio volatility decreases by including a higher number of assets and the concept of risk-return profile of an investor. This represents the foundation of passive investment and many economists formalized relaxed version of the MV-efficient portfolio, like the basic model of long-term investments for multi-period investments of Samuelson(1969) and Merton(1969,1971). The fundamen-

tal critic to the basic MV-framework is that the model simplifies the real world decision of investor in asset allocation. In the MV-framework, the investors share the same beliefs on expectation, variance and correlation of financial assets. Relaxing this assumption leads to a more realistic representation of the real financial world and it introduces possible market anomalies.

Active investors believe market to be inefficient. They focus on trying to beat the market and get reward by trading against market inefficiencies. Active investment has been object of controversial studies and opinions. Sharpe(1991) focused on the relationship between returns and fees, and concluded that if the market return is calculated as the average between passive and active investments, passive investment return will always be higher than the active investment due to the higher level of fees for the latter strategy. Pedersen(2016) points out that the arithmetic of Sharpe does not hold because he does not consider the difference between the difference between informed and uninformed managers. Passive investors need to rebalance their positions if the market portfolio composition changes, and Sharpe assumed they can trade for free. In real world, this is does not hold due to market frictions and capital structuring decisions like IPOs and share repurchases. Academia common perception of active investment during last century was that active mutual funds have no skills and are not able to create substantial returns over the market, stated in Jensen (1968) and Fama(1970). During the last two decades, studies showed how some managers have been able to pick well performing stocks and create superior alpha performance, as pointed out in Fama(2010), Kosowski(2006), Kosowski(2007), Jagannathan(2010).

Even if agents are assumed to be rationale and to develop decision-making processes based a pure utility criteria, real word decisions suffer by biases. In 2017, the Nobel prize in economic sciences has been awarded to professor Richard H. Thaler for his contributions to behavioural economics. Central part of Thaler studies has been the incorporation realistic assumptions into decision processes of individuals. In his findings, he highlights how a human being is not completely a homo oeconomicus, and that his decisions are not completely rationale. If decisions are not completely rational, it means they contain errors. If these errors are applied to financial decisions having effects to the financial markets, they can have positive or negative effects on prices, but since these actions are not rational it becomes difficult (if not impossible) to understand them. In financial markets, these anomalies are defined market anomalies. Market anomalies are defined events moving away

prices from their fair value. Fair value prices can be defined as efficient-market prices, in the sense of Fama(1970). However, empirical data shows how market can be defined as efficiently inefficient market[1]. Markets are enough efficient to discourage uninformed asset managers to undertake active investment strategies, but also enough inefficient to permit able investment managers with superior information to beat the market and create positive alpha performance. Informed managers are able to beat the market, which means exploit market inefficiencies to profit. In Fama framework, this profit should always be zero and investors should not try to catch it since developing these active investment strategies carry opportunity costs. During the last two decade, the market efficiency has been heavily discussed. While promoters of efficient prices still believes market to be fully efficient, it seems that informed investors are able to beat the market and gain active investment profits, but the markets are efficient enough to discourage uninformed and unskilled managers to undertake active investment strategies.

[1]For reference, Pedersen(2016).

3 Asset Pricing Model

3.1 The Capital Asset Pricing Model

Optimal investment decision for an individual investor relies on various inputs. One of these is the expected return of the assets available for investors. In investment theory the expected return of a specific asset is referred to as the first moment of the asset's return distribution. This means that this quantity need to be estimated in a certain way. The most famous used tool for asset pricing is called Capital Asset Pricing Model, sometimes called just CAPM[2]. This asset pricing technique builds on the Markowitz mean-variance portfolio theory (MV). CAPM and MV are both essential models for the research area of financial markets and investment theory, and they both represent the basics of any kind of more sophisticated asset pricing model and portfolio strategy. To understand the basic concept of the CAPM is necessary to introduce few key concepts. The CAPM relies and refers to the market portfolio, which is defined as the portfolio of all the risky assets available in the financial market. This portfolio is created by incorporating all the assets in the market at a certain point in time. The value of the market portfolio is created as the sum of the single value of all the pooled assets. This has an important implications for the expected return and variance of this portfolio. Given a specific portfolio, its expected return and variance can be computed as:

$$\mathrm{E}[r] = \sum_{i}^{n} w_i r_i$$

$$\mathrm{Var}[r] = w^{'} \Sigma w$$

where

$$\sum_{i}^{n} w_i = 1$$

$$w_i = \frac{MV_i}{\sum_i^n MV_i}$$

From this specification of the first two moments of the return distribution of the market portfolio is possible to see how the portfolio weights are set depending on the market capitalization MV_i of every asset $i = 1, .., n$. With a total number of assets equal to n, the portfolio balances the weight so that the sum of all the weights is equal to 1.

[2]For reference, Treynor (1961), Sharpe (1964), Lintner (1965), Mossin (1966).

The Sharpe ratio of a market portfolio is always positive. By plotting the relationship of asset return r_i with the standard deviation σ_i it is possible to define another important concept for asset pricing, the Capital Market Line. This line connects the risk free r_f return/risk combination with the return/risk combination of the market portfolio. Since the risk free has zero risk by definition, its volatility is equal to zero. The slope of the equation can be easily calculated and is equal to:

$$\frac{E[r_m] - r_f}{\sigma[r_m]}$$

Given a universe of stocks and a risk free asset with non constrainted investors with mean-variance preferences, if they all share the same beliefs on the asset returns and the market is in equilibrium, the followings statements hold:

- the tangency portfolio is equal to the market portfolio of all risky assets
- each investor combines risk-free and market portfolio for the optimal investment strategy
- for any risk asset i, the following equation holds:

$$E[r_i] - r_f = \beta_i(E[r_m] - r_f)$$

where

$$\beta_i = \frac{Cov[r_i, r_m]}{Var[r_m]}$$

3.2 Implications of the Capital Asset Pricing Model

The main result of the CAPM is that β_i is the determinant of any positive risk premium of a certain asset. By looking at how beta is constructed, it is easily to see that it is just a ratio between two covariances. At the numerator, the covariance refers to the relationship between a specific asset and the market portfolio, while the denominator is the covariance of the market portfolio with itself, which represents its variance. Covariances are measures of variability, and in this case variability is the key aspect determining the risk premium of a specific asset. An asset i which has a positive covariance with the market portfolio should provide high returns with the market portfolio is high and low returns when the market return is already down. The risk premium of this asset will be higher. An asset j

with negative covariance will instead provide negative returns with positive returns of the market portfolio but positive returns when the market portfolio is in a depression state. The risk premium of the asset is expected to be lower. These two results come straight from the math above but can be also explained by economic theory. An asset which moves with the market will be unable to protect an investor in case of downturn and investors will not be interested in it. To make it interesting, it needs to provide a positive risk premium over the risk free rate. On the other side, investors will be attracted from an asset which provide returns in turmoil periods, with lower or even negative risk premium.

3.3 The Security Market Line

As described above, the result of the CAPM is the link between the risk premium of an asset return and its beta, described as the ratio between the asset covariance with the market portfolio and the variance of the latter. This relationship is not always respected because of the tight restrictions set at the beginning of the model and because of the complexity of financial markets and investors decisions. When the model does not hold, a new variable needs to be put inside this equation. For example, if the expected risk premium is higher than the one predicted from the model, the CAPM equation is:

$$E[r_i] - r_f > \beta_i(E[r_m] - r_f)$$

The market excess return of the asset i is higher than the premium predicted by the CAPM. If the opposite is true, the relationship is:

$$E[r_i] - r_f < \beta_i(E[r_m] - r_f)$$

In this case the CAPM predicted a higher risk premium than the market excess return. In both case the disequality can get back to an equation by inserting a new variable called α which refers to the misspricing of the expected return predicted by the CAPM. The equation goes:

$$E[r_i] - r_f = \alpha_i + \beta_i(E[r_m] - r_f)$$

From both academic and pratiction point of views, this new variable is often referred as the abnormal excess return of the asset i. In asset pricing theory, the main goal is to get a measure of the price of an asset by analysing the risk it carries and the possible gains

for picking such an investment strategy. If the real return of this asset is equal to the estimated one, there is no possibility of making an extra profit: the return an investor will get will be equal to the one she estimated before undertaking such investment. This concept is crucial for active investors like hedge fund managers which seek for abnormal profits. While passive investors rely on asset pricing techniques and choose to invest capitals in portfolios by believing in the efficiency of financial markets, active investors seek for market misspricings and arbitrage opportunities in order to achieve 'abnormal returns' or, in jargon terms, they look for 'alphas'.

The Security Market Line SML is the line linking risk free and market portfolio return in a Cartesian plane with CAPM Betas in the x-axis and expected returns in the y-axis. This representtion is the most important for CAPM because it directly show possible inefficiencies of the model, the so-called 'abnormal returns' or 'alphas'. If an asset is correctly priced, the couple $(\beta_i, E[r_i])$ lays over the SML, otherwise there is a misspricing. If the point plots below (above) the SML, the asset is overpriced (underpriced) according to the CAPM. The vertical distance between the point and the line is the alpha of the stock return.

3.4 Critics

The general idea of CAPM is that it is unable to price the return of assets and that the Security Market Line is flatter than predicted[3], leading to underestimated expected returns of low-beta stocks and overestimated expected returns of high-beta stocks, having a robust effect on the price of stocks. The final conclusion is that the model need to adjust for other explanatory variables in order to provide more robust estimations of expected returns.

3.5 Factor Models

Factor models assume that the covariation in asset returns comes from their linear dependence with common factors. The idea comes from the fact that economic variables affect all the financial assets and that the magnitudes of the effects change for every different asset. Since CAPM suffers by having just one explanatory variable, factor models can be useful to compare its result with the one of models with more explanatory variables.

[3]For reference Black(1972), Fama(1973).

3.6 Basic Properties of Factor Models

A one-factor model represents the general belief that a system as the following exist:

$$r_i = E[r_i] + \beta_i(F - E[F]) + e_i, \quad i = 1, 2, ..N$$

The model can be reformulated in terms of excess return as following:

$$r_i - r_f = E[r_i - r_f] + \beta_i(F - E[F]) + e_i, \quad i = 1, 2, ..N$$

By assuming the total number of stocks in the current universe equal to N, the one-factor model implies that the return of every stock is linked to its expected value plus an error $e_i \sim N(0, \sigma_{e_i})$ and a premium from the factor F. Properties are:

$$Cov[F, e_i] = 0, \quad i = 1, 2, ..N$$

$$Cov[e_i, e_j] = 0 \quad i, j = 1, 2, ..N \quad i \neq j$$

$$Cov[e_i, e_i] = \sigma_{e_i}$$

The factor F affecting all the N asset returns is orthogonal to errors and errors between different assets are orthogonal to each other. From these properties:

$$Cov[r_i, F] = \beta_i Var[F]$$

which defines β_i as

$$\beta_i = \frac{Cov[r_i, F]}{Var[F]}$$

This result should be familiar for the reader. It is not far from the general result of the CAPM described above. Now, the first two moments of the one-factor model can be defined as following:

$$E[r_i] = E[r_i] + \beta r_i + \beta_i(F - E[F])$$

$$Var[r_i] = \beta_i^2 Var[F] + Var[e_i]$$

$$Cov[r_i, r_j] = \beta_i \beta_j Var[F]$$

The results come from the combination of all the previous properties and the general

formula for calculating the variance of two random variables. It is easy to see how the covariances of every asset is defined on the basis of its relationship to the factor F. The return variance can be decomposed into two parts, the first part is the systematic risk while the second one is the idiosyncratic risk, and its covariance with another asset completely depend on the two betas and the factor variance $Var[F]$. Assets' sensitivity to the common factor F become the crucial variable for the parametrization of covariance matrices. A major advantage of factor model is the reduced number of parameters required to estimate the covariance matrix of a portfolio optimization problem. For example, a covariance matrix with N finacial assets requires estimating N variances and $\frac{N(N-1)}{2}$ covariances while imposing a factor model will require $2N + 1$ parameters.

If returns follow a one-factor model, the expected returns are of the form

$$E[r_i] = r_f + \beta_i P$$

where P is equal to the risk premium associated with the factor F. Combining this result with the one-factor model representation of an asset return in terms of excess return will lead to

$$r_i - r_f = \beta_i(r_p - r_f) + e_i$$

where

$$\beta_i = \frac{Cov[r_i, r_P]}{Var[r_P]}$$

3.7 The Single-Index Model

This one-factor model has been developed by Sharpe(1960) and it represents a basic one-factor model where the factor F is the return of a market portfolio. For estimation, a stock market index is taken as the benchmark for the market portfolio. The model is

$$r_i - r_f = \alpha_i + \beta_i(r_m - r_f) + e_i$$

For estimation, it is common to apply a least-square method like OLS in order to estimate the parameters.

3.8 Multi-Factor Models

In the case of a multifactor model, the return of an asset depends on how it varies respect to a certain number of factors plus an error term. The model goes as following:

$$r_i = E[r_i] + \boldsymbol{\beta}_i(\boldsymbol{F} - E[\boldsymbol{F}]) + e_i \quad i = 1, 2, ..., N$$

where

$$\boldsymbol{\beta}_i = \begin{pmatrix} \beta_{i,1} \\ \vdots \\ \beta_{i,J} \end{pmatrix} \quad \mathbf{F} = \begin{pmatrix} F_1 \\ \vdots \\ F_K \end{pmatrix} \quad E[\boldsymbol{F}] = \begin{pmatrix} E[F_1] \\ \vdots \\ E[F_K] \end{pmatrix}$$

3.9 The Fama-French models

The best known multifactor model is the Fama-French three factor model developed by Fama-French(1992). The model is an extension of the Single-Index factor model where the two economists added two new factors. The three factors are:

1. Market factor: the return of the market portfolio
2. Small-minus-Big SMB factor: the return of a portfolio of stocks going long on companies with small market value and shorting companies with high market value
3. High-minus-Low HML factor: the return of a portfolio of stocks going long on companies with high book-to-market B/M value and shorting the stocks with low B/M

The rationale behind these two new factors is based on economic theory and empirical results. Many studies showed that companies with lower market capitalization are able to produce higher return than companies with high level of equity. The main reason for the value factor relies on the intrinsic value of a stock and on behavioural biases of investors.

When a stock has a high B/M, it means that its book value is high if compared with its market value, and opposite for low B/M. The main argument relies on the fact that a stock with high B/M is said to be cheap since its intrinsic value is high if compared with the market value. Stock price is volatile and suffers by behavioural biases caused by herd behaviour and uninformed investors, while the intrinsic value is a more stable measure of the real value of a stock. The investment strategy which focuses on this relationship between book and market value of equity is called value investing. Rosenberg Reid and Lanstein(1985) showed the value effect. Stocks of companies with high book to market

have higher average returns than predicted by CAPM while low book to market have lower return than the one suggested by betas.

Similarly, Banz(1981) shows the size effect and how empirical average returns on small size listed companies are higher than the predicted returns from CAPM. On another formulation of the model, Fama and French add two more factors for the Fama-French five factors model. These two factors are:

1. Robust-minus-Weak RMW factor: the return of a portfolio of stocks with long position on the most profitable stocks and short positions on the least profitable stocks

2. Conservative-minus-Aggressive factor: the return of a portfolio of stocks with long position in companies investing conservatively and shorting companies with more aggressive investment policies

4 Bayesian Statistics

The Bayesian approach differs from the classic frequentist approach by considering the concept of probability from another perspective. The focus of Bayesian statistics relies on the Bayes rule and conditional probability properties. Let's define two events, A and B, and define the joint probability of these two event by:

$$P(A \cap B) = P(A|B)P(B)$$

where $P(A|B)$ is the conditional probability of event A conditioned on event B and $P(B)$ is the marginal probability of event VB. By basic properties of probability, the joing probability can also be written as:

$$P(A \cap B) = P(B|A)P(A)$$

From this equality, it is possible to define the Bayes rule. The Bayes rule develops in the following formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

This concept is simple but still powerful, and it represents the basic idea behind Bayesian statistics. Let's make a simple example. Suppose a biologist is interested in analysing the impact of the steady increase of temperatures on the ozone depletion. The model will contain some measurements, for example the degrees in various cities in a particular period of time, but it will also contain an unknown parameter, for example the impact of electricity usage on temperature rise. While the biologist can measure the past temperatures, it cannot directly evaluate the impact of electricity usage impact. Let's define B as the observed event and A as the event of interest. The researcher can sample a total number of temperatures and model this sample with a certain probability law. Let's define Y as a random variable modelling the observed temperatures and θ an unknown parameter, which the research need to estimated. The Bayes rule goes as following:

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)}$$

where $\pi(y)$ refers to the density function of Y and

1. $\pi(y)$ is the marginal density of the observed temperatures

2. $\pi(\theta)$ is called prior distribution

3. $\pi(y|\theta)$ is the conditional distribution of the observed values

For a certain realization of the random variable $Y = y$, $\pi(y)$ represents the marginal density of the observed temperatures, that is the unconditioned probability law governing this phenomena. It is possible to retrieve marginal probabilities by integrating out the conditional probability for the conditioned variables. In practice, the marginal probability can be computed as:

$$\pi(y) = \int \pi(y|\theta)\pi(\theta)d\theta$$

The prior density $\pi(\theta)$ represents the unconditioned probability law governing the unknown parameter. As concerned from the model, this is related to an initial hypothesis or idea about the density structure of this parameter. It is common to utilize preexisting data or information to set an initial hypothesis for this distribution. The conditional distribution $\pi(y|\theta)$ represents an additional probability law set by the researcher given the unknown parameter θ. The so-called "Bayesian learning process" relies on this basic rule and to these conditional probabilities. In this simple example, the learning process concerns studying the conditional distribution of the unknown given a specific observation of the variable Y by taking advantage of the conditional probabilities set by the researcher and by the Bayes rule.

Bayesian statistics can be applied to various sciences, and economics is one of these. In particular, Bayesian statistics is a useful alternative for time series analysis. The focus of this important area of statistical studies relies on the study of how random variables develop over time and how to forecast their values. In a Bayesian framework, the Bayes rule and its properties can be applied to update and adjust the knowledge of the probabilistic structure of θ_t and $_t$. Let's assume the biologist decides to specify the observed temperature over time as a time series. The Bayes rule becomes:

$$\pi(\theta|y_{1:t}) = \frac{\pi(y_{1:t}|\theta)\pi(\theta)}{\pi(y_{1:n})}$$

Where $\pi(y_{1:t})$ is the joint distribution of $y_1, ...y_t$.

The major change is about the probability distribution of the random variable Y_t. Now, the biologist needs to define the density function of all the observed temperatures for the whole observed period $1, ..., t$, which represents the time when the researchers develops the

model. This is a joint probability distribution, it represents the probability structure of a set of t random variables where $Y_1 = y_1$, $Y_2 = y_2$,.., $Y_t = y_t$, where upper case letters refer to random variables while lower case letters refer to specific realization of variables. Striclty speaking, these t variables are no more "random" since time is passed and the biologist has been able to observed them. The joint probability refers to the odds of the having these t variables Y_i with these specific realizations y_i, for $i = 1, .., t$. The issue here is how to model the structure of these joint probabilities. A basic property of statistics called 'independence' states that if an event A and an event B are independent, the joint probability of these two events is:

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

It is easy to understand the point of this rule. If A and B are independent, it means that knowing the probability structure of event A will not help to understand the probability structure of B, and viceversa. This turns to:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

This simplifies the probability framework of joint probabilities. In theory if events are independent, their joint density function is represented by the product of the marginal density functions. In practice, evaluating independency between random variables is still a major area of study of statistics and a problem that every statisticians face during studies. Assuming independency is a simple assumption which facilitates model estimation and calculation.

A property of Dynamic Linear Models refers to the dependence structure of $Y_1, ..., Y_n$ by saying that these random variables are conditional independent and identically distributed given θ. Given this characteristic, the conditional joint probability can be computed as:

$$\pi(y_{1:t}|\theta) = \prod_{i=1}^{t} \pi(y_i|\theta)$$

By combining the previous formula for the Bayesian rule with this property of observables in DLMs, it comes that:

$$\pi(\theta|y_{1:t}) = \frac{\pi(y_{1:t}|\theta)\pi(\theta)}{\pi(y_{1:n})} = \frac{\prod_{i=1}^{t}\pi(y_i|\theta)\pi(\theta)}{\pi(y_{1:n})}$$

This shows the argument for the recursion in DLMs and in particular for the filtering procedure. The posterior density of the parameter θ can be computed recursively since the conditional independence of the observables Y_i. In Bayesian statistics is used to described and write posterior densities with a special symbol which means "proportional to". This a way of eliminating all the density elements which are independent from the interest variable. For example:

$$\pi(\theta|y_{1:t}) \propto \prod_{i=1}^{t}\pi(y_i|\theta)\pi(\theta)$$

The missing element is the joint distribution of the observables. This is independent of any realization of the unknown parameter and so it is common to say that it represents a mere proportional factor for the final posterior density. The conditional independence property facilitates the computations of the posterior density function. The learning process for the posterior density only requires the information at the previous state in time. More specifically, at time $t-1$ the prior for time t is:

$$\pi(\theta|y_{1:t-1}) \propto \prod_{i=1}^{t-1}\pi(y_i|\theta)\pi(\theta)$$

The likelihood is $\pi(y_t|y_{1:t-1},\theta)$ but can be simplified to $\pi(y_t|\theta)$ due to the conditional independence of Y_i. The Bayes rule gives:

$$\pi(\theta|y_{1:t-1},y_n) \propto \prod_{i=1}^{t-1}\pi(y_i|\theta)\pi(y_n)\pi(\theta)$$

The conditional independence property simplifies the estimation of models working with conditional independence by reducing the computational workload. At every point in time when a new observation is retrived, the new posterior density just requires the *i.i.d.* probability densities of the singular Y_i conditioned to the unknown parameter θ.

4.1 Prior and Model Specification

In order to start the learning process, the researcher needs to specify the model and to set the prior for the interested event. This means setting $\pi(y|\theta)$ and $\pi(\theta)$. The first density is given and decided by the model of the researcher, the second density function should refer

to a genuine and honest espression of the phenomena the researcher wants to study. Here, it is common to utilize other studies as starting point for evaluating a model. A major concern for this distribution is about computational issues which can be in part avoided by using conjugate priors. The choice of a prior refers to setting both $\pi(y|\theta)$ and $\pi(\theta)$ and the latter is said conjugate for a model specified by the former if the density family representing $\pi(\theta)$ is the same also for the posterior density $\pi(\theta|y)$. For example, setting a Gaussian density function for $\pi(\theta)$ triggers a Gaussian posterior density function, which means that the Gaussian family is conjugate to the model specified by $\pi(y|\theta)$. To check if a prior is conjugate for the desidered model, it is required to check the resulting likelihood function of the posterior density resulting from the Bayes rule.

4.2 Bayesian Inference for Unknown Quantities

The two main approaches for estimating unknown parameters are maximum likelihood estimation and Bayesian inference. The former is the classic approach for parameter estimation of frequentist statistics. After observing a set of n random variables distributed with a specific probability law, for example Gaussian probability distribution, the maximum likelihood estimation consists in creating the joint probability distribution of the observed variables and optimizing the function. The optimization procedure will maximize the joint probability: strictly speaking, this means to maximize the chances of observing the n observed values for the n random variables. The optimal argument of the maximum will represent the unknown parameters of the model. The latter approach is the Bayesian approach for parameter estimation. After imposing an initial idea of the unknown values, the Bayesian learning procedure and a sampling algorhythm will update the estimation of the unknown parameters. For Bayesian statistics, unknown parameters are considered as random quantities, and every time a new observation comes in, the algorithm updates the best guess for these random quantities. The most important drawdown of Bayesian inference is the computational issue underlying the posterior densities calculations. Bayesian inference is computationally demanding if models are complex but Markov Chain Monte Carlo methods help to reduce the intensive task of posterior distribution calculations.

4.3 Markov Chain Monte Carlo

In Bayesian statistics an unknown parameter is treated as an unknown quantity. To gather information about the quantity, it is common to estimate the posterior density function of

the unknown by applying the Bayes rule and Bayesian learning procedure above described. Unfortunately, posterior distributions are not always analytically treatable and numerical methods need to be applied. A way to overcome this problem is to rely on the law of large number underpinning the concept of Monte Carlo simulation. For example, if a model requires an estimate of a specific unknown random variable ψ, it can be simulated by running a Monte Carlo simulation with significant number of simulations. Then, the sample mean of the simulated sample will be referred to as an approximated value of the desider variable ψ. This works also for functions of ψ, like:

$$E_\pi(g(\psi)) \approx N^{-1} \sum_{j=1}^{N} g(\psi_j) \tag{1}$$

where $\pi(.)$ is the posterior distribution from which the sample is generated and N refers to the number of simulations. The problem relies on the statistical properties of the sample. In theory, to have a reliable and independent estimate of a parameter it is required to collect an i.i.d. sample. In practise, i.i.d. assumption is strong and sometimes unfeasible. This turns out to dependent sample, rather than independent. As described above, Markov chains are stochastic processes where each state of the chain depends only on the previous state but it is independent of the others. This property of Markov chains is called memorylessness and can be applied in the sampling procedure without damaging the law of large number property of Monte Carlo simulations. By combining these two concepts, this kind of sampling procedures are called Markov Chain Monte Carlo (MCMC).

4.4 Gibbs Sampler

Markov Chain Monte Carlo methods are powerful tools which combine the law of large numbers and the memoryless property of Markov chains. Among the various methods, the Gibbs sampler is widely applied for sampling the posterior distribution of a multivariate random variables. The idea of Gibbs sampling relies on basic assumptions of MCMC and leads to a robust sample of the multivariate posterior distribution. Suppose the unknown quantity is still ψ and it is a multivariate random variable. This means that $\psi = [\psi^{(1)}, \psi^{(2)}, .., \psi^{(m)}]$. If ψ evolves as a Markov chain, it respects the memorylessness principle. By imposing an initial arbitrary point of ψ, the Gibbs sampler will draw each of the m values one by one, by conditioning every draw just on the previous draw. In this way, the final sample will respect the memorylessness property of Markov chain and

will generate a path dependent multivariate sample which can be used to create a proper guess for the unknown quantity ψ by Monte Carlo method.

4.5 Diagnostics

After running the simulation for the Markov chain, it is important to check if it converged. The output of the Gibbs sampler needs to reach the stationary distribution and this can be assessed by checking the autocorrelation function of parameters or by checking if the sample mean is stable once a substantial number of simulations are cut as burn in. The burn-in part of the simulated value is intended as a total number of simulations required by the Markov-chain to reach its equilibrium distribution. Strictly speaking, after plugging some prior hypothesis for the unknown parameters, the Markov chain needs to wash out the information of a possible poor prior. By burning a part of the simulated sample, the simulation is able to assess a better Monte Carlo estimate of unknown parameters.

5 State Space Model

5.1 Definition

A state space model is a mathematical model representing a particular physical system. It links a set of input, output and state variables which are able to specify and define the system over time. The model represents the law underpinning the evolution of the system and allows to study how the input and state variables affect the output values. In control engineering, input and output variables are related by differential equations represented by matrices and vector spaces. Even tough this is a tool proper of control engineering, it has been applied to many other sciences. In economics, the analysis of how data develop over time has been an area of great significance and interest for researchers, policy makers, investors and many other institutions. State space models can be used to model univariate and multivariate time series and it lends itself a elegant tool for econometricians. More specifically, econometricians need powerful and flexible tools able to model and understand data, whether they are data on GDP of a major economy or irregular economic data from an emerging market. A first major credit to SSM concern nonstationary time series. The classic tools for time series analysis are the ARMA models estimated by the Box-Jenkins method where nonstationary time series need a preliminary transformation. State space models can be defined in a specific way in order to achieve the same results of an ARMA model. In this case the model is defined as the state space model representation of this particular ARMA model. The power of SSM is that it can properly handle nonstationary data and show possible instability in the mean level and in the variance of a specific series, analyze structural breaks or even proceed with missing observations.

5.2 Properties of a State Space Model

First step is to consider a specific series of data, in this case a time series $(Y_t)_{t\geq 1}$. For simplicity, it is usual to define the dependence of the observations over time as Markovian. The reason lies within the elegant proprieties of this process. $(Y_t)_{t\geq 1}$ is a Markov chain if and only if for any $t > 1$,

$$\pi(y_t|y_{1:t-1}) = \pi(y_t|y_{t-1})$$

A main characteristic of a Markov chain is that the probability of each event only depends on the probability of the previous event. This is referred to as the 'memoryless' property

of a Markov process. In this setup, it can be argue that the conditional probability of the random variable Y_t conditioned on all the past realizations $(y_1, .., y_{t-1})$ is equal to the conditional probability of the random variable Y_t conditioned on the last value y_{t-1}. Stricly speaking, this means that the information up to time $t-1$ is the same information as the one carried just at time $t-1$. This characteristic simplifies the definition of the joint probability distribution of the observable time series (Y_t) and allows it to be treatable, even if the Markovian structure is not always able to properly describe the evolution of the underlying system. A State space model is defined after two time series: $(Y_t : t = 1, 2..)$ representing the observable time series of data and $(\theta_t : t = 0, 1, ..)$ representing the unobservable time series of state variables. The two main characteristics of a SSM are:

- θ_t is a Markov chain
- Conditionally to θ_t, (Y_t) dependes only on θ_t and not on other values of Y_t at any other point in time

These two proprieties mean that the SSM is fully defined by three density function:

1. $\pi(\theta_0)$: Initial distribution of the state variable θ_t
2. $\pi(\theta_t|\theta_{t-1})$: Conditional distribution of the state variable with the previous value
3. $\pi(y_t|\theta_t)$: Conditional density of the observed variable with the state variable

These results come from the structure of the SSM. At any point in time, the observed variable Y_t depends only on the state variable θ_t and not on any future or past values of itself. In term of conditional probability distribution, this means that the probability distribution of Y_t conditioned to all its past realizations $(Y_1, ..., Y_{t-1})$ and $(\theta_0, ..., \theta_{t-1})$ reduces to $\pi(y_t|\theta_t)$.

A general SSM can be defined by two equations. The first equation is called 'observation equation' and concerns the evolution of the observable variable Y_t. The second equation is called 'state equation' and explains the structure of the unobservable variable θ_t, also called state variable. The general model can be written as following:

$$Y_t = h_t(\theta_t, v_t)$$

$$\theta_t = g_t(\theta_{t-1}, w_t)$$

In this very general formulation, the system is defined by two functions h_t and g_t. These two functions can be defined as linear or nonlinear and it is important to properly fit the evolution scheme of the two variables of the system, Y_t and θ_t. However, the nonlinearity of the system dramatically increases the computational requirement of the fitting recursion of the model.

5.3 State Estimation

The main point of state space modelling is to study the evolution of an observable variable with the help of a defined system. This system lies on the observation and state equation linking observable and unobservable variables. In a more mathematical point of view, the word 'state space' refers to the finite-dimensional vector space over the set of real numbers R. Every single state of the system can be represented as a vector within the collection of vectors forming the vector space. Since the vectors are all defined over the set of real numbers, the state space is an Euclidean Space. For a defined state space model, the aim is to combine all the properties of the model in order to make inference on the unobserved variable and to forecast the evoulution of the system. It is also possible to look backward to its evolution, by starting from a certain point in time and going back toward the starting point. In state space modelling, it is usual to distinguish between three type of state estimation: filtering, forecasting and smoothing.

5.4 Dynamic Linear Model

Dynamic linear models turn to be very powerful tools for forecasting and many other statistical applications on the basis of the Markovian property of the variables and statistical properties of the Gaussian distribution. A specific and important class of SSM is represented by the Dynamic Linear Model class, also called Gaussian linear state space model. The main characteristic of this class concerns the normality of the random vectors and the initial distribution of the state variable θ_t. More specifically, a general time-variant DLM can be written as following:

$$\underset{mx1}{Y_t} = \underset{(m\times p)(p\times 1)}{F_t}\ \underset{}{\theta_t} + \underset{(m\times 1)}{v_t} \qquad v_t \sim N(0, V_t)$$

$$\underset{(p\times 1)}{\theta_t} = \underset{(p\times p)(p\times 1)}{G_t}\ \theta_{t-1} + \underset{(p\times 1)}{w_t} \qquad w_t \sim N(0, W_t)$$

where m referes to the total number of time series treated in the model and p refers to

the total number of unobservable variables in the model. The two series of random errors are Gaussian white noise vectors and are assumed to be uncorrelated with each other and with any other variables of the model. In particular, this means that the covariance matrices of the random errors can be defined as following:

$$E(v_t v_s^T) = \begin{cases} V_t & \text{if } t = s \\ 0 & \text{otherwise} \end{cases}$$

$$E(w_t w_s^T) = \begin{cases} W_t & \text{if } t = s \\ 0 & \text{otherwise} \end{cases}$$

The matrices F_t, G_t define the evolution of the system. The first matrix links the state variable θ_t with the observation variable Y_t, while the second matrix links the state variable θ_t with its first lagged value. For simplicity, the matrices are assumed to be known, which means that the system evolution is known, but this is not always the case. If treated as unknowns, these two matrices can be estimated from data.

5.5 Kalman Filter

A state space model is defined by two types of variables, where the state vector is made by unobservable quantities. As defined before, the characteristics of the SSM and the model specification allow to infer the values of these unobservable quantities. The starting point for the study of a DLM is to set the initial distribution of the state vector, which is defined as:

$$\theta_0 \sim N(m_0, C_0)$$

Once this distribution is parametrized, the model can recursively compute the forecast of the state vector and the observation vector at any point in time. Let's suppose that the time span of interest goes from time $t = 1$ to $t = T$. At any point in time, the model can be used to calculate the one-step-ahead predictive distribution of the state vector θ_t and the observation vector Y_t. Once these two distributions are retrieved, the model is able to update the inference of the state vector. This is the filtering recursion: at any point in time, the recursion computes the one-step-ahead distributions of Y_t and θ_t, and updates the inference on the state vector once the new observation is available. This is the main aim of SSM: creating a forecast for an unobserved vector of variables and adjust the

forecast once a new set of observations are available. Once the system is set, this becomes and extremely powerful tool. More specifically, the filtering recursion requires:

1. $\pi(\theta_t|y_{1:t-1})$: the one-step-ahead predictive distribution of the state vector conditioned on all the past values of the observable vector

2. $\pi(y_t|y_{1:t-1})$: the one-step-ahead predictive distribution of the observable vector conditioned on all the past values of the observable vector

The result of the filtering recursion is the filtering distribution of the state vector $\pi(\theta_t|y_{1:t})$.

These three density functions are defined as following:

$$\pi(\theta_t|y_{1:t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|y_{1:t-1})d\theta_{t-1}$$

$$\pi(y_t|y_{1:t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})d\theta_t$$

$$\pi(\theta_t|y_{1:t-1}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})}{\pi(y_t|_{1:t-1})}$$

The derivation of these three density functions relies on the Markovian property of the state vector, the conditional independence of the observable vector and the Bayes rule for conditional probability.

5.6 Kalman Filter for Dynamic Linear Models

One of the main reason of the popularity of dynamic linear models is the computational ease granted by the powerful properties of the Gaussian distribution. In particular, the computations for the probability distributions described above start to be computationally heavy once the dimension of the vectors and the time span increase. By talking about computational requirement, the filtering recursion of dynamic linear models are considerably simplified because of the Gaussian property of the model. In particular, the solution to the filtering problem of DLMs is given by the Kalman filter. The proof of the result relies on the general properties of the multivariate Gaussian distribuion and on the properties of linear state space models. Let's define again a classic DLM:

$$Y_t = F_t\theta_t + v_t \quad v_t \sim N(0, V_t)$$
$$\theta_t = G_t\theta_{t-1} + w_t \quad w_t \sim N(0, W_t)$$

By following the initial definition of the starting distribution of the state vector $\theta_0 \sim N(m_0, C_0)$, it is easy to see that:

$$\theta_{t-1} \sim N(m_{t-1}, C_{t-1})$$

As described above, the filtering procedure requires three different density functions. The main advantage of the DLM properties is that the three density functions are defined in closed form by the first two moments of the their distributions. Since the model contains Gaussian random variables, the first and the second moment of the distribution are sufficient statistics. Let's suppose to have information until time $t-1$. This means that the last observation is represented by the realized vector of observed values y_{t-1}. The first step for the filtering recursion is to forecast the one-step-ahead distributions of the two variables of interest, the observed variable Y_t and the unobserved state variable θ_t. In this case, the Gaussian property of the DLM helps the filtering recursion by allowing the two distributions to be fully specified with their first two moments. In particular, this leads to the following distribution:

$$\theta_{t|t-1} \sim N(a_t, R_t)$$
$$a_t = E(\theta_t|y_{1:t-1}) = G_t m_{t-1}$$
$$R_t = Var(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t^{'} + W_t$$

$$y_{t|t-1} \sim N(f_t, Q_t)$$
$$f_t = E(y_t|y_{1:t-1}) = F_t a_t$$
$$Q_t = Var(y_t|y_{1:t-1}) = F_t R_{t-1} F_t^{'} + V_t$$

This elegant solution relies on the combination of the Markovian property of the state vector θ_t, the conditional independence of Y_t and the Gaussian property of the DLM. More specifically, the joint distribution of $(\theta_0, \theta_1, .., \theta_t, Y_1, Y_2, ..Y_t)$ is defined as:

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \prod_{j=1}^{t} \pi(\theta_j|\theta_{j-1}) \pi(y_j|\theta_j)$$

This is a multivariate normal density function representing the joing distribution of the all the state variables and all the observed variables until time t. [talk about integrate

out to get marginal distributions] The one-step-ahead forecast distribution of the state vector is the best forecast of the the unobserved variables for the next step in the time span, in this case time t. The filtering recursion is not done yet, since to conclude the recursion from time $t-1$ to time t the model needs to incorporate the new observation of the variable Y_t at time t and update the prediction of the unobserved state variables for time t. Stricly speaking, once the new observation Y_t kicks in, the recursion is able to produce the filtered distribution of θ_t. The filtered distribution is defined as following:

$$\theta_{t|y_{1:t}} \sim N(m_t, C_t)$$
$$m_t = E(\theta_t|y_{1:t}) = a_t + R_t F_t^{'} Q_t^{-1}(Y_t - F_t a_t)$$
$$C_t = Var(\theta_t|y_{1:t}) = R_t - R_t F_t^{'} Q_t^{-1} F_t R_t$$

By summing up, the Kalman filter is a recursion algorithm which allows to recursively project the prediction distribution of observed/unobserved variables and update the projection of the unobserved variables once a new observation kicks in. The estimation-correction philosophy of the Kalman filter recursion is the same as the linear inference problem of a Bayesian system.

5.7 Kalman Smoother for Dynamic Linear Models

The most attractive feature of the Kalman filter on DLMs is the sequential estimation of forecasted and filtered values for observable variables and unobservable states. At every point in time, the filter elaborates the one-step forward forecast of the observables and unobservables, $y_{t|t-1}$ and $\theta_{t|t-1}$. When the new observation y_t is retrived, the forecasted values are corrected with the Kalman gain matrix, leading to the noise-free estimates of the two categories of variables, the so-called filtered values. Another interesting feature of the filter activates once all the observations are gathered. When time time variable t hits the end of the time series, T, it is possible to compute a backward estimate of the unobservable states. This procedure is called smoothing, because it compute the noise-free estimate of the unobservables by conditioning the expectation on joint probability distribution of the entire sample of observation. For a DLM described as above, the Kalman smoother recursion starts at $t = T$ and computes the conditional distributions of the states θ_t given the entire set of information for y_t. The smoothing distribution is

defined as following:

$$\theta_{t|y_{1:T}} \sim N(h_t, H_t)$$
$$h_t = E(\theta_t|y_{1:T}) = m_t + C_t G'_{t+1t} R^{-1}_{t+1}(h_{t+1} - a_{t+1})$$
$$H_t = Var(\theta_t|y_{1:T}) = C_t - C_t G'_{t+1} R^{-1}_{t+1}(R_{t+1} - H_{t+1}) R^{-1}_{t+1} G_{t+1} C_t$$

Inside this formulas it can be seen the filtered estimates for time t and also the predictive estimates for time t. It is important to highlight that since this recursive algorhythm runs backwards, it requires a starting point from the last observation at time T. By comparing the first smoothed estimate and last filtered estimate for time $t = T$, it is easy to see that these quantities are the same, so that:

$$h_T = m_T$$
$$H_T = C_T$$

6 Model

6.1 Seemingly Unrelated Regression

In statistics, linear regression is a tool used to model relationships between a dependent variable y and one or more independent variables $x_1, .., x_n$. A large number of different estimation methods have been developed over time and they all differ in computational procedures and theoretical assumptions. In its most common estimation, a univariate or multivariate linear regression is estimated by applying a least square method called Ordinary Least Squares method. The procedure for OLS estimation is to minimize the sum of squared residuals SSE and create closed form solution for the unknown parameters. For example, the basic asset pricing tool of the CAPM is a simple univariate regression model estimated with OLS technique. The dependent variable is represented by the interested stock excess return time series, and the unique independent variable is the market portfolio return time series. By estimating the β^M, the model is calculating a ratio between two variability measures. In particular, $\beta^M = \frac{Cov(r_i,r_m)}{Var(r_m)}$ represents a ratio between the a measurement of the joint variability of stock i with the market portfolio, and the total variability of the market portfolio. In this setting, the economic intuition of the CAPM relies on the correlation between the stock and the market portfolio. In case more stocks are observed, the analysis will require to compute a regression model as CAPM for every stock. In this case, the regression models can be estimated equation by equation in a joint setting of OLS. In particular, this joint system assumes orthogonality among error series. This means that cross-equation error terms are uncorrelated, and the covariance matrix of the system error terms is a diagonal matrix. In order to relax this assumption and model cross-equation correlation among error terms, Seemingly unrelated model can be applied. This category represents an extension of the basic system of multivariate regressions since it allows the error terms to be correlated and remove the orthogonality principle. The gain in estimation is reflected by a superior and more informative structure of the covariance matrix. This category of econometrical models refers to the Generalized least squares methods (GLS), which represents the techniques for estimating unknown parameters in a linear setting where residuals are correlated among each other. The classic reasoning for GLS is that neither ordinary nor weighted least squares are able to produce statistically efficient inferences on the unknown parameters. A further analysis can be done with simultaneous equations model. This category of models refer to a set of linear simultaneous

equations modelled by Generalized Method of Moments (GMM) of Instrumental Variable (IV). Modelling error term correlation among different OLS regressions seems to achieve a higher specification of the underlying system. A standard linear regression model for modelling the relationships between two time series is:

$$Y_t = \alpha + \beta X_t + e_t$$

The two variables have only a temporal dimension since their values change over time. The regression is equivalent to:

$$\underset{(nx1)}{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}} = \underset{(n\times 2)}{\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}} \underset{(2x1)}{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}} + \underset{(nx1)}{\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}}$$

The two coefficients for the regression are estimated by minimizing the sum of squared residuals. A more interesting case would be to consider dynamic regressions, where the coefficients are evaluated in their temporal evolution. The regression would be:

$$Y_t = \alpha_t + \beta_t X_t + e_t$$

6.2 Dynamic Linear Modelling of Seemingly Unrelated Regression

A solution for this problem can be found through dynamic linear modelling. The most common solution is to create two state equation, one for α and one for β, and model them as independent random walks. This means create two AR(1) processes with $\Phi = 1$ and assume the two processes to be orthogonal. The following notation is atipical since this DLM is based on an univariate observation variable, but its understanding will ease to understand more complex DLMs. This univariate DLM becomes:

$$y_t = \underset{(1\times 2)}{\begin{pmatrix} 1 & x_t \end{pmatrix}} \underset{(2\times 1)}{\begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix}} + e_t \qquad e_t \sim N(0, \sigma_e)$$

$$\underset{(2\times1)}{\begin{pmatrix}\alpha_t\\ \beta_t\end{pmatrix}} = \underset{(2\times2)}{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}} \underset{(2\times1)}{\begin{pmatrix}\alpha_{t-1}\\ \beta_{t-1}\end{pmatrix}} + \underset{(1x1)}{\begin{pmatrix}w_t^{\alpha}\\ w_t^{\beta}\end{pmatrix}} \qquad \begin{pmatrix}w_t^{\alpha}\\ w_t^{\beta}\end{pmatrix} \sim N(\mathbf{0}, W)$$

where

$$W = \begin{pmatrix}\sigma_{w^{\alpha}} & 0\\ 0 & \sigma_{w^{\beta}}\end{pmatrix}$$

The indexes for the variables help to have the correct intuition on how DLMs work. Every observation of the vector Y_t and the OLS information matrix X can be seen as a row by row slice of the temporal dimension of the observed dependent and independent variables. At every point in time, the variables are treated in the DLM in order to model the evolution of the state variables, α_t and β_t. At every point in time, the Kalman filter will forecast $y_{t+1|t}$, $\alpha_{t+1|t}$ and $\beta_{t+1|t}$[4]. Then, when the filtered values are estimated, the DLM will roll to the next observation period.

If the number of coefficients increase, nothing needs to be changed and this presented model still holds, but a new operator needs to be insert if the number of modelled observed time series. Suppose the considered SUR model has now two stock return time series. The SUR model for these two stocks becomes:

$$\underset{(2\times1)}{\begin{pmatrix}y_{1,t}\\ y_{2,t}\end{pmatrix}} = \Big(\underset{(2\times1)\times(2\times4)}{\begin{pmatrix}1 & r_t^M & r_t^S & r_t^H\end{pmatrix}} \otimes I_2 \Big) \underset{(8\times1)}{\begin{pmatrix}\alpha_t\\ \beta_t^M\\ \beta_t^S\\ \beta_t^H\end{pmatrix}} + \underset{(2x1)}{\begin{pmatrix}v_{1,t}\\ v_{2,t}\end{pmatrix}} \qquad \begin{pmatrix}v_{1,t}\\ v_{2,t}\end{pmatrix} \sim N(\mathbf{0}, V)$$

$$\underset{(8x1)}{\begin{pmatrix}\alpha_t\\ \beta_t^M\\ \beta_t^S\\ \beta_t^H\end{pmatrix}} = \left(\underset{(4\times2)\times(4\times2)}{\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}} \otimes I_2 \right) \underset{(8x1)}{\begin{pmatrix}\alpha_{t-1}\\ \beta_{t-1}^M\\ \beta_{t-1}^S\\ \beta_{t-1}^H\end{pmatrix}} + \underset{(8x1)}{\begin{pmatrix}w_t^{\alpha}\\ w_t^{\beta^M}\\ w_t^{\beta^S}\\ w_t^{\beta^H}\end{pmatrix}} \qquad \begin{pmatrix}w_t^{\alpha}\\ w_t^{\beta^M}\\ w_t^{\beta^S}\\ w_t^{\beta^H}\end{pmatrix} \sim N(\mathbf{0}, W)$$

[4] $t+1|t$ and $t|t-1$ have the same meaning.

where

$$\alpha_t = \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix} \quad \beta_t^M = \begin{pmatrix} \beta_{1,t}^M \\ \beta_{2,t}^M \end{pmatrix} \quad \beta_t^S = \begin{pmatrix} \beta_{1,t}^S \\ \beta_{2,t}^S \end{pmatrix} \quad \beta_t^H = \begin{pmatrix} \beta_{1,t}^H \\ \beta_{2,t}^H \end{pmatrix}$$

$$w_t^\alpha = \begin{pmatrix} w_{1,t}^\alpha \\ w_{2,t}^\alpha \end{pmatrix} \quad w_t^{\beta^M} = \begin{pmatrix} w_{1,t}^{\beta^M} \\ w_{2,t}^{\beta^M} \end{pmatrix} \quad w_t^{\beta^S} = \begin{pmatrix} w_{1,t}^{\beta^S} \\ w_{2,t}^{\beta^S} \end{pmatrix} \quad w_t^{\beta^H} = \begin{pmatrix} w_{1,t}^{\beta^H} \\ w_{2,t}^{\beta^H} \end{pmatrix}$$

The Kronecker product $\otimes$ is a linear algebra operator which represents the generalization of the outer product from vectors to matrices. This operator is not equal to the matrix operator, which is a completely different operation. In particular:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{11}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{pmatrix}$$

The intuition for the Kronecker product relies on the basic relationship between dependent and independent variables in classic multivariate regression models. In such category of models, it is common to regress a vector of observations over a multivariate collection of regressors. The system described above is a collection of more regression models sharing the same independent variables. The Kronecker product combined with an identity matrix of order m activates the regressors for all the m regressions, where m refers to the number of regressions stack in the system.

In order to model a higher number of stock returns time series, a ten stocks DLM representation of the Fama French three factors regression is estimated. The model presented above will be extended in order to capture ten time series. This will lead to a significant increase of the dimension of the vectors and matrices required for the model to work. In particular, the model is:

$$\underset{(10\times1)}{\begin{pmatrix} y_{1,t} \\ \vdots \\ y_{10,t} \end{pmatrix}} = \Bigg(\underset{(10\times1)\times(10\times4)}{\begin{pmatrix} 1 & r_t^M & r_t^S & r_t^H \end{pmatrix}} \otimes I_{10} \Bigg) \underset{(40\times1)}{\begin{pmatrix} \alpha_t \\ \beta_t^M \\ \beta_t^S \\ \beta_t^H \end{pmatrix}} + \underset{(10\times1)}{\begin{pmatrix} v_{1,t} \\ \vdots \\ v_{10,t} \end{pmatrix}} \qquad \begin{pmatrix} v_{1,t} \\ \vdots \\ v_{10,t} \end{pmatrix} \sim N(0, V)$$

$$\underset{(40\times1)}{\begin{pmatrix}\alpha_t\\ \beta_t^M\\ \beta_t^S\\ \beta_t^H\end{pmatrix}} = \underset{(10\times4)\times(10\times4)}{\left(\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\otimes I_{10}\right)} \underset{(10\times4)}{\begin{pmatrix}\alpha_{t-1}\\ \beta_{t-1}^M\\ \beta_{t-1}^S\\ \beta_{t-1}^H\end{pmatrix}} + \underset{(40\times1)}{\begin{pmatrix}w_t^\alpha\\ w_t^{\beta^M}\\ w_t^{\beta^S}\\ w_t^{\beta^H}\end{pmatrix}} \qquad \begin{pmatrix}w_t^\alpha\\ w_t^{\beta^M}\\ w_t^{\beta^S}\\ w_t^{\beta^H}\end{pmatrix} \sim N(0,W)$$

where

$$\alpha_t = \begin{pmatrix}\alpha_{1,t}\\ \vdots\\ \alpha_{10,t}\end{pmatrix} \quad \beta_t^M = \begin{pmatrix}\beta_{1,t}^M\\ \vdots\\ \beta_{10,t}^M\end{pmatrix} \quad \beta_t^S = \begin{pmatrix}\beta_{1,t}^S\\ \vdots\\ \beta_{10,t}^S\end{pmatrix} \quad \beta_t^H = \begin{pmatrix}\beta_{1,t}^H\\ \vdots\\ \beta_{10,t}^H\end{pmatrix}$$

$$w_t^\alpha = \begin{pmatrix}w_{1,t}^\alpha\\ \vdots\\ w_{10,t}^\alpha\end{pmatrix} \quad w_t^{\beta^M} = \begin{pmatrix}w_{1,t}^{\beta^M}\\ \vdots\\ w_{10,t}^{\beta^M}\end{pmatrix} \quad w_t^{\beta^S} = \begin{pmatrix}w_{1,t}^{\beta^S}\\ \vdots\\ w_{10,t}^{\beta^S}\end{pmatrix} \quad w_t^{\beta^H} = \begin{pmatrix}w_{1,t}^{\beta^H}\\ \vdots\\ w_{10,t}^{\beta^H}\end{pmatrix}$$

The complete model can be found in the appendix. The same conditions stated for the initialization of the two stocks model are valid and holds also for ten stocks with the adjuments for the matrices dimension.

6.3 Kalman Filter Recursion

The Kalman filter recursion is an iterated algorhythm and it requires an arbitrary starting point for the first observation. The common practice is to start by assigning the initial state vector equal to a prior hypothesis of their values. This can be seen as the analyst's best guess of the initial values of the unobservable states. In order to express complete ignorance about the states, it is common to set the initial value of the state equation equal to a vector of zero, so that:

$$m_0 = \begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}$$

For the initial covariance matrix C_0, setting an initial value starts to be more involving.

As defined by Hamilton(1994), if the eigenvalues of the observation model matrix $F_t \otimes I_m$ are all inside the unit circle, the state equation θ_t is covariance-stationary. In this case, the covariance matrix of θ_0 is set as:

$$\text{vec}(C_0) = [I_{m\times m} - (G \otimes G)]^{-1}\text{vec}(W)$$

The problem in here is that the model treated in this document relies on an identity matrix state system matrix G. This means that this definition of the initial state covariance matrix is not suitable. To correct this issue, the initial matrix C_0 is set in order to express complete ignorance on the real covariance structure of the state variables. The final specification for starting the Kalman filter recursion is:

$$m_0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \qquad C_0 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

The formula presented above for the general definition of a Seemingly Unrelated Regression holds for the model. Since the database has been set up with daily observation, a series of numerical instability problems have been encountered. A major problem of the recursion has been connected to the positive definiteness and symmetry of iterated covariance matrices. A non-positive definiteness matrix is singular and cannot be inverted, and the Kalman recursion heavily relies on inverted matrices. To overcome this numerical issue, it is required to apply a sort of numerical optimization able to increase the numerical stability of the recursion. A robust algorhythm able to achieve this stability is the Singular Value Decomposition[5]. SVD is a matrix factorization which is here applied to deal with the vector space of the posterior covariance matrix, and with the orthonality issues arising from the numerical instability of the algorhythm. In short, the SVD helps to increase the stability of the numerical optimization and to reduce the possibility of creating singular covariances matrices. The posterior covariance matrix C_t defined through SVD is: $C_t^{-1} = F_t^{'}V^{-1}F_t + R_t^{-1}$ The final step is to invert the matrix.

[5]For reference, Wang(1992).

6.4 Markov Chain Monte Carlo for Multivariate Random Variables

The covariance matrices of observables and unobservables can be evaluated with Markov Chain Monte Carlo methods. This means modelling Bayesian inferences for unknown matrices. In case of univariate random variable, the basic assumption is to impose the prior of the unknown variance as an inverse-Gamma prior. In case of multivariate observations, the priors of variance-covariance matrices are modelled as inverse-Wishart priors[6], a multivariate estension of Gamma distribution. In particular, for the above described SUR model the Gibbs sampler works to draw a sample from estimating the posterior density function $\pi(\theta_{0:T}, \psi | y_{1:T})$

$$\Phi_0 = V^{-1} \quad \Phi_i = W_i^{-1} \quad \text{for} \quad i = 1, .., 4$$
$$W = \text{blockdiag}(W_1, W_2, W_3, W_4)$$

where $\psi = [\Phi_0, \Phi_1, \Phi_2, \Phi_3, \Phi_4]$ represents the list of unknown parameters which need to be estimated through MCMC methods.

The model assumes $\Phi_0, .., \Phi_4$ to be independently distributed as Wishart random variables, $\Phi_i \sim W_i(v_i, S_i)$ for $i = 0, 1, .., 4$

The sampling can be divided in two parts. In the first part, the states $\theta_{0:T}$ are sampled iteratively by using an algorithm called Forward filtering backward sampling FFBS. In the second part, the sampler computes the full conditional distribution of the unknown parameters.

6.5 Gibbs Sampler and FFBS Algorithm

In the first part of the sampling procedure, the FFBS algorithm runs the Kalman filter recursion for updating the filtered estimates of the unobservable states, and then runs the Kalman smoother to get the total period joint probability distribution of the unobservable states. The idea of FFBS is to gather the filtered unobservable states from time $t = 0$ to time $t = T$ by running the filter, then compute the smoothed values of the unobservable states once all the observations are available, in order to create a Markov chain of the state variables.

In the second part of the sampling procedure, the sampler needs to draw simulated values

[6]This is so because it is usual to work with precision instead of variances, where precisions are merely their inverses.

of the unknown covariances from the Markov chain simulated series of unobservable states created in the first part of the sampling procedure. The sampler will draw a random value from the distribution of interest at every iteration, and estimate an expected value for the unknown by Monte Carlo method once the iterations are concluded. The posterior density function of the model is proportional to:

$$\prod_{t=1}^{T} N_{y_t}(F_t\theta_t, \Phi_0^{-1}) N_{\theta_t}(G\theta_{t-1}, \Phi^{-1}) N_{\theta_0}(m_0, C_0) W_{\Phi_0}(v_0, S_0) \prod_{i=1}^{h} W_{\Phi_i}(v_i, S_i)$$

where

$$\Phi = W^{-1}$$

and the index on every distribution refers to the random variable it specifies.

In particular, the full conditional distribution of Φ_0 is

$$W(v_0 + \frac{T}{2}, S_0 + \frac{SS_y}{2})$$

where $SS_y = \sum_{t=1}^{T}(y_t - F_t\theta_t)(y_t - F_t\theta_t)'$

The full conditional distributions of the other four covariances are

$$W_i(v_i + \frac{T}{2}, S_i + \frac{SS_i}{2})$$

where

$$SS_i = \sum_{t=1}^{T} SS_{ii,t}$$

$$SS_t = \begin{pmatrix} SS_{11,t} & \cdots & SS_{1h,t} \\ \vdots & \vdots & \ddots \\ SS_{h1,t} & \cdots & SS_{hh,t} \end{pmatrix}$$

To facilitate the simulation, it is convenient to express the Wishart hyperparameters as[7]:

$$v_0 = (\delta_0 + m - 1)/2 \quad v_j = (\delta_j + p_j - 1)/2 \quad \text{for} \quad j = 1, .., 4$$

and

$$S_0 = \frac{V_0}{2} \quad S_j = \frac{W_{j,0}}{2} \quad \text{for} \quad j = 1, .., 4$$

[7]The modus operandi of fixing hyperparameters has been inspired by Petris(2009).

Where V_0 and $W_{j,0}$ represent the starting hypothesis for the simulation, m is the total number of observed series for y_t and p_j is the total number of state variables treated by the covariance matrix j. Since every covariance matrix W_j calculates the variance-covariance measures for one of the four coefficients in the Fama French regression among the observed series, in this model $m = p_j$ for all the covariances matrices. In this way:

$$E(V) = \frac{1}{\delta_0 - 2} V_0 \quad E(W_j) = \frac{1}{\delta_j - 2} W_{j,0} \quad for \quad j = 1, .., 4$$

The full conditional distribution are then:

$$W\Big(\frac{\delta_0 + m + T}{2}, \frac{V_0}{SS_y}\Big)$$

$$W_j\Big(\frac{\delta_j + p_j + T}{2}, \frac{W_{j,0}}{SS_j}\Big)$$

To respect the properties of MCMC and create a robust sample from the posterior density function, it is common to discard a total number of simulation as so-called burn-in and evaluate the unknown quantities through Monte Carlo methods for the rest of the sample. In this document the total number of simulations for the two stocks SUR is 30000, and for the ten stocks is 10000. The burn-in is set to 80% of simulated values. The starting values of V and W are set as uninformative prior in order to express a total ignorance of the possible values of the covariances. Then, for every simulation the sampler draws a Markov chain of unobservable states and samples five covariance matrices. The covariances are saved as global values for V and W, in order to start the following recursion with the new drawn covariance matrices. As described above, the MCMC needs to converge the its stationary distribution. When drawn covariances are set as new initial values for the MCMC recursion, the FFBS will filter and smooth a new Markov chain with the previously drawn covariance matrices. After several simulations, this procedure should force the MCMC to converge to its stationary distribution. The ergodic means of the diagonals of V and W are plotted in the appendices.

As for the Kalman filter, the Markov Chain Monte Carlo is a recursive process and it requires a starting point for simulating the error covariances. To start it, it is required to set values for the variables in the Wishart hyperparameters. After several simulations,

the parameters have been set equal to:

Table 1: Variables for Wishart hyperparameters

Parameters	
δ_0	300
δ_1	30
δ_2	30
δ_3	30
δ_4	30

These parameters express the prior uncertainty of the covariances. It is difficult to express honest priors, and it is common to use data-dependent priors or model them after several simulations of the system. The initial covariance matrices are set with huge initial variables in order to express an extreme high level of uncertainty about the correct values of the unknown quantities. All the initial matrices are set as diagonal matrices where all the values are set equal to 100,000.

$$V_0 = (\delta_0 - 2)\begin{pmatrix} 100,000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 100,000 \end{pmatrix} \qquad W_1 = (\delta_1 - 2)\begin{pmatrix} 100,000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 100,000 \end{pmatrix}$$

$$W_2 = (\delta_2 - 2)\begin{pmatrix} 100,000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 100,000 \end{pmatrix} \qquad W_3 = (\delta_3 - 2)\begin{pmatrix} 100,000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 100,000 \end{pmatrix}$$

$$W_4 = (\delta_4 - 2)\begin{pmatrix} 100,000 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 100,000 \end{pmatrix}$$

Even if the numbers on the diagonal axis seem to be extremely high and not relatable for the daily covariances of unobservable errors, the Markov Chain Monte Carlo simulation quickly adjust for this completely uninformative initial value.

6.6 Momentum Investment Strategy

Quantitative equity investing strategies are model driven active investments. These strategies try to discover market anomalies by apply computational statistics on financial models. The focus of these strategies is on the correct computer implementation of the investment strategy, since these investment decisions are completely computer based. The positive aspect of quantitative investing is the completely absence of behavioural issues, since decisions are just based on model results. Data accuracy becomes prominent for correct investments, and strategy are usually backtested on historical data. During the last two decades, these strategies have been heavily deployed by hedge funds trying to create alpha returns. Quantitative investing can be categorized into fundamental quant, statistical arbitrage, high frequency trading. The difference among these three categories is on the rebalancing timing and information signals.
Among the various active investment strategies, in the last two decades a particular strategy became central in market efficiency debate. In finance, momentum is the empirical persistence of past price movements into future adjustments. If prices grew in the past 3,6 or 12 months, empirical data seem to confirm that they will grow further, and same goes for prices falling down. Momentum is a market anomaly, since predicting future stock prices is impossible and forecasting their values can be just done with a certain degree of uncertainty by looking at companies' term sheets and future contracts.

Momentum investing is widely diffused among hedge fund managers and can be considered into two different dimensions. Cross-sectional momentum considers a certain universe of stocks, and uses past prices are signals for investment positions. This strategy will be long positive-performing stocks and short with negative-performing peers. In this way, the strategy will try to exploit the market anomaly of price movement persistence. Time-series momentum considers single stocks in their time dimension and sets positions only based on their single performances over time.

This market anomaly is still a debated argument among financial economists and seem to be caused by behavioural biases like:

1. Price underreaction/overreaction: when the intrinsic value of a stocks shifts, market price does not immediately shift to its new intrinsic value due to contrarian investors which do not share the same idea of intrinsic value, leading to a slow adjustment of market price which can be exploited by informed investors

2. Herd behaviour: overreacting investors can move their portfolios based to irrational exuberance and increase the volatility of prices

3. Financial institutions seeking to stabilize prices and wages instead of maximizing expected return

4. Liquidity problems or risk management rules can tie up investment policy of investors

5. Anchoring biases lead investors to be overconfident on their investment decisions even if wrong

6. Confirmation bias lead investors to increase their position when prices decrease in order to survive the prices decrease, moved also by anchoring biases

Even if momentum strategies are usually based on the past movement of stock prices in cross-sectional or time-series dimension, a forecasting algorhythm can be exploited in order to develop an hybrid active investment strategy which is a combination of fundamental quantitative equity strategy and momentum investing. The Kalman filter for the Fama French DLM allows to forecast asset return. In particular, the filtering recursion is able to predict asset return one-step before the observation date. A quantitative investment strategy can be developed and based on the Kalman filter forecast recursion for the observable variable y_t. The rebalance signals for the investment strategy are set as the one-step forecasted returns resulting from the filtering recursion. At every point in time, the model will provide the forecast for the following daily stock returns. The investment strategy will take long positions on assets with positive forecasted returns and short positions on assets with negative returns. Every time the filter runs, it will automatically decide where to invest. Once the investment rule is set, it is necessary to set a rebalance strategy. The portfolio needs to modify its composition in time by following the signals, but this cannot be done every day otherwise the portfolio will suffer by major losses caused by transaction costs. Also, it is not realistic to rebalance a portfolio every day. The rebalancing timespans for this strategy are set to 3,5,7 and 10 days. They will set the rebalancing dates. The filtering recursion will be evaluated among the whole considered period and the rebalancing will be triggered only during rebalancing dates.

7 Data

Dynamic linear models can be applied to univariate or multivariate time series. In this document, the observed variables are represented by the financial stock return of listed companies. These data can be easily downloaded from public sources or from data providers like Bloomberg or Reuters. For this document, the data have been gathered from Yahoo Finance, Bloomberg, and imported in Python. The downloaded data refer to the time series of adjusted closing prices. For every stock, the series of adjusted prices refer to the entire period of public trading of the stocks. This means that all the series have different starting points. For simplicity, before starting the recursion, the data have been cleaned so that the considered periods for the two and ten stocks SUR DLM refer only to the period when all the stocks were traded. Then, daily returns are calculated as percentage variation between opening and closing prices. The factors required for the Fama French factor analysis are provided by professor French in his personal website. The downloaded factors refer to the US research return data directly managed by professor French. Once the database is done, the analysis can start. All the required operations have been self-developed in Python through scientific packages for numerical analysis, data management, visualization and statistical tools. Every calculation and every graph has been self-produced on Python.

7.1 Dynamic Linear Model - 2 stocks

The two selected stocks for the first SUR model are Apple and IBM. From the historical data of a stock, it is possible to capture its sample distribution and sample moments. If the returns are assumed to be distributed as Gaussian Normal random variables, sample mean and sample variance are sufficient statistics for the true unknown population distribution[8]. These two statistics are enough to describe the Gaussian family of probability distribution. To gather more information about the tails and how centered a distribution is, it is useful to compute also the third and fourth moment of the sample distribution. However, skewness and kurtosis are not treated in this analysis.

Both stocks have long series of prices. IBM has a longer one but it has been truncated in order to consider only the period when both stocks were listed[9]. Apple price was steady

[8] For an exhaustive definition of sufficient statitics refer to Mukhopadhyay(2000).

[9] Even if not applied in the analysis, Kalman filter can handle missing observation. For more, Koopman(2012)

until 2007, which is exactly the year when the company presented the first iPhone, then the price had a huge bump until nowadays. For IBM, it is possible to see the effect of the Dot-com bubble of the second half of nineties and the downturn of the first years of the new century. Since prices are non-stationary process, it is useful to analyze stock returns.

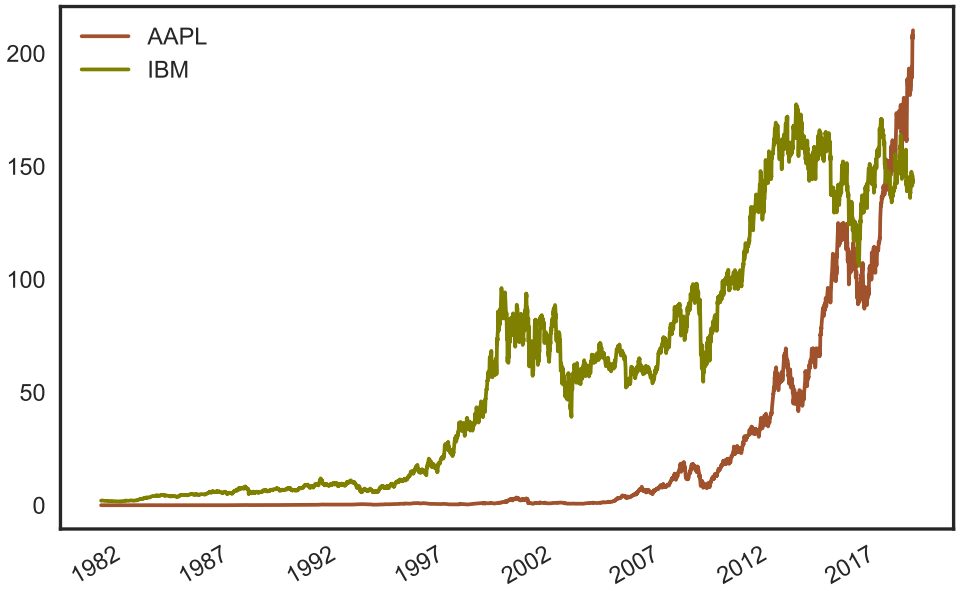

Figure 1: Historical price - two stocks SUR

It is easy to see that Apple return has been more volatile than IBM among the analyzed period of time. Both series seem to be stationary since they oscillate around a mean of zero, even if there have been some major movements like around 2002 for Apple and around 1990 for IBM. The historical distributions are plotted with their Gaussian kernel density estimate. The historical distributions of daily returns of Apple and IBM are both non-exactly normal. The non-Gaussian distribution of the historical returns is one of the critics of the model which will be discussed at the end of the document.

For a statistical analysis of return predictability, the classic Box-Jenkins approach can be applied to investigate the autocorrelation function. From the figure, the null hypothesis of statistical zero autocorrelation is rejected just for the first lag for both stocks.

In statistical inference, it is usual to work with stationary processes since non-stationarity can cause problems. For example, when dealing with empirical data, it is common to assume that samples are made of independent and identically distributed random variables. Independency is a powerful assumption which relaxes the issues related with statistical modelling. Another example is spurious regression. When performing a regression, if the

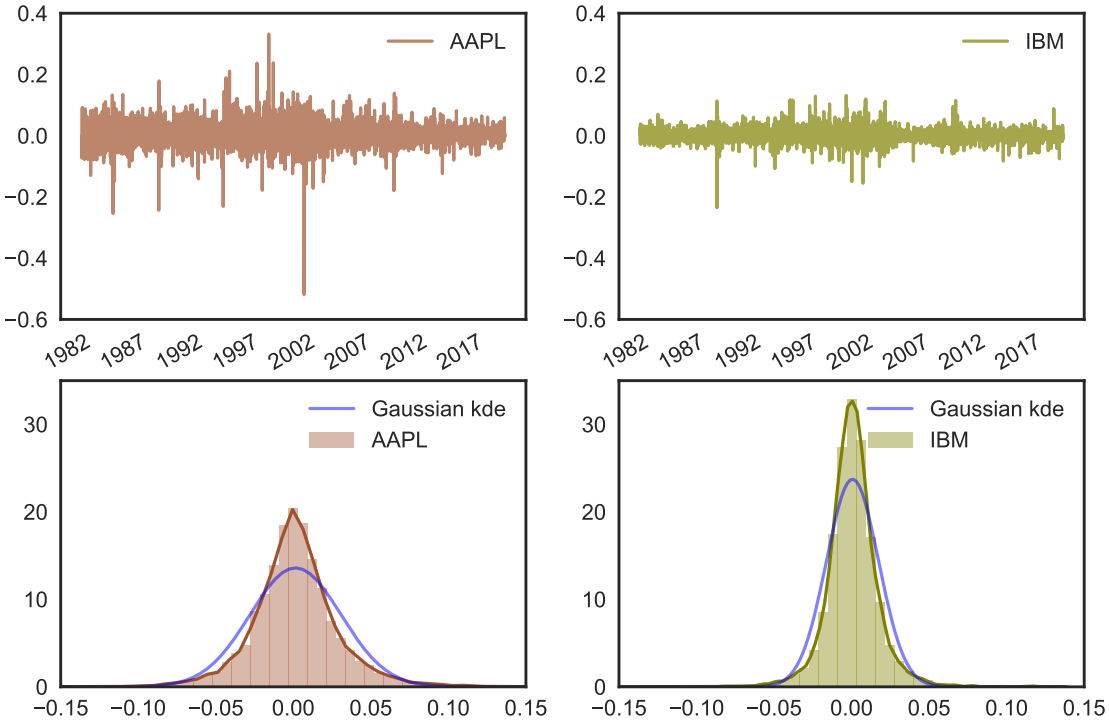

Figure 2: Historical returns - two stocks SUR

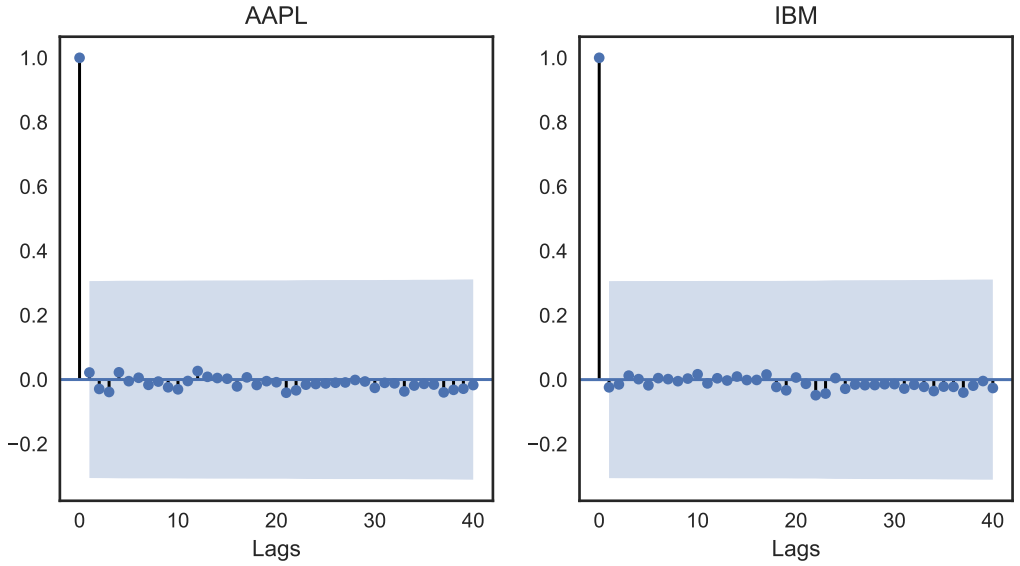

Figure 3: Autocorrelation function

residual series is non-stationary, the regression is spurious. This is usually the case when the regression shows high value of R^2 but there is no real relationship between dependent and independent variables. A process is non-stationary if it has a unit root among the roots of the characteristic equation. This equation is a polynomial form upon which depends the solution of the ordinary differential equation specifying the process. If one of

Table 2: Descriptive statistics - two stocks SUR

	AAPL	IBM
Observations	9501	9501
μ	0.001	0.001
σ	0.029	0.017
Min	-0.519	-0.235
Q1	-0.013	-0.008
Q2	0.000	0.000
Q3	0.015	0.009
Max	0.332	0.132

the roots is equal to one, the process has a unit root and it is not covariance-stationary. A basic test for the presence of a unit root in a time series is the Augmented Dickey Fuller test. In this test, the null hypothesis is that the series has a unit root. In the current analysis, both return series are tested and results suggest they both are covariance-stationary. As it is commonly shown in literature, stock prices show trends. The log transformation of stock prices is a time series with order of integration one, I(1), so taking the difference of log transformed stock prices leads to stationary measurement of stock returns. Another way of de-trending the series is to compute the daily return series, as it is done in this document. As described above, SSM can handle nonstationary series and is in this sense more flexible than more classic tools like Box-Jenkinks ARMA models.

Table 3: Unit root - two stocks SUR

Augmented Dickey Fuller		
	AAPL	IBM
Test statistic	-4.00	-21.43
p-value	0.001	0.000
used lag	16	22
Critical Values		
1%	-3.431	-3.431
5%	-2.862	-2.862
10%	-2.567	-2.567

The first basic tool for asset pricing is the Capital Asset Pricing Model, so-called CAPM. It is represented by a simple univariate regression where the dependent variable is given by the excess return of the asset under analysis, while the independent variable is

the market excess return. For the analysis, the regression is computed for both stocks and the market excess return is given by the SP500 excess return. Both stocks show positive and significant market beta, which means that both stocks co-move with the market in both rise and fall periods. Apple variations are higher than the SP500 since its market beta is higher than one, while IBM moves less proportionally to the market index. The constant is significant for both stocks: in asset pricing this means that part of the returns of Apple and IBM are not statistically explained by the CAPM. There is a part of return which can be, called abnormal return, which is not captured by the model. There is space for improvement.

Table 4: CAPM - two stocks

	AAPL	IBM
β^M	1.2732***	0.9468***
	(0.025)	(0.013)
α	0.001***	0.0003**
	(0.000)	(0.000)
Observation	9448	9448
R^2	0.215	0.362
Adj-R^2	0.215	0.362
F-Test	2581	5363
Log-Likelihood:	21053	27306

A more sophisticated model is the Fama French 3 factors model. The idea is to capture a higher part of co-movement of returns by passing to a multivariate regression. As described above, the model controls for Small/Big and Value/Growth portfolio returns. Besides the size coefficient of Apple and IBM's measure of unexplained return, all the other coefficients are significant. The unexplained return measure of IBM is represented by α and it is not statistically different from zero, meaning that the three factors model is able to capture and explain IBM's return and it is more reliable than the CAPM. For Apple this does not hold, since α is still statistically different from zero. The two new coefficients allow to better understand the nature of the underlying stock returns. Negative size effect refers to companies with high market cap, and negative value refers to growth stocks.

7.2 Dynamic Linear Model - 10 stocks

For the ten stock model, the selected stocks are: Apple, IBM, Boeing Co, Boston Scientific Corp, Walt Disney, FedEx, General Electric, Johnson Johnson, JP Morgan and Nike. As

Table 5: Fama French - two stocks

	AAPL	IBM
β^M	1.203***	0.8969***
	(0.025)	(0.013)
β^S	-0.0201	-0.338***
	(0.046)	(0.024)
β^H	-0.8387***	-0.3382***
	(0.048)	(0.025)
α	0.0011***	0.0004
	(0.000)	(0.000)
Observation	9448	9448
R-squared	0.239	0.385
Adj-R-squared	0.239	0.385
F-Test	989.6	1969
Log-Likelihood:	21203	27477

for the two stocks model, the data analysis relies on asset pricing tools, unit root tests, descriptive statistics and price plots. The histogram returns and return time series are omitted since they all resemble the previously presented plots for two stocks. All the returns seem to be non-normally distributed and move around a mean of zero. The autocovariance function has no significant lag after the first one.

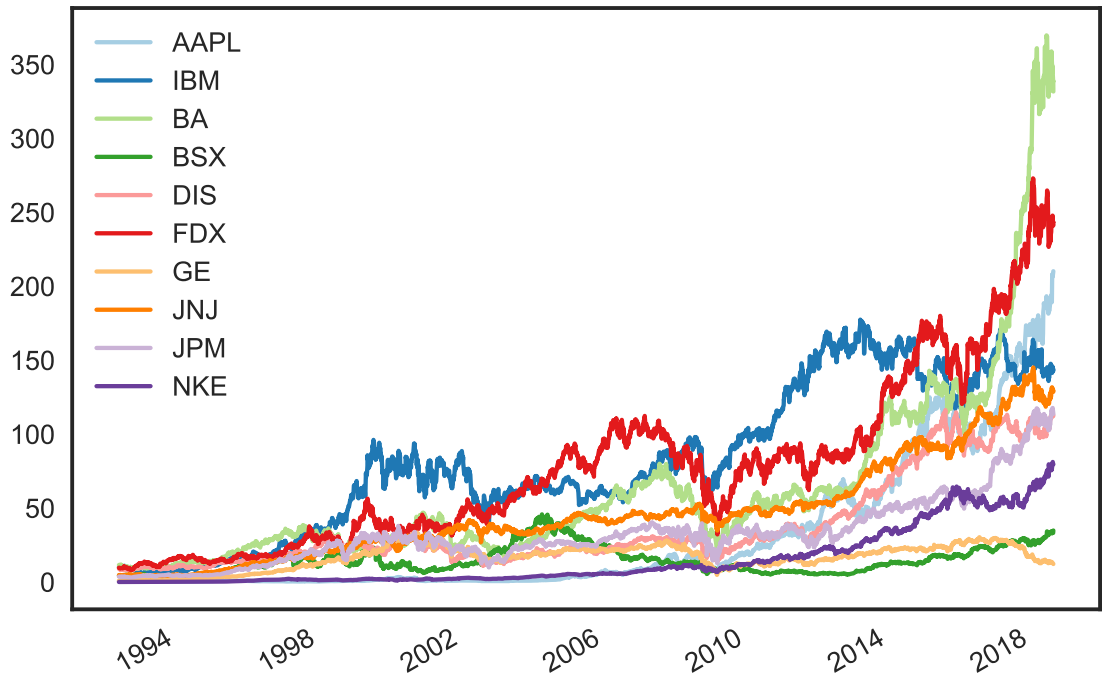

Figure 4: Stock prices - 10 stocks SUR

Table 6: Descriptive statistics - 10 stocks

	AAPL	IBM	BA	BSX	DIS
Observation	6558	6558	6558	6558	6558
μ	0.001	0.001	0.001	0.001	0.001
σ	0.029	0.018	0.019	0.025	0.019
Min	-0.519	-0.155	-0.176	-0.267	-0.184
Q1	-0.013	-0.008	-0.009	-0.012	-0.009
Q2	0.000	0.000	0.000	0.000	0.000
Q3	0.014	0.009	0.010	0.012	0.009
Max	0.332	0.132	0.155	0.261	0.160
	FDX	GE	JNJ	JPM	NKE
Observation	6558	6558	6558	6558	6558
μ	0.001	0.001	0.001	0.001	0.001
σ	0.019	0.018	0.013	0.021	0.024
Min	-0.145	-0.128	-0.158	-0.198	-0.207
Q1	-0.009	-0.008	-0.006	-0.009	-0.010
Q2	0.000	0.000	0.000	0.000	0.000
Q3	0.010	0.009	0.007	0.010	0.011
Max	0.139	0.197	0.122	0.174	0.251

Table 7: Augmented Dickey Fuller - 10 stocks

Augmented Dickey Fuller					
	AAPL	IBM	NKE	JPM	JNJ
Test statistic	-18.13	-17.82	-49.09	-14.93	-26.06
p-value	0	0	0	0	0
used lag	20	22	22	35	10
Critical Values					
1%	-3.431	-3.431	-3.431	-3.431	-3.431
5%	-2.862	-2.862	-2.862	-2.862	-2.862
10%	-2.567	-2.567	-2.567	-2.567	-2.567
	GE	FDX	DIS	BA	BSX
Test statistic	-15.11	-60.02	-33.36	-14.58	-23.91
p-value	0	0	0	0	0
used lag	26	1	6	35	12
Critical Values					
1%	-3.431	-3.431	-3.431	-3.431	-3.431
5%	-2.862	-2.862	-2.862	-2.862	-2.862
10%	-2.567	-2.567	-2.567	-2.567	-2.567

As expected, the sample distributions of all the stocks are very similar. Every stock return moves around a mean of zero with almost same volatility and presents no unit root.

Table 8: CAPM - 10 stocks

	AAPL	IBM	BA	BSX	DIS
β^M	1.1445***	0.9009***	0.9323***	0.943***	1.0408***
	(0.028)	(0.016)	(0.017)	(0.025)	(0.016)
α	0.001***	0.0002	0.0004**	0.0003	0.0002
	(0)	(0)	(0)	(0)	(0)
Observation	6558	6558	6558	6558	6558
R-squared	0.202	0.329	0.314	0.175	0.39
Adj-R-squared	0.201	0.329	0.314	0.175	0.39
F-Test	1655	3210	3002	1391	4188
Log-Likelihood:	14734	18475	18031	15434	18401
	FDX	GE	JNJ	JPM	NKE
β^M	0.9868***	1.134***	0.5668***	0.8389***	1.4994***
	(0.017)	(0.014)	(0.013)	(0.02)	(0.018)
α	0.0003*	0.0002	0.0004***	0.001***	0.0003
	(0)	(0)	(0)	(0)	(0)
Observation	6558	6558	6558	6558	6558
R-squared	0.327	0.491	0.229	0.206	0.506
Adj-R-squared	0.327	0.491	0.229	0.206	0.506
F-Test	3185	6336	1947	1698	6712
Log-Likelihood:	17852	19196	19875	16855	17553

The CAPM is able to explain the excess return of few stocks like IBM, BSX, DIS, GE and NKE since it captures and explain all their excess returns. The alpha coefficients of these stocks are not statistically different from zero. For the rest, the part of the return series remains unexplained since alphas are statistically different from zero. Note that the CAPM results for AAPL and IBM are here different since the total number of considered days is equal to 6558 and it is lower than the number of observations for the two stocks SUR model. This is so because the ten stocks SUR consider only the daily observations of all the joint stocks, and drops all the observations for which at least one stock was not listed.

The three factors model is able to improve the asset pricing of unexplained returns but AAPL, JNJ and JPM alpha coefficients remain unexplained.

Table 9: Fama French - 10 stocks SUR

	AAPL	IBM	BA	BSX	DIS
β^M	1.1274***	0.8951***	0.937***	0.9415***	1.0437***
	(0.028)	(0.016)	(0.017)	(0.025)	(0.016)
β^S	0.0758	-0.2243***	-0.0681*	-0.0148	-0.1233***
	(0.053)	(0.03)	(0.033)	(0.049)	(0.031)
β^H	-0.6973***	-0.3519***	0.1711***	0.049	0.069*
	(0.052)	(0.029)	(0.032)	(0.047)	(0.03)
α	0.001**	0.0003	0.0004	0.0003	0.0002
	(0)	(0)	(0)	(0)	(0)
Observation	6558	6558	6558	6558	6558
R-squared	0.224	0.346	0.318	0.175	0.392
Adj-R-squared	0.224	0.346	0.318	0.175	0.392
F-Test	632.4	1157	1019	464.4	1409
Log-Likelihood:	14829	18563	18050	15435	18414
	FDX	GE	JNJ	JPM	NKE
β^M	0.9929***	1.1437***	0.5682***	0.8403***	1.5297***
	(0.017)	(0.014)	(0.012)	(0.02)	(0.016)
β^S	0.1**	-0.3426***	-0.4859***	0.0001	-0.3406***
	(0.033)	(0.027)	(0.024)	(0.039)	(0.031)
β^H	0.3066***	0.2631***	-0.1604***	0.0607	1.1468***
	(0.032)	(0.026)	(0.023)	(0.038)	(0.03)
α	0.0003	0.0002	0.0004***	0.001***	0.0002
	(0)	(0)	(0)	(0)	(0)
Observation	6558	6558	6558	6558	6558
R-squared	0.336	0.515	0.276	0.206	0.607
Adj-R-squared	0.336	0.514	0.276	0.206	0.607
F-Test	1107	2315	832.6	566.8	3380
Log-Likelihood:	17898	19348	20081	16856	18308

8 Results

8.1 Filtered and Forecasted State Variables - 2 stocks

To avoid problems linked to in-sample performance measurements for the portfolios, the Markov Chain Monte Carlo recursions for the Bayesian inference of the covariance matrices of observed and unobserved error terms have been computed for a restricted amount of time in order to show ignorance of the future stock prices and evaluate an out-of-sample performance of each portfolio. Once the MCMC recursion has converged, the Monte Carlo estimate of the observed and unobserved error covariances are plugged in the model to run the Kalman filter and Kalman smoother. As described above, after running the recursion from time $t = 0$ to time $t = T$, the model is able to filter out the unobservable states from noisy observations. In the presented model, the unobservables states are represented by the relevant coefficients of the Fama French regression.

The Markov Chain Monte Carlo estimated covariances matrices are:

$$V = \begin{pmatrix} 0.000132 & 0.00000298 \\ 0.00000298 & 0.000128 \end{pmatrix}$$

$$W_1 = \begin{pmatrix} 0.00131 & 0.0000061 \\ 0.0000061 & 0.0013 \end{pmatrix} \quad W_2 = \begin{pmatrix} 0.0013 & 0.00000349 \\ 0.00000349 & 0.00132 \end{pmatrix}$$

$$W_3 = \begin{pmatrix} 0.00131 & -0.0000153 \\ -0.0000153 & 0.00127 \end{pmatrix} \quad W_4 = \begin{pmatrix} 0.00137 & 0.0000117 \\ 0.0000117 & 0.00136 \end{pmatrix}$$

where $W = \text{blockdiagonal}(W_1, W_2, W_3, W_4)$.

The following graphs refer to the last three years values for filtered and forecasted unobservable states. This facilitates to visualize the properties of data. The plots for the full sample can be found in the appendix.

The first coefficient α_t is a measure of the excess abnormal return over the asset pricing model. This is the most important measure for active investors like hedge fund managers. These investors seek to gain reward in any possible state of the market and they try to have a zero exposure to the market movement, which translates into a total β of zero. In a basic model as CAPM, the expected excess return of a portfolio is:

$$E[r_i] = \alpha + \beta E[r_m] = \alpha$$

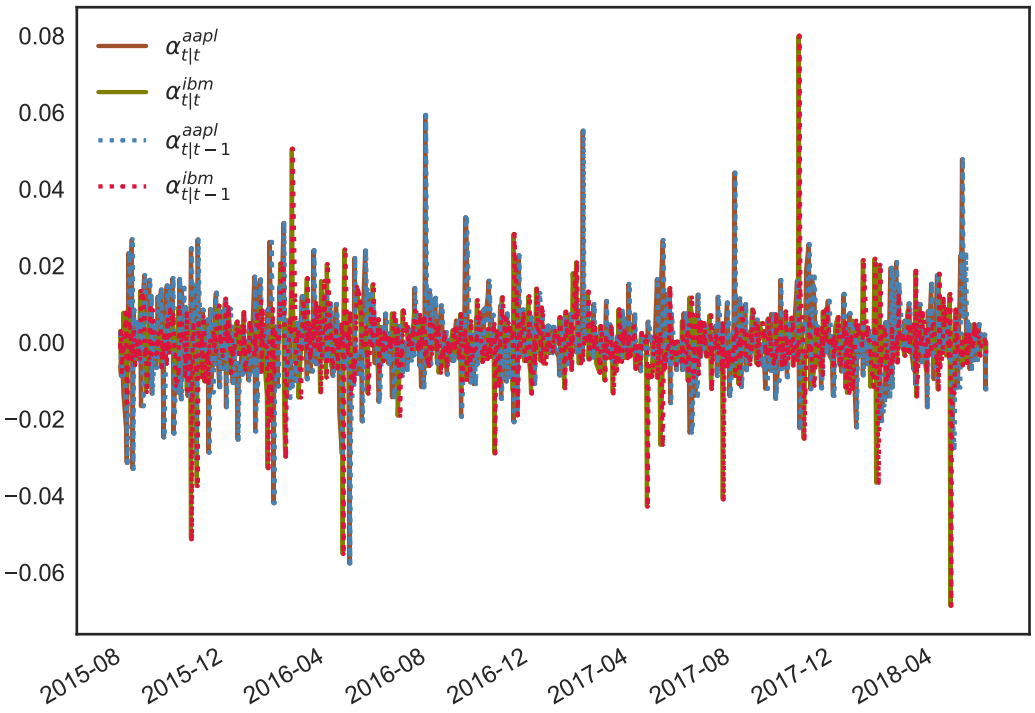

Figure 5: Filtered and one-step forecasted α_t

Generally, creating alpha returns is difficult and it can be entitled to hedge fund manager skills, and sometimes luck. The result from the filtering recursion generates the one-step forecast value $\alpha_{t|t-1}$ and the filtered estimate $\alpha_{t|t}$. The forecasted value $\alpha_{t|t-1}$ is the optimal forecast of α_t created at time $t-1$. The filtered value $\alpha_{t|t}$ is the noise-free estimate of α_t calculated at time t right after the new observation of the empirical returns y_t comes in. The one-step forecast series is perfectly following the filtered value, signaling that the forecasted value for α is a robust estimate of its filtered estimate. The estimates evolve around a mean of zero, resembling a stationary process, but has some spikes and downturn among the whole series. The time series of filtered α is similar to the empirical stock returns. This graph seems to confirm the non-predictability of alpha returns. Since the time series resemble a stationary process, it seems not possible to predict future development of abnormal excess returns.

The sensitivity to the market portfolio is an important measurement for investors and risk managers. This coefficient is crucial for create robust hedging strategies, evaluate the market risk of a portfolio of stocks, assess how market shocks affect portfolio performances and many other arguments. Market movement impacts on a specific stock is expected to vary over time. This is linked to the equity premium returned by that stock. A high value

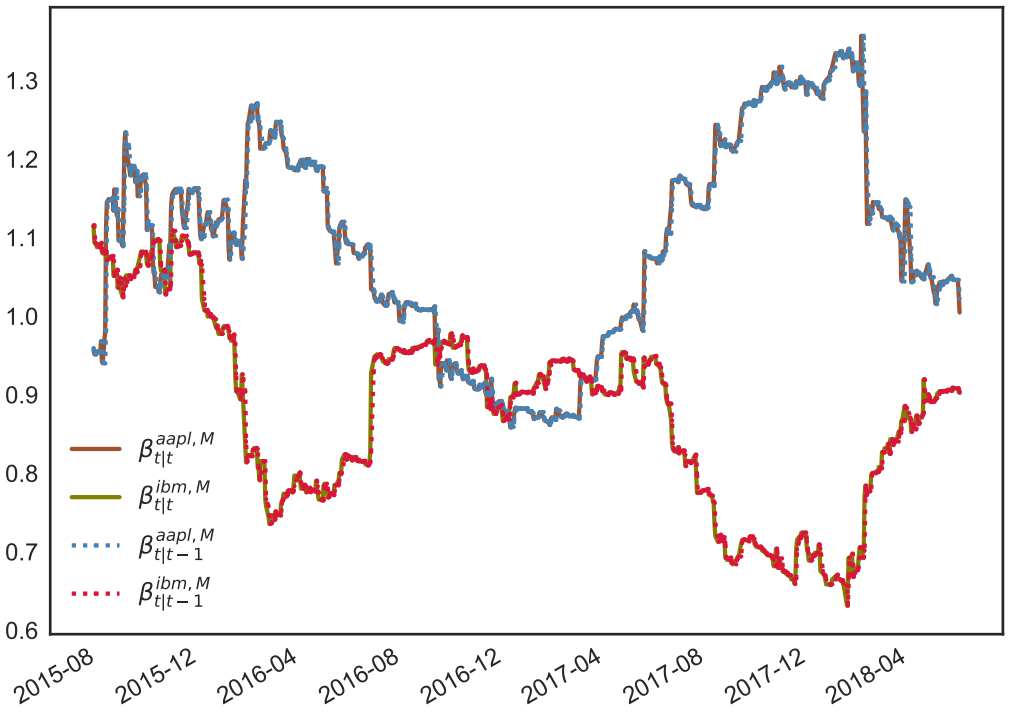

Figure 6: Filtered and one-step forecasted β_t^M

of β^M means that positive market movements positively affect the excess return of the stock. The last three years filtered values of the market sensitivity are highly volatile and diverging. The two underlyings have been affected in opposite directions. The series seem to be cyclical, since there is no major trend.

The size factor is related to the market capitalization of companies. Values are mostly negative, since both Apple and IBM are companies with high market values of equity. The series present high volatility with huge jumps.

The value factor is related to the relationship between the book value of equity and the listed price. The filtered estimates seem to be less volatile and more stable than the other two factor series. There are no major jumps like the size factor series. This could be related to the behavioural biases affecting price opinions. Anchoring biases, underreaction/overreaction on prices and the slow adjustment of price opinion can be the reason why the filtered estimate variations are less volatile than the other factors.

Forecasted prices are plotted with real price series. The forecasted prices use one-step ahead predicted returns from the Kalman recursion algorhytm. The plotted values for the forecasted prices are plotted for the following day, in order to compare the forecasted price

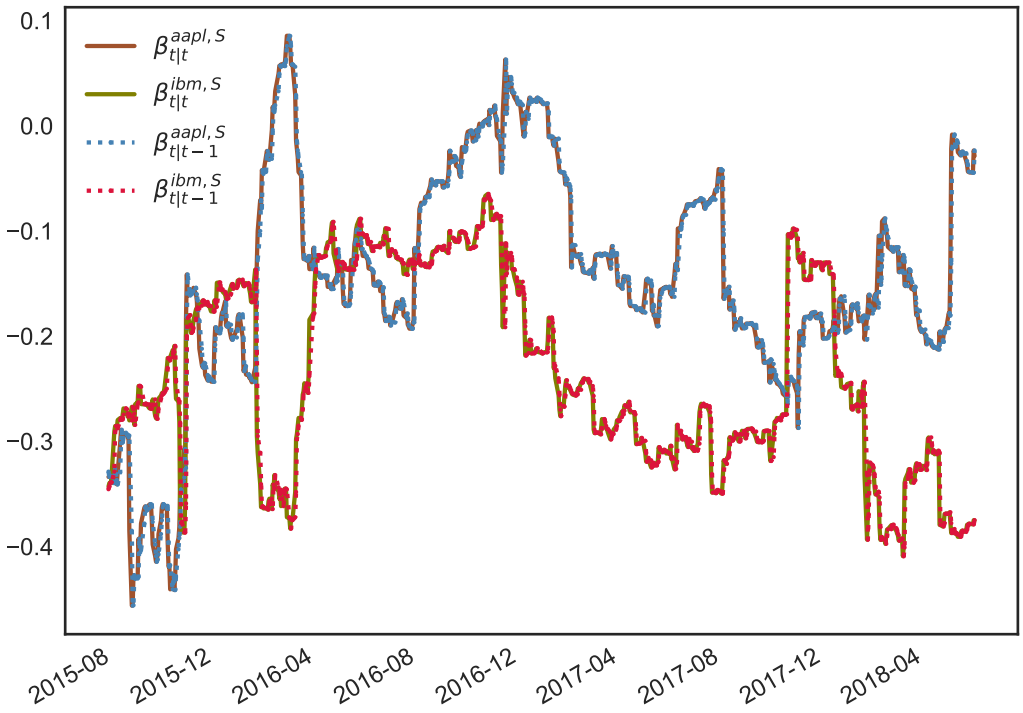

Figure 7: Filtered and one-step forecasted β_t^S

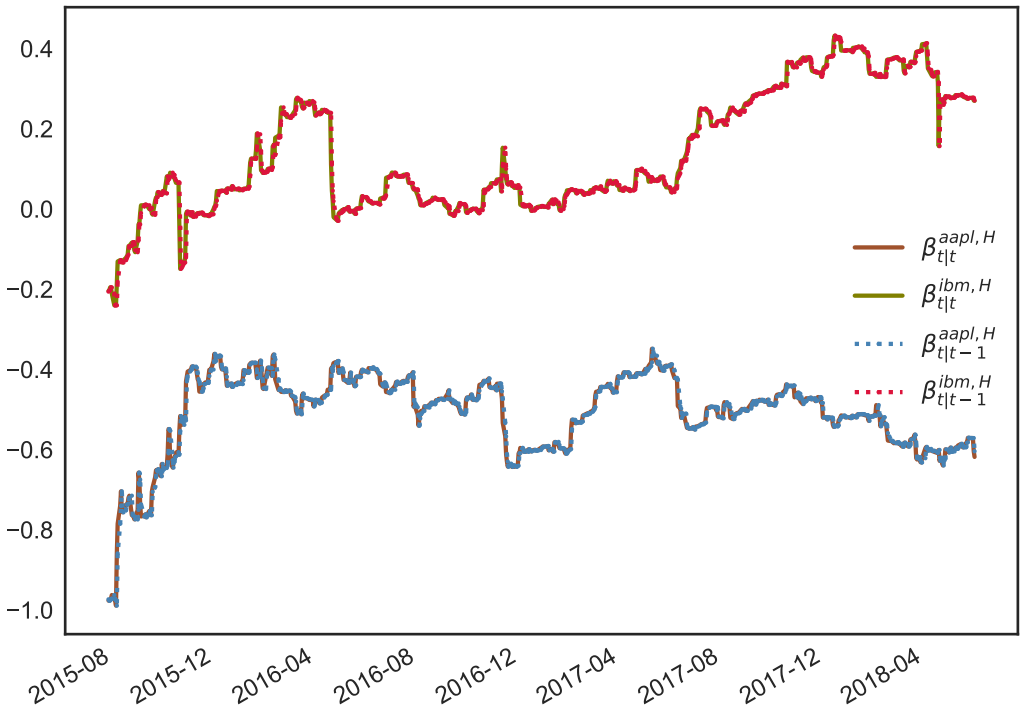

Figure 8: Filtered and one-step forecasted β_t^H

with the real price. This means that, for example, at time $t = 1000$ the real price is the traded price at $t = 1000$ while the forecasted price refers to $t|t-1 = 1000|999$. This is the

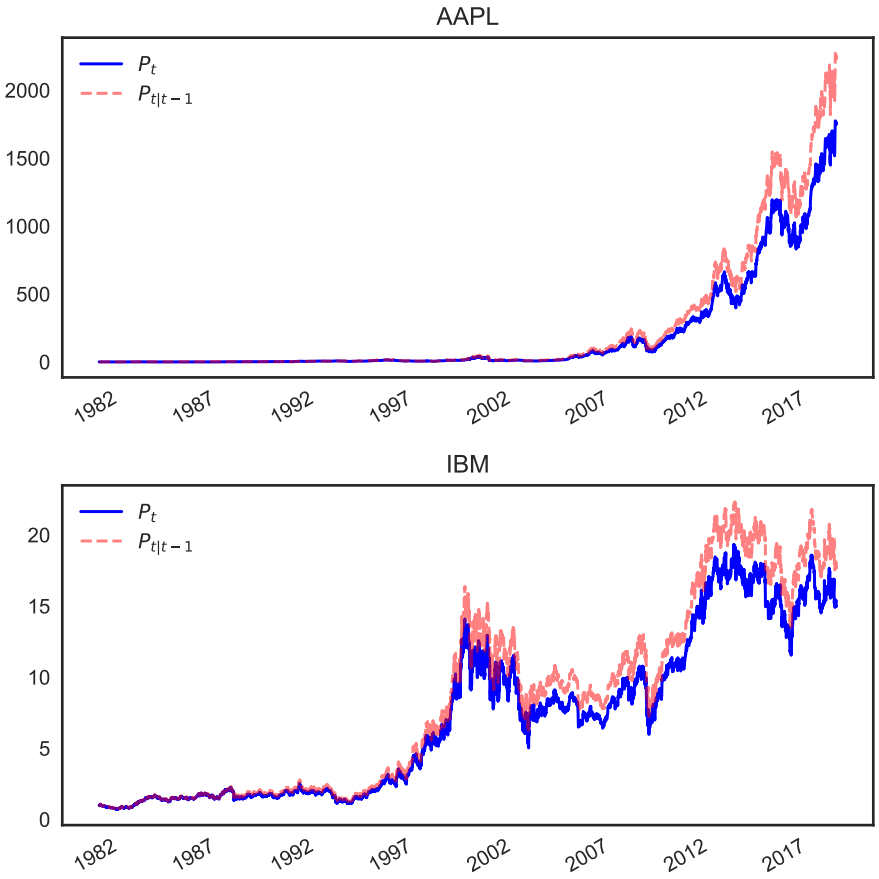

Figure 9: Stock price forecasting for two stocks

forecast computed at time $t = 999$.

Forecasted prices are able to follow the real price series, but at some point they start to create overestimations. This can be explained by the decision of running the Gibbs sampler for a restricted period of time. To create an out-of-sample measurement of the performances of the investment strategies, the inference for unknown parameters has been restricted. Strictly speaking, the Gibbs sampler evaluates the unknowns on a reduced set of observations, accounting to 10 years for the two stocks SUR and 5 years for the ten stocks SUR. Since the overestimations refer to the last periods of observed returns, it seems reasonable to expect error covariances to be slightly biased.

8.2 Filtered and Forecasted State Variables - 10 stocks

In the analysis of the quantitative strategy, it will become clear that two stocks are not enough to create a portfolio with an acceptable volatility. In order to assess how the strategy may perform with a higher number of stocks, and reduce the risk measure of the

strategy, the previously presented two stocks SUR is applied to a total of ten stocks. The eight new stocks are: Boeing (BA), Boston Scientific Corporation (BSX), Walt Disney (DIS), FedEx (FDX), General Electric (GE), Johnson Johnson (JNJ), JP Morgan (JPM) and Nike (NKE). In this case, the total number of estimated coefficients are fourty and data visualization becomes cumbersome. Filtered estimates plots can be found in the appendix. Also, rolling regressions have been computed and rolling coefficients can be compared to filtered estimates. The rolling regressions accounted for 365 observations, with a window of 1 day. Asset forecasted prices for the ten stocks SUR seem to confirm that the model is reliable since the filtering recursion is able to produce sound forecasts for the one-step forecasted asset return. However, the forecasted prices of some stocks are significantly higher than the listed prices. This may be connected with the decision of evaluating the errors covariance matrices with Bayesian inference over a reduced amount of time. The reason for this choice is to evaluate the quantitative strategy performances in a out-of-sample framework. The Gibbs sampling procedure has been concentrated over a reduce period of observed returns, and the performances have been calculated within the remaining period only. This should improve the attendability of the performances and wipe in-sample issues.

Modelling ten stocks leads to five covariances matrix of dimension 10x10. The convergence of the simulation can be assessed by plotting the Monte Carlo estimate of the entries of each covariance matrix. In order to represent a feasible and more understable visualization of the estimates, the following plots will just consider the diagonals of the five covariances matrices.

8.3 Momentum Strategy - 2 stocks

The first quantitative investment portfolio contains positions on Apple and IBM. The first day of observation has been December 12th 1980. The above described MCMC recursion for Bayesian inference has been iterated for 30,000 times over a timespan of 10 years. After the two covariance matrices have been estimated by Monte Carlo methods, the Kalman recursion provides the one-step ahead forecast of the observed variables which represent the information signal for computer based rebalances. At every point in time, the algorhythm will evaluate the position for Apple and IBM based on the one-step ahead forecast and modify the position once the rebalancing date comes. Here, the simulated portfolio has been analyzes with four rebalancing dates: 3 days, 5 days, 7 days and 10

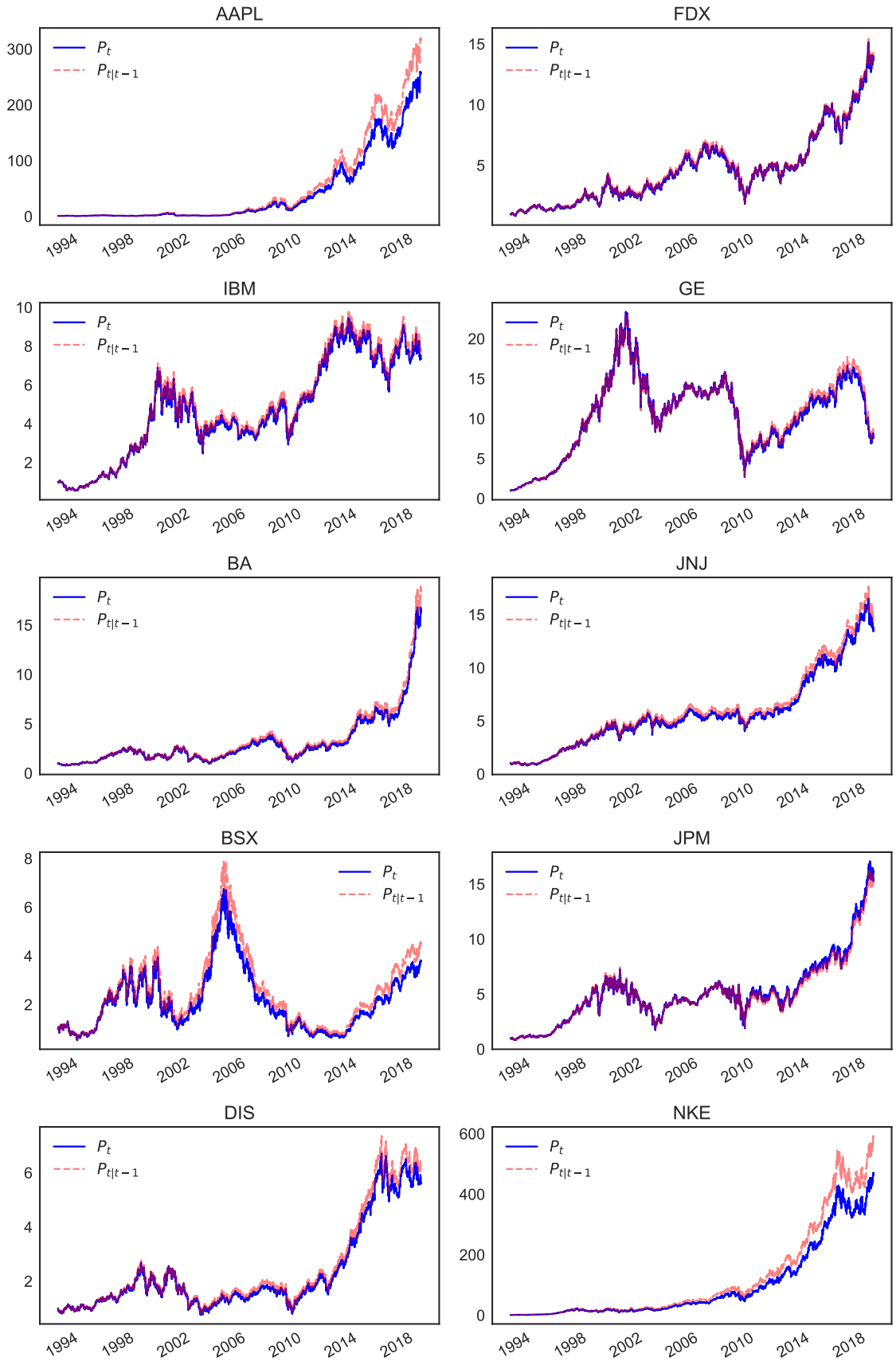

Figure 10: Stock price forecasting for two stocks

days. The perfomance in log-linearized scale is the following.

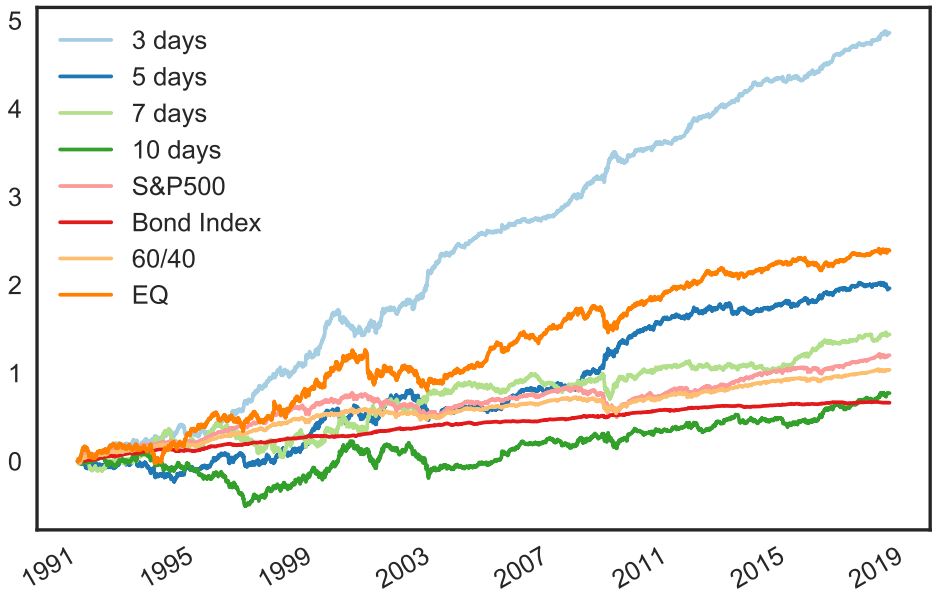

Figure 11: Quantitative strategy with two stocks

As expected, the total number of days between two rebalances is crucial for the portfolio performance. The 10 days portfolio has the worst performance and it is never able to bet the market portfolio. The 3 days portfolio has the best performance, with a market capitalization of almost twice the 5 days portfolio in log-scaled measure. The EQ portfolio is an equally weighted portfolio of Apple and IBM. It is clear that the quantitative strategy for the two stocks SUR is not optimal for any rebalancing period higher than 3 days, since the equally weighted portfolio is able to provide a higher capitalization of the final portfolio. The 60/40 portfolio is created with a weight of 60% on SP500 and 40% on the bond index, where the bond index refers to the Barclays US Unhedged Bond Index.

This gives a higher picture of the portfolio performance. Expected returns increase for every portfolio and for every rebalancing timespan, but the volatility seems to be constant. Volatility is a major concern when strategies are not enough diversificated. A higher number of stocks might help to reduce the portfolio variability, while the expected return, and so the Sharpe ratio, is linearly connected to the number of rebalances in a fixed amount of time.

Hedge fund managers are not always interested in earning an investment return greater than that of the SP500. This investment point of view should be proper of a passive

Table 10: Performance for the two stocks quantitative strategy

3 days	AAPL	IBM	Portfolio	*7 days*	AAPL	IBM	Portfolio
Expected return	44.2%	42.7%	41.3%	Expected return	18.3%	13.6%	13.7%
Volatility	46.6%	26.6%	27.5%	Volatility	46.7%	26.7%	30.3%
Sharpe Ratio	0.949	1.605	1.501	Sharpe Ratio	0.392	0.511	0.451

5 days	AAPL	IBM	Portfolio	*10 days*	AAPL	IBM	Portfolio
Expected return	23.9%	15.6%	15.9%	Expected return	8.1%	7.7%	5.7%
Volatility	46.7%	26.7%	25.9%	Volatility	46.7%	26.7%	25.6%
Sharpe Ratio	0.513	0.582	0.616	Sharpe Ratio	0.173	0.287	0.223

investor: he will try to optimize a portfolio selection of attractive stocks and to gain equity premiums for his stock picking activity. Here, the concept of opportunity cost kicks in. If the passive investor is not able to create a positive extra return his passive investment benchmark (SP500), economic theory says he should invest his money on an index tracking this benchmark and employ his time in another utility-increasing activity. This concept does not hold for an hedge fund manager, because his goal is not to beat the SP500 but to make money in every market condition. A good hedge fund manager is able to provide positive returns during both the upside and downside of the business cycle. Even if the market is plunging, an informed investor is able to capture market anomalies and trade against them in order to provide positive returns to his clients. Even if the SP500 is not the right benchmark for an hedge fund manager, it stills represent a good measure of the total amount of money an investor can make in financial markets without having any kind of information about the picked stocks. To have a wider picture of the quantitative investment portfolio here described, a one sample t-test for excess return over the SP500 can be performed to assess if the portfolio is able to create an average return statistically different from zero. This basic statistical procedure is used to determine if a sample of observations can have been sampled from a population with a specific true mean. The statistic for test is calculated as

$$\text{t-stat} = \frac{\bar{y} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}}$$

where $\bar{y}$ is the sample mean, μ is the tested true mean of the population, $\hat{\sigma}$ is the sample standard devation and n is the total number of observation. This statistics is distributed as a Standard normal distribution and if its value is bigger than 2 or lower than -2 it

means that the p-value of the test is lower than the significance level of 5

Table 11: t-test for two stocks quantitative strategy

	3 days	5 days	7 days	10 days
t-stat	5.38	1.38	0.7	-0.29
p-value	0.000	0.160	0.470	0.770

From the t-statistics for different rebalancing data is clear that the 3 days rebalancing portfolio is able to generate a positive and statistically significant extra profit over the market portfolio. From 5 days on, the timespan of rebalancing is too wide and the portfolios are no able to statistically outperform the SP 500. A two stocks investment strategy does not seem to be reliable. Rebalancing every 3 days is quite unrealistic for most investors. Also, the controlled environment of complete absence of market frictions enhanced the performances. The 5,7 and 10 days rebalance strategies are not optimal. The 10 days rebalancing portfolio does not outperform the SP500. The 5 and 7 days portfolios are able to rise more than the SP500 after 2008. This sounds unrealistic since financial crisis of subprimes in 2008 created a shortage of liquidity in financial markets. Transaction costs linked to liquidity became a huge problem for investors. In the current setting, the computer based rebalanced was neither constraint nor affected by costs linked to bid-ask spread and other liquidity measures. Even if an additional analysis should be undertaken, it seems clear that the strategy needs to include a higher number of stocks to diversify its composition

8.4 Momentum Strategy - 10 stocks

Since volatility was almost constant for every portfolio created out of two stocks, an example with a higher number of stocks can be useful to evaluate if it is possible to diversify a quantitative investment portofolio. In this case, the eight new stocks are: Boeing (BA), Boston Scientific Corporation (BSX), Walt Disney (DIS), FedEx (FDX), General Electric (GE), Johnson Johnson (JNJ), JP Morgan (JPM) and Nike (NKE). The joint observations for these ten stocks start on May 19th 1992 and the MCMC estimates the error covariances over a timespan of 5 years in order to reduce the computational requirements of the simulations. The portfolio performance analysis starts on May 19th 1997 in order to cut the first 5 years of information computed in the Bayesian inference.

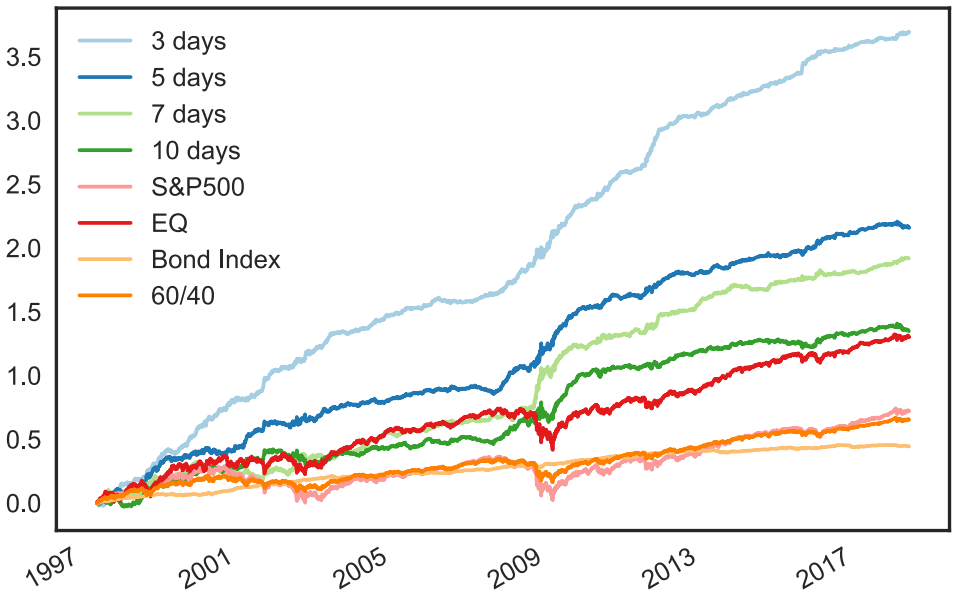

Figure 12: Quantitative strategy with ten stocks without fees

As for two stocks and as expected, a shorter timespan for rebalancing leads to superior performance for the quantitative investment portofolio. It is clear how the increased number of stocks enhances the performance of all the portfolios since the SP500 is the worst performing portfolio among the five considered. The market capitalization of the 3 days portfolio is 1.5 times the 5 days portfolio in log-scaled measurements. An interesting feature of the portfolios is their performance during the financial crisis of 2008. This could again be explained by the controlled environment for their simulated paths. The complete absence of market frictions like liquidity problems and margin calls could explain how they outperformed the SP500 in this simulated environment. It is interesting to note how the quantitative strategy starts now to be much more profitable than the equally weighted portfolio. The diversification benefit is responsible for this major improvement of the general strategy.

Table 12: Performance for 10 stocks quantitative strategy

	Portfolio			
Average	35.0%	18.9%	17.4%	10.4%
Volatilty	15.6%	15.4%	15.6%	15.3%
SR	2.24	1.22	1.11	0.66
Rebalance	3 days	5 days	7 days	10 days

Including eight more stocks in the portfolio creates a massive shift in the risk-adjusted

performance measure. The 3 days portfolio is able to create a SR of more than two while the volatility decreased for almost 30 percent for every rebalancing timespan. This huge shift in the profitability of the strategy is due to the significant gains in diversification. Portfolio volatilities are now around 15 percent. As before, a t-test for statistically significant excess return over the SP500 can create a reliable benchmark measure of the portfolios performances.

Table 13: t-test for 10 stocks quantitative strategy

	3 days	5 days	7 days	10 days
t-stat	5.38	2.54	2.14	1.07
p-value	0.000	0.011	0.032	0.280

The 10 days portofolio is the only portfolio not able to deliver a statistically significant positive extra return. The increased number of stocks modelled through the Kalman filter recursion and treated in the strategy lead to a superior performance able to statistically beat the SP500 from a 3 to 7 days rebalancing timespan. A higher number of stocks improved the strategy from a risk-adjusted point of view.

8.5 Momentum Strategy and Proportional Fees - 10 stocks

In order to relax the strong assumption of absence of market frictions, the portfolio performances has be modelled with a proportional fee on investments. A proportional fee hits whenever the investor changes position on a particular asset, and it is proportional to the invested value. For computations, a proportional fee can be treated as a negative return of the invested amount. Every type the position on an asset changes, the return for the following period will be:

$$r_{i,t+1} = \frac{P_{i,t+1} - P_{i,t}}{P_{i,t}} - \tau \mathbf{1}_{\{sign(pos_t \times pos_{t+1}) < 0\}}$$

where τ is the percentage of proportional fees and $\mathbf{1}_{\{sign(pos_t \times pos_{t+1}) < 0\}}$ is an indicator function which is equal one if the position on asset i changes from time t to $t+1$.

The graphics are presented for three different level of fees: 0.05%, 0.5% and 1%. Starting from the top, a small fee does not modify the previous performance order since

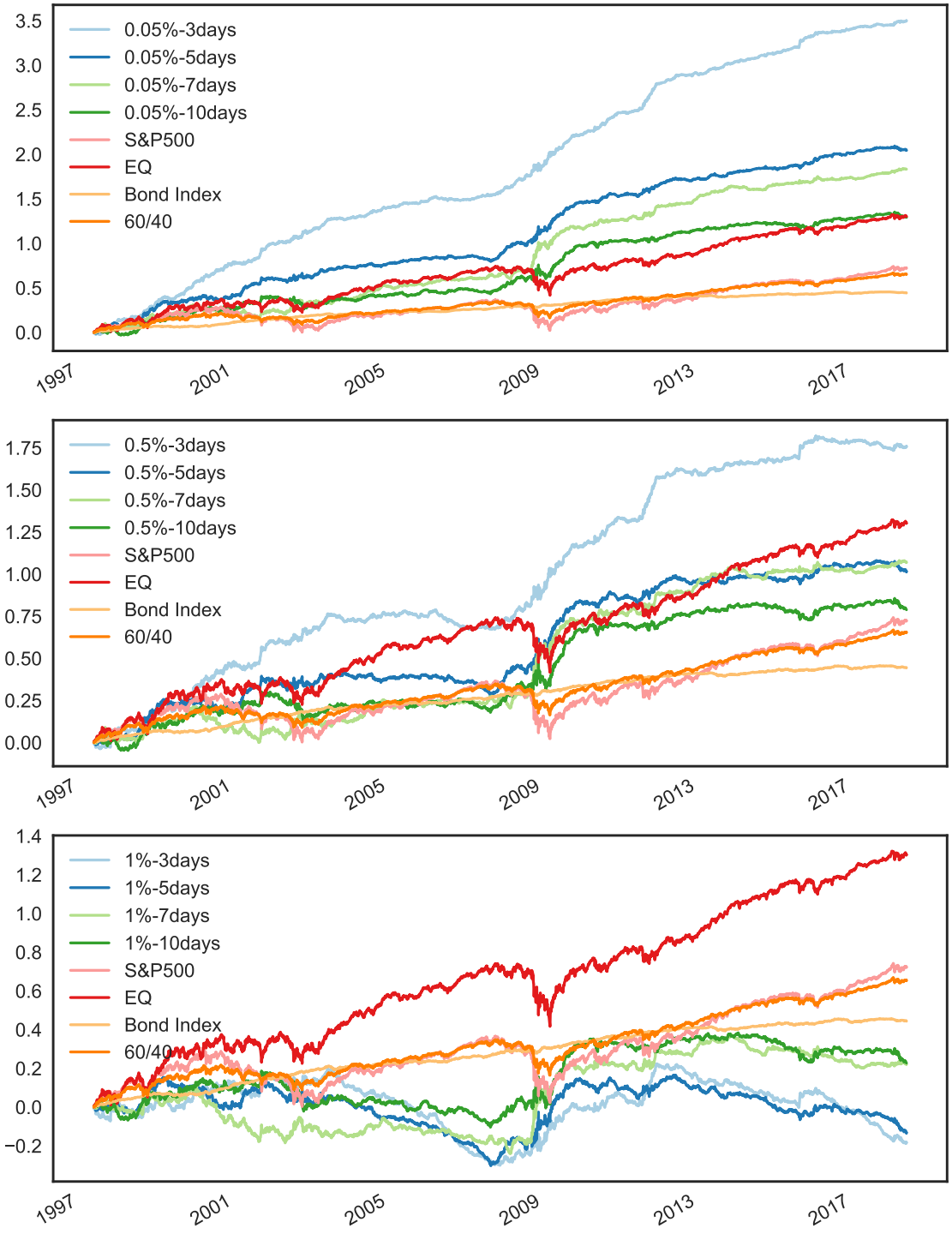

Figure 13: Quantitative strategy with ten stocks and fees

all the portfolios are able to outperform the SP500 in log-scaled measurements. Once the fee increases by 10 times to 0.5%, there is a significant downshift for the curves but they still gain more than the market portfolio. Once the fee fixes at 1%, there is no portfolio able to deliver more than the SP500. It is interesting to note how the ripples of the curves

change. The most significant ripples can be seen during the last financial crisis and their magnitude increase as the fees increase. For 1% fee, there is a huge bounce back after the financial crisis which stops around 2013. As fees increase, all the observed portfolio start to be more noisy. At every point in time, the performances sway and oscillate among trends. As expected, a longer period among rebalances can help to dampen down the performance drag. For a 1% proportional fee, the 10 days portfolio is the most acceptable investment choice, but it is still unable to deliver more than the market option.

Regarding the equally weighted portfolio, it is highly inefficient when fees are kept low but it starts to gain positions once fees increase. When fees are at 1%, it becomes the best pick for maximizing the market capitalization of invested capital.

Table 14: Performane for 10 stocks quantitative strategy with fees

3 days	Portfolio			*7 days*	Portfolio		
Average	32.8%	13.6%	-7.8%	Average	16.5%	8.1%	-1.1%
Volatilty	15.6%	15.2%	15.1%	Volatilty	15.5%	15.4%	15.3%
SR	2.11	0.89	-0.51	SR	1.05	0.53	-0.07
Fees	0.05%	0.50%	1.00%	Fees	0.05%	0.50%	1.00%

5 days	Portfolio			*10 days*	Portfolio		
Average	17.6%	6.2%	-6.4%	Average	9.8%	4.2%	-2.0%
Volatilty	15.4%	15.2%	15.2%	Volatilty	15.3%	15.1%	15.1%
SR	1.14	0.41	-0.42	SR	0.64	0.27	-0.13
Fees	0.05%	0.50%	1.00%	Fees	0.05%	0.50%	1.00%

Performances start to fall down as fees increase, leading also to negative expected returns once fee touches 1%.

There is no portfolio able to statistically overperform the market portfolio except for the 3 days portfolio when proportional fees amount to 0.05% of financial returns. Including fees showed and confirmed the downside of the strategy.

Table 15: t-test for 10 stocks quantitative strategy with fees

Fee 0.05%	3 days	5 days	7 days	10 days
t-stat	5.036	2.33	1.98	0.9706
p-value	0.000	0.019	0.047	0.330

Fee 0.5%	3 days	5 days	7 days	10 days
t-stat	1.826	0.44	0.55	0.022
p-value	0.060	0.650	0.576	0.981

Fee 1%	3 days	5 days	7 days	10 days
t-stat	-1.78	-1.677	-1.03	-1.03
p-value	0.073	0.093	0.299	0.301

9 Conclusion

This document focused on the implementation and developement of a quantitative momentum strategy. Instead of just considering the past months movements of stock prices, the investment signals were based on forecast estimations of asset prices. The predictive power has been borrowed from the Fama French three factors model in a state space representation of their asset pricing regression. State space modelling allowed to consider regression coefficients in their time dimensions and the Kalman filter recursion has able to extrapolate the unobservable factors of the Fama French three factors and filter them out from noisy observations.

The final result of performance analysis showed how two stocks are not enough to create a reliable and profitable strategy in a controlled environment with no market frictions like fees, unless the rebalancing time span is kept short. This seems to be unrealistic for most investors. Volatility is another major concern. To diversificate and improve the risk-adjusted performance, the ten stocks strategy has been evaluated on a frictions-free environment and in presence of proportional fees. Increasing the total number of stocks for the quantitative strategy created a massive shift in the risk-adjusted evaluation. The major improvement is on the volatility measure of the strategy. Diversification is a renowed concept for passive investment strategies and it can also be applied in the current example. A one sample t-test showed how the investment strategy for ten stocks is able to overperform the market portfolio with statistical significance. When fees are

introduced, performance start to suffer major losses. If fees are low, the portfolios are still able to outperform the market portfolio with statistical significance. When fees increase, the strategies start to be unprofitable. An equally weigthed portofolio is able to create a major utility gain when fees start to increase. This reduces the profitability of this active strategy.

The strategy is suitable for investors able to minimize the impact of transaction costs and are able to trade within short horizons. However, a more concerned analysis accounting for more transaction costs and market imperfections might help to assess a more reliable analysis of the strategy performances. For example, bid-ask spread can be used as a proxy of the liquidity of an asset in order to pick the most liquid stocks. Including liquid stocks in the quantitative strategy will reduce the impact of liquidity costs. A supplementary analysis of the selected stocks might also help to improve the performances of the strategy. Diversification should help to reduce the volatility of the strategy.

To summarize the results found in the document, the Fama French three factors dynamic linear model is capable of producing reliable signals for developing a quantitative equity strategy based on momentum and the Kalman filtering recursion. The strategy is profitable in a controlled environment. However, transaction costs have a massive impact on the performance of the strategy.

10 Further Studies and Criticisms

10.1 Model checking

The model applied in this document is a linear state space model representation of a factor model regression where the regression coefficients are estimated through the kalman filter recursion. As seen before, the DLM one-step forecasted value for the observable is $E(Y_t|y_{1:t-1})$. A way of checking the underlying model is to check the process property of the innovations, defined as:

$$e_t = y_t - E(Y_t|y_{1:t-1}) = y_t - F_t a_t$$

This is defined as the forecast error of the forecast observable variable, which has been crucial for the rebalancing process of the quantitative investment strategy. In DLM, e_t is expected to be a Gaussian process. To check this property, it is required to standardize e_t in order to model it as a standard Gaussian multivariate distribution. Since e_t is a vector of forecast error, the basic univariate standardization process is not precise. To standardardize the vectors, the process becomes:

1. Calculate $Var(e_t) = F_t C_t F_t^{'} + V$
2. Decompose $Var(e_t)^{-1}$ by Cholesky decomposition so that $Var(e_t)^{-1} = G_t G_t^{'}$
3. Standardize the error vectors $G_t^{'} e_t \sim N(0, \Sigma)$ where Σ is an identity matrix
4. Create a univariate sample of standardized errors by collecting every univariate error in every error vector for the entire observated time

Errors are far to be related to a $N(0, 1)$. They resemble a t-student distribution. This may be because the historical distributions of the asset returns are not normally distributed, but they are distributed as t-student random variables. This leads to fatter tails and higher probability mass of extremely high and extremely low returns. This finding has been enlighted in financial economics literature and it is a major drawdown of most of financial economics theories. To properly deal with this property of empirical returns, it would have been required to develop a nonlinear State Space Model where the observation variable is distributed as a t-student random variable. This has not be done because a nonlinear State Space Model is no more a DLM, and all the property of the Kalman filter

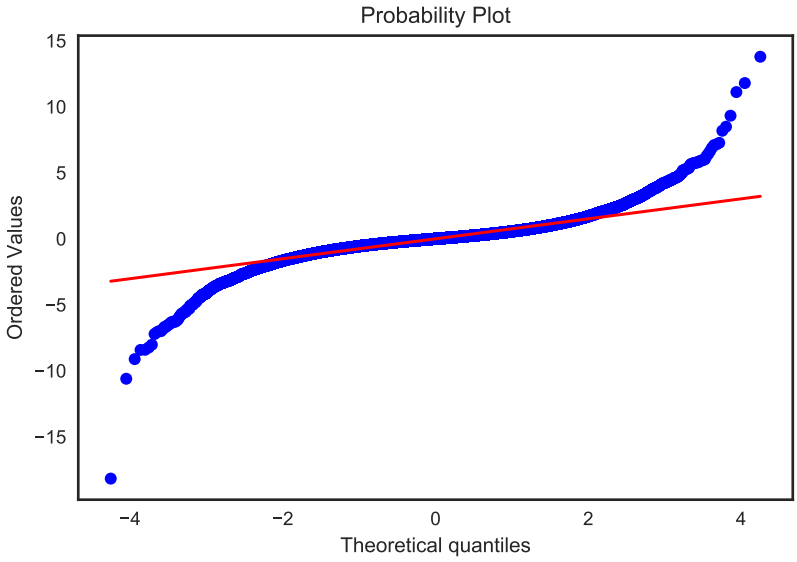

Figure 14: Filtered and smoothed β_t^M

are no more applicable. The computational requirements for such SSM are much more involving and high than the once required to produce this document. Describing empirical asset returns with normal distributions has been a widely and accepted approximation which helps to model the economics underpinning financial econometrics. By concluding, a possible extension of this thesis would be expanding the analysis to a nonlinear SSM for factor investing.

10.2 Out-of-sample Investment Performance

The quantitative investment strategy has been developed by estimating the error covariances V and W for a restricted period, in particular the timespan for their inference was 10 years for the two stocks model and 5 years for the ten stocks model. The difference has been decided to ease the simulations, as described above. Evaluating the investment strategy for the remaining periods has been a robust approximation for assessing an out-of-sample measurement of the investment performance. The unique drawdown of the out-of-sample property is found in the calculation of $y_{t|t-1}$. It is calculated as the best estimate of the future asset return for time t, conditioned on being at time $t-1$, but in its formulation $E(y_{t|t-1}) = F_t a_t$ meaning that the prediction incorporates the future development of the regressors of the Fama French model. To have a completely robust out-of-sample measurement of the investment strategy, the matrix F_t needs to be estimated.

10.3 Monte Carlo Particle Filter

As stated above, in order to parametrize the unknown covariances through Bayesian inference, the Gibbs sampler has to simulate a complete Markov chain of unobservable states θ_t over the time dimension. This powerful tool for estimating the chains required to sample the covariances are highly inefficient for sequential updating. Stricly speaking, every time a new observation period t kicks in, it forces the Gibbs sampler to create a new Markov chain of unobservables. In other words, if the sampler simulates a chain from time $t = T, T-1, .., 0$ and a new observation hits at time $t = T+1$, the previously simulated chain is useless since the starting point of the memoryless process starts now at time $t = T + 1$. In the last decade, new methods for a more efficient sequential update have been developed. Recently, a new simulation approach called Sequential Monte Carlo has been adopted by econometricians. Sequential Monte Carlo provides an alternative to the Gibbs sampler recursion[10]. This simulation method is related to the family of particle filter, a family of Monte Carlo algorhytm applied in signal processing and Bayesian statistics. The Bayesian inference for the errors covariances can be estimated with this approach and a possible estension could be comparing the two methods and the effects on filtered estimates.

[10] Liu(2001).

11 Appendix

11.1 Graphs for SUR and MCMC diagnostics

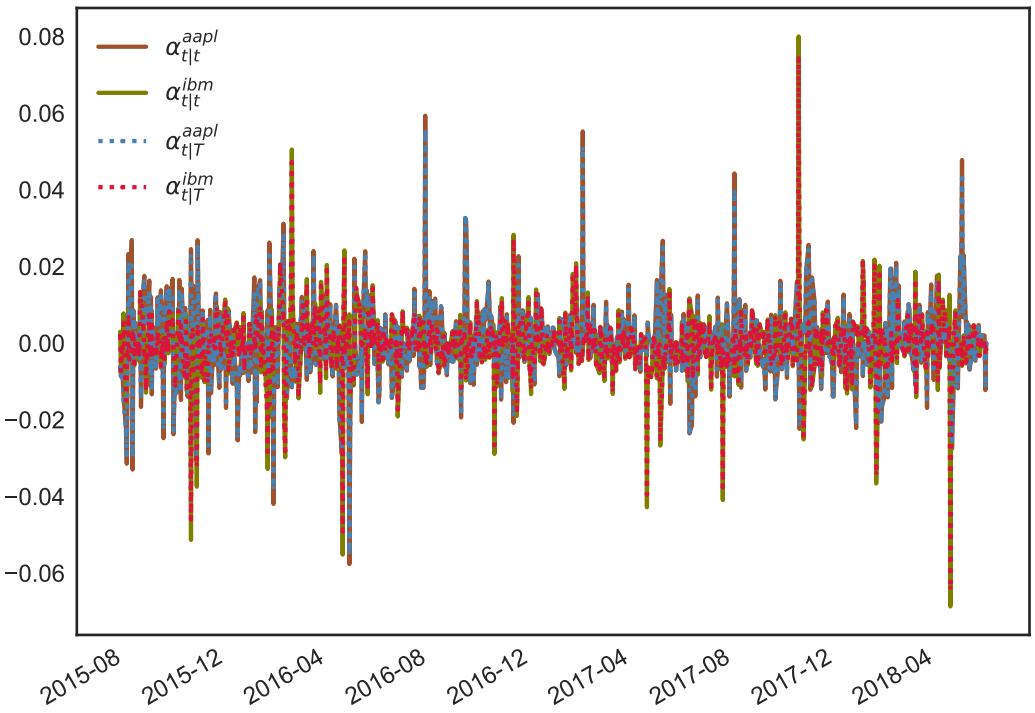

Figure 15: Filtered and smoothed β_t^M

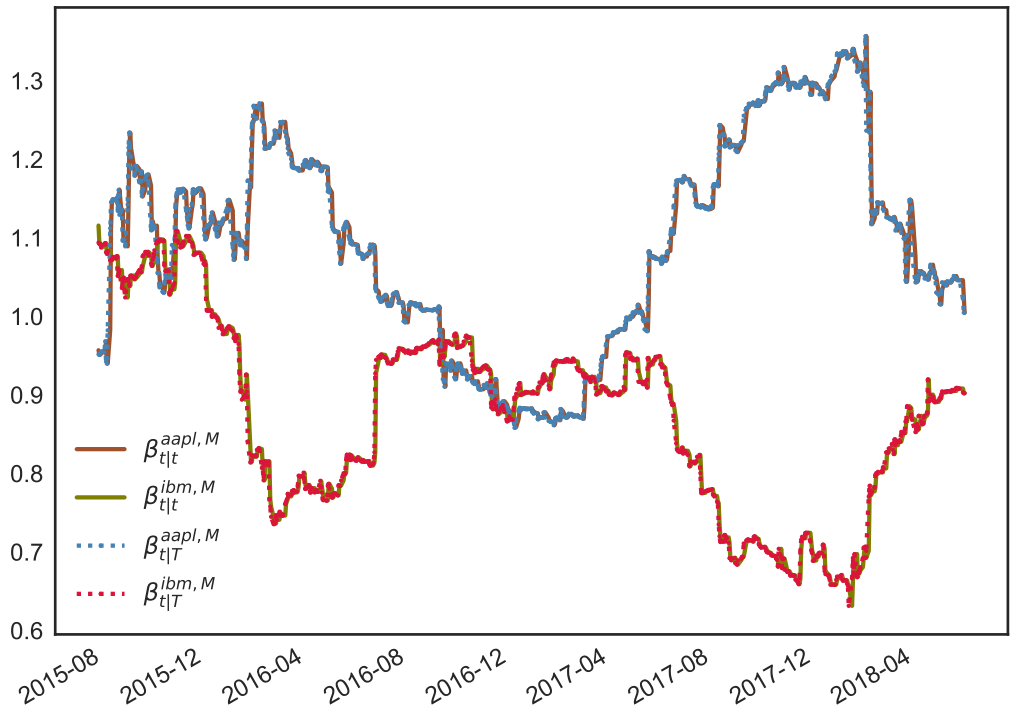

Figure 16: Filtered and smoothed β_t^M

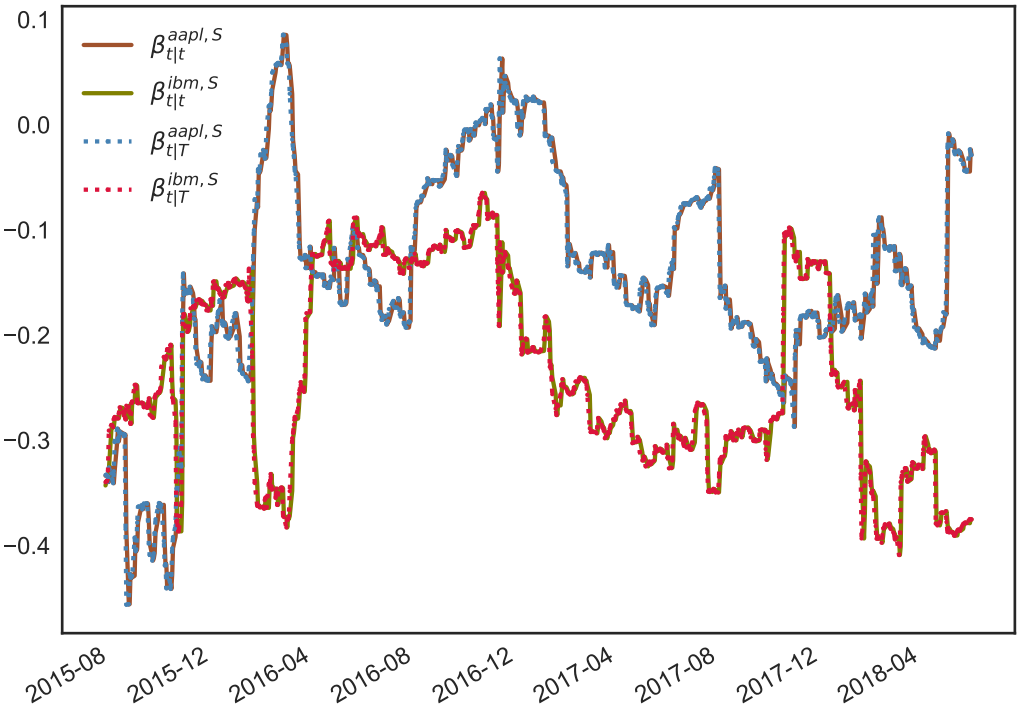

Figure 17: Filtered and smoothed β_t^S

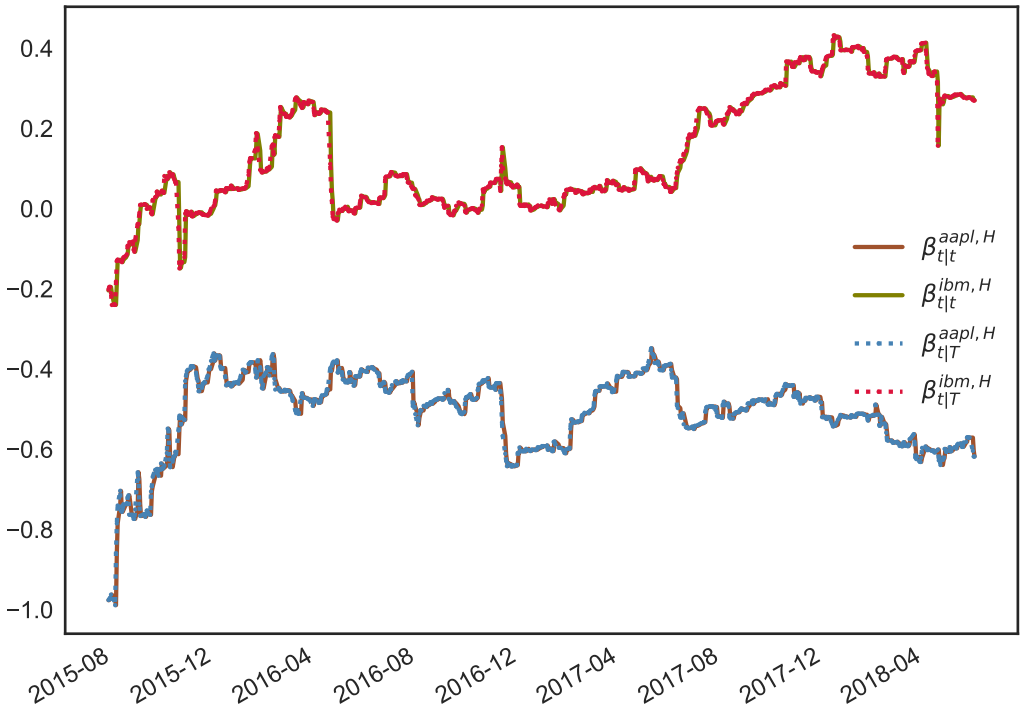

Figure 18: Filtered and smoothed β_t^H

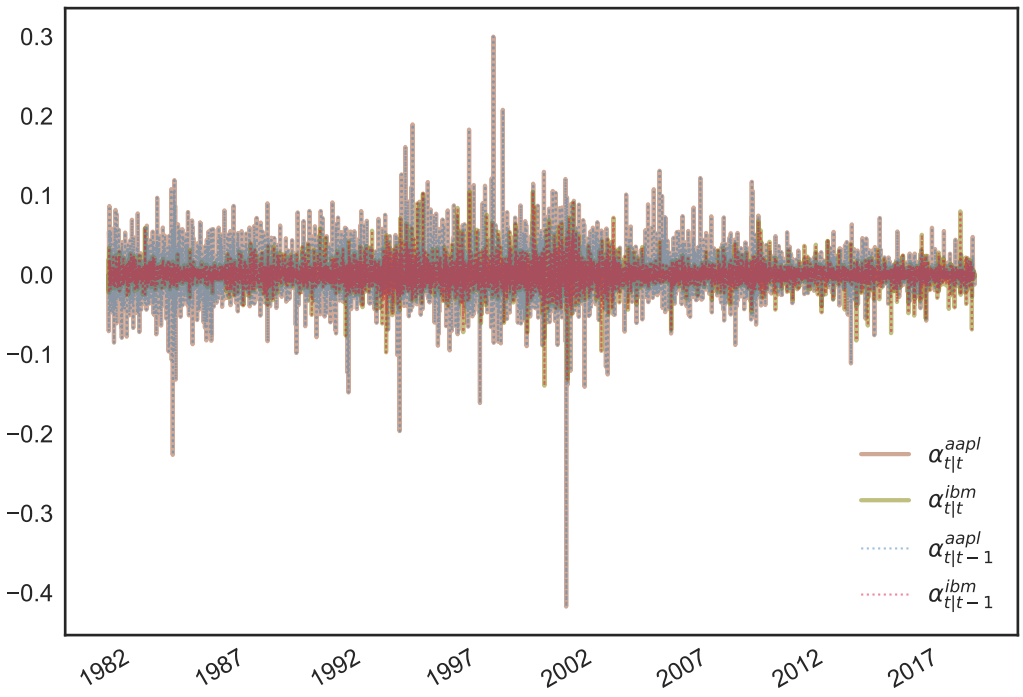

Figure 19: Filtered and smoothed α_t

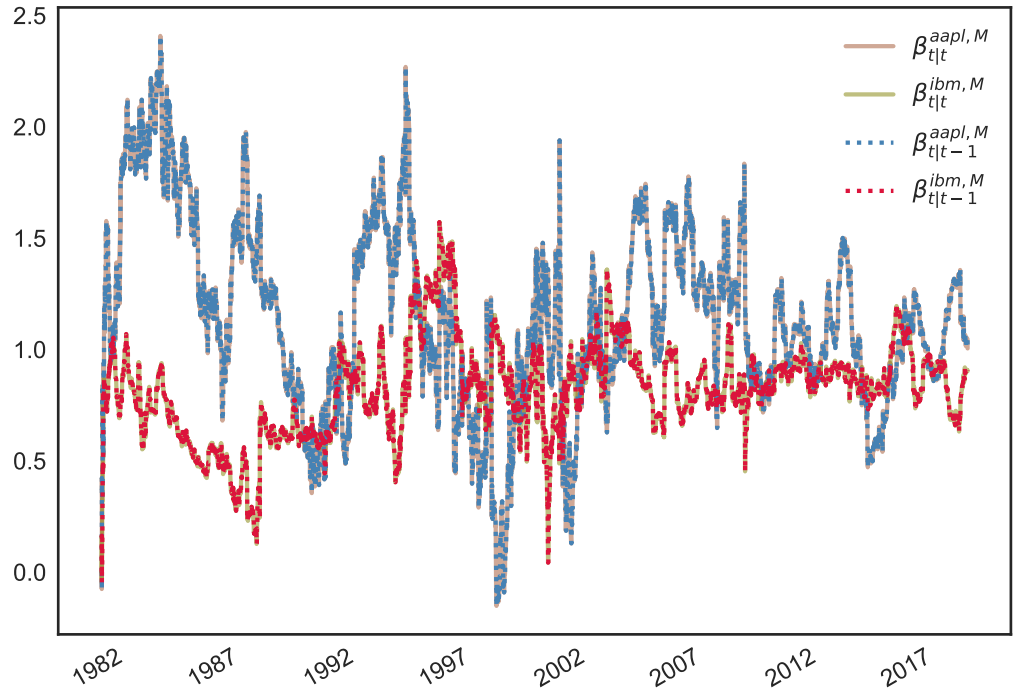

Figure 20: Filtered and smoothed β_t^M

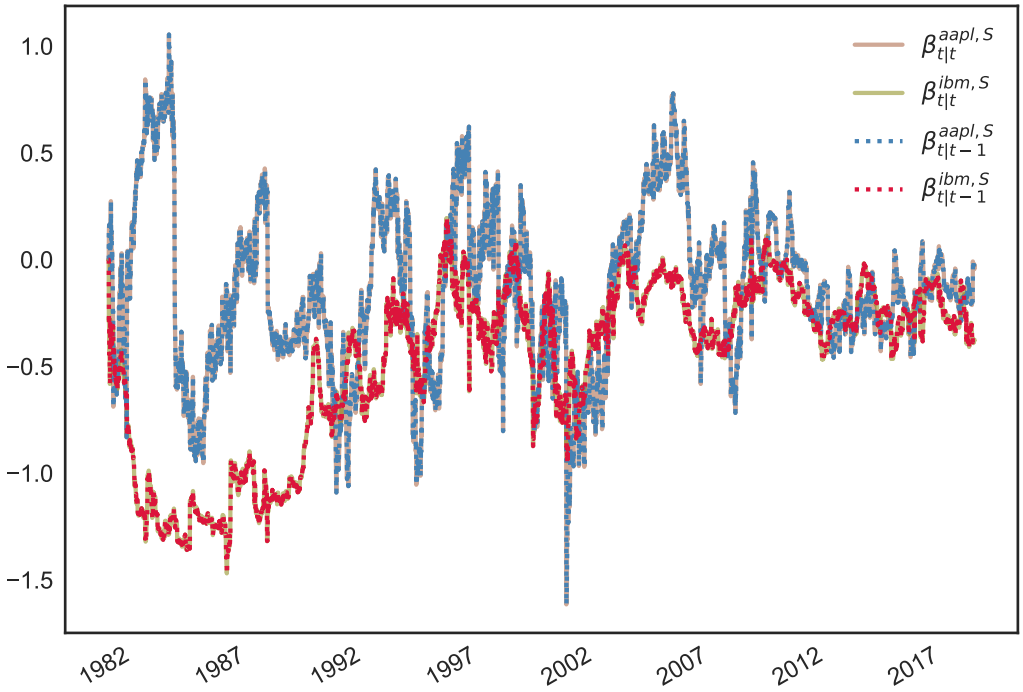

Figure 21: Filtered and smoothed β_t^S

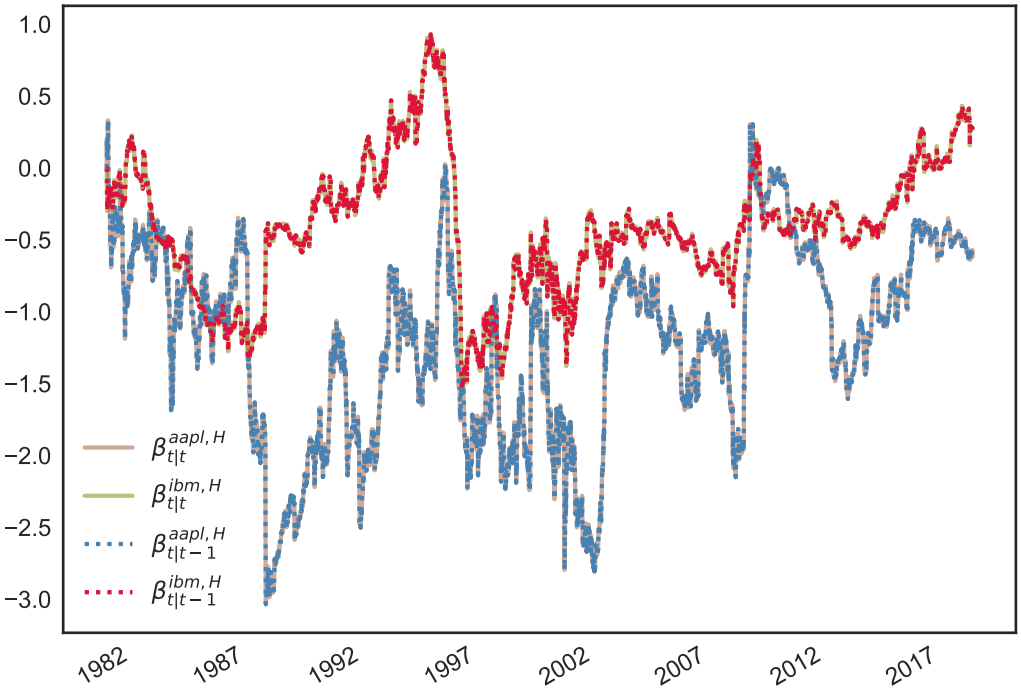

Figure 22: Filtered and smoothed β_t^H

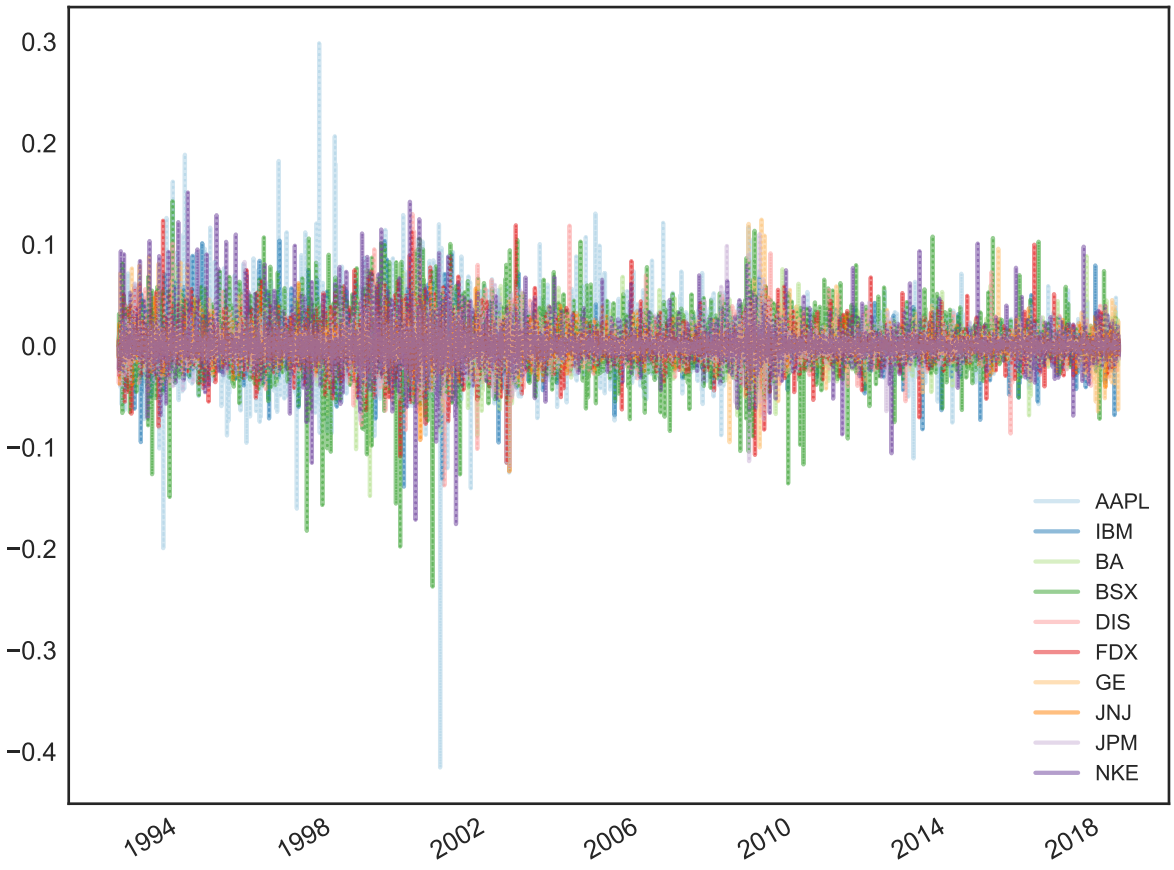

Figure 23: Filtered and smoothed α_t

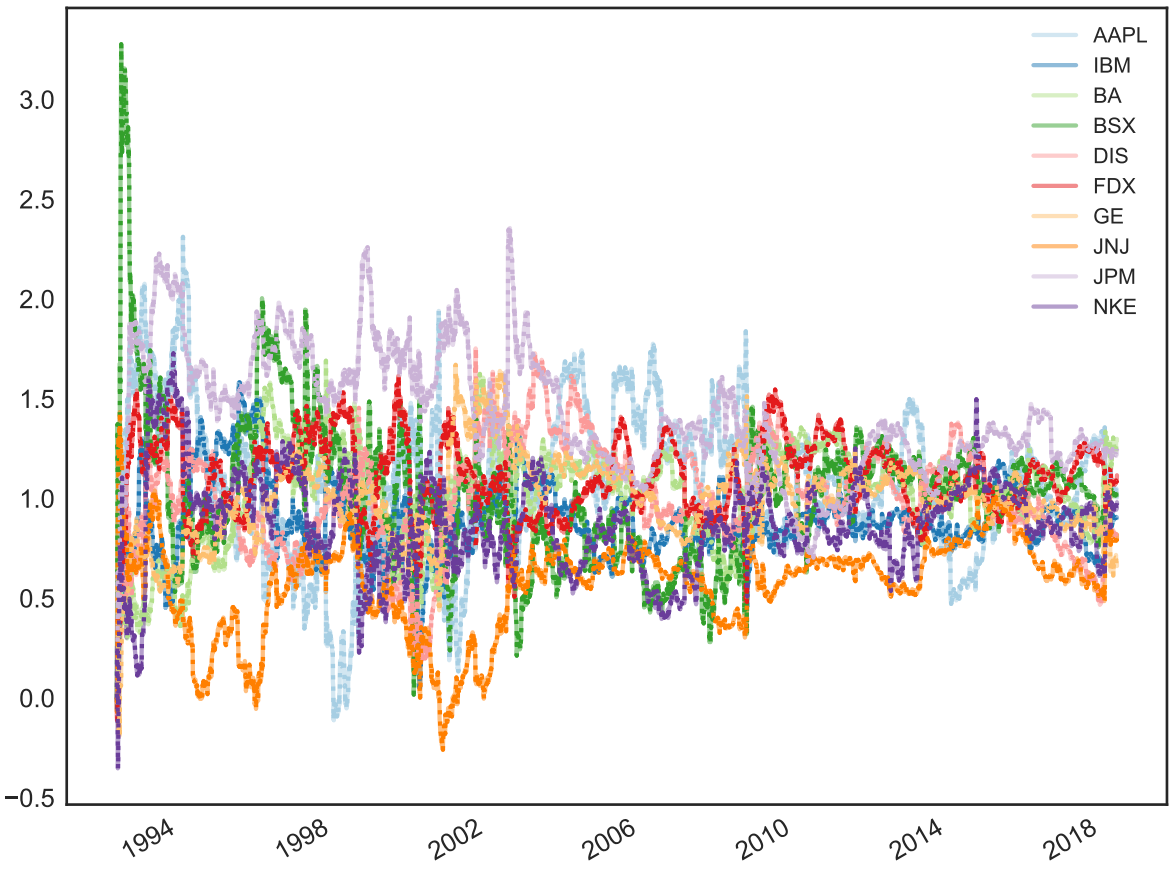

Figure 24: Filtered and smoothed β_t^M

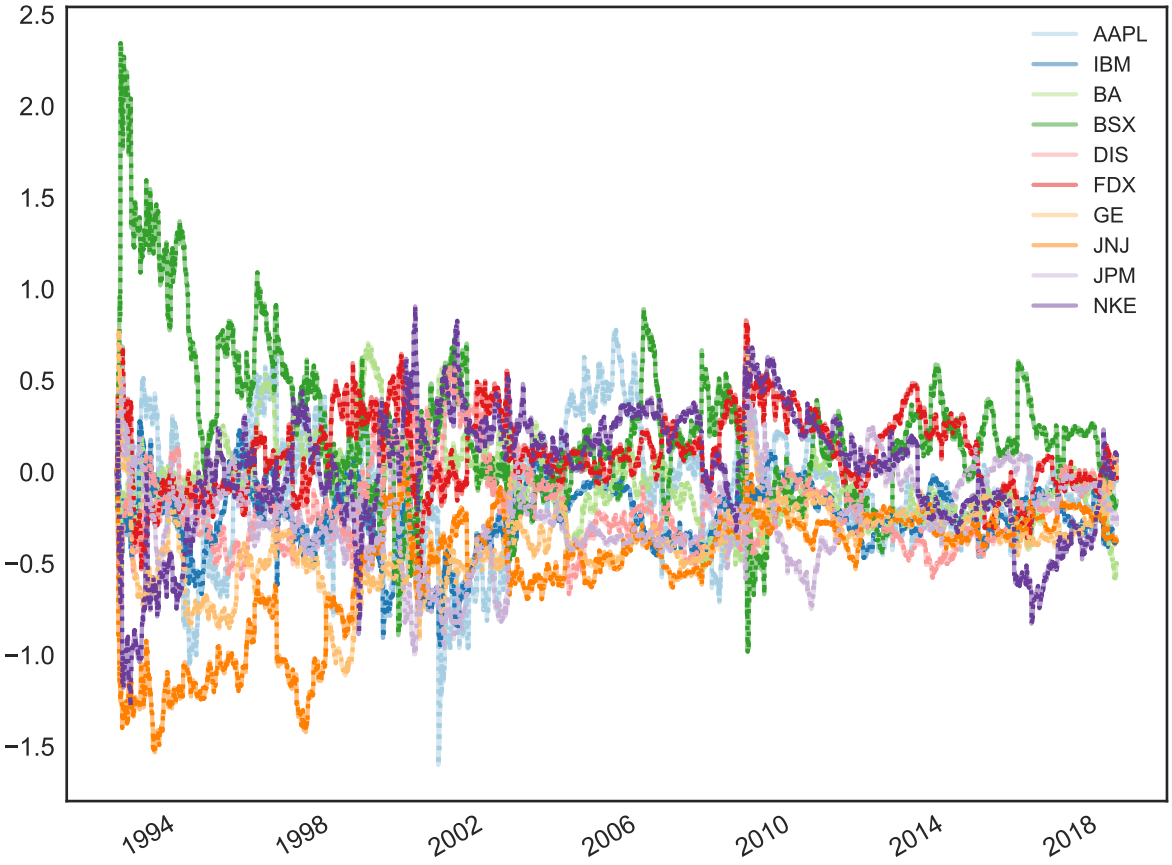

Figure 25: Filtered and smoothed β_t^S

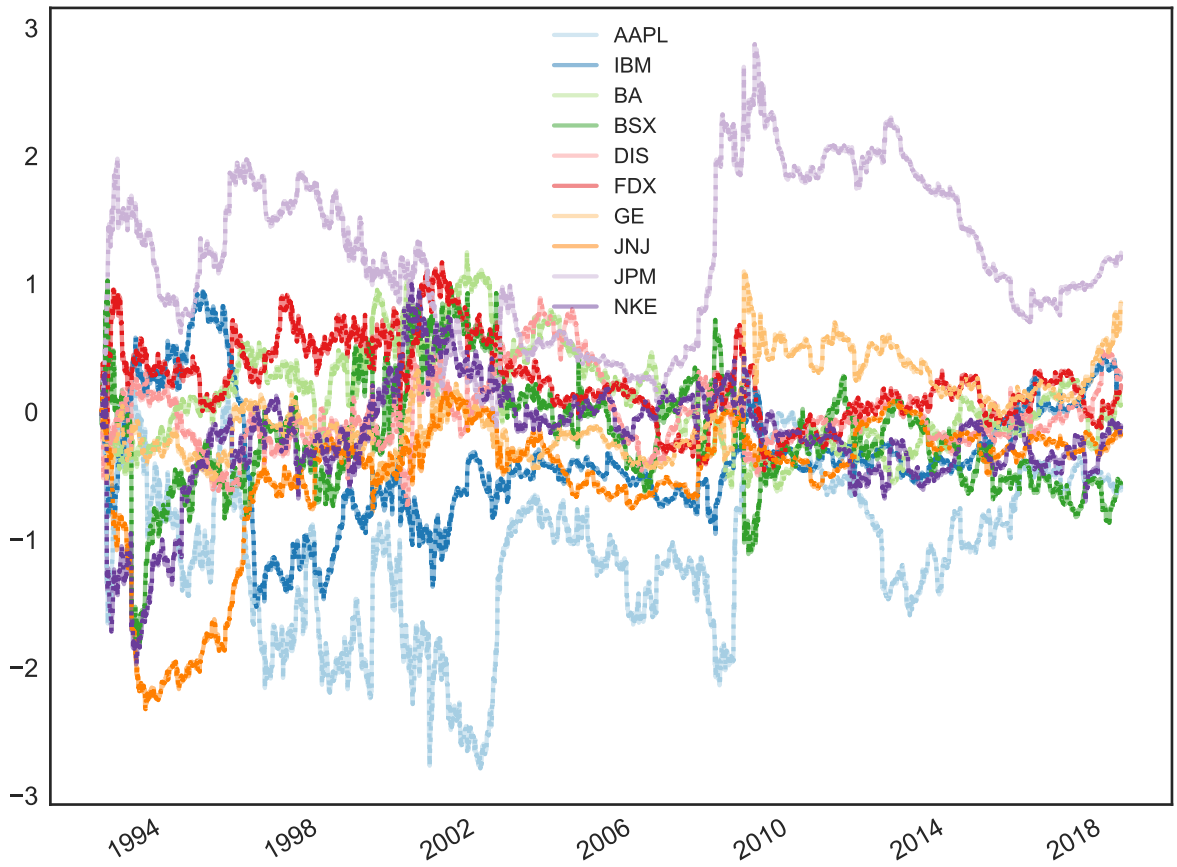

Figure 26: Filtered and smoothed β_t^H

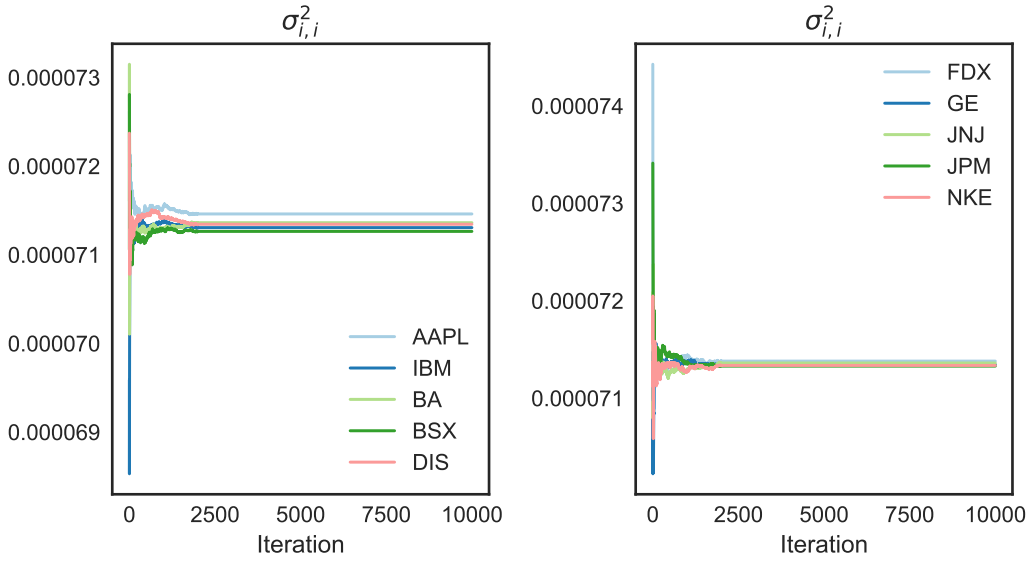

Figure 27: MCMC diagnostics for observable error covariances v_i

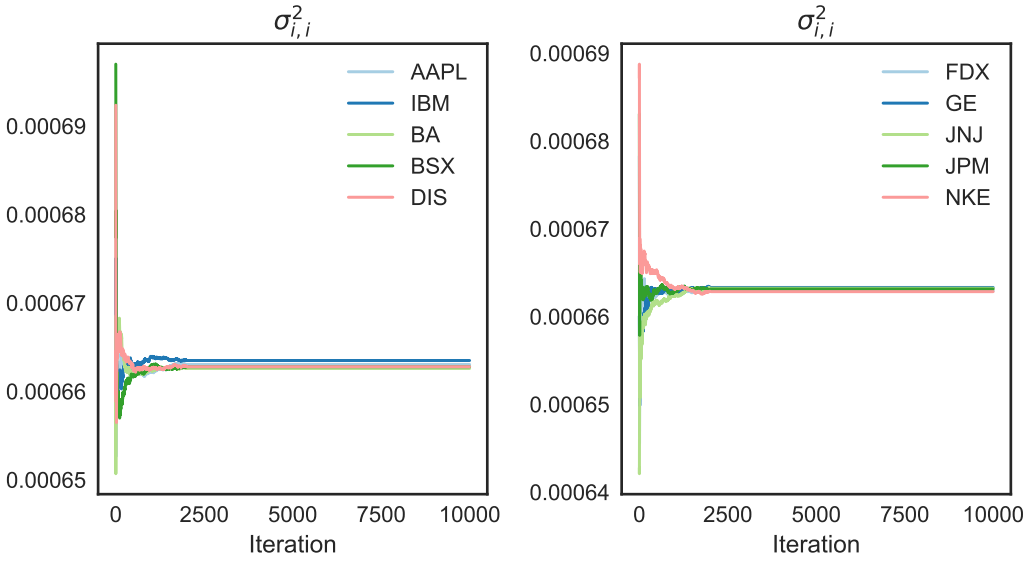

Figure 28: MCMC diagnostics for unobservable error covariances for α

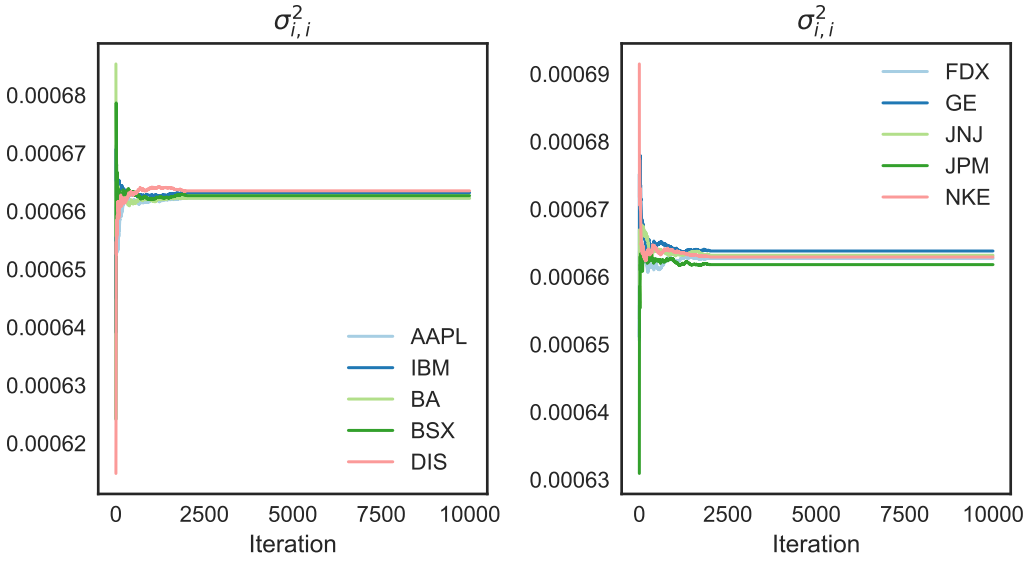

Figure 29: MCMC diagnostics for unobservable error covariances for β^M

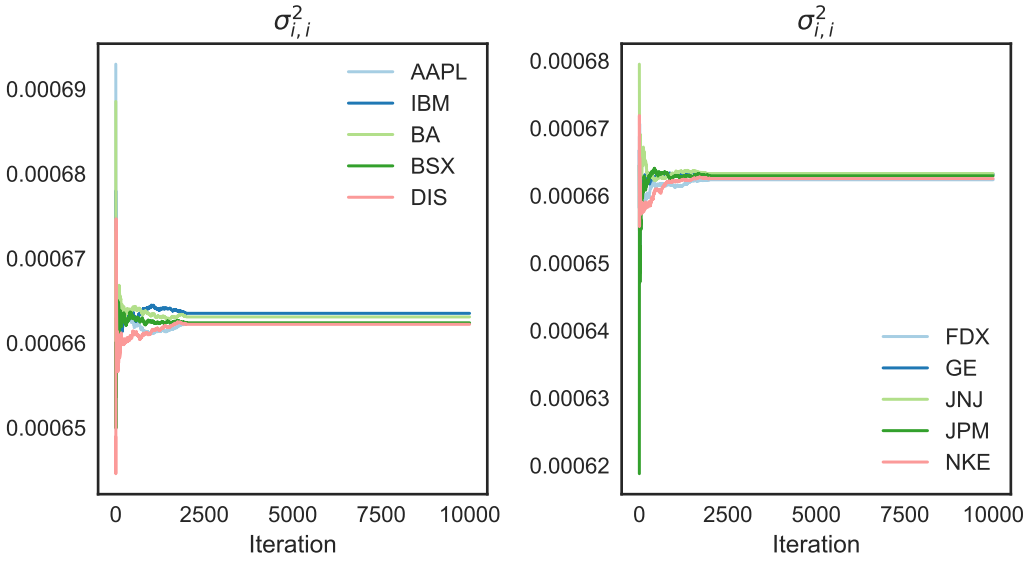

Figure 30: MCMC diagnostics for unobservable error covariances for β^S

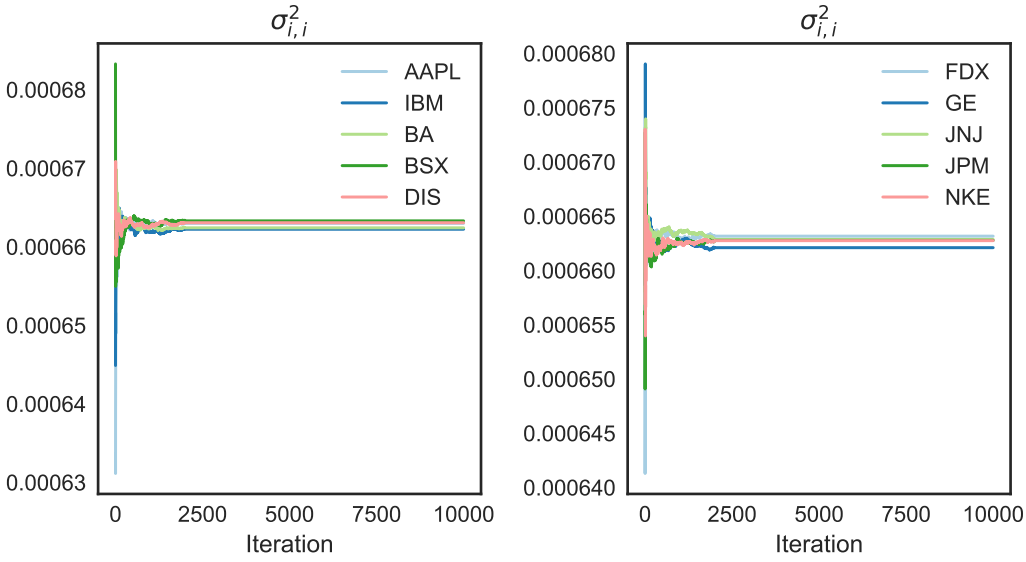

Figure 31: MCMC diagnostics for unobservable error covariances for β^H

Figure 32: Rolling coefficients 10 stocks - Part 1

Figure 33: Rolling coefficients 10 stocks - Part 2

11.2 Dynamic Linear Model - 10 Time Series

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \\ y_{6,t} \\ y_{7,t} \\ y_{8,t} \\ y_{9,t} \\ y_{10,t} \end{pmatrix} = \left(\begin{pmatrix} 1 & r_t^M & r_t^S & r_t^H \end{pmatrix} \otimes I_{10} \right) \begin{pmatrix} \alpha_t \\ \beta_t^M \\ \beta_t^S \\ \beta_t^H \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \\ v_{4,t} \\ v_{5,t} \\ v_{6,t} \\ v_{7,t} \\ v_{8,t} \\ v_{9,t} \\ v_{10,t} \end{pmatrix} \qquad \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \\ v_{4,t} \\ v_{5,t} \\ v_{6,t} \\ v_{7,t} \\ v_{8,t} \\ v_{9,t} \\ v_{10,t} \end{pmatrix} \sim N(\mathbf{0}, V)$$

$$\begin{pmatrix} \alpha_t \\ \beta_t^M \\ \beta_t^S \\ \beta_t^H \end{pmatrix} = \left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \otimes I_{10} \right) \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1}^M \\ \beta_{t-1}^S \\ \beta_{t-1}^H \end{pmatrix} + \begin{pmatrix} w_t^{\alpha} \\ w_t^{\beta^M} \\ w_t^{\beta^S} \\ w_t^{\beta^H} \end{pmatrix} \qquad \begin{pmatrix} w_t^{\alpha} \\ w_t^{\beta^M} \\ w_t^{\beta^S} \\ w_t^{\beta^H} \end{pmatrix} \sim N(\mathbf{0}, W)$$

where

$$\alpha_t = \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \\ \alpha_{4,t} \\ \alpha_{5,t} \\ \alpha_{6,t} \\ \alpha_{7,t} \\ \alpha_{8,t} \\ \alpha_{9,t} \\ \alpha_{10,t} \end{pmatrix} \quad \beta_t^M = \begin{pmatrix} \beta_{1,t}^M \\ \beta_{2,t}^M \\ \beta_{3,t}^M \\ \beta_{4,t}^M \\ \beta_{5,t}^M \\ \beta_{6,t}^M \\ \beta_{7,t}^M \\ \beta_{8,t}^M \\ \beta_{9,t}^M \\ \beta_{10,t}^M \end{pmatrix} \quad \beta_t^S = \begin{pmatrix} \beta_{1,t}^S \\ \beta_{2,t}^S \\ \beta_{3,t}^S \\ \beta_{4,t}^S \\ \beta_{5,t}^S \\ \beta_{6,t}^S \\ \beta_{7,t}^S \\ \beta_{8,t}^S \\ \beta_{9,t}^S \\ \beta_{10,t}^S \end{pmatrix} \quad \beta_t^H = \begin{pmatrix} \beta_{1,t}^H \\ \beta_{2,t}^H \\ \beta_{3,t}^H \\ \beta_{4,t}^H \\ \beta_{5,t}^H \\ \beta_{6,t}^H \\ \beta_{7,t}^H \\ \beta_{8,t}^H \\ \beta_{9,t}^H \\ \beta_{10,t}^H \end{pmatrix}$$

$$w_t^{\alpha} = \begin{pmatrix} w_{1,t}^{\alpha} \\ w_{2,t}^{\alpha} \\ w_{3,t}^{\alpha} \\ w_{4,t}^{\alpha} \\ w_{5,t}^{\alpha} \\ w_{6,t}^{\alpha} \\ w_{7,t}^{\alpha} \\ w_{8,t}^{\alpha} \\ w_{9,t}^{\alpha} \\ w_{10,t}^{\alpha} \end{pmatrix} \quad w_t^{\beta^M} = \begin{pmatrix} w_{1,t}^{\beta^M} \\ w_{2,t}^{\beta^M} \\ w_{3,t}^{\beta^M} \\ w_{4,t}^{\beta^M} \\ w_{5,t}^{\beta^M} \\ w_{6,t}^{\beta^M} \\ w_{7,t}^{\beta^M} \\ w_{8,t}^{\beta^M} \\ w_{9,t}^{\beta^M} \\ w_{10,t}^{\beta^M} \end{pmatrix} \quad w_t^{\beta^H} = \begin{pmatrix} w_{1,t}^{\beta^S} \\ w_{2,t}^{\beta^S} \\ w_{3,t}^{\beta^S} \\ w_{4,t}^{\beta^S} \\ w_{5,t}^{\beta^S} \\ w_{6,t}^{\beta^S} \\ w_{7,t}^{\beta^S} \\ w_{8,t}^{\beta^S} \\ w_{9,t}^{\beta^S} \\ w_{10,t}^{\beta^S} \end{pmatrix} \quad w_t^{\beta^H} = \begin{pmatrix} w_{1,t}^{\beta^H} \\ w_{2,t}^{\beta^H} \\ w_{3,t}^{\beta^H} \\ w_{4,t}^{\beta^H} \\ w_{5,t}^{\beta^H} \\ w_{6,t}^{\beta^H} \\ w_{7,t}^{\beta^H} \\ w_{8,t}^{\beta^H} \\ w_{9,t}^{\beta^H} \\ w_{10,t}^{\beta^H} \end{pmatrix}$$

Bibliography

[1] Albert J. , (2007), Bayesian Computation with R , Springer

[2] Banz R. , (1981), The Relationship between Return and Market Value of Common Stocks , Journal of Financial Economics 9(1), 3–18

[3] Banz R. , (1981), The Relationship between Return and Market Value of Common Stocks , Journal of Financial Economics 9(1), 3–18

[4] Black F. , (1972), Capital Market Equilibrium with Restricted Borrowing. , Journal of Business 45(3) 444–454

[5] Black F., Jensen M.C., Scholes M. , (1972), The Capital Asset Pricing Model: Some Empirical Tests , Studies in the Theory of Capital Markets, pp. 79–121

[6] Brandimarte P. , (2014), Handbook in Monte Carlo simulation: Applications in financial engineering, risk management, and economics , Wiley

[7] Campbell J. Y., Shiller R. J. , (1989), The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors , The Review of Financial Studies, Volume 1, number 3, pp. 195-228

[8] Davidson-Pilon C. , (2015), Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference , Addison-Wesley Professional

[9] Del Moral P. , Doucet A., Jasra A. , (2011), An adaptive sequential Monte Carlo method for approximate Bayesian computation , Springer

[10] Downey, A. , (2013), Think Bayes: Bayesian Statistics in Python , O'Reilly

[11] Fama E. , (1969), Efficient Capital Markets: A Review of Theory and Empirical Work , The Journal of Finance, Vol. 25, No. 2, pp. 383-417

[12] Fama E. , (1970), Efficient Capital Markets: A Review of Theory and Empirical Work. , Journal of Finance 25(2) 383–417

[13] Fama E., French K. , (2010), Luck versus Skill in the Cross-Section of Mutual Fund Returns , The Journal of Finance

[14] Fama E., French K. , (1992), The Cross-Section of Expected Stock Returns , Journal of Finance 47(2) 427–465

[15] Fama E., MacBeth J. D. , (1973), Risk, Return, and Equilibrium: Some Empirical Tests , Journal of Political Economy 81(3) 607–636

[16] Gârleanu N., Pedersen L. , (2015), Efficiently Inefficient Markets for Assets and Asset Management , NBER Working Paper No. 21563

[17] Geweke J. , (2005), Contemporary Bayesian Econometrics and Statistics , Wiley

[18] Geweke J.,Koop G., van Dijk H. , (2011), The Oxford handbook of Bayesian econometrics , Oxford University Press

[19] Greene W. , (2008), Econometric analysis , Pearson

[20] Hamilton, J. D. , (1994), Time Series Analysis , Princeton University Press

[21] Harvey A. C. , (1990), Forecasting, Structural Time Series Models and the Kalman Filter , Cambridge University Press

[22] Jagannathan R., Malakhov A., Novikov D. , (2010), Do Hot Hands Exist among Hedge Fund Managers? An Empirical Evaluation , The Journal of Finance, Volume 65

[23] Jensen M. , (1968), Problems in Selection of Security Portfolios , American Finance Association

[24] Kahneman D. , Tversky A. , (1979), Prospect Theory: An Analysis of Decisions under Risk. , Econometrica 47(2), 263–291

[25] Kalman R., Bucy R. , (1961), New Results in Linear Filtering and Prediction Theory , Journal of Basic Engineering

[26] Kosowski R., Naik N., Teo M. , (2007), Do hedge funds deliver alpha? A Bayesian and bootstrap analysis , Journal of Financial Economics, vol. 84, issue 1, 229-264

[27] Koopman S. J., Durbin J., (2012), Time Series Analysis by State Space Methods, Oxford Statistical Science Series, 38

[28] Kosowski R., Timmermann A., Wermers R., White H. ,(2006), Can Mutual Fund "Stars" Really Pick Stocks? New Evidence from a Bootstrap Analysis , The Journal of Finance

[29] Lintner J. , (1965), The Valuation of Risky Assets and the Selection of Risky Investment in Stock Portfolios and Capital Budgets. , Review of Economics and Statistics 47(1), 13-37

[30] Liu J. , (2008), Monte Carlo Strategies in Scientific Computing , Springer

[31] Markowitz H , (1952), Portfolio selection , The Journal of Finance, Vol. 7, No. 1. (Mar., 1952), pp. 77-91

[32] McKinney W. , (2017), Python for data analysis , O'Reilly

[33] Merton R. C. , (1969), Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case , The Review of Economics and Statistics, Vol. 51, No. 3 (Aug., 1969), pp. 247-257

[34] Merton R. C. , (1971), Optimum Consumption and Portfolio Rules in a Continuous-Time Model , Journal of Economic Theory 3, 373-413 (1971)

[35] Mossin, J. , (1966), Equilibrium in a Capital Asset Market. , Econometrica 34(4), 768-783

[36] Pedersen L. , (2015), Efficiently Inefficient , Princeton University Press

[37] Pedersen L. , (2018), Sharpening the Arithmetic of Active Management , Princeton University Press

[38] Petris G., Petrone S., Campagnoli P. , (2009), Dynamic Linear Models in R , Springer

[39] Rosenberg B. K. Reid R. Lanstein , (1985), Persuasive Evidence of Market Ine- ciency. , Journal of Portfolio Management 11, 9–16

[40] Samuelson P. , (1969), Lifetime Portfolio Selection By Dynamic Stochastic Programming , The Review of Economics and Statistics, Vol. 51, No. 3 (Aug., 1969), pp. 239-246

[41] Sargent T. , (1987), Dynamic macroeconomic theory , Harvard University Press

[42] Sharpe W. , (1964), Capital Asset Prices: A Theory of Market Equilibrium under Con- ditions of Risk , Journal of Finance 19(3), 425–442.

[43] Sharpe W. , (1964), Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk , The Journal of Finance, vol. 19, 425-442

[44] Sharpe W. , (1991), The Arithmetic of Active Management , Financial analyst journal, volume 47

[45] Suess, Eric A., Trumbo, Bruce E. , (2010), Introduction to Probability Simulation and Gibbs Sampling with R , Springer

[46] Timmermann A., Elliott G. , (1961), Economic forecasting , Princeton University Press

[47] Treynor J. , (2016), Market Value, Time, and Risk , Unpublished original document

[48] Treynor J., Black F. , (1973), How to Use Security Analysis to Improve Portfolio Selection. Journal of Business 46(1), 66–86. , Journa of Business 46(1), 66-86