Three Essays on Modeling Asset Returns: A Nonparametric Approach

THREE ESSAYS ON MODELING ASSET RETURNS: A NONPARAMETRIC APPROACH

BY

AZAM SHAMSI ZAMENJANI, M.Sc., B.Sc.

A thesis

submitted to the degroote school of business

and the school of graduate studies

of McMaster university

in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

Ph.D. Dissertation (2018) McMaster University

(DeGroote School of Business) Hamilton, Ontario, Canada

TITLE:	Three Essays on Modeling Asset Returns: A Nonparametric Approach
AUTHOR:	Azam Shamsi Zamenjani (DeGroote School of Business) McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. John Maheu
COMMITTEE:	Dr. Ronald Balvers
COMMITTEE:	Dr. Jeffrey Racine
NUMBER OF PAGES:	xxiii, 223

To my parents

Abstract

This thesis examines three important topics in empirical finance: The nonparametric conditional beta, the propagation of risks between markets, and the predictability of the market return density using economics and financial variables.

First, I introduce a model that links the conditional beta to the second order conditional moments of returns. A key insight of my approach is that if the joint distribution of stock returns and market returns are correctly specified, then it follows that their contemporaneous pricing relationship is completely determined by the associated conditional distribution. The model that I propose is able to study the effects of a big shock in the market on the beta of an individual stock or a portfolio. This approach allows the beta to be a flexible function of the sign and size of the market portfolio which is absent in the finance literature. My results demonstrate that beta does depend on the market portfolio in a nonlinear manner. This carries implications for systematic risk measurement significantly different than what we have in a fixed parametric model. My model nests the Gaussian and Student-t distribution as special cases but importantly allows for deviations from the elliptic family of distributions. This includes asymmetric distributions. I extend the model to include more assets and provide a test to see if other factors are priced. My empirical results illustrate that in the event of big shocks in the market the firm's beta coefficient can go down or shoot up, depending on the market conditions (volatility).

Second, I extend the literature on the spillover effects or contagion effects that focus on

the transmission of shocks through moments to spillover effects on the conditional density and seek to shed greater light on the contemporaneous information transmission between markets. The objective is to develop a joint model of returns on two different markets governed by an infinite mixture model from which the conditional density of the first market given the returns of the other market can be derived. This enables us to study how a shock in one market influences the contemporaneous and future (one-day-ahead, one-week-ahead, and one-month ahead) conditional density of the other market. This makes it possible to explore the contemporaneous spillover effects of big shocks in one market on various features of the density of the other market such as conditional expected return, volatility, skewness and kurtosis, and value at risk.

Third, I investigate the predictability of the market return density using financial and macroeconomic variables. The objective is to develop a Bayesian nonparametric model of the distribution of market returns where the weights of the mixture change over time. Available information on financial and macroeconomic variables is employed to predict the weights of the mixture components in the predictive density of market returns over time. This permits the density of market returns to be unknown and to change over time. Moreover, the proposed model examines whether certain financial and macroeconomic variables convey any useful information about the predictive density which extends the literature on an important question in empirical finance, the predictability of market return. Using the proposed stick-breaking model, instead of focusing only on the conditional mean or variance of market return, I investigate whether these variables are useful in predicting the entire density of stock returns through predicting the time-varying weights of the mixture density. I seek to predict features of the market return density in addition to what is captured by the first and second moments. This matters in empirical distributions in finance and economics where we might have heavy tails and asymmetry. I provide evidence that these features can be of great value in applications such as asset allocation and risk management. I evaluate the

incremental improvement in the predicted density of market returns statistically by the log predictive likelihood criteria and economically by a portfolio selection application.

Acknowledgement

I would never have been able to finish my dissertation without the constructive guidance of my advisor and committee members, the tremendous help from my friends, and the ceaseless support from my beloved family.

I would first and foremost like to thank my supervisor, Professor John Maheu, for being a tremendous mentor, and for his immense help with this dissertation and developing my research and writing skills. I have learned a great deal from him about finance and academic research in general. I have had his full support in all phases of my Ph.D. studies. He has been encouraging me in my research projects and taught me great lessons on teaching strategies.

I would also like to express my special appreciation to my committee members, Dr. Ron Balvers and Dr. Jeffrey Racine. They have generously given their time and expertise to review my dissertation at every stage. Their insightful comments and brilliant suggestions have helped me to improve the quality of this dissertation. I would like to thank Professors Clarence Kwan, Peter Miu, Jiaping Qiu, Narat Charupat, Trevor Chamberlain, and Anna Danielova who have taught me Finance courses and attended the in-house seminars at which I presented my research at various stages during the past five years.

I would like to thank my husband, Alireza, for his support and enthusiasm. He has been my source of strength and inspiration. Last but certainly not least, I take this opportunity to express my appreciation to all members of the DeGroote School of Business. I have had great pleasure in working and living with them. My special thanks go to Deb Randall Baldry and my Ph.D. fellows, Hadi, Jia, Robin, and Samira, for creating a fun atmosphere and helping me get through difficult times. I wish my fellow Ph.D. candidates every success and thank them for their support. This has been a wonderful place to pursue knowledge.

Chapter 2 is based on joint work with Professor John Maheu.

Contents

List of Figures

Chapter 1

Introduction

This thesis examines three important topics in financial econometrics with a particular focus on estimating an accurate model for financial asset returns using Bayesian nonparametric methods. These topics are of interest to individual investors, managers and policy makers. Informed decisions must take into account the possibility that asset values may increase or decrease as well as the likelihood of these events. This work aims to advance the methodology of modeling asset returns that allows the innovation distribution to be unknown and to change over time. Most of the literature has concentrated on modeling the conditional (co)variance of returns in forms of stochastic volatility and (multivariate) generalized autoregressive conditional heteroskedasticity (GARCH) models with a fixed parametric innovation density. This means that once volatility dynamics are removed, the innovations distribution is constant over time.

This research considers the estimation of an unknown density for the return innovations rather than making a specific assumption. The Bayesian nonparametric approach relies on estimating an infinitely countable mixture of distributions to approximate the true underlying density and yields exact finite sample inference. This results in more accurate predictive densities for individual assets and portfolios of assets. It also provides better measures of risk.

In the first essay, I design and estimate a multivariate Bayesian nonparametric model to study the conditional beta in the capital asset pricing model (CAPM) framework. This study extends recent work that assumes beta changes over time but is fixed at time t as a function of the contemporaneous value of market return. This model is able to study how big shocks in the market affect the beta of an individual stock or a portfolio. There is a large literature that derives testable implications from conditional CAPM. One route to a conditional CAPM is from Hansen and Richard (1987). This model links the second order conditional moments of returns to the conditional beta which determines the risk premium of an asset. One strand of the literature (Bollerslev et al., 1988; Bali and Engle, 2010; Engle, 2016) has investigated this model through a multivariate GARCH model (MGARCH). Recently Engle (2016) proposes a multivariate normal GARCH model from which the conditional distribution defines the dynamic beta coefficient. This directly links time-varying second moments to the time-varying beta in a consistent fashion. The parametric pricing relationship holds more generally for the elliptic family of distributions. This is an attractive approach but may be limiting if the parametric distributional assumptions are not valid.

A key insight of my approach is that if the joint distribution of stock returns and market returns are correctly specified then it follows that their contemporaneous pricing relationship is completely determined by the associated conditional distribution. I semiparametrically model the conditional distribution as a countably infinite mixture of normals (Jensen and Maheu, 2013) using a Dirichlet process prior (Ferguson, 1973). Each normal component in the mixture has a conditional covariance directed by a MGARCH process. The mixing is over both the mean vector and the covariance matrix. Parsimony is guaranteed by the almost surely discrete Dirichlet process prior used for the mixture components. To overcome the infinite-dimensionality of the problem, I apply slice sampler introduced in Walker (2007).

My model nests the Gaussian and Student-t distribution as special cases but importantly allows for deviations from the elliptic family of distributions. This includes asymmetric distributions. The proposed approach allows the beta to be a flexible function of the sign and size of the market portfolio. Although the time series of the realized conditional betas from the semiparametric model are similar to the benchmark model, I find that beta coefficient does depend on the market portfolio nonlinearly, hence implications for systematic risk measurement is very different than in a fixed parameter model.

In the parametric model, beta is constant at each time and does not depend on the contemporaneous value of market return, while with the proposed semiparametric model, I find that beta changes when we have a significant market correction or crash versus modest moves in the market. I extend the model to include more assets and provide a test to check if other factors are priced. When the market is highly volatile, beta is not affected by unexpected shocks in the market return. While in a calm market, beta can change dramatically from unexpected shocks. For stocks which are highly correlated with the market, an unexpected shock during calm periods increases the beta coefficient. The effect is the reverse for the stocks with low correlation with the market. In other words, when an asset is highly correlated with the market, large moves in a stable market increase the conditional covariance between the market and the asset more than they increase the conditional variance of the market, resulting in a significant increase in the beta coefficient. When an asset has low conditional correlation with the market, large moves in a stable market increase the conditional variance of the market more than they increase the conditional covariance between the market and the asset, leading to a drop in conditional beta. These are important contemporaneous nonlinear dynamics that are absent in other models.

In an attempt to properly estimate the shock spillovers among markets, in the second essay, I develop a joint model of returns on two different markets governed by an infinite mixture model from which the conditional density of one market given the returns of the other market can be derived. This enables us to study how a shock in one market influences the contemporaneous and future (one-day-ahead, one-week-ahead, and one-month ahead) conditional density of the other market. The proposed model extends the literature on "spillover effects" or "contagion effects" that focus on transmission through conditional moments to spillover effects on the entire density.

There is a vast body of literature on the topic of return spillover and volatility spillover among different markets. One strand of literature estimates the lagged volatility spillover among markets mostly by working in a multivriate GARCH-BEKK (Engle and Kroner, 1995) setting (Darbar and Deb, 1997; Reyes, 2001; Caporale et al., 2006; Gardebroek and Hernandez, 2013; Mensi et al., 2014). The BEKK model captures any significant lagged information transmission among asset volatilities. Another line of studies investigates the contemporaneous spillover effect of a shock in one market on the expected return of the second market. This is usually done by an ad hoc multi-step estimation procedure; first a model is estimated for the returns of the first market, and then the residuals are used as regressors for the returns of the second market (Ng, 2000; Christiansen, 2007; Balli et al., 2015; Apergis and Miller, 2008). This multi-step procedure assumes that the shocks to the first market are external drivers for the second market, which is not necessarily true. This approach also assumes that the contemporaneous shocks to the first market affect the expected return of the second market linearly.

I propose an infinite mixture model for capturing the contemporaneous spillover effects without imposing any assumption on the linearity of spillover effects or the externality of the shocks. With the speed of information transmission among markets these days, estimating the contemporaneous spillovers is of considerable importance. I model the return density of both markets jointly using Dirichlet process mixture model, allowing an unconstrained interaction of markets. Each normal component in the mixture has a conditional covariance directed by a MGARCH-BEKK process. The mean vector and the covariance matrix of the assets are assumed to be component-specific. By estimating the joint density and then deriving the conditional density, I am able to study the spillover effects of the lagged and contemporaneous shocks in either of the markets on the shape of the entire density of the other market as opposed to the extant literature that mainly focuses on spillovers effects on the first and second moments. This enables us to explore the contemporaneous spillover effects of big shocks in one market on different features of the density of the other market such as conditional expected return, volatility, kurtosis, value at risk, etc.

The proposed model can be used to assess risk transmission among any class of assets. For example, it can estimate how a stock market crash in the US market will affect the Canadian markets. More generally, it can measure financial stress. For instance, conditional on a 3% drop in the oil market, what is the probability that the US market will drop by more than 3%? Density intervals and other probability statements can be directly derived from this model.

I apply the proposed model to a dataset on S&P 500 and oil returns, and also a dataset on S&P TSX returns and oil returns, studying contemporaneous and lagged spillover effects of a shock in oil market on Canada market and the US market. The results show that the contemporaneous transmission of shocks from oil market to both Canada market and the US market are more significant than the lagged spillovers. I provide empirical evidence that a big positive or negative shock in oil market can affect the shape of the entire conditional density of the Canada market returns and the US market returns, increasing the value at risk of an investments in these markets in the event of both positive and negative shocks.

The third essay extends the literature on predictability of market return. I develop a Bayesian nonparametric model of the distribution of market return where the weights of the mixture change over time. Available information on financial and macroeconomic variables is employed to predict the weights of the mixture components in the predictive density of market return over time. This permits the density of the innovations to be unknown and to change over time. Moreover, the proposed model examines whether certain financial and macroeconomic variables convey any useful information about the predictive density of market return which extends the literature on an important question in empirical finance, predictability of market return.

To estimate the predictive density of market returns, I consider a Bayesian nonparametric mixture model for the distributions over time that allows the mixing distribution to change with time. In general, this can be done by allowing time-variation in the weights of the mixture, the atoms, or both. See for example MacEachern (1999); De Iorio et al. (2004); Gelfand et al. (2005); Teh et al. (2006); Griffin and Steel (2006); Duan et al. (2007); Rodriguez and ter Horst (2008).

Here, I assume that atoms are constant, and variation of the mixing distribution over time comes from the time-varying weights. To determine the weights at each time, I use available information on financial and macroeconomic variables. I propose a new nonparametric mixture model based on the dependent stick-breaking prior of Rodriguez and Dunson (2011) where the time-varying sticks are determined by probit transformations of linear combination of the predictors. This permits the density of innovations to change over time. Moreover, this is a novel and flexible approach by which I examine whether the financial/economic variables carry any useful information to predict the market return's time-varying density.

Prediction of stock market returns and being able to profit by timing the market using available information has been an interesting topic in finance literature. Predictability of the market returns can be exploited in areas such as asset allocation and risk management. There are contradicting conclusions on whether available information on the financial and macroeconomic variables can improve the out-of-sample predictability of market returns. Rozeff (1986); Fama and French (1988); Campbell and Shiller (1988); Cutler et al. (1991); Hodrick (1992); Rapach and Wohar (2006) and Lettau and Ludvigson (2001) report that certain financial variables explain a significant fraction of movements in stock returns. On the other hand, many studies have criticized the forecasting power of the financial and macroeconomic variables (see for example Goetzmann and Jorion (1993); Nelson and Kim (1993); Bossaerts and Hillion (1999); Ferson et al. (2003); Engstrom (2003); Goyal and Welch (2003); Lettau and Ludvigson (2005); Ang and Bekaert (2006) and Welch and Goyal (2008)). These studies either conclude that there is no strong evidence supporting the out-of-sample forecasting power of these variables or show that the statistics for many of these predictors are spuriously significant.

The studies mentioned above along with many others only investigate the predictability of mean returns. For an extensive review see Rapach and Zhou (2013). However, out-of-sample point forecasts of market returns do not convey any information about the uncertainty of the realized returns nor the spread and shape of the return distribution, while this information may be valuable for investors. The economic value of a return forecast depends not only on the point forecasts but the entire return distribution.

Although the value of predictors for the predictive mean is unclear, there is evidence that predictors help explain other features of the predictive distribution. Schwert (1989), Engle et al. (2012), Christiansen et al. (2012), Paye (2012), and Asgharian et al. (2013) use various financial and macroeconomic factors to predict the volatility of stock returns. Allowing for predictability of the conditional mean and the conditional variance of stock returns, Cenesizoglu and Timmermann (2012) examine different forecasting models with different predictors using Welch and Goyal (2008) data. They show that using these predictors can improve the economic performance of the prediction models upon constant mean and variance, although this is not the case when the focus is only on statistical measures.

The literature on the predictability of the entire density of market returns using financial and economic predictors is scarce. Some studies investigate this matter through predictability of the conditional mean and conditional variance, as mentioned above. Cenesizoglu and Timmermann (2008) propose a quantile approach to capturing predictability in the distribution of stock returns and find that many of the economic variables studied in the literature are useful in predicting the tails of the return distribution but not necessarily its center. Instead of focusing only on the conditional mean, conditional variance, or a few quantiles of market excess returns, I investigate whether these financial and macroeconomic variables are useful in predicting the one-period-ahead full density of monthly US stock returns. I attempt to predict features of the market return density in addition to what is captured by the first and second moments. This is particularly important in financial and econometric applications where we usually have heavy tails and asymmetry. I provide evidence that these features can be important in applications such as portfolio selection.

To estimate the proposed model, I apply a Gibbs sampler along with the slice sampler of Walker (2007), the collapsed sampler of Rodriguez and Dunson (2011), and Metropolis-Hastings sampler. The findings add to the growing literature on predictability of market returns. My results show that certain predictors are useful in predicting the density of market returns. I evaluate the incremental improvement in the predicted density of market return statistically by a log-predictive likelihood criteria (Bayes factors) and economically by a portfolio optimization application. The statistical and economic performance of the proposed model is superior over different benchmark models. The strongest predictive ability is found from the kitchen sink case which includes all predictors under study, stock variance, dividend payout ratio, earning price ratio, and dividend price ratio.

The rest of the dissertation proceeds as follows. Chapter 2 reviews the Bayesian methods and nonparametric approaches that I use in all three essays. Chapter 3 examines how the beta coefficient of a firm depends on the market return. Chapter 4 focuses on the contemporaneous spillover of shocks in one market into the entire density of other markets. Chapter 5 studies the predictability of the market return density using financial and economics variables. Chapter 6 concludes.

Chapter 2

Bayesian Inference and Nonparametric Approaches

The asset return models in all three chapters are extended in a Bayesian nonparametric framework. The Bayesian nonparametric approach relies on estimating an infinitely countable mixture of distributions to approximate the true underlying density and yields exact finite sample inference. To avoid iterating the same literature in each section, this section reviews the definitions and different techniques in Bayesian inference and nonparametric methods with some simple examples and applications.

2.1 Bayesian Inference

Bayesian inference is an approach to decision making in the presence of uncertainty (in the data, in the statistical model and the parameters of the model). In the classic approach, statistical inference can only be defined for experiments that are repeatable. Many experiments in the social sciences are one time events (e.g., the market closing at tomorrow) that do not fit a repeated experiment design. The classic approach must assume a fictitious world in which tomorrow's stock market close occurs again and again under identical conditions.

If these conditions cannot be met (which is usually the case) then the unknown future event is not a random variable and cannot be interpreted by classic statistical methods. Unlike classic inference, Bayesian approach does not require repeated experiments to define probability. Bayesian approaches also view unknown events with uncertainty and express this uncertainty using a probability distribution. The unknown events do not have to be subject to repetition. Both Bayesian and classic approaches view the unknown model parameters fixed, but the Bayesians express their uncertainty about these parameters using probability.

Bayesian analysis is simply applying Bayes Theorem and updating a prior distribution to a posterior distribution. Bayes rule shows us how our prior beliefs are changed once we observed an event. The result is the posterior probability.

Using Bayes Theorem we can obtain the posterior distribution of a set of parameters, θ, after observing the data from time 1 to T, $p(\theta|y_{1:T})$, as follows

$$p(\theta|y_{1:T}) = \frac{p(y_{1:T}|\theta)p(\theta)}{\int p(y_{1:T}|\theta)p(\theta)d\theta} \propto p(y_{1:T}|\theta)p(\theta)$$

where $p(\theta)$ denotes the prior density, and $p(y_{1:T}|\theta)$ is often called the likelihood. The denominator is the integrating constant and called the marginal likelihood. Therefore, the posterior is proportional to the likelihood times the prior. That is, we update the prior beliefs to posterior beliefs after observing the data. All inference about θ is conducted through the posterior and conditions on the data, $y_{1:T}$. Only the information contained in the likelihood function matters for inference. Therefore, Bayesian inference is based on what did occur as opposed to classical inference that is concerned about what might have happened. Bayesian approach is ex post and Bayesian inference is conditional on the observed data. The Bayesian approach can be very useful for real time estimation in which new data continually arrives and model estimated need to be updated because this approach to learning is sequential in nature and involves repeated application of Bayes' rule.

Once we have the posterior density of the parameters, θ, we can calculate the posterior expectation of any function of the parameters, $h(\theta)$ as

$$E(h(\theta)|y_{1:T}) = \int h(\theta)p(\theta|y_{1:T})d\theta \tag{2.1}$$

For example, the posterior mean of the parameters can be calculated as

$$E(\theta|y_{1:T}) = \int \theta p(\theta|y_{1:T})d\theta \tag{2.2}$$

As in the classical statistical approach, a Bayesian believes that there exists a true but unknown model parameter. However, the difference is that the Bayesian can use a distribution to express uncertainty. A prior is used to summarize the uncertainty regarding θ before we observe the data and learn from it. One way of selecting priors is to plot the pdf of the observed data and conjecture the values of the parameter that are more or less likely. When it is possible to identify the form of a posterior distribution, we say that the prior and data density are conjugate. That is, if F is a class of data densities and P a class of priors, then P is conjugate for F if for any likelihood of data that belongs to F and any prior of the parameters that belongs to P, the posterior distribution of the parameters also belongs to P. In most cases we assume independence between parameters. Thus the parameters' joint prior density can be written as the product of the prior densities of the parameters separately. This expedites sampling from the posterior density.

We can obtain the posterior distribution, either analytically or we obtain a series of draws from it which can be used to estimate a number of features of the unknown parameter's distribution such as the posterior mean and standard deviation, the posterior median and other quantiles, $100(1-\alpha)\%$ density regions. Most often, the posterior distribution does not have a standard and known form. Therefore, it is difficult or impossible to obtain the features like the posterior mean analytically. Development of Markov chain Monte Carlo (MCMC) technology has successfully solved the problem of computational complexity. The simulation and numerical methods provide a sample from the posterior. Then we can estimate posterior moments and other features of the density using this sample. There are several approaches to simulating from a distribution such as Importance Sampling, Gibb Sampler where we have conditionally conjugate priors, and Metropolis-Hastings where the conditional distributions are of unknown forms.

The predictive density for the model $\mathcal{M}$ is defined as the conditional distribution of y_{t+1} given $y_{1:t} = \{y_1, ..., y_t\}$ where all parameter uncertainty has been integrated out.

$$p(y_{t+1}|y_{1:t}, \mathcal{M}) = \int p(y_{t+1}|\theta, y_{1:t}, \mathcal{M}) p(\theta|y_{1:t}, \mathcal{M}) d\theta \tag{2.3}$$

This is the key quantity used in forecasting and model comparison. Suppose we have another model $\mathcal{M}'$ and evaluate the predictive density at the data y_{t+1} using $\mathcal{M}'$. If we find that $p(y_{t+1}|y_{1:t}, \mathcal{M}) > p(y_{t+1}|y_{1:t}, \mathcal{M}')$ then this means that model $\mathcal{M}$ is better able to account for the observation y_{t+1}. We can compare models on a subset of data which tends to minimize an effect the prior distribution has.

The criteria that I use to compare different models in the next chapters is the value of the predictive likelihood. For each particular model $\mathcal{M}$, the predictive likelihood for $r_{L:T}$, $L < T$ is expressed in terms of the one-step-ahead predictive likelihoods,

$$m(r_{L:T}|r_{1:L-1}, \mathcal{M}) = \Pi_{t=L}^{T} p(r_t|r_{1:t-1}, \mathcal{M}) \tag{2.4}$$

where $L > 1$ is chosen to eliminate the influence of the priors on model comparison. Usually in computing these quantities, which tend to be small, it is best to work with the log-predictive likelihoods.

Assume that data $y_{1:T}$ follows one of two models $\mathcal{M}_1$ or $\mathcal{M}_2$ with densities $p(y_{1:T}|\mathcal{M}_1)$

and $p(y_{1:T}|\mathcal{M}_2)$. The posterior odds ratio is

$$\frac{p(\mathcal{M}_1|y_{1:T})}{p(\mathcal{M}_2|y_{1:T})} = \frac{p(y_{1:T}|\mathcal{M}_1)p(\mathcal{M}_1)}{p(y_{1:T}|\mathcal{M}_2)p(\mathcal{M}_2)}$$

If the prior probabilities are equal, $p(\mathcal{M}_1) = p(\mathcal{M}_2)$, then the posterior odds equal the Bayes factor. If the Bayes factor is 10, then it suggests that $\mathcal{M}_1$ is 10 times better at explaining the data compared to $\mathcal{M}_2$. In comparing two models ($\mathcal{M}_1$ versus $\mathcal{M}_2$) a rough guide is that a log-Bayes factor bigger than 3 is a strong evidence for outperformance of $\mathcal{M}_1$ and a log-Bayes factor bigger than 5 is a very strong evidence for outperformance of $\mathcal{M}_1$ against $\mathcal{M}_2$ (Kass and Raftery (1995)). Bayesian model comparisons favours parsimonious model specifications. Complex models are only chosen if they provide an improved description of the data. We can compare nested or non-nested models using predictive likelihoods.

2.2 Dirichlet Process Mixture (DPM) Model

The financial literature has recognized for some time (Fama, 1965; Mandelbrot, 1997) that return distributions exhibits fat tails and asymmetry which are significantly different than the normal distribution. This phenomena is more evident in daily and higher-frequency returns and has led many researchers to use different models to explain skewness and excess kurtosis. Mixture models provide a flexible approach to capture these features of the return distribution.

Mixture models can be divided into two categories in finance. Continuous mixtures have a mixing variable which follows a continuous distribution. The stochastic volatility model of Clark (1973) and the subsequent literature (Taylor, 1994) is a good example of this specification. The second class of models is discrete mixtures. Examples include a finite mixture of normals and Markov switching models. Here, I focus on discrete mixture models used in finance.

In the mixture model analysis, how to determine the number of mixtures is a fundamental problem. The traditional method including model selection based on different number of clusters may over-fit or under-fit the data. While a Bayesian non-parametric model side-steps this part, it allows the data set to determine its own clusters without actually doing the model selection. Suppose the observation follows an unknown distribution which we try to investigate, the Bayesian method defines a known prior distribution for the underlying unknown distribution. The prior distribution in the non-parametric approach is the space of distributions.

A typical model for y_t in the Bayesian nonparametric framework is

$$f(y) = \int f(y|\theta)dH(\theta) \tag{2.5}$$

$$H(\cdot) = \sum_{g=1}^{\infty} \eta_g \delta_{\theta_g}(\cdot) \tag{2.6}$$

where H is almost surely discrete, $f(.|.)$ is a kernel density, and $\eta = (\eta_1, \eta_2, ...)$ is a vector of the weights such that $\eta_g \geq 0$ for all g, $\sum_{g=1}^{\infty} \eta_g = 1$ and $\delta_x(\cdot)$ denotes the distribution degenerate at x. $\theta_1, \theta_2, ...$ are assumed to be draws from some known distribution. H is the unknown discrete mixing distribution and from a Bayesian perspective requires a prior assumption.

An infinite mixture model of this form can flexibly represent a wide range of continuous distributions. This nonparametric model is often embedded in a richer time-series model in the finance literature since it is able to capture the fat tails and skewness existing in the financial data but not able to capture time dependencies.

Before discussing the model in (2.5)-(2.6) I review selection of a prior for H and some of its properties. By far the most popular prior for H is the Dirichlet process (DP) prior. The DP, introduced by Ferguson (1973), is a distribution of probability measures over a measurable space. A draw from a DP, say H, is a discrete distribution with a countable number of atoms. The DP is indexed by two parameters: a positive scalar α, called the concentration parameter, and a base distribution H_0. The draw $H \sim DP(\alpha, H_0)$ is centered on the base distribution which serves as a prior guess. The concentration parameter controls how close we believe the unknown H is to H_0. The larger α, the stronger belief in H_0 and the more distinct atoms we have with significant mass. This prior is convenient in that we can center it around a well known family of parametric distributions such as the Gaussian, and impose some measure of parsimony on the predictive density of y_t through α.

A key feature of the DP is that for any finite partition $\{A_1, ..., A_K\}$ of the parameter space,

$$H(A_1), ..., H(A_K)|\alpha, H_0 \sim \mathcal{D}(\alpha H_0(A_1), ..., \alpha H_0(A_K)), \tag{2.7}$$

where $\mathcal{D}(a_1, ..., a_K)$ denotes a Dirichlet distribution with parameter vector $(a_1, ..., a_K)$. This property implies $E[H(A_i)] = H_0(A_i)$ and $\mathrm{V}(H(A_i)) = \frac{H_0(A_i)(1-H_0(A_i))}{(1+\alpha)}$.

There are several different representations for the DP: the Chinese restaurant representation, the Polya-urn process, and the stick-breaking representation. A *Chinese Restaurant Process* (CRP), displayed in Figure 2.1, is a distribution over partitions. Imagine a restaurant with an infinite number of tables that have a unique dish $\theta_g \sim H_0$ that is eaten at each table. A new customer ϕ_n (the nth customer) chooses an empty table with probability $\frac{\alpha}{n-1+\alpha}$ which is served the new dish $\theta \sim H_0$ or decides to sit at one of the occupied tables that have dish θ_g, $g = 1, \ldots, K$ with probability $\frac{c}{n-1+\alpha}$ where c is the number of people sitting at that table.

Another closely related representation for a draw from $H \sim DP(\alpha, H_0)$, is the *Polya Urn Scheme*. Imagine an urn with α black balls. Whenever a black ball is drawn, a new color is generated from H_0 and the new color ball along with the black ball is placed back in the urn. If the drawn ball is non-black, it is placed back together with a ball with the same color.

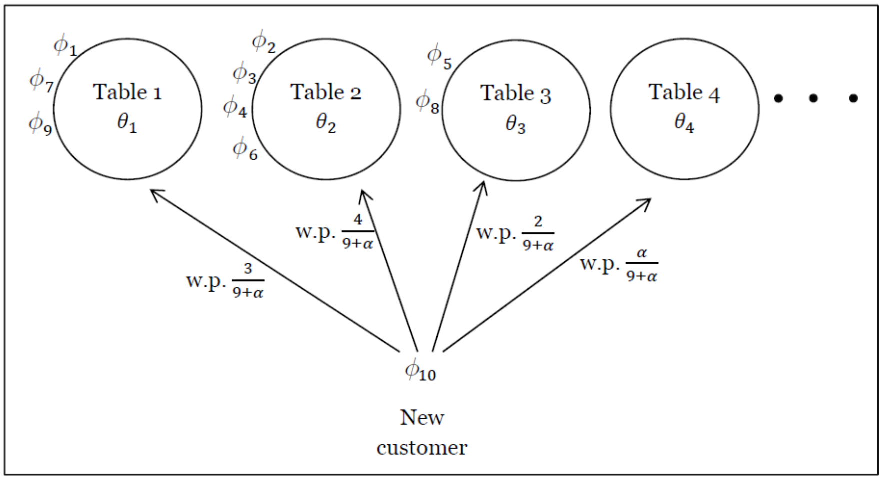

Figure 2.1: Chinese restaurant process.

Thus, the nth new ball that we put in the urn is the same color as one of the existing colors, say color g, in the urn with probability $\frac{n_g}{n-1+\alpha}$ where n_g is the number of existing balls with color g, or has a new color with probability $\frac{\alpha}{n-1+\alpha}$. Blackwell and MacQueen (1973) prove that the distribution of colors after n draws converges to a discrete distribution that follows a DP. Let s_n be an indicator and record the color of the nth ball. If the first ball is $s_1 \sim H_0$, thereafter ball colors are generated as

$$s_{n+1}|s_{1:n} \sim \sum_{g=1}^{K} \frac{n_g}{\alpha+n}\delta_{s_n,g} + \frac{\alpha}{\alpha+n}\delta_{s_n,K+1} \tag{2.8}$$

where $\delta_{h,g}$ denotes the Kronecker delta and K is the current number of distinct colors.

The observations from a Polya urn are exchangeable and this key point means we can always rearrange our sample to assume we are dealing with the last observation and apply the Polya urn (conditional distribution) results. This is the basis of the Gibbs sampler we discuss below.

Another, constructive definition of the DP, is due to Sethuraman (1994) and often referred to as a *Stick-Breaking Representation* due to the construction of the weights. A draw from the DP, $H \sim DP(\alpha, H_0)$ almost surely has the form

$$
\begin{aligned}
H(\cdot) &= \sum_{g=1}^{\infty} \eta_g \delta_{\theta_g}(\cdot) && (2.9)\\
\eta_1 &= v_1, \eta_g = v_g \prod_{l=1}^{g-1}(1-v_l), \quad g \geq 2 && (2.10)\\
v_l &\stackrel{iid}{\sim} \mathcal{B}(1,\alpha), \; l = 1, 2, ... && (2.11)\\
\theta_g &\stackrel{iid}{\sim} H_0, \; g = 1, 2, ... && (2.12)
\end{aligned}
$$

This formulation and the CRP/Polya Urn provide different avenues to posterior sampling.

Note that any draw H from a $DP(\alpha, H_0)$ is a discrete probability measure (Ferguson, 1973) and, therefore, it is possible for a sample from H to have repeated elements. Moreover, the DP can not be used as a prior for distribution of continuous variables. To deal with this issue, Lo (1984) introduces the Dirichlet process mixture (DPM) model and suggests working with a hierarchical model. The DPM model takes the following form

$$
\begin{aligned}
y_t|\phi_t &\sim F(\phi_t), \; t = 1, ..., T && (2.13)\\
\phi_t &\stackrel{i.i.d.}{\sim} H && (2.14)\\
H|\alpha, H_0 &\sim DP(\alpha, H_0) && (2.15)
\end{aligned}
$$

where $F(\phi_t)$ is the distribution associated with density $f(y_t|\phi_t)$. If $F()$ is a continuous distribution then so is the DPM model. The parameters, ϕ_t, are assumed to be distributed according to an unknown distribution H which follows the Dirichlet process prior.

Equivalently, the model in the form of Equation (2.5) with H from Equations (2.9)-(2.12) becomes

$$
p(y_t) = \sum_{g=1}^{\infty} \eta_g f(y_t|\theta_g) \tag{2.16}
$$

where θ_g represent the unique draws from H_0[1] while ϕ_t in Equation (2.14) represent draws from H which may have repeats.

With the advent of Markov chain Monte Carlo (MCMC) methods Escobar (1994) and Escobar and West (1995) provide tractable approaches to posterior simulation. This is the first generation of posterior samplers.

Here, I discuss the Gibbs sampling method for the model in Equations (2.13)-(2.15). Denote the set of distinct values of all ϕ_ts by $\theta = \{\theta_1, \theta_2, ..., \theta_K\}$ of size $K \leq T$. Conditional on K, we introduce indicators $s_t = g$ if $\phi_t = \theta_g$ so that, given $s_t = g$ and θ, $y_t \sim F(\theta_g)$. The configuration set $S = \{s_1, ..., \ s_T\}$ partitions the data $y_{1:T} = \{y_1, ..., \ y_T\}$ into K distinct groups so that the $n_g = \#\{t : s_t = g\}$ observations in group g have the same parameter value θ_g. Also define $I_g = \{t : s_t = g\}$ for the set of indices of observations in group g and $Y_g = \{y_t : s_t = g\}$ as the corresponding group of observations. Define $\phi_{-t} = \{\phi_1, ..., \phi_{t-1}, \phi_{t+1}, ..., \phi_T\}$, $S_{-t} = \{s_1, ..., s_{t-1}, s_{t+1}, ..., s_T\}$, $K^{(t)} =$ as the number of distinct values in ϕ_{-t} and $n_g^{(t)} = \#\{s_t \in S_{-t} : s_t = g\}$, then West et al. (1994) show that the Gibbs sampling steps are

- $\phi_t | \alpha, \phi_{-t}, y_{1:T}$, $t = 1, \ldots, T$,
- $\theta | \alpha, S, y_{1:T}$,
- $\alpha | S, y_{1:T}$.

The first step is a consequence of samples from the DP being exchangeable and is based on the following

$$(\phi_t | \phi_{-t}, \ S_{-t}, \ K^{(t)}, y_{1:T}) \sim \frac{\alpha}{T-1+\alpha} h(y_t) H(d\theta | y_t) + \frac{1}{T-1+\alpha} \sum_{g=1}^{K^{(t)}} n_g^{(t)} f(y_t | \theta_g) \delta_{\theta_g} \quad (2.17)$$

where $h(y_t) = \int f(y_t|\theta) dH_0(\theta)$ is the predictive density derived from the prior evaluated at

[1] Assuming H_0 is a continuous distribution.

y_t and $H(d\theta|y_t) \propto f(y_t|\theta)dH_0(\theta)$ the posterior distribution based on one observation. This stage generates a new configuration by sequentially sampling indicators from the posteriors

$$s_t = \begin{cases} g & \text{w.p. } \frac{n_g^{(t)}}{T-1+\alpha} f(y_t|\theta_g), \\ K^{(t)}+1 & \text{w.p. } \frac{\alpha}{T-1+\alpha} h(y_t), \end{cases}$$

for any index t. If $s_t = K^{(t)} + 1$, draw a new ϕ_t from $H(d\theta|y_t)$. The second sampling step $\theta|\alpha, S, y_{1:T}$, is not necessary but it tends to improve the mixing of the Markov chain. The first step only changes one parameter at a time conditional on all others while the second step allows for an update of all unique θ at one time. The final sampling step of α follows the Gibbs step from Escobar and West (1995).

West et al. (1994) derive the conditional predictive density for future data y_{T+1},

$$p(y_{T+1}|\theta, S, \alpha, y_{1:T}) = \frac{\alpha}{\alpha+T} \int f(y_{T+1}|\theta_{K+1})dH_0(\theta_{K+1}) + \sum_{g=1}^{K} \frac{n_g}{\alpha+T} f(y_{T+1}|\theta_g) \quad (2.18)$$

where θ_{K+1} is a new independent draw from H_0. Note that Equation (2.18) represents a potentially infinite mixture model since for each new observation there is a possibility (proportional to α) to introduce a new parameter θ.

The final estimate of the predictive density is obtained by integrating out the parameter uncertainty. Given M Gibbs draws of $\phi^{(m)}$, $S^{(m)}$ and $\alpha^{(m)}$ the estimate is

$$p(y_{T+1}|y_{1:T}) \approx \frac{1}{M} \sum_{m=1}^{M} p(y_{T+1}|\theta^{(m)}, S^{(m)}, \alpha^{(m)}, y_{1:T}). \quad (2.19)$$

There are important alternatives to the Gibbs sampler. Equation (2.9) represents the DP as a countably infinite sum of atomic measures, which leads to the infinite mixture for the density of y_t in Equation (2.16). The second generation of posterior sampling methods allow for inference on H and use Equation (2.16) directly. One approach is to truncate this infinite mixture to a finite number large enough to approximate the true model. Ishwaran and James (2001) provide details on the approximation and block sampling methods. A second approach to deal with the infinite number of parameters in Equation (2.16) is the slice sampler method of Walker (2007). He introduces a latent variable that randomly truncates the model to a finite mixture model, but whose marginal distribution preserves the original model. This turns an infinite dimensional sampling problem into a finite one.

Once we have the sample from the posterior distribution of the unknown parameters, θ, we can approximate the posterior mean of any function $h(\theta)$ as

$$E(h(\theta)|y_{1:T}) = \int h(\theta)p(\theta|y_{1:T})d\theta \approx \frac{1}{M}\sum_{g=1}^{M} h(\theta^{(g)}) \tag{2.20}$$

where $\theta^{(g)}$, $g = 1, ..., M$ is the sample drawn from the posterior density of θ using MCMC algorithms. Often a burn-in sample is considered and the random draws at the beginning of the MCMC algorithm are dropped from the sample in order to ensure the convergence of the random draws is achieved.

An application of the DPM to stock returns To illustrate how an infinite mixture model is able to capture the fat tails in a data series, I nonparametrically estimate the density of gold returns, applying the DPM model and using the Gibbs sampler discussed above. The data are the monthly log-returns from January 1970 to November 2012 with $T = 515$ observations. Table 2.1 displays summary statistics of the dataset. This table shows that monthly returns of gold display skewness and excess kurtosis.

Total number	Mean	Variance	Skewness	Kurtosis
515	0.327	4.6	8.1	1.26

Table 2.1: Summary statistics of gold returns. The data set is the monthly returns from January 1970 to November 2012.

In this example, $\theta = (\mu, h)$, and the base measure, H_0, is assumed to have a conjugate prior of

$$H_0(\mu_t, h_t) \equiv \mathcal{NG}(\mu_t, h_t|\mu_0, \rho, \frac{\alpha_0}{2}, \frac{\beta_0}{2})$$

$$\mu_0 = \ 0,\ \ \rho = \ 4,\ \ \alpha_0 = \ 4,\ \ \beta_0 = \ 20,\ \ \alpha \sim \mathcal{G}(a, b)$$

where $\mathcal{NG}(.)$ is the normal-gamma distribution[2] and $\mathcal{G}(a, b)$ is the gamma distribution[3]. Using the results above we have

$$h(y_t) \quad \propto \quad t_{\alpha_0}\left(y_t|\mu_0, \frac{\beta_0(1+\rho)}{\alpha_0\ \rho}\right)^4. \tag{2.21}$$

$$H(d\theta|y_t) \quad \propto \quad \mathcal{NG}\left(\mu, h|\bar{\mu_t}, (1+\rho)^{-1}, \frac{\alpha_0+1}{2}, \frac{\omega_t^*}{2}\right) \tag{2.22}$$

in which

$$\bar{\mu_t} \quad = \quad \frac{y_t + \rho\mu_0}{1+\rho},\ \omega_t^* = -(1+\rho)\bar{\mu_t}^2 + y_t^2\ + \rho\mu_0^2 + \beta_0. \tag{2.23}$$

$t_\nu(\mu, \sigma^2)$ denotes a Student-t density with location μ, scale parameter σ^2 and degree of freedom ν and $t_\nu(x|\mu, \sigma^2)$ is the associated density function evaluated at x.

The predictive density given parameter draws is a mixture of normals and Student-t distributions,

$$\begin{aligned} &p(y_{T+1}|\theta, S, \alpha) = \\ &= \frac{\alpha}{\alpha+T}\int f(y_{T+1}|\mu_{k+1}, h_{k+1})dH_0(\mu_{k+1}, h_{k+1}) + \Sigma_{g=1}^k\ \frac{n_g}{\alpha+T} f(y_{T+1}|\mu_g, h_g) \\ &= \frac{\alpha}{\alpha+T}\ t_{\alpha_0}\left(y_{T+1}|\mu_0, \left(\frac{\beta_0}{\alpha_0}\right)\left(1+\frac{1}{\rho}\right)\right) + \sum_{g=1}^{k}\frac{n_g}{\alpha+T}\ \phi(y_{T+1}|\mu_g, h_g). \end{aligned} \tag{2.24}$$

[2] $(X, T) \sim \mathcal{NG}(\mu, \rho, \alpha, \beta)$, with density $f(x, \tau|\mu, \rho, \alpha, \beta) = \frac{\beta^\alpha\sqrt{\rho}}{\Gamma(\alpha)\sqrt{2\pi}}\tau^{\alpha-1}e^{-\tau\beta}e^{\frac{-\rho\tau(x-\mu)^2}{2}}$.

[3] $X \sim \mathcal{G}(a, b)$, with density $f(x|a, b) = \frac{b^a}{\Gamma(a)}x^{a-1}e^{-xb}$. Then we have $E(x) = \frac{a}{b}$.

The predictive density is obtained by averaging this result over the MCMC draws as in Equation (2.19).

After running the MCMC algorithm for $M = 7000$ iterations and dropping the first $M_0 = 1000$ draws as burn-in, I estimate predictive density. Figure 2.2 shows the estimated nonparametric predictive density of monthly returns with $\alpha \sim \mathcal{G}(1, 24)$ (the first panel) as well as its logarithmic scale plot (the second panel) in comparison to the normal distribution with mean and variance equal to sample mean and sample variance, respectively. The nonparametric model captures significant deviations from the normal distribution by using approximately 5 components, on average. The log-density of the nonparametric model displays thicker upper tails meaning large increases in gold prices are more likely than large drops. The posterior mean of the precision parameter α is 0.17.

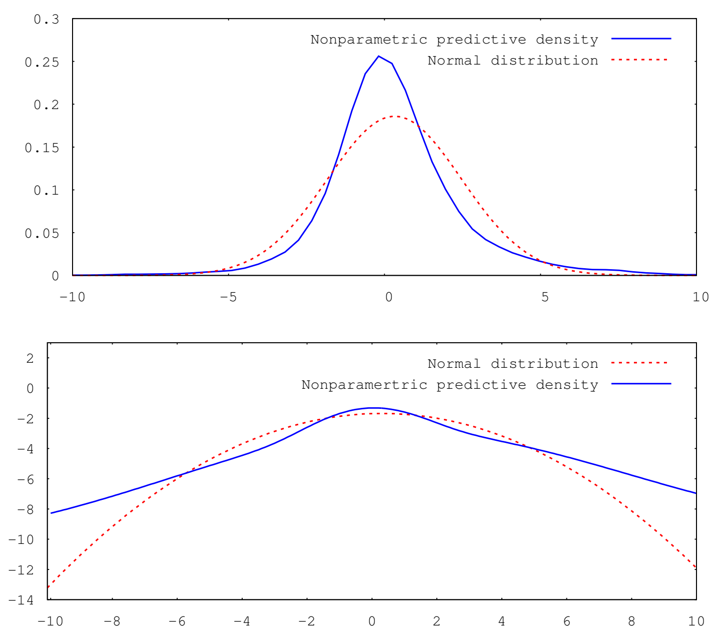

Figure 2.2: Comparison of nonparametric predictive density with the normal distribution. The second panel shows the associated log-densities

Pooling data into different clusters is a distinct advantage of the DPM model. For instance, Fisher et al. (2015) use the DPM in the context of a linear factor structure to classify the skill of mutual funds into sub-populations.

The DPM model is suitable for modeling an unknown i.i.d. distribution such as the unconditional distribution of returns. Without modification it cannot deal with the pronounced volatility dynamics in return data. For this reason the DPM is often embedded in a more sophisticated time-series specification of returns. I turn to some examples next.

GARCH-DPM and SV-DPM

To make the model richer and more suitable for financial data, recent research combines the DPM model with a time series model of volatility such as GARCH or stochastic volatility.

GARCH The first class of studies incorporates the GARCH (Bollerslev, 1986) functional form into the infinite mixture model. From the perspective of GARCH literature, these models impose an ARMA structure on squared innovations along with a parametric Normal, Student-t, or in some cases (Bauwens et al., 2007; Galeano and Ausin, 2010) a finite mixture of normals for the innovation density. By extending GARCH models to a Bayesian semiparametric setting, the functional form of the conditional density of the data can be estimated nonparametrically in the same framework as DPM model (2.13)-(2.15) but with a GARCH factor entering into each of the components of this mixture. For general reviews on multivariate GARCH (MGARCH) models, see Bauwens et al. (2006), and Virbickaite et al. (2015).

Kalli et al. (2013) define an infinite mixture model with GARCH components to estimate a unimodal and asymmetric conditional return distribution. The unknown distribution H has a stick-breaking prior (SBP) with a standard exponential distribution as the base distribution H_0. The SBP is a more general case of the DP for which there exists less tractable results. With the SBP, the rate of weights decay is controlled by potentially an infinite number of parameters, instead of only one parameter, α, which is the case for DP. Kalli et al. (2013) look at the daily log returns of three stock indices and find evidence in favor of their semiparametric model against GARCH, EGARCH, and GJR-GARCH models. Other work on Bayesian nonparametric GARCH models are Lau and Cripps (2012) and Ausín et al. (2014).

Jensen and Maheu (2013) propose a Bayesian nonparametric modeling approach for the MGARCH model,

$$y_t|\mu_t, B_t, \Sigma_t \sim \mathcal{N}(\mu_t, \Sigma_t^{1/2} B_t^{-1} (\Sigma_t^{1/2})'), \; t = 1, ..., T, \tag{2.25}$$

$$\Sigma_t = \Gamma_0 + \Gamma_1 \odot y_{t-1} y_{t-1}' + \Gamma_2 \odot \Sigma_{t-1}, \tag{2.26}$$

$$\mu_t, B_t|H \overset{iid}{\sim} H, \tag{2.27}$$

$$H|\alpha, H_0 \sim DP(\alpha, H_0), \tag{2.28}$$

$$H_0(\mu_t, B_t) \equiv \mathcal{N}(\mu_0, V_0) - \mathcal{W}(P, \nu + d - 1), \quad \nu \geq 1, \tag{2.29}$$

where $y_t = (y_{1t}, ..., y_{dt})'$ is the d-dimensional vector of returns and the symbol $\odot$ denotes the Hadamard product. They assume a parametric model for the dynamics of the conditional covariance matrix of returns (Equation (2.26)) proposed by Ding and Engle (2001), and a DPM prior for the multivariate distribution of the returns Equations (2.27)-(2.29). In the Sethuraman (1994) representation of the DPM model, this model can be cast as an infinite mixture of multivariate normals with mixing over both the location and scale matrix of the normal components,

$$f(y_t|\Sigma_t) = \sum_{g=1}^{\infty} \eta_g \phi(y_t|\mu_g, \Sigma_t^{1/2} B_g^{-1} (\Sigma_t^{1/2})') \tag{2.30}$$

$$\eta_1 = v_1, \eta_g = v_g \prod_{l=1}^{g-1} (1 - v_l), \quad g \geq 2 \tag{2.31}$$

$$v_l \overset{iid}{\sim} \mathcal{B}(1, \alpha), \; l = 1, 2, ... \tag{2.32}$$

$$\theta_g \equiv (\mu_g, B_g) \overset{iid}{\sim} H_0, \; g = 1, 2, ... \tag{2.33}$$

This nests the Gaussian case when $\eta_1 = 1$ and $\eta_g = 0$, $g > 1$. When $\mu_g = \mu$, $\forall g$ and $\alpha \to \infty$, $H \to H_0$ and we obtain a Student-t distribution.

Using the conditionally conjugate priors (normal and Wishart) as the base measure, H_0, slice sampling methods from Walker (2007) and Kalli et al. (2009) can be used for posterior simulation. In contrast to the previous example that was based on the Polya urn and integrated the unknown H out, slice sampling works directly with the infinite mixture representation in Equation (2.30). In the following sampling steps $S = \{s_1, ..., s_T\}$ is the configuration set where $s_t = g$ if the tth observation uses the gth component's parameters (μ_g, B_g), and $n_g = \#\{t : s_t = g\}$. $u_1, \dots, u_T$ are the auxiliary variables introduced by Walker (2007) to make the sampling more tractable.

Step 1 The posterior distribution of (μ_g, B_g), $g = 1, ..., K$: Using the transformation $\mathfrak{z}_t = \Sigma_t^{-1/2} y_t$, and making use of conditionally conjugate prior we have

$$B_g | y^T, S, \mu_g, \Sigma^T \sim W\left(P^{-1} + \sum_{t:s_t=g} (\mathfrak{z}_t - \Sigma_t^{-1/2}\mu_g)(\mathfrak{z}_t - \Sigma_t^{-1/2}\mu_g)')^{-1}, n_g + d - 1 + \nu_0\right)$$

$$\mu_g | y^T, S, B_g, \Sigma^T \sim \mathcal{N}(\bar{\mu}, \bar{V}) \tag{2.34}$$

in which

$$\bar{V} = \left(\sum_{s_t=g} \Sigma_t^{-1/2'} B_g^{-1} \Sigma_t^{-1/2} + V_0^{-1}\right)^{-1}, \ \bar{\mu} = \bar{V}\left(\sum_{s_t=g} \Sigma_t^{-1/2'} B_g^{-1} \mathfrak{z}_t + V_0^{-1}\mu_0\right). \tag{2.35}$$

Step 2 Updating v_g, $g = 1, ..., K$: By the conjugacy of the generalized Dirichlet distribution to multinomial sampling (Ishwaran and James, 2001) we have

$$v_g | S \sim \mathcal{B}\left(1 + \sum_{t=1}^{T} 1(s_t = g), \alpha + \sum_{t=1}^{T} 1(s_t > g)\right). \tag{2.36}$$

Then update η according to Equation (2.31).

Step 3 Updating $u_t,\ t = 1, ..., T$ (Walker, 2007):

$$u_t|S \sim U(0, \eta_{s_t})$$

Then we update K such that $\sum_{g=1}^{K} \eta_g > 1 - \min\{u_t\}_{t=1}^{T}$. Additional η_g and (μ_g, B_g), will need to be generated from the priors if K is incremented.

Step 4 Updating S (Walker, 2007): For each $t = 1, ..., T$,

$$p(s_t = g|y^T, \Sigma^T) \propto \mathbf{1}(\eta_g > u_t)\phi(y_t|\mu_g, \Sigma_t^{1/2} B_g^{-1} (\Sigma_t^{1/2})'), g = 1, ..., K \tag{2.37}$$

Step 5 Updating GARCH parameters $\Gamma = \{\Gamma_0, \Gamma_1, \Gamma_2\}$: Assuming prior $p(.)$ for Γ

$$p(\Gamma|\mu, B, S, y^T, \Sigma^T) \propto p(\Gamma) \times \prod_{t=1}^{T} \phi(y_t|\mu_{s_t}, \Sigma_t^{1/2} B_{s_t}^{-1} (\Sigma_t^{1/2})') \tag{2.38}$$

which is not of standard form, and the Metropolis Hastings sampler can be applied.

Given a large number of draws from the sampling steps above, the predictive density of y_{T+1} given y^T can be approximated using M draws of the posterior as follows:

$$f(y_{T+1}|y^T) \approx \frac{1}{M} \sum_{m=1}^{M} \phi(y_{T+1}|\mu^{(m)}_{s^{(m)}_{T+1}}, \Sigma^{(m)^{1/2}}_{T+1} B^{(m)-1}_{s^{(m)}_{T+1}} \Sigma^{(m)^{1/2'}}_{T+1}) \tag{2.39}$$

where $x^{(m)}$ is the mth draw of parameter x, and $s^{(m)}_{T+1}$ at each iteration m is one of the $K^{(m)}$ components, say component g, with probability $\eta_g^{(m)}$, or is a new component with probability $1 - \sum_{g=1}^{K^{(m)}} \eta_g^{(m)}$. Note that we are able to compute $\Sigma^{(m)}_{T+1}$ at each iteration of the algorithm from:

$$\Sigma^{(m)}_{T+1} = \Gamma_0^{(m)} + \Gamma_1^{(m)} \odot y_T y_T' + \Gamma_2^{(m)} \odot \Sigma_T^{(m)}.$$

This is recursively computed from $t = 1$ to $t = T + 1$.

Jensen and Maheu (2013) consider two datasets, equity return and foreign exchange rate, to estimate the model employing Polya urn and stick-breaking sampling schemes. Comparison of the Bayes factors and density forecasts with parametric GARCH models (Gaussian and Student-t innovations) support the flexible semiparametric approach, particularly in the case of asymmetric distributions and during high volatile periods.

Working with a univariate version of the model Ausín et al. (2014) estimate the return density of Hang Seng Index and Bombay Stock Exchange index and carry out Bayesian prediction of the value-at-risk.[5] They compare the results of the semiparametric model with a Gaussian, a Student-t, and a mixture of two zero mean Gaussian distributions, and find significant differences in the return predictive distribution, particularly in the tails.

Stochastic Volatility The second class of studies combine stochastic volatility with a DPM model. The main difference of the SV model with GARCH, is that conditional on time t information, the conditional variance is stochastic and can be thought of as the impact of an unobserved news flow process.

Jensen and Maheu (2010) extend the standard SV specification that has parametric return innovations to the following semiparametric (SV-DPM) setting

$$y_t \sim \mathcal{N}(\mu_t, \lambda_t^{-2} \exp(h_t)) \tag{2.40}$$

$$h_t | h_{t-1} \sim \mathcal{N}(\delta h_{t-1}, \sigma_v^2) \text{ and } h_t \perp y_t,\ |\delta| < 1 \tag{2.41}$$

$$(\mu_t, \lambda_t^2)' \sim H \tag{2.42}$$

$$H \sim DP(H_0, \alpha) \tag{2.43}$$

$$H_0(\mu_t, \lambda_t^2) \equiv \mathcal{N}(m, (\tau \lambda_t^2)^{-1}) - \mathcal{G}(\nu_0/2, s_0/2) \tag{2.44}$$

[5] Value-at-risk indicates the potential loss associated with an unfavorable movement in market prices over a given time period at a certain confidence level.

The latent log volatility h_t follows a parametric, stationary, first-order autoregressive (AR) process defined with the AR parameter δ but the rest of the model is nonparametric inasmuch no assumption is made about the underlying distribution of return innovations. Note that, assuming $h_t \perp y_t$, Jensen and Maheu (2010) remove any leverage effect (Jacquier et al., 2004). Equations (2.42) and (2.43) assume the mixture's probabilities and parameters follow the DP prior. The base distribution, H_0, is a conjugate conditional normal-gamma distribution. The Sethuraman (1994) representation for this semiparametric model is

$$y_t \sim \sum_{g=1}^{\infty} \eta_g \mathcal{N}(c_g, d_g^{-2} \exp(h_t)) \tag{2.45}$$

with the mixture weights distributed as $\eta_1 = v_1$, $\eta_g = v_g \prod_{l=1}^{g-1}(1 - v_l)$, $g > 1$, where $v_l \sim \mathcal{B}(1, \alpha)$. The mixture parameters (c_g, d_g^2) have the $\mathcal{NG}$ distribution (Equation (2.44)). They construct an MCMC sampling method for the model and apply it to a long series of daily market returns and find strong deviations from both normal and Student-t distributions.

To take into consideration the correlation between returns and volatility innovations (Jacquier et al., 2004; Nakajima and Omori, 2009), Jensen and Maheu (2014) extend the univariate algorithm of the semiparametric stochastic volatility model above by lifting the assumption of $y_t \perp h_t$. To model the unknown joint distribution of return and volatility innovations (u_t and v_t in the following expressions), they choose a DPM model (ASV-DPM) as below,

$$y_t = \mu + \exp(h_t/2)u_t \tag{2.46}$$

$$h_t = \delta h_{t-1} + v_t \tag{2.47}$$

$$(u_t, v_t)' \sim \mathcal{N}_2(0, \Lambda_t) \tag{2.48}$$

$$\Lambda_t \sim H \tag{2.49}$$

$$H \sim DP(H_0, \alpha) \tag{2.50}$$

They compare the ASV-DPM model with the parametric SV models by daily predictive Bayes factors.[6] The empirical experiment provides evidence favoring the nonparametric asymmetric stochastic volatility model more often than the parametric version.

Delatola and Griffin (2013) extend the parametric, linearized stochastic volatility model of Omori et al. (2007) to capture the leverage effect in an infinite mixture model. They include a constant leverage effect and nonparametrically model the distribution of the log-squared returns. The semiparametric stochastic volatility model with leverage effect in this case is specified as

$$y_t^* = h_t + z_t^* \tag{2.52}$$

$$h_t = \delta h_{t-1} + d_t \rho \sigma_v \exp(\mu_t/2)[a^* + b^*(y_t^* - h_t - \mu_t)] + \sigma_v(1-\rho^2)\epsilon_t^* \tag{2.53}$$

$$z_t^* \sim \mathcal{N}(\mu_t, \delta\sigma_z^2) \tag{2.54}$$

$$\mu_t \sim H \tag{2.55}$$

$$H \sim DP(\alpha, H_0) \tag{2.56}$$

where $a^* = \exp(\delta\sigma_z^2/8)$, $b^* = 0.5a^*$, $d_t = sign(y_t)$, and same as in Nakajima and Omori (2009) and Jacquier et al. (2004), $y_t^* = \log(y_t^2 + c)$ and $z_t^* = \log(u_t^2) \sim F$ which can be approximated by a mixture of normal distributions.

Kalli and Griffin (2015) define a Bayesian nonparametric model for cross-sectional aggregation of AR(1) models (Robinson, 1978; Granger, 1980) to account for long memory in volatility. Suppose that we have d time series, $\{h_{i1}, ..., h_{iT}\}_{i=1}^d$, where each time series

[6] The predictive likelihood is defined as,

$$m(y_L, \ldots, y_T | y^{L-1}, M) = \sum_{t=L}^{T} f(y_t | y^{t-1}, M) \tag{2.51}$$

where M denotes the particular model, and $L < T$ is chosen to eliminate the influence of the priors. The one-step-ahead predictive likelihoods $f(y_t|y^{t-1}, M)$, can be estimated by computing the sample mean of the likelihood using MCMC draws of the unknown parameters and latent variables conditional on the data history y^{t-1}.

follow an AR(1) process with persistence parameter $\Phi_i \overset{iid}{\sim} F_\Phi$. The finite aggregate process is written as

$$h_t = \sum_{i=1}^{d} \frac{h_{it}}{d}, \; t = 1, ..., T. \tag{2.57}$$

An infinite aggregate process assumes that d in Equation (2.57) goes to infinity. Kalli and Griffin (2015) assume that the stochastic volatility follows an infinite aggregate process,

$$y_t = \beta \exp(h_t/2) u_t, \tag{2.58}$$

$$h_t = \sum_{g=1}^{\infty} h_t(\Phi_g, \sigma^2 \pi_g), \tag{2.59}$$

where $u_t \sim \mathcal{N}(0,1)$ and π_g is the proportion of the variation in h_t explained by the gth process associated with AR parameter Φ_g. This parametrization flexibly models the dependence of the volatility process in financial time series. The persistence parameter of each AR(1) process is independently drawn from a distribution F_ϕ which is estimated nonparametrically following a DP prior. They apply the linearized model of SV (Harvey et al., 1994), and also use a finite approximation to the DP (Ishwaran and Zarepour, 2000; Neal, 2000; Kim and Shephard, 1998). The empirical results in the daily returns of HSBC and Apple Inc. show significant difference in the distributions of the volatility persistence, suggesting different lasting effect of the information in these two sets of return.

Mixture modeling appears in many other areas of finance. Mixture models underpin a great deal of research in empirical finance. Development and use of these modeling methods will continue to be a central component of empirical work in finance.

Chapter 3

Essay 1: Nonparametric Dynamic Conditional Beta

3.1 Introduction

This chapter nonparametrically estimates the dynamic conditional beta of a stock using a Bayesian semiparametric multivariate GARCH model. This extends Engle's (2016) parametric version of dynamic conditional beta to the case of an unknown general continuous distribution. In this setting the whole distribution can affect the compensation for risk.

Researchers have long studied the beta coefficient of a stock which represents the nondiversifiable risk arising from exposure to market movements. Traditional approaches estimate the beta coefficient by regressing excess stock returns on the excess market return as in the one-factor Capital Asset Pricing Model (CAPM, Sharpe (1964) and Lintner (1965)), or exploiting more empirically supported asset pricing models, such as Fama-French three-factor model, which incorporate additional explanatory variables (Fama and French (1993)). Our multivariate model nests both cases, but allows for time variation in the conditional second moments. There is a large literature based on multivariate GARCH (MGARCH) models that link a time varying beta to the conditional second moments. Some examples include Bollerslev et al. (1988),Giannopoulos (1995), McCurdy and Morgan (1992) and Choudhry (2002).

Recently Bali et al. (2016) proposes a multivariate normal GARCH model from which the conditional distribution defines the dynamic beta coefficient. This directly links time-varying second moments to the time-varying beta in a consistent fashion. The parametric pricing relationship holds more generally for the elliptic family of distributions. This is an attractive approach but may be limiting if the parametric distributional assumptions are not valid.

A key insight of our approach is that if the joint distribution of excess stock returns and market returns are correctly specified then it follows that their contemporaneous pricing relationship is completely determined by the associated conditional distribution. Therefore, we semiparametrically model the conditional distribution as a countably infinite mixture of normals. Each normal component in the mixture has a conditional covariance directed by a MGARCH process. Our model nests the Gaussian and Student-t distribution as special cases but importantly allows for deviations from the elliptic family of distributions. This includes asymmetric distributions which the elliptic family omit being only symmetric. The mixing is over both the mean vector and covariance matrix.

We follow Jensen and Maheu (2013) to implement a Bayesian semi-parametric MGARCH model and extend it to allow for asymmetric shocks in volatility. The data strongly support the semiparametric MGARCH specification over Gaussian and Student-t distributional alternatives.

We use a new approach to selecting the number of factors in a model. Since spefications with a different number of factors are not comparable by the usual Bayes factors due to different dimensions of the dependent variable we select the number of factors based on the marginal predictive likelihood. This relies on the marginal predictive likelihood of the individual stock return derived from models with different dimensions and is directly comparable across specifications. Empirically, the one factor model is strongly supported for all stocks compared to specifications with Fama-French factors and momentum.

In this framework, the conditional distribution of stock returns given the market excess return (and possibly other factors) can be represented as an infinite mixture with weights written as functions of the value of the market excess return. Consequently, the beta coefficient of a security at each time will depend nonparametrically on the contemporaneous value of market return, as opposed to the beta derived from existing models which is insensitive to the contemporaneous value of the market return.

Although the time series of the realized conditional betas from the semiparametric model are similar to the benchmark model, we find significant nonlinear dependence in beta as a function of the contemporaneous value of the market excess return. In the parametric models, beta is constant as a function of the market excess return.

When the market is highly volatile, beta is not affected by unexpected shocks in the market return. While in a calm market, beta can change dramatically from unexpected shocks. For stocks which are highly correlated with the market, an unexpected shock during calm periods increases the beta coefficient. The effect is the reverse for the stocks with low correlation with the market. In other words, when an asset is highly correlated with the market, large moves in a stable market increase the conditional covariance between the market and the asset more than they increase the conditional variance of the market, resulting in a significant increase in the beta coefficient. When an asset has low conditional correlation with the market, large moves in a stable market increase the conditional variance of the market more than they increase the conditional covariance between the market and the asset, leading to a drop in conditional beta. These are important contemporaneous nonlinear dynamics that are absent in other models.

The remainder of the chapter is structured as follows. We begin by reviewing the benchmark model which is an MGARCH model with Student-t innovations. Section 3.3 provides a general theoretical setting of the multivariate model used in this study, key features of the semiparametric MGARCH model, and the use of the Dirichlet process prior. Posterior sampling is detailed in Section 3.4. The derivation of the nonparametric dynamic conditional beta is presented in Section 3.5. Data is introduced in Section 3.6, and Section 3.7 assesses models with different number of factors and compares the performance of the proposed model to the benchmark model. Applications of the semiparametric model are found in Section 3.8, and Section 3.9 provides some implications of the semiparametric model in finance. Section 3.10 concludes and an Appendix defines distributions and collects the detailed derivations.

3.2 Benchmark Model

Our benchmark model is a straightforward extension of Bali et al. (2016). Bali et al. (2016) defines dynamic conditional beta using a multivariate GARCH (MGARCH) model assuming a multivariate normal distribution as the joint density of stock returns and factors. We replace the normal distribution with a Student-t to accommodate the fat-tails in the data. Let the excess stock return on asset i be $r_{i,t}$ and a vector of regressors (factors) including the excess market return be $r_{f,t} = (r_{f_1,t}, r_{f_2,t}, ..., r_{f_q,t})'$. $r_t = (r_{i,t}, r'_{f,t})'$ is assumed to follow the MGARCH-t

$$r_t | r_{1:t-1} \sim t(\mu, H_t, \nu), \tag{3.1}$$

$$H_t = \Gamma_0 + \Gamma_1 \odot (r_{t-1} - \eta)(r_{t-1} - \eta)' + \Gamma_2 \odot H_{t-1}, \tag{3.2}$$

where $t(\mu, \Sigma, \nu)$ denotes a t-distribution (see appendix) with mean vector μ, scale matrix Σ and degree of freedom ν and $r_{1:t-1} = \{r_1, \ldots, r_{t-1}\}$ is the information set available at time

$t-1$. The scale matrix, H_t, is based on the vector-diagonal multivariate GARCH model of Ding and Engle (2001) but other MGARCH formulations could be used. The symbol $\odot$ denotes the Hadamard product. The parameter is $\Gamma = \{\Gamma_0, \Gamma_1, \Gamma_2, \eta\}$, with the symmetric positive definite matrices parameterized as $\Gamma_0 = \Gamma_0^{1/2}(\Gamma_0^{1/2})'$, $\Gamma_1 = \gamma_1(\gamma_1)'$, and $\Gamma_2 = \gamma_2(\gamma_2)'$ where Γ_0 is a lower triangular $(q+1)\times(q+1)$ matrix and γ_1, γ_2 and η are $(q+1)$-vectors. η permits a nonlinear asymmetric response to shocks and can be considered a multivariate version of the asymmetric GARCH model (Engle and Ng, 1993).

Partition $r_t = (r_{1,t}', r_{2,t}')'$ into a k_1 and k_2 ($k_1+k_2 = q+1$) vector and similarly $\mu = (\mu_1', \mu_2')'$ and

$$H_t = \begin{bmatrix} H_{11,t} & H_{12,t} \\ H_{12,t} & H_{22,t} \end{bmatrix}.$$

Applying the properties of the Student-t distribution (Roth, 2013) the conditional distribution of $r_{1,t}$ given $r_{2,t}$ is

$$\begin{aligned} r_{1,t}|r_{2,t} &\sim t(\mu_{1|2}, H_{t,1|2}, \nu_{1|2}), && (3.3)\\ \mu_{1|2} &= \mu_1 + H_{12,t}H_{22,t}^{-1}(r_{2,t}-\mu_2), && (3.4)\\ H_{t,1|2} &= \frac{\nu + (r_{2,t}-\mu_2)'H_{22,t}^{-1}(r_{2,t}-\mu_2)}{\nu + k_2}(H_{11,t} - H_{12,t}H_{22,t}^{-1}H_{12,t}'), && (3.5)\\ \nu_{1|2} &= \nu + k_2, && (3.6) \end{aligned}$$

where the conditional mean is $\mu_{1|2}$, the conditional scale matrix is $H_{t,1|2}$ and the degree of freedom $\nu_{1|2}$.

This is a useful result in that it tells us how the conditional distribution of $r_{1,t}$ reacts to any value of $r_{2,t}$. For instance, if $r_{1,t} \equiv r_{i,t}$ and conditioning on one factor, the excess market return, $r_{2,t} \equiv r_{m,t}$, substituting into Equation (3.4) directly gives a dynamic risk premium for asset i as

$$E[r_{i,t}|r_{m,t}, H_t] = \mu_i + H_{12,t}H_{22,t}^{-1}(r_{m,t} - \mu_m). \tag{3.7}$$

This tells how the expected excess return of asset i reacts to any value of the market. If the market shock is zero ($r_{m,t} = \mu_m$) then the expected value is μ_i but for all other realizations the market shock impacts the expected return of the asset. Engle identifies the dynamic conditional beta that arises from the joint relationship as

$$\beta_t = H_{12,t} H_{22,t}^{-1}. \tag{3.8}$$

This is the derivative of Equation (3.7) with respect to $r_{m,t}$. A conditional pricing relationship is obtained by setting $r_{2,t} \equiv E[r_{m,t}|r_{1:t-1}]$ and substituting into Equation (3.7).

There are several advantages to modeling excess returns in this way. First, it confronts the simultaneous nature of the asset return and the factors that price the risk premium. Rather than specifying a single equation partial equilibrium relationship the model begins with the full joint dynamics. Second, the joint distribution of the asset and the factors directly pins down the conditional distribution and the implications for the risk premium. The dynamic beta is a function of the conditional covariance matrix. This is a general result that holds for the elliptic family of distributions.

We estimate the model from a Bayesian perspective. The posterior density has the non-standard form

$$p(\mu, \Gamma, \nu | r_{1:T}) \propto p(\nu) p(\mu) p(\Gamma) \times \prod_{t=1}^{T} t(r_t | \mu, H_t, \nu), \tag{3.9}$$

where $t(r_t|\mu, H_t, \nu)$ is the density of the Student-t distribution, and $p(\nu)p(\mu)p(\Gamma)$ is the prior density for μ, Γ, ν. Posterior draws of the parameters vector are simulated with a Metropolis-Hastings sampler.

Although attractive, the conditional distribution in Equation (3.3) has some drawbacks. The conditional beta derived from MGARCH-t model, at each time, is constant with respect to the contemporaneous value of market return (Equation 3.8), and consequently, the conditional expected return of the stock is a linear function of the factor returns. This pricing relationship will not hold for more general distributions not belonging to the elliptic family. The elliptic family of distributions are symmetric about their mean and do not account for asymmetry observed in financial returns.

This model imposes a strong assumption on the functional form of the joint distribution of the data. In this chapter, I remove this restrictive assumption by employing a Dirichlet process mixture (DPM) to model the unknown joint distribution of returns. This results in a potentially non-constant conditional beta and a nonlinear conditional expected return of the stock as a function of the contemporaneous value of the market return.

3.3 MGARCH-DPM Model

Unlike the benchmark model that assumes a specific parametric joint distribution for the individual asset returns and the factors, we model this joint distribution nonparametrically by an infinite mixture of normal distributions which can approximate any continuous multivariate distribution. Recall that $r_t = (r_{i,t}, r_{f_1,t}, ..., r_{f_q,t})'$ represents the excess return vector of an individual stock and q factors at time t. The infinite mixture representation can be written as

$$r_t | H_t, \mu, B, W \sim \sum_{j=1}^{\infty} \omega_j N(\mu_j, (H_t^{1/2}) B_j (H_t^{1/2})'). \tag{3.10}$$

where $H_t^{1/2}$ is the Cholesky decomposition of H_t, $\mu = \{\mu_1, \mu_2, \dots\}$, $B = \{B_1, B_2, \dots\}$ and $W = \{\omega_1, \omega_2, \dots\}$ is the vector of the weights, such that $\omega_j \geq 0$ for all j and $\sum_{j=1}^{\infty} \omega_j = 1$. The mixing is over the mean vector and the component B_j of the covariance matrix. The second component, H_t of the covariance matrix captures volatility clustering through time but is not a function of j. B_j is a symmetric positive definite matrix which scales H_t to yield a better estimate of the joint density of data.

The conditional mean can be derived in exactly the same way as in the benchmark model except it will follow an infinite mixture of conditional normal distributions. If $r_{ft} = (r_{f_1,t}, ..., r_{f_q,t})'$ then the conditional density of $r_{i,t}$ given $r_{f,t}$ is a mixture distribution as well and the conditional expectation can be written as the following weighted mixture

$$E(r_{i,t}|r_{f,t}, H_t) = \sum_{j=1}^{\infty} q_j(r_{f,t}) E(r_{i,t}|r_{f,t}, \mu_j, B_j, H_t). \tag{3.11}$$

The weights, $q_j(r_{f,t})$ are a function of the factors and affect how much each conditional expectation, $E(r_{i,t}|r_{f,t}, \mu_j, B_j, H_t)$, in the mixture contributes. The details on the derivations will be explained later but for now it is important to see that unlike the parametric model the conditional expectation is not a linear function of the factors. To obtain the nonparametric conditional beta, we take the derivative of Equation (3.11) with respect to the desired factor. The conditional beta is not constant in general but it changes as the contemporaneous value of the corresponding factor changes. The next section introduces the Dirichlet process prior to estimate this model. In Section 3.5 we derive the nonparametric conditional beta.

In Bayesian inference the Dirichlet process (DP) prior (Ferguson, 1973) is a standard prior used for infinite dimensional objects such as Equation (3.10). A draw from a DP, $G \sim DP(\alpha, G_0)$, is almost surely a discrete distribution and is governed by two parameters. The concentration parameter α, a positive scalar and a base distribution G_0. The nonparametric distribution G is centered on the base distribution G_0, which can be considered as the prior guess; $E(G) = G_0$. The concentration parameter measures the strength of belief in G_0. The larger α, the stronger belief in G_0 and the more distinct elements we have with non-negligible mass. Lo (1984) introduces Dirichlet process mixture (DPM) model in which G is the mixing measure over a continuous kernel. This has become a standard Bayesian approach to nonparametric estimation of an unknown continuous distribution. In this chapter, G is the unknown distribution that governs the mixing over the mean vector and covariance matrix of the normal kernel in our mixture model.

The model (MGARCH-DPM) is an extension of Jensen and Maheu (2013) and allows for asymmetry in the MGARCH process from shocks to volatility and fat tails without making any restrictive assumption. The hierarchical form of the model is,

$$
\begin{aligned}
r_t \quad & |\phi_t, H_t \sim N(\xi_t, H_t^{1/2}\Lambda_t(H_t^{1/2})'),\ t = 1, ..., T && (3.12)\\
\phi_t \quad & \equiv \{\xi_t, \Lambda_t\}|G \sim G, && (3.13)\\
G| \quad & \alpha, G_0 \sim DP(\alpha, G_0), && (3.14)\\
G_0 \quad & \equiv N(\mu_0, D) \times \mathcal{W}^{-1}(B_0, \nu_0), && (3.15)\\
H_t \quad & = \Gamma_0 + \Gamma_1 \odot (r_{t-1} - \eta)(r_{t-1} - \eta)' + \Gamma_2 \odot H_{t-1}. && (3.16)
\end{aligned}
$$

In this model ξ_t is a $(q+1)$-vector and Λ_t is a symmetric positive definite matrix and H_t follows the same MGARCH specification as the benchmark parametric model. $\mathcal{W}^{-1}(B_0, \nu_0)$ represents an inverse Wishart distribution (see appendix) with scale matrix B_0 and degree of freedom ν_0.

Sethuraman (1994) characterizes a stick-breaking representation of the DP. Combining this with the normal kernel gives the associated stick breaking representation of the MGARCH-DPM density as

$$
\begin{aligned}
p \quad & (r_t|\mu, B, W, H_t) = \sum_{j=1}^{\infty} \omega_j N(r_t|\mu_j, H_t^{1/2} B_j (H_t^{1/2})'), && (3.17)\\
\omega_1 \quad & = v_1,\ \omega_j = v_j \prod_{l=1}^{j-1}(1 - v_l),\ j > 1, && (3.18)\\
v_j \quad & \overset{iid}{\sim} \text{Beta}(1, \alpha), && (3.19)\\
\mu_j \quad & \overset{iid}{\sim} N(\mu_0, D), \quad B_j \overset{iid}{\sim} \mathcal{W}^{-1}(B_0, \nu_0), && (3.20)
\end{aligned}
$$

where $N(r_t|\mu_j, H_t^{1/2} B_j (H_t^{1/2})')$ denotes the multivariate normal density with mean μ_j and covariance $H_t^{1/2} B_j (H_t^{1/2})'$ evaluated at r_t. Note that μ and B are the set of unique points of support in the discrete distribution G while ξ_t and Λ_t denote draws from G in Equation (3.13), with the possibility of repeated draws of μ_j and B_j.

The model nests several special cases. First, the Gaussian model is obtained when $\alpha \to 0$ as $\omega_1 = 1$, $\omega_j = 0, \forall j > 1$ and $B_1 = I$. The Student-t model results from μ_j being constant for all j and $\alpha \to \infty$, since $G \to G_0$, the inverse Wishart distribution.

3.4 Posterior Sampling

To estimate the unknown parameters in Equations (3.12)-(3.16), we apply an MCMC sampler along with the slice sampler of Walker (2007). Slice sampling introduces a latent variable, $u_t \in (0,1)$, to elegantly convert an infinite sum to a finite mixture model which makes the sampling feasible. Estimating the joint posterior density of u_t and other model parameters and then integrating out the slice variable u_t recovers the desired posterior density. In practice, this means jointly sampling all parameters including the slice variable but then discarding u_t. Define u_t such that the joint density of (r_t, u_t) given $(W, \Theta \equiv (\mu, B))$ is given by

$$f(r_t, u_t | W, \Theta) = \sum_{j=1}^{\infty} \mathbf{1}(u_t < \omega_j) N(r_t | \mu_j, (H_t^{1/2})' B_j H_t^{1/2}). \tag{3.21}$$

Let $s_{1:T} = \{s_1, ..., \ s_T\}$ be the configuration set that partitions the data $r_{1:T}$ into c distinct clusters such that observation r_t is assigned parameter $\theta_{s_t} = (\mu_{s_t}, B_{s_t})$. Let $n_j = \{\#t | s_t = j\}$ be the number of observations allocated to state j. The full likelihood is

$$p(r_{1:T}, u_{1:T}, s_{1:T} | W, \Theta) = \Pi_{t=1}^{T} \mathbf{1}(u_t < \omega_{s_t}) N(r_t | \mu_{s_t} + \Gamma_{s_t} r_{t-1}, (H_t^{1/2}) B_{s_t} (H_t^{1/2})'), \tag{3.22}$$

and the joint posterior is proportional to

$$p(W_{1:K})\Pi_{j=1}^{K}p(\mu_j, B_j)\Pi_{t=1}^{T}\mathbf{1}(u_t < \omega_{s_t})N(r_t|\mu_{s_t} + \Gamma_{s_t}, (H_t^{1/2})B_{s_t}(H_t^{1/2})') \tag{3.23}$$

where K is the smallest natural number that satisfies the condition $\sum_{j=1}^{K}\omega_j > 1 - \min\{u_t\}_{t=1}^{T}$ and $W_{1:K}$ denotes the finite set of W and similarly for other parameters $\mu_{1:K}$ and $B_{1:K}$. Having defined the notation, the steps of the MCMC algorithm are discussed next.

Steps of MCMC algorithm for MGARCH-DPM Model

1. The posterior distribution of $\theta_j = (\mu_j, B_j),\ j = 1, ..., K$: Using the transformation $z_t = H_t^{-1/2} r_t$, and applying the results of conditionally conjugate priors for the linear regression model we have

$$B_j|r_{1:T}, s_{1:T}, \mu_j, \Gamma \sim \mathcal{W}^{-1}\left(n_j + \nu_0, B_0 + \sum_{s_t=j}(z_t - H_t^{-1/2}\mu_j)(z_t - H_t^{-1/2}\mu_j)'\right) \tag{3.24}$$

$$\mu_j|r_{1:T}, s_{1:T}, B_j, \Gamma \sim N(\bar{\mu}, \bar{D}) \tag{3.25}$$

in which

$$\bar{D}^{-1} = D^{-1} + \sum_{t|s_t=j} H_t^{-1/2'} B_j^{-1} H_t^{-1/2},\ \bar{\mu} = \bar{D}\left(\sum_{t|s_t=j} H_t^{-1/2'} B_j^{-1} z_t + D^{-1}\mu_0\right). \tag{3.26}$$

2. Updating $v_j,\ j = 1, ..., K$.

$$v_j|S \sim \text{Beta}\left(1 + \sum_{t=1}^{T}\mathbf{1}(s_t = j), \alpha + \sum_{t=1}^{T}\mathbf{1}(s_t > j)\right). \tag{3.27}$$

Then we update $W_{1:K}$ based on Equation (3.18).

3. Updating $u_t,\ t = 1, ..., T$. $u_t|s_{1:T} \sim \mathcal{U}(0, \omega_{s_t})$. Then update K such that $\sum_{j=1}^{K}\omega_j >$

$1-\min\{u_t\}_{t=1}^{T}$. Additional ω_j and θ_j will need to be generated from the priors if K is incremented.

4. Updating $s_{1:T}$. For each $t=1,...,T$,

$$p(s_t=j|r_{1:T}) \propto \mathbf{1}(\omega_j>u_t)N(r_t|\mu_j, H_t^{1/2}B_j(H_t^{1/2})'), j=1,...,K. \tag{3.28}$$

5. Updating α: Assuming a gamma prior $\alpha \sim \mathcal{G}(a_0, b_0)$ (see appendix) α can be sampled following the two steps below (Escobar and West, 1995). Recall that c is the number of alive clusters defined as the number of clusters in which at least one observation is allocated. Note that $c \leq K$. Then the sampling steps are as follows.

   (a) $(\tau|\alpha, c) \sim \text{Beta}(\alpha+1, T)$.

   (b) Sample α from

   $$\alpha|\tau \sim \pi_\tau \mathcal{G}(a_0+c, b_0-\log(\tau)) + (1-\pi_\tau)\mathcal{G}(a_0+c-1, b_0-\log(\tau)),$$

   where π_τ is defined by $\frac{\pi_\tau}{1-\pi_\tau} = \frac{a_0+c-1}{T(b_0-\log(\tau))}$.

6. Updating GARCH parameters $\Gamma = (\Gamma_0^{1/2}, \gamma_1, \gamma_2, \eta)$. The conditional posterior is

$$p(\Gamma|\mu, B, S, r_{1:T}, H_{1:T}) \propto p(\Gamma) \times \prod_{t=1}^{T} N(r_t|\mu_{s_t}, H_t^{1/2}B_{s_t}(H_t^{1/2})') \tag{3.29}$$

which is not of standard form, and we apply a Metropolis-Hastings sampler. Given the current value Γ of the chain, the proposal Γ' is sampled $\Gamma' \sim N(\Gamma, \widehat{V})$. The draw is accepted with probability

$$\min\{p(\Gamma'|\mu, B, S, r_{1:T}, H_{1:T})/p(\Gamma|\mu, B, S, r_{1:T}, H_{1:T}), 1\},$$

and otherwise rejected. $\widehat{V}$ is proportional to the inverse Hessian matrix of $\ell = \log[p(\Gamma|\mu, B, S, r_{1:T}, H_{1:T})]$ evaluated at its posterior mode, $\widehat{\Gamma}$, which is computed once at the start of estimation. $\widehat{V}$ is scaled to achieve an acceptance rate between 0.2 and 0.5. In this chapter we apply Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to approximate the posterior mode of ℓ.

3.5 Nonparametric Dynamic Conditional Beta

To study the behaviour of the conditional beta of an individual stock, we first consider a special case of our model, $r_t = (r_{i,t}, r_{m,t})$ where $r_{i,t}$ and $r_{m,t}$ represent an individual stock's excess return and the market excess return, respectively. Applying the posterior sampling algorithm, we sample model parameters for many iterations and after dropping a set of burn-in draws we have the following set of sampled parameters:

$$\{(\mu_j^{(g)}, B_j^{(g)}), v_j^{(g)}, j = 1, ..., K^{(g)}\}, \{s_t^{(g)}, u_t^{(g)}, t = 1, ..., T\}, H_{1:T}^{(g)} = \{H_1^{(g)}, ..., H_T^{(g)}\}, \quad (3.30)$$

for $g = 1, ..., M$ where M is the number of MCMC iterations. At each iteration $g = 1, ..., M$ of the algorithm, a draw of $G|r_{1:T}$, can be written as

$$G^{(g)} = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)} \delta_{\theta_j^{(g)}} + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) G_0(\theta), \quad (3.31)$$

where $\theta_j^{(g)} = (\mu_j^{(g)}, B_j^{(g)})$ and $\delta_{\theta_j^{(g)}}$ is a mass point at $\theta_j^{(g)}$.

Combining this with the normal kernel gives the predictive density for the generic return

$(\mathbf{r}_{i,t}, \mathbf{r}_{m,t})$ conditional on $G^{(g)}$ as

$$p(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|\theta) G_0(\theta) d\theta, \tag{3.32}$$

where $f(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|\theta)$ is the multivariate normal density.

To assess the nonlinear regression function $E(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T})$, or the conditional beta of the individual stock i, we require the conditional density derived from this predictive (joint) density of $(\mathbf{r}_{i,t}, \mathbf{r}_{m,t})$. Therefore,

$$\begin{aligned} p(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T}, G^{(g)}) &= \frac{p(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|r_{1:T}, G^{(g)})}{p(\mathbf{r}_{m,t}|r_{1:T}, G^{(g)})} \qquad (3.33) \\ &= \frac{p(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|r_{1:T}, G^{(g)})}{\sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f_2(\mathbf{r}_{m,t}|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f_2(\mathbf{r}_{m,t}|\theta) G_0(\theta) d\theta} \\ &= \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t}) f(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, \theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})\right) f(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, G_0), \qquad (3.34) \end{aligned}$$

where

$$q_j^{(g)}(\mathbf{r}_{m,t}) = \frac{\omega_j^{(g)} f_2(\mathbf{r}_{m,t}|\theta_j^{(g)})}{\sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f_2(\mathbf{r}_{m,t}|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f_2(\mathbf{r}_{m,t}|\theta) G_0(\theta) d\theta} \tag{3.35}$$

and $f_2(\mathbf{r}_{m,t}|\theta_j^{(g)})$ is the marginal (normal) density of $\mathbf{r}_{m,t}$ and $f(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, G_0)$ is the conditional distribution using the base measure. The terms $q_j^{(g)}(\mathbf{r}_{m,t})$ determine which components in the mixture receive more weight. Clusters that have a marginal density $f_2(\mathbf{r}_{m,t}|\theta_j^{(g)})$ that has a higher likelihood value for $\mathbf{r}_{m,t}$ will receive larger weights. The marginal density, and hence relative weight of clusters, will change with $\mathbf{r}_{m,t}$ as well as over time through the MGARCH component, H_t. These features will determine the relative weights on the cluster specific conditional expectations which we derive next.

Our focus is on the conditional mean of $\mathbf{r}_{i,t}$ given $\mathbf{r}_{m,t}$. Using the properties of the normal distribution the conditional mean directly comes from Equation (3.34) and is

$$E(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})[\mu_{j,1}^{(g)} + \beta_{jt}^{(g)}(\mathbf{r}_{m,t} - \mu_{j,2}^{(g)})] + \left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})\right) \frac{\int [\mu_1 + \beta_t(\mathbf{r}_{m,t} - \mu_2)] N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}{\int N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}. \tag{3.36}$$

The cluster specific beta is defined as

$$\beta_{jt}^{(g)} = \frac{(H_t^{(g)^{1/2}} B_j H_t^{(g)^{1/2'}})_{12}}{(H_t^{(g)^{1/2}} B_j H_t^{(g)^{1/2'}})_{22}} \tag{3.37}$$

where the subscript (i, j) on $()_{ij}$ denotes element (i, j) of the matrix and β_t in the second line of Equation (3.36) is defined as $\beta_{jt}^{(g)}$ except B_j is replaced with B. The numerator and denominator in the last term of Equation (3.36) can be approximated by simulation.

Integrating all parameter and distributional uncertainty results in an estimate of the predictive conditional mean as

$$E(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T}) \approx \frac{1}{M} \sum_{g=1}^{M} E(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T}, G^{(g)}). \tag{3.38}$$

The predictive mean of the conditional beta is the derivative of this conditional expectation of $\mathbf{r}_{i,t}$ given $\mathbf{r}_{m,t}$, Equation (3.38) with respect to $\mathbf{r}_{m,t}$. This is,

$$b_{m,t}(\mathbf{r}_{m,t}) = \frac{\partial E(r_{i,t}|r_{m,t}, r_{1:T})}{\partial r_{m,t}} \Big|_{r_{m,t}=\mathbf{r}_{m,t}}. \tag{3.39}$$

Full details on this derivative and estimate are provided in the appendix.

In the case that we have more than one factor, we follow the same process. We first estimate the joint model and back out the conditional distribution of the stock return $r_{i,t}$ given all factors. The nonparametric conditional beta in this case is a vector. It is defined analogously to Equation (3.39) as the partial derivative with respect to the factor. For instance in the case of the Fama-French 3-factor model with size factor ($r_{SMB,t}$), value factor ($r_{HML,t}$), and market factor (Fama and French, 1993), beta for size factor is defined as

$$b_{SMB,t} = \frac{\partial E(r_{i,t}|r_{m,t}, r_{SMB,t}, r_{t,HML}, r_{1:T})}{\partial r_{SMB,t}}\Big|_{\substack{r_{m,t}=\mathbf{r}_{m,t} \\ r_{SMB,t}=\mathbf{r}_{SMB,t} \\ r_{HML,t}=\mathbf{r}_{HML,t}}} \tag{3.40}$$

with a similar expression for the other factor coefficients $b_{m,t}$ and $b_{HML,t}$.

3.6 Data

We use the value-weighted index constructed by the Center of Research in Security Prices (CRSP) as a proxy for market returns. Daily market excess returns as well as five individual stock excess returns for IBM, General Electric or GE, Exxon or XOM, Amgen or AMGN, and bank of America or BAC are obtained from 2000/01/03 to 2013/12/31 (3521 daily observations). Excess returns are derived after subtracting the risk-free return approximated by the three-month Treasury bill rate. All returns are scaled by 100. Figure 3.1 displays the data and Table 3.1 reports summary statistics. All individual stocks display skewness and excess kurtosis. Figure 3.1 shows that returns with absolute large (small) value tend to be followed by other large (small) absolute returns reflecting volatility clustering. Daily data for the size factor, $r_{SMB,t}$, value factor, $r_{HML,t}$, and momentum factor, $r_{MOM,t}$, are obtained from Kenneth French's website.

3.7 Model Performance

The criteria that we use to compare different models is the value of the log-predictive likelihood. For each particular model $\mathcal{M}$ (i.e., MGARCH-t or MGARCH-DPM), the predictive likelihood for $r_{L:T}$, $L < T$ is expressed in terms of the one-step-ahead predictive densities,

$$m(r_{L:T}|r_{1:L-1}, \mathcal{M}) = \Pi_{t=L}^{T} p(r_t|r_{1:t-1}, \mathcal{M}) \tag{3.41}$$

where $L > 1$ is chosen to eliminate the influence of the priors on model comparison. We can approximate the one-step-ahead predictive likelihoods, $p(r_t|r_{1:t-1}, \mathcal{M})$, by averaging the data density over draws of the unknown parameters conditional on the data history $r_{1:t-1}$. This integrates out parameter and distributional uncertainty as

$$\begin{aligned} p(r_t|r_{1:t-1}, \mathcal{M}) &= \int p(r_t|\theta, r_{1:t-1}, \mathcal{M}) p(\theta|r_{1:t-1}, \mathcal{M}) d\theta \\ &\approx \frac{1}{M} \sum_{g=1}^{M} p(r_t|\theta^{(g)}, r_{1:t-1}, \mathcal{M}) \end{aligned} \tag{3.42}$$

where $\theta^{(g)}$ is a posterior draw from $p(\theta|r_{1:t-1}, \mathcal{M})$ and $p(r_t|\theta^{(g)}, r_{1:t-1}, \mathcal{M})$ is the data density given $\theta^{(g)}$ and $r_{1:t-1}$ for model $\mathcal{M}$.

The following priors are used in estimation. In the MGARCH-t model, $\nu \sim \mathcal{U}(2, 100)$, and $\mu \sim N(0, 0.1I)$ for both models. For each of GARCH parameters in both models, we set $\Gamma_{0,ij}^{1/2} \sim N(0, 100)\mathbf{1}_S$, $\gamma_{1,i} \sim N(0, 100)\mathbf{1}_S$ and $\gamma_{1,i} \sim N(0, 100)\mathbf{1}_S$, $i = 1, \ldots, q+1$, $j \leq i$ as prior distribution where S denotes the following restriction: $\text{diag}(\Gamma_0^{1/2}) > 0, \gamma_{11} > 0, \gamma_{22} > 0$ to impose identification. For the concentration parameter $\alpha \sim \mathcal{G}(0.1, 0.3)$. The prior on α controls the number of the distinct components in the mixture model, although with a large number of observations the effect of the prior is diminished. For the hyper-parameters of the base measure G_0, we set $B_0 = (\nu_0 - q - 1)I$ which makes $E(B) = I$ and centers the conditional covariance of r_t at H_t. $\nu_0 = 8$, but other values for ν_0 do not change our conclusions.

Based on Equation (3.42), the predictive likelihoods for the two models are estimated as

$$p(r_t|r_{1:t-1}, \text{MGARCH-t}) \approx \frac{1}{M}\sum_{g=1}^{M} t(r_t|\mu^{(g)}, H_t^{(g)}, \nu^{(g)}), \tag{3.43}$$

$$p(r_t|r_{1:t-1}, \text{MGARCH-DPM}) \approx \frac{1}{M}\sum_{g=1}^{M} N(r_t|\mu^{(g)}_{s_t^{(g)}}, H_t^{(g)^{1/2}} B^{(g)}_{s_t^{(g)}} H_t^{(g)^{1/2'}}). \tag{3.44}$$

Note that we are able to compute $H_t^{(g)}$ at each iteration of the MCMC since we have $H_{t-1}^{(g)}$ and GARCH parameters: $H_t^{(g)} = \Gamma_0^{(g)} + \Gamma_1^{(g)} \odot (r_{t-1} - \eta^{(g)})(r - \eta^{(g)})'_{t-1} + \Gamma_2^{(g)} \odot H_{t-1}^{(g)}$. In MGARCH-DPM model, at each iteration g, $s_t^{(g)}$

To determine the factors to be used, we compare the values of the *marginal* predictive likelihood of the individual stock return derived from each model, using different factors. The predictive likelihoods discussed above are directly comparable but when comparing a model with 2 factors versus 3 factors the independent variable r_t is 2 dimensional and 3 dimensional, respectively. These predictive likelihood values are not comparable. Instead we compare the marginal predictive likelihood for the individual stock return only. This is obtained from each full model after integrating out the factors in each model. For instance, for excess stock return i we compare the one-factor model against the two-factor model with $p(r_{i,t}|r_{i,1:t-1}, r_{f_1,1:t-1})$ and $p(r_{i,t}|r_{i,1:t-1}, r_{f_1,1:t-1}, r_{f_2,1:t-1})$. These marginal predictive likelihoods are derived from the full predictive likelihood. For example, the first one is obtained from $p(r_{i,t}, r_{f_1,t}|r_{i,1:t-1}, r_{f_1,1:t-1})$ by marginalizing out $r_{f_1,t}$. This can be done directly on the terms (3.43) and (3.44) by selecting the associated univariate marginal density from the multivariate Student-t and normal on the right hand side of these equations.

We first compare the performance of the MGARCH-DPM model with different factors. These factors include market excess return, size factor and value factor from the Fama-French 3-factor model, and the momentum factor. The set of factors can be extended to include any factor. Table 3.2 reports the marginal log-predictive likelihood of IBM, BAC, GE, XOM and

AMGN, for the MGARCH-DPM model, for 2012/03/12 to 2013/12/31 (500 observations) when we use different factors. The table shows that, for all stocks under study, 1-factor model with market excess return as the only factor results in a better marginal predictive likelihood compared to the 3 and 4-factor models. The evidence for one factor is very strong. For instance, the log-predictive Bayes factor for the 1 factor IBM model against the 3 factor version is 243.35. Therefore the remainder of the empirical results focus on the the 1-factor model.

Table 3.3 reports the log-predictive likelihoods for the 1-factor MGARCH-t and MGARCH-DPM models, and the log-Bayes factor over 2012/03/12 to 2013/12/31. Bivariate models based on daily excess returns on IBM, GE, XOM, AMGN and BAC each with excess market returns are considered. The results strongly support our semi-parametric model relative to the benchmark model. For instance, log-Bayes factors are all greater than 211. This is very strong evidence of significant deviations from the Student-t MGARCH model.

Figure 3.2 displays the time-series of the market and IBM excess returns as well as the difference in the log-predictive likelihood of the two models using

$$\log p(r_t|r_{1:t-1}, \text{MGARCH-DPM}) - \log p(r_t|r_{1:t-1}, \text{MGARCH-t}). \tag{3.45}$$

Positive values favour the MGARCH-DPM specification. This figure shows that the MGARCH-DPM model almost always outperforms MGARCH-t model. There are large differences when the market or IBM returns are extreme.

3.8 Applications of Semiparametric Conditional Beta

This section presents empirical estimates of the nonparametric dynamic conditional beta from the MGARCH-DPM model for several individual stocks and compares them with the corresponding counterpart from the parametric MGARCH-t model. Not only does the beta computed in this way change over time, but also the time-varying conditional beta is sensitive to the contemporaneous value of excess market return. This implies that the value of the systematic risk of an asset at each time depends on the level of the market return.

The model is applied to derive a nonparametric conditional beta (calculated in Section 3.5) using excess returns on a single stock and on the market return ($q = 1$). This results in a conditional expected return of the individual stock comparable to the conditional CAPM model.

The analysis reported here is based on 25000 iterations of the MCMC algorithm. The first 15000 draws were dropped as burn-in and the following 10000 used for inference. The average acceptance rate of GARCH parameters is about 20% and about 30% for parametric and nonparametric models, respectively.

Tables 3.4-3.8 report the posterior mean and the 0.95 probability density intervals of the fixed parameters for both models and for different stocks. The estimated MGARCH parameters from the two models are consistent. The tables report c, the number of components in the mixture used to estimate the unknown density. On average, the bivariate joint density of IBM, XOM, GE, and BAC with the market is estimated using about 3.6-6.3 components but the density intervals indicate considerable uncertainty. However, for AMGN and the market, about 15 components are used, showing that this joint density is far more complex than the others. These results are compatible with the small degree of freedom estimated in the benchmark models. Estimates of η_1 and η_2 are consistently positive indicating a larger response to the conditional covariance from negative shocks.

Figures 3.3-3.7 compare the posterior mean of the *realized* beta over time derived from both models for each of the stocks. For MGARCH-t model, the posterior mean of Equation (3.8) is reported while for the MGARCH-DPM model the posterior mean of Equation (3.39) is evaluated at the realized excess market return value for time t. As seen in the figures, both models result in very similar time series for the conditional beta.

Figures 3.8-3.12 illustrate posterior mean of each stock's conditional beta as a function of the contemporaneous market excess return using Equation (3.39) at several dates. These figures show that beta is changing over time and, more importantly, at each time the value of beta is sensitive to the contemporaneous value of the market excess return. For each stock there are dates that beta is a constant function of the market return which would be consistent with the MGARCH-t model. However, each stock has dates in which beta is nonlinearly dependent on the market return. Moreover, often beta is asymmetrically related to the market; when the market excess return increases (large positive values), conditional beta drops more significantly (Figures 3.8-3.10).

The nonlinear relationship between beta and the market transfers directly into the conditional expected excess return. For example, Figure 3.13 displays the posterior mean of the conditional expected excess return of IBM given different values of the contemporaneous market excess return, derived from Equation (3.38), for dates for which the conditional betas are illustrated in Figure 3.8. This figure clearly shows how the nonlinear conditional beta results in the nonlinear conditional expected return.

To investigate the significance of this nonlinear relationship Figures 3.14-3.18 display the posterior mean of the nonparametric conditional beta as a function of the market excess return as well as the 0.90 density intervals for selected dates for each stock. Beta derived from the MGARCH-t model is included and is a constant function at each time. It is clear from these figures that there are significant departures in beta from the constant beta from the MGARCH-t model.

Finally, Figures 3.19-3.23 provide a three dimensional version of Figures 3.8-3.12 for each stock. In some periods beta is essentially flat and consistent with the MGARCH-t model while in other times beta is very sensitive to the market return.

3.8.1 Summary of Empirical Results

As the empirical results illustrate, the conditional beta is time-varying and at each time depends on the contemporaneous market excess return, as opposed to the constant beta of the benchmark model.

The previous results show some periods in which the conditional beta is insensitive to the value of $r_{m,t}$ (beta is almost constant with respect to $r_{m,t}$) while in other time periods beta changes significantly with $r_{m,t}$. To measure the sensitivity of $b_{m,t}(\mathbf{r}_{m,t})$ to $\mathbf{r}_{m,t}$ at each time t consider the following measure

$$d_t = \max_{\mathbf{r}_{m,t}} b_{m,t}(\mathbf{r}_{m,t}) - \min_{\mathbf{r}_{m,t}} b_{m,t}(\mathbf{r}_{m,t}), \tag{3.46}$$

where $b_{m,t}(\mathbf{r}_{m,t})$ is defined in Equation (3.39). Large values of d_t indicate that $b_t(\mathbf{r}_{m,t})$ is strongly sensitive to $\mathbf{r}_{m,t}$, while a $d_t = 0$ indicates no sensitivity. The MGARCH-t model has a $d_t = 0$ for all t. Figure 3.24 illustrates this d_t over time for all individual stocks. Among these four stocks, the dynamic conditional beta for IBM and BAC have the most sensitivity and XOM has the least sensitivity to $\mathbf{r}_{m,t}$. What is apparent is that during relatively high volatility periods such as 2002-03, 2009 and 2011:6-2012, d_t attains its smallest values over the sample. In these periods shocks to the market are expected to be large. During lower volatility periods large shocks to the market and firms are unexpected and the conditional beta adjusts accordingly.

To investigate how $b_{m,t}(\mathbf{r}_{m,t})$ changes with different market conditions Figures 3.25-3.29 show the broad trends that we find in all stocks. When the market is highly volatile, an individual stock's conditional beta is less affected by unexpected shocks in the contemporaneous market return. While in a calm market, the conditional beta changes remarkably from unexpected shocks to the market. However, the changes depend on the stocks correlation with the market.

When the market is calm, an unexpected shock increases the conditional beta for a stock that is highly correlated with the market, while this effect is completely the reverse for stocks with low correlation with the market. In other words, when an asset is highly correlated with the market, a large move in a stable market increases the conditional covariance between the market and the asset more than it increases the conditional variance of the market, resulting in a significant increase in the conditional beta. When an asset is less correlated with the market, a large move in a stable market increases the conditional variance of the market more than it increases the conditional covariance between the market and the asset, leading to a drop in the conditional beta.

It is often the case that the effect on $b_{m,t}(\mathbf{r}_{m,t})$ from $\mathbf{r}_{m,t}$ is asymmetric. Frequently $b_{m,t}(\mathbf{r}_{m,t})$ is more sensitive to large positive values of $\mathbf{r}_{m,t}$ compared to negative values. In addition, when the market is calm, we see both u-shape and inverse u-shape patterns for the conditional beta of all stocks.

3.9 Implications of the semiparametric model in Finance

From Equation (3.34), we are able to examine the whole conditional density of the stock given factors. Therefore we are able to study the individual stock's conditional expected return under different risk scenarios. For example, the semiparametric model allows us to study the effect of big shocks in the factors (i.e., market return) on stock's expected return and risk measures such as value-at-risk.

Consider the predictive conditional expected return of IBM at time t derived from the 1-factor model, $E[r_{IBM,t}|r_{m,1:t-1}, r_{IBM,1:t-1}]$. Using the semiparametric model, this value is a nonlinear function of $r_{m,t}$. Therefore, when a large shock is expected to the market, this shock affects our expectation of the IBM return nonlinearly. While in the benchmark model this effect is linear. For instance, Figure 3.30 illustrates IBM's predictive conditional expected return for a specific date (2000-11-30). From this figure we can assess the expected impact of a large positive or negative shock to the market on the value of IBM's return.

Consider a second example of a large realized market shock in 2008-10-28. We can study how the semiparametric model is able to predict the effect of this shock on IBM's expected return. Figure 3.31 shows IBM's predictive **conditional** expected return for this day derived from the benchmark model and the semiparametric model. The realized market return and IBM on this day are %9.77 and %9.56, respectively. This point is illustrated on the graph as well. It is clear how the nonlinearity resulting from the semiparametric model reduces the prediction error. This shows how we can benefit from the semiparametric model in the events that we expect big positive or negative shocks in the market (i.e., a political event, a new financial policy).

Figure 3.31 shows only one specific date. We looked at all dates that the market has realized a big shock (more than 6%) and compared the performance of the semiparametric model with the benchmark. The root mean squared error of the prediction for the benchmark and the semiparameric model is 8.394 and 8.131, respectively, showing the outperformance of the semiparametric model by 3.2% improvement in prediction.

In addition to the expected return, the semiparametric model enables us to study the effect of big shocks in the market on IBM's whole conditional density and different risk measures. Figure 3.32 illustrates the effect of +5% and −5% shocks in the market return on IBM's predictive conditional density on 2012-05-17 derived from the semiparametric 1-factor model. The value at risk of investment in IBM when we have no shock in the market is 2.306%. A +5% shock in the market return decreases the value at risk to 0.180%, while a −5% shocks in the market return increases the value at risk to 4.471%. Therefore, we can carry out different risk scenario analyses in order to indicate the effect of big shocks in the market on our investment in a specific firm.

3.10 Conclusion

This chapter derives a dynamic conditional beta representation using a Bayesian semiparametric multivariate GARCH model. We show how to select the number of factors and that the predictive Bayes factors strongly support this semiparametric model over a multivariate GARCH with Student-t innovations. Empirically we find the time-varying beta from our model nonlinearly depends on the contemporaneous value of excess market return. In highly volatile markets, beta is almost constant, while in stable markets, the beta coefficient can depend asymmetrically on the contemporaneous value of the market excess return.

3.11 Appendix

3.11.1 Distributions

If $\mathbf{r} \sim t(\mu, \Sigma, \nu)$ then the density function of the Student-t (Bauwens et al., 2000) is

$$f(\mathbf{r}|\nu, \mu, \Sigma) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}}|\Sigma|^{-1/2}\left[1 + \frac{1}{\nu}(\mathbf{r}-\mu)^T\Sigma^{-1}(\mathbf{r}-\mu)\right]^{-(\nu+p)/2}, \nu > 0.$$

The $q \times q$ matrix B follows an inverse Wishart density with a symmetric positive definite scale matrix B_0 and degree of freedom $\nu_0 \geq q+1$, if its pdf can be written as

$$f(B|B_0, \nu_0) = \frac{|B_0|^{\nu_0/2}}{2^{\frac{q\nu_0}{2}}\pi^{\frac{q(q-1)}{4}}\Pi_{i=1}^{q}\Gamma(\frac{\nu_0+1-i}{2})}|B|^{-\frac{\nu_0+q+1}{2}}\exp\left[-\frac{1}{2}tr(B^{-1}B_0)\right],$$

with $E(B) = \frac{1}{\nu_0-q-1}B_0$.

The pdf of the Gamma distribution $\mathcal{G}(a, b)$ with shape parameter a and scale parameter b is written as

$$f(x|a,b) = \frac{b^a}{\Gamma(a)}x^{a-1}e^{-xb},\ x \in (0, \infty),\quad E(x) = \frac{a}{b}.$$

3.11.2 Derivation of the nonparametric conditional beta

$$E(\mathbf{r}_{i,t}|\mathbf{r}_{i,t}, r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})[\mu_{j,1}^{(g)} + \beta_{jt}^{(g)}(\mathbf{r}_{m,t} - \mu_{j,2}^{(g)})] + \tag{3.47}$$
$$\left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})\right) \frac{\int [\mu_1 + \beta_t(\mathbf{r}_{m,t} - \mu_2)] N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}{\int N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}.$$

Let

$$A_1 = \int [\mu_1 + \beta_t(\mathbf{r}_{m,t} - \mu_2)] N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB, \tag{3.48}$$

$$A_2 = \int N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB. \tag{3.49}$$

A_1 and A_2 can be easily approximated by Monte Carlo simulation as follows

$$A_1 \approx \frac{1}{N} \sum_{n=1}^{N} [\mu_{n,1} + \beta_{n,t}^{(g)}(\mathbf{r}_{m,t} - \mu_{n,2})] N(\mathbf{r}_{m,t}|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22}) \tag{3.50}$$

$$A_2 \approx \frac{1}{N} \sum_{n=1}^{N} N(\mathbf{r}_{m,t}|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22}) \tag{3.51}$$

where μ_n and $B_n,\ n = 1, ..., N$ are i.i.d draws from the prior $p(\mu, B)$ which in our model is $N(\mu|\mu_0, D)$ and $\mathcal{W}^{-1}(B|B_0, \nu_0)$, and

$$\beta_{nt}^{(g)} = \frac{(H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{12}}{(H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22}}. \tag{3.52}$$

Now we obtain the posterior mean of the nonparametric conditional beta by taking the derivative of Equation (3.47):

$$b_{m,t}(\mathbf{r}_{m,t}) = \frac{1}{M} \sum_{g=1}^{M} b_{m,t}(\mathbf{r}_{m,t}, G^{(g)}) = \frac{1}{M} \sum_{g=1}^{M} \frac{\partial E(r_{i,t}|r_{m,t}, r_{1:T}, G^{(g)})}{\partial r_{m,t}}\Big|_{r_{m,t}=\mathbf{r}_{m,t}}. \tag{3.53}$$

After replacing A_1 and A_2 with their approximations we have

$$\frac{\partial E(r_{i,t}|r_{m,t}, r_{1:T}, G^{(g)})}{\partial r_{m,t}} \approx \sum_{j=1}^{K^{(g)}} q_j^{(g)}(r_t^m)\beta_{jt}^{(g)} \tag{3.54}$$
$$+ \sum_{j=1}^{K^{(g)}} q_j'^{(g)}(r_t^m)[\mu_{j,1}^{(g)} + \beta_{jt}^{(g)}(r_t^m - \mu_{j,2}^{(g)})]$$
$$- \sum_{j=1}^{K^{(g)}} q_j'^{(g)}(r_t^m)\frac{\sum_n[\mu_{n,1} + \beta_{tn}^{(g)}(r_t^m - \mu_{n,2})]N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}$$
$$+ \left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(r_t^m)\right)\{\frac{\sum_n \beta_{tn}^{(g)} N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}$$
$$+ \frac{\sum_n[\mu_{n,1} + \beta_{tn}^{(g)}(r_t^m - \mu_{n,2})]N'(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}$$
$$- \frac{[\sum_n[\mu_{n,1} + \beta_{tn}^{(g)}(r_t^m - \mu_{n,2})]N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})] \sum_n N'(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{[\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})]^2}\}$$

where $\beta_{jt}^{(g)}$, $\beta_{nt}^{(g)}$, and $q_j^{(g)}(r_t^m)$ are defined in Equations (3.37), (3.52), and (3.35), respectively, and $N'(x|.)$ is the derivative of the pdf of Normal distribution with respect to x. In the case that we have more than one factor (say q factors), the derivations follow similarly but the derivative will be a vector of size q, each element of which is the coefficient of the corresponding factor.

Stock	Mean	Variance	Skewness	Kurtosis	Max	Min
Market	0.017	1.744	-0.070	7.067	11.350	-8.950
IBM	0.028	3.070	0.230	7.834	13.019	-15.567
GE	-0.003	4.277	0.323	8.397	19.702	-12.797
XOM	0.032	2.672	0.367	11.163	17.180	-13.950
AMGN	0.034	4.758	0.508	5.907	15.090	-13.437
BAC	0.031	10.701	0.891	23.399	35.261	-28.969

Table 3.1: Summary statistics of the daily excess returns on the market portfolio, IBM, GE, XOM and AMGN, BAC from 2000/01/03 to 2013/12/31 (3521 observations).

Stock	Log Marginal Predictive Likelihood 1-factor model	3-factor model	4-factor model
IBM	-774.37	-1017.72	-1240.97
AMGN	-917.05	-1200.30	-1347.57
BAC	-1071.36	-1235.12	-1451.62
XOM	-671.59	-889.56	-1074.24
GE	-774.43	-970.31	-1123.54

Table 3.2: This table reports the marginal log-predictive likelihood for MGARCH-DPM model, for the last 500 observations, from 2012/03/12 to 2013/12/31. Data are daily excess market, HML, and SMB returns coupled with excess returns on IBM, BAC, and AMGN from 2000/01/03 to 2013/12/31.

	log-predictive likelihood				
Model	IBM	GE	XOM	AMGN	BAC
MGARCH-DPM	-983.27	-964.99	-875.47	-1140.12	-1473.11
MGARCH-t	-1353.67	-1369.03	-1300.21	-1571.32	-1684.72
log-Bayes factor vs MGARCH-t	370.40	404.04	424.74	431.20	211.61

Table 3.3: This table reports the log-predictive likelihood for the bivariate MGARCH-t and MGARCH-DPM models and the log-Bayes factors, for the last 500 observations, from 2012/03/12 to 2013/12/31. Bivariate data are daily excess market returns coupled with excess returns on IBM, GE, XOM, AMGN, and BAC from 2000/01/03 to 2013/12/31.

IBM	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.102	$(0.055, 0.146)$	0.023	$(0.015, 0.037)$
γ_{02}	-0.043	$(-0.081, 0.003)$	-0.042	$(-0.053, -0.034)$
γ_{03}	0.020	$(0.001, 0.053)$	0.020	$(0.002, 0.048)$
γ_{11}	0.247	$(0.199, 0.307)$	0.150	$(0.144, 0.160)$
γ_{12}	0.267	$(0.232, 0.313)$	0.224	$(0.210, 0.233)$
γ_{21}	0.971	$(0.965, 0.977)$	0.975	$(0.971, 0.977)$
γ_{22}	0.953	$(0.945, 0.961)$	0.955	$(0.951, 0.961)$
μ_1			0.025	$(0.016, 0.046)$
μ_2			0.041	$(0.022, 0.074)$
ν			5.37	$(5.01, 5.54)$
c	5.6	$(3.00, 11.0)$		
α	0.571	$(0.070, 1.61)$		
η_1	0.570	$(0.349, 0.714)$	0.807	$(0.776, 0.864)$
η_2	0.533	$(0.434, 0.618)$	0.507	$(0.451, 0.644)$

Table 3.4: IBM Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on IBM and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

XOM	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.141	$(0.108, 0.182)$	0.110	$(0.012, 0.200)$
γ_{02}	0.014	$(-.003, 0.030)$	0.016	$(-0.058, 0.073)$
γ_{03}	0.014	$(0.001, 0.041)$	0.032	$(0.001, 0.082)$
γ_{11}	0.250	$(0.223, 0.283)$	0.228	$(0.165, 0.310)$
γ_{12}	0.238	$(0.198, 0.287)$	0.228	$(0.175, 0.288)$
γ_{21}	0.956	$(0.947, 0.965)$	0.958	$(0.935, 0.977)$
γ_{22}	0.960	$(0.953, 0.969)$	0.958	$(0.939, 0.974)$
μ_1			0.025	$(-0.076, 0.129)$
μ_2			0.022	$(-0.050, 0.092)$
ν			9.89	$(6.16, 13.90)$
c	3.6	$(2.00, 9.00)$		
α	0.324	$(0.011, 1.15)$		
η_1	0.480	$(0.345, 0.591)$	0.436	$(-0.051, 0.775)$
η_2	0.524	$(0.436, 0.613)$	0.514	$(0.279, 0.708)$

Table 3.5: XOM Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on XOM and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

GE	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.061	$(0.023, 0.093)$	0.031	$(0.012, 0.056)$
γ_{02}	-0.033	$(-0.054, -0.014)$	-0.029	$(-0.039, -0.008)$
γ_{03}	0.018	$(0.001, 0.052)$	0.036	$(0.022, 0.052)$
γ_{11}	0.196	$(0.174, 0.216)$	0.170	$(0.145, 0.188)$
γ_{12}	0.204	$(0.181, 0.225)$	0.180	$(0.168, 0.192)$
γ_{21}	0.974	$(0.967, 0.981)$	0.974	$(0.970, 0.981)$
γ_{22}	0.964	$(0.957, 0.970)$	0.971	$(0.967, 0.974)$
μ_1			0.004	$(-0.034, 0.029)$
μ_2			0.049	$(0.015, 0.071)$
ν			6.47	$(5.35, 7.05)$
c	5.04	$(3.00, 10.0)$		
α	0.501	$(0.060, 1.42)$		
η_1	0.554	$(0.414, 0.707)$	0.633	$(0.555, 0.785)$
η_2	0.464	$(0.395, 0.539)$	0.463	$(0.416, 0.561)$

Table 3.6: GE Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on GE and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

AMGN	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.137	$(0.089, 0.171)$	0.084	$(0.065, 0.106)$
γ_{02}	-0.011	$(-0.031, 0.012)$	-0.028	$(-0.044, -0.007)$
γ_{03}	0.016	$(0.001, 0.039)$	0.034	$(0.015, 0.059)$
γ_{11}	0.211	$(0.182, 0.239)$	0.165	$(0.156, 0.175)$
γ_{12}	0.188	$(0.172, 0.211)$	0.228	$(0.195, 0.242)$
γ_{21}	0.965	$(0.945, 0.958)$	0.973	$(0.971, 0.976)$
γ_{22}	0.951	$(0.945, 0.958)$	0.956	$(0.950, 0.965)$
μ_1			0.002	$(-0.014, 0.035)$
μ_2			0.038	$(0.024, 0.070)$
ν			5.81	$(5.56, 6.08)$
c	15	$(7.00, 28.0)$		
α	2.41	$(0.500, 5.21)$		
η_1	0.508	$(0.428, 0.596)$	0.768	$(0.686, 0.876)$
η_2	0.542	$(0.459, 0.630)$	0.479	$(0.443, 0.566)$

Table 3.7: AMGN Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on AMGN and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

BAC	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.064	$(0.024, 0.115)$	0.065	$(0.045, 0.091)$
γ_{02}	-0.031	$(-0.062, 0.012)$	-0.022	$(-0.036, -0.008)$
γ_{03}	-0.007	$(-0.061, 0.042)$	0.029	$(0.004, 0.051)$
γ_{11}	0.284	$(0.235, 0.351)$	0.219	$(0.206, 0.238)$
γ_{12}	0.212	$(0.183, 0.260)$	0.213	$(0.197, 0.229)$
γ_{21}	0.962	$(0.954, 0.968)$	0.962	$(0.956, 0.966)$
γ_{22}	0.955	$(0.945, 0.963)$	0.963	$(0.956, 0.969)$
μ_1			0.000	$(-0.033, 0.033)$
μ_2			0.040	$(0.021, 0.072)$
ν			6.37	$(5.992, 6.839)$
c	6.24	$(3.000, 12.000)$		
α	0.658	$(0.092, 1.800)$		
η_1	0.436	$(0.326, 0.547)$	0.472	$(0.400, 0.554)$
η_2	0.414	$(0.333, 0.501)$	0.432	$(0.354, 0.558)$

Table 3.8: BAC Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on BAC and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

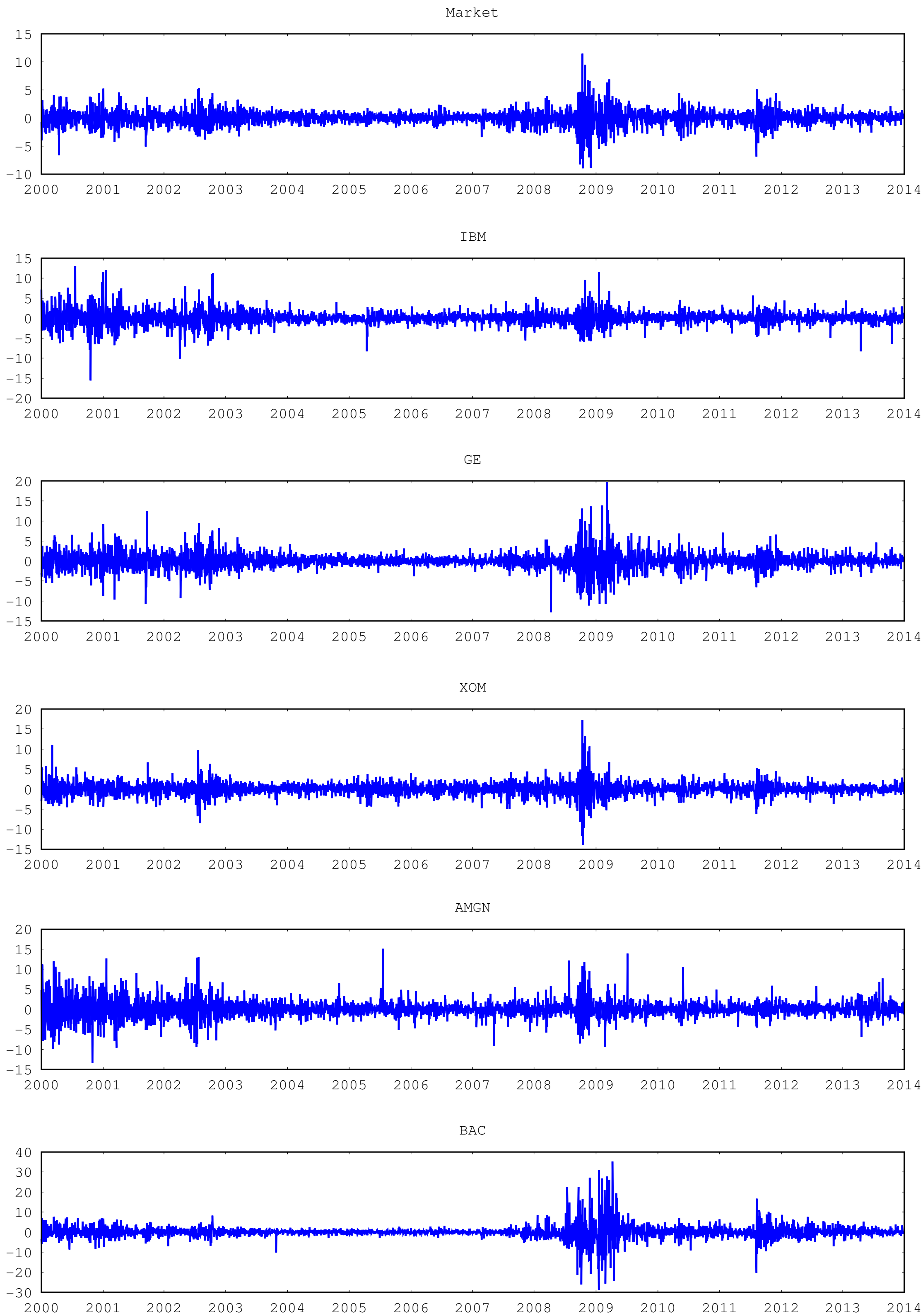

Figure 3.1: Daily excess returns on the market, IBM, GE, XOM, AMGN and BAC.

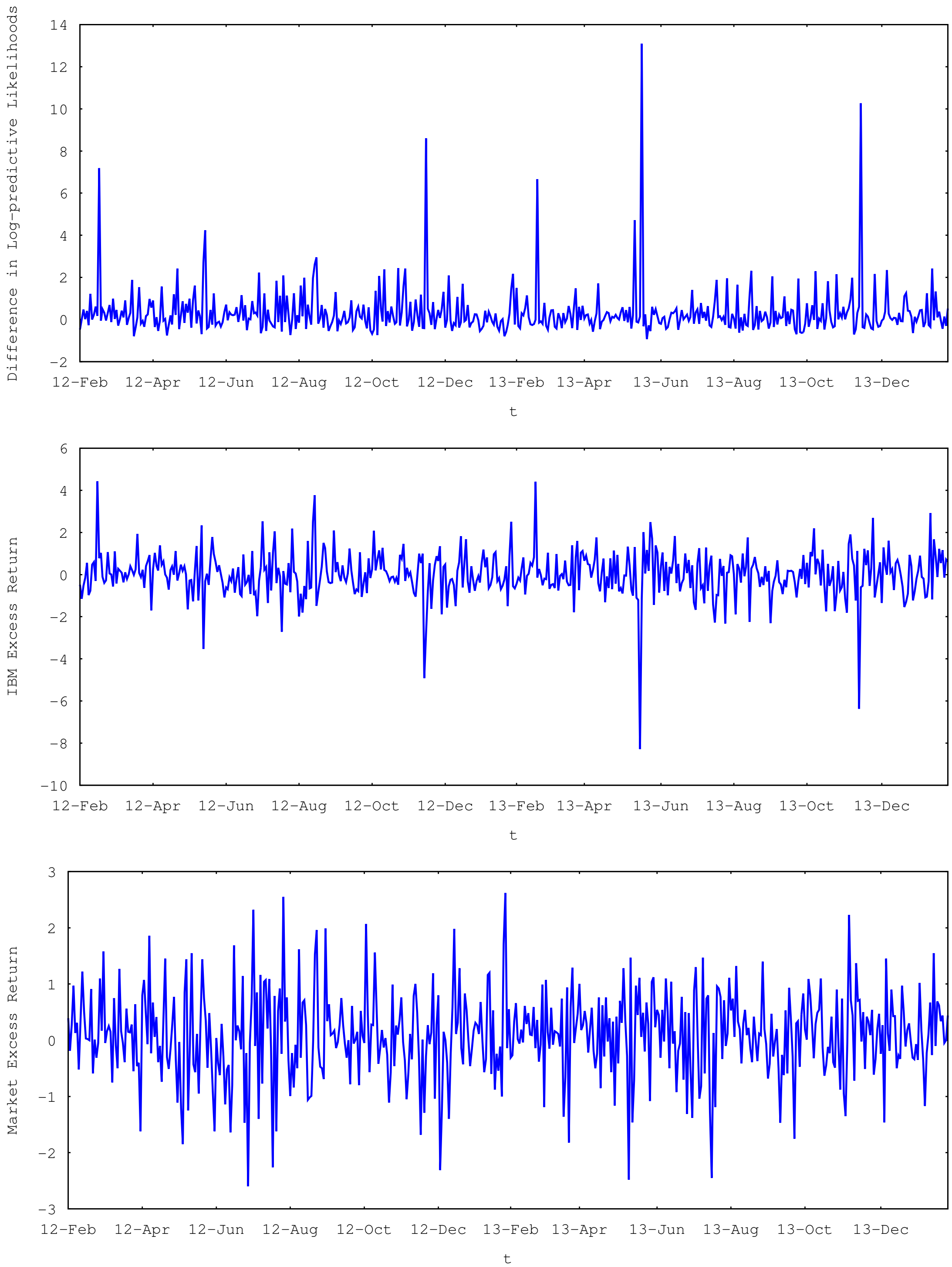

Figure 3.2: The first panel indicates the difference of log predictive likelihood of the two models corresponding to each of the last 500 observations, from 2012/01/05 to 2013/12/31, for MGARCH-t and MGARCH-DPM. The second and third panel illustrate the time series returns on IBM and the market.

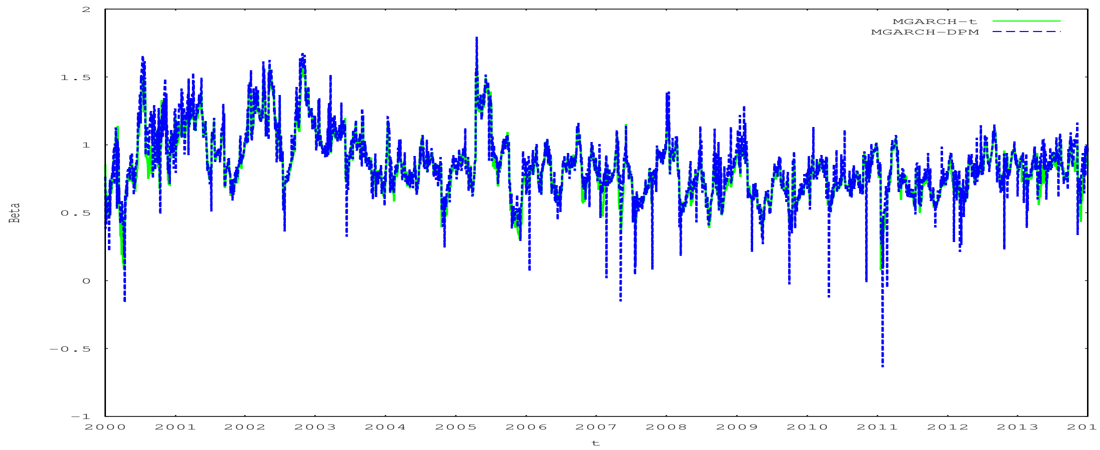

Figure 3.3: IBM: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

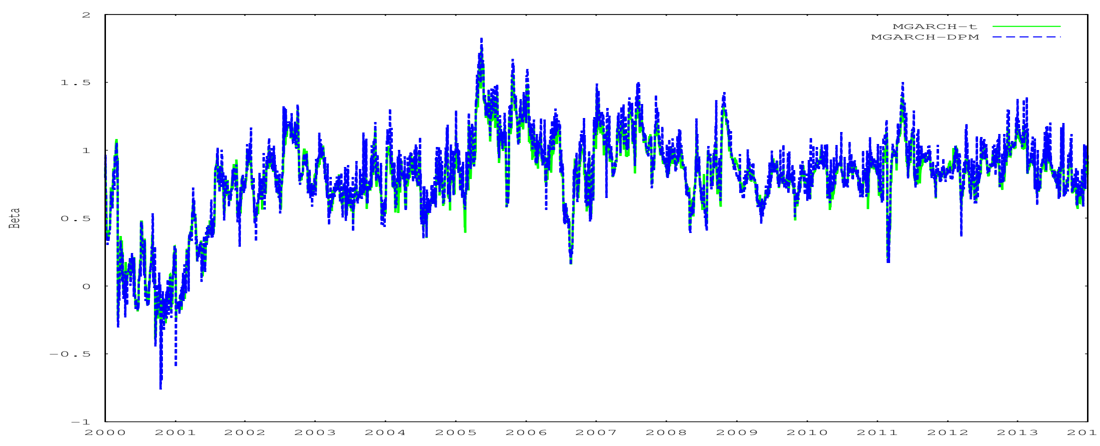

Figure 3.4: XOM: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

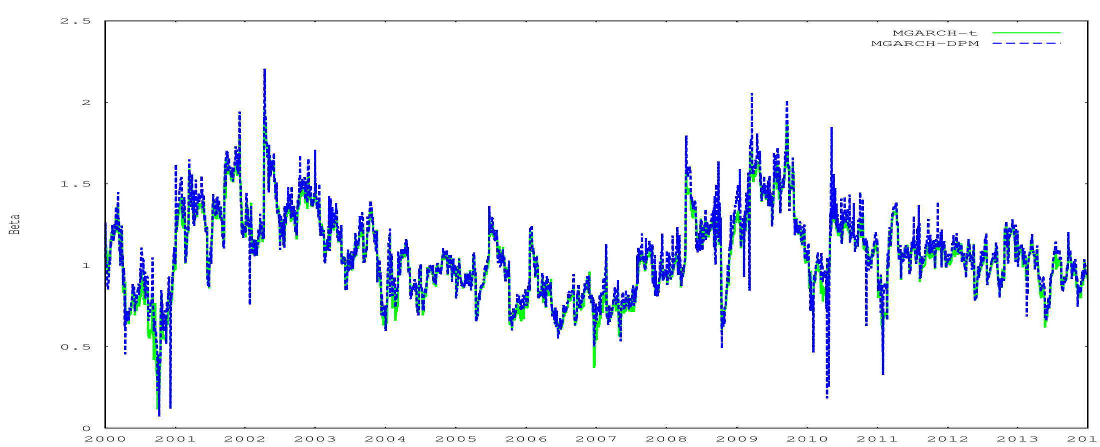

Figure 3.5: GE: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

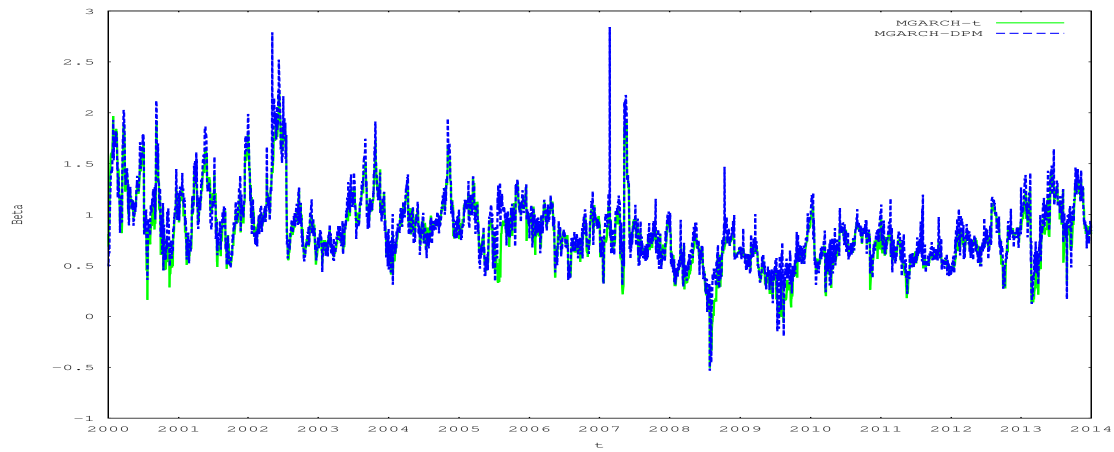

Figure 3.6: AMGN: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

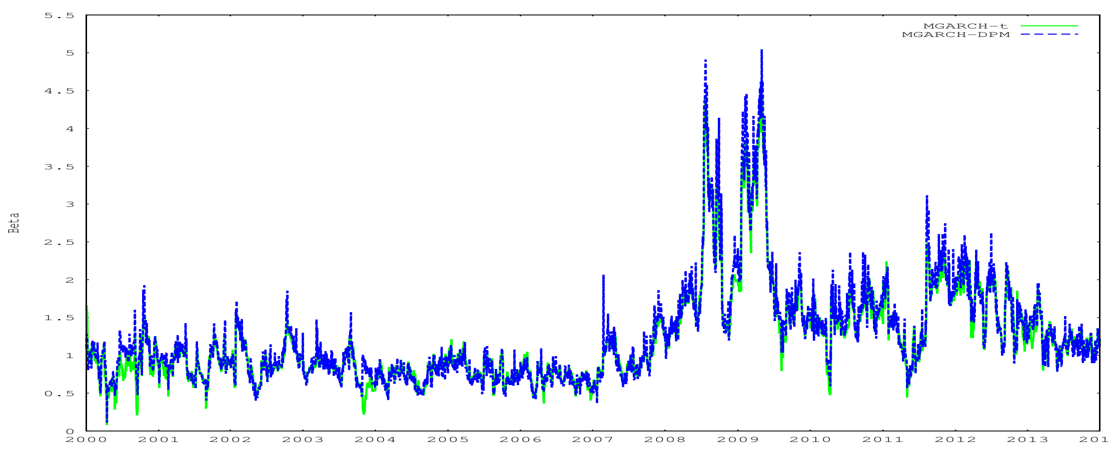

Figure 3.7: BAC: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

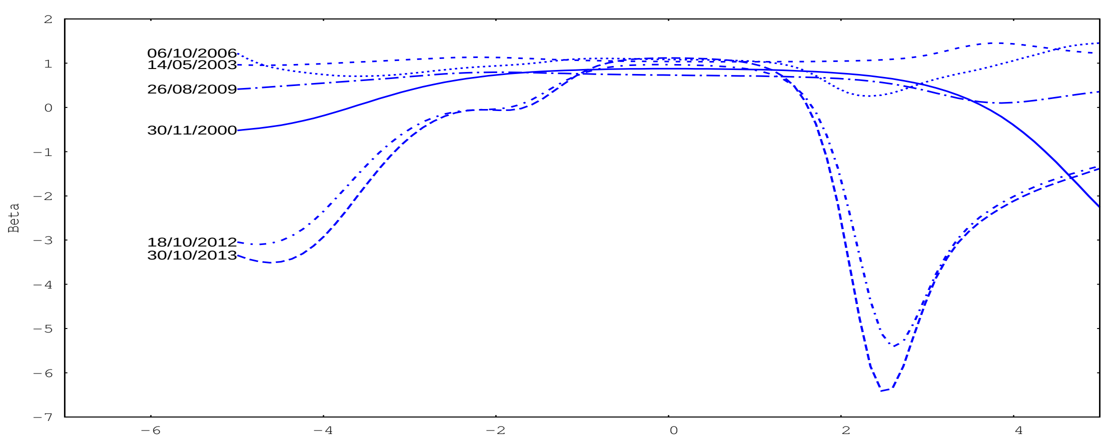

Figure 3.8: IBM: posterior mean of conditional beta as a function of the market excess return for different dates.

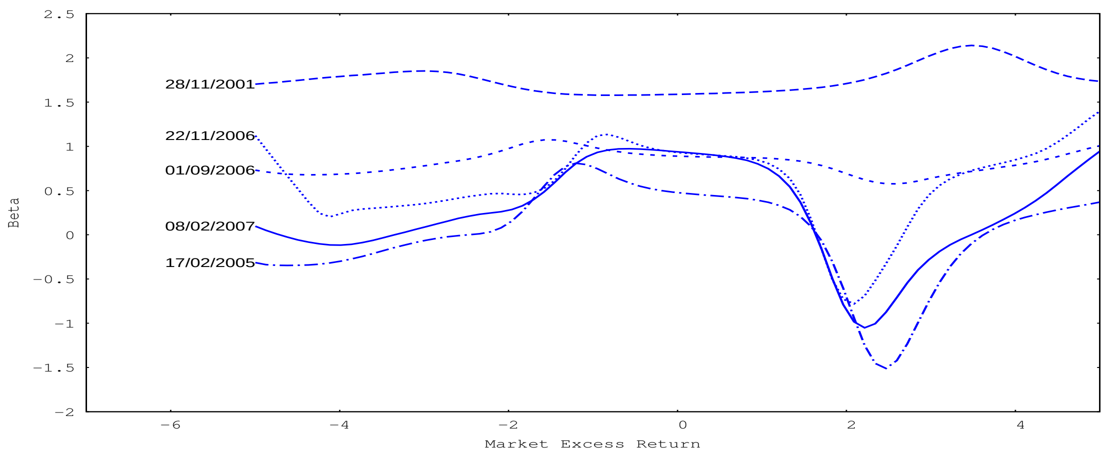

Figure 3.9: XOM: posterior mean of conditional beta as a function of the market excess return for different dates.

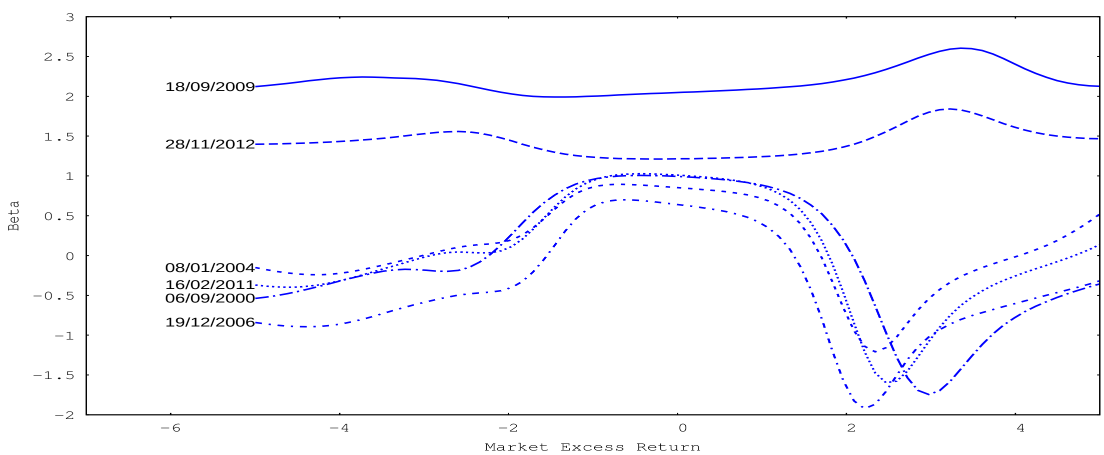

Figure 3.10: GE: posterior mean of conditional beta as a function of the market excess return for different dates.

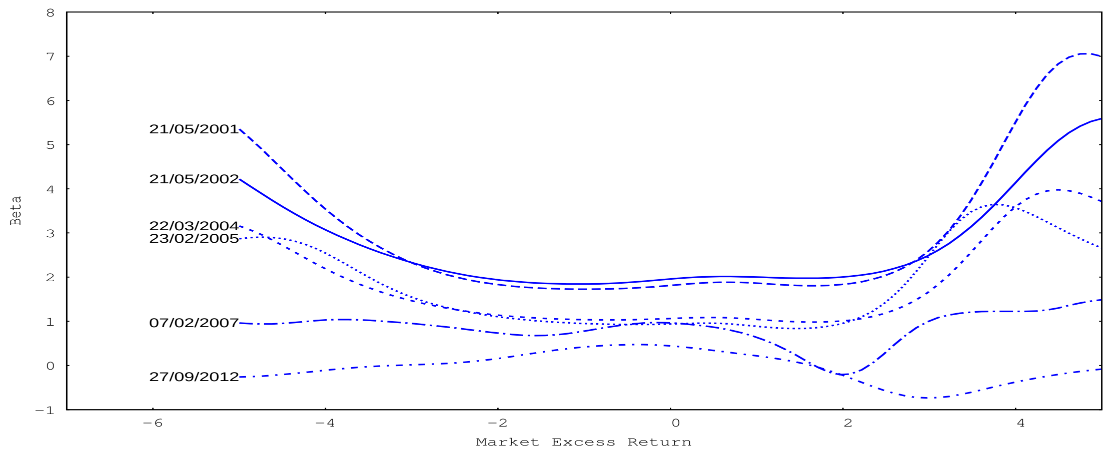

Figure 3.11: AMGN: posterior mean of conditional beta as a function of the market excess return for different dates.

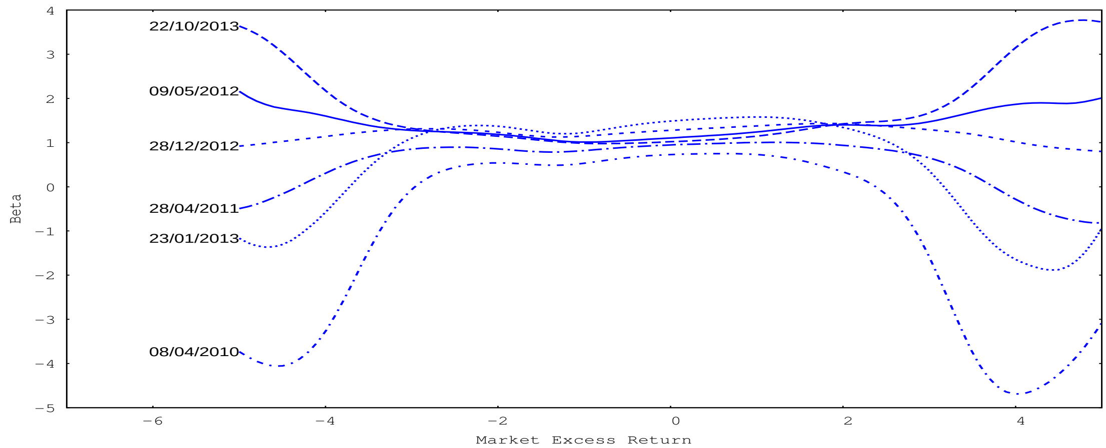

Figure 3.12: BAC: posterior mean of conditional beta as a function of the market excess return for different dates.

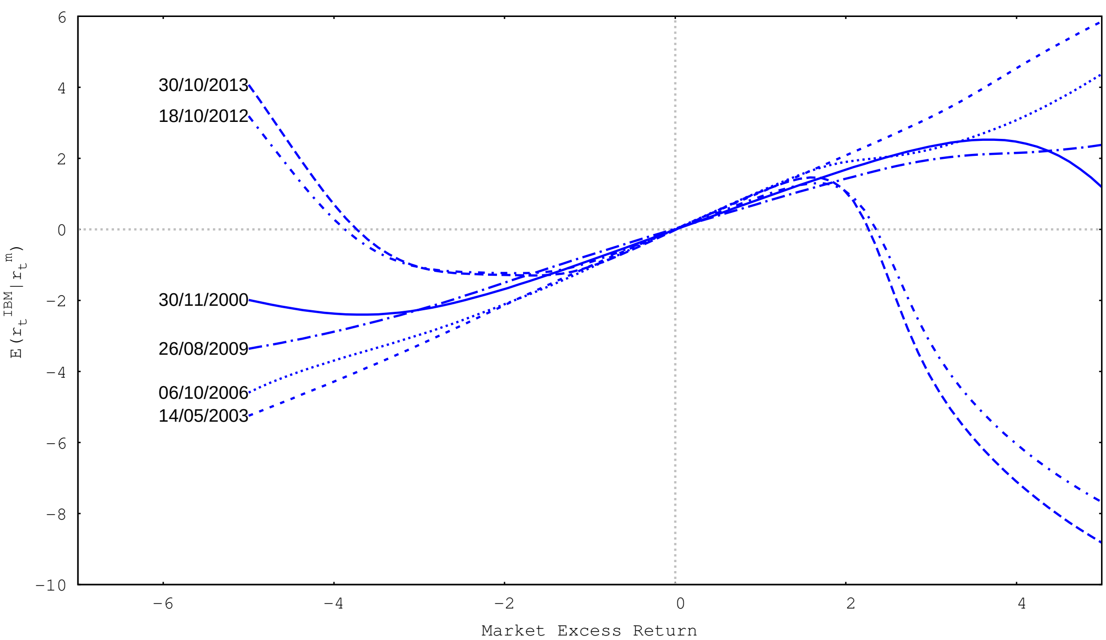

Figure 3.13: IBM: posterior mean of the conditional expected excess return of IBM given different values of the contemporaneous market excess return for different dates.

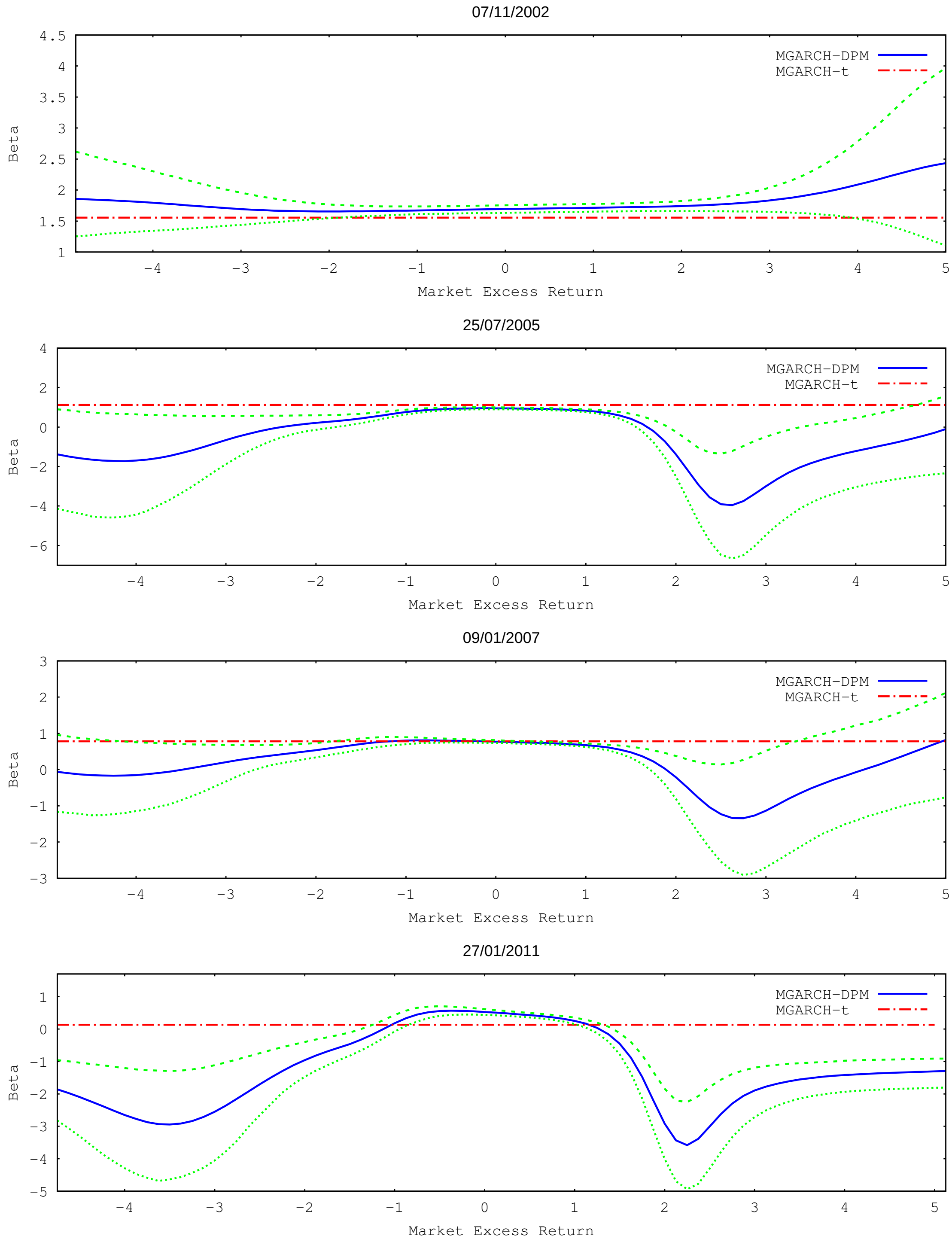

Figure 3.14: The posterior mean and 0.90 density intervals of IBM's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

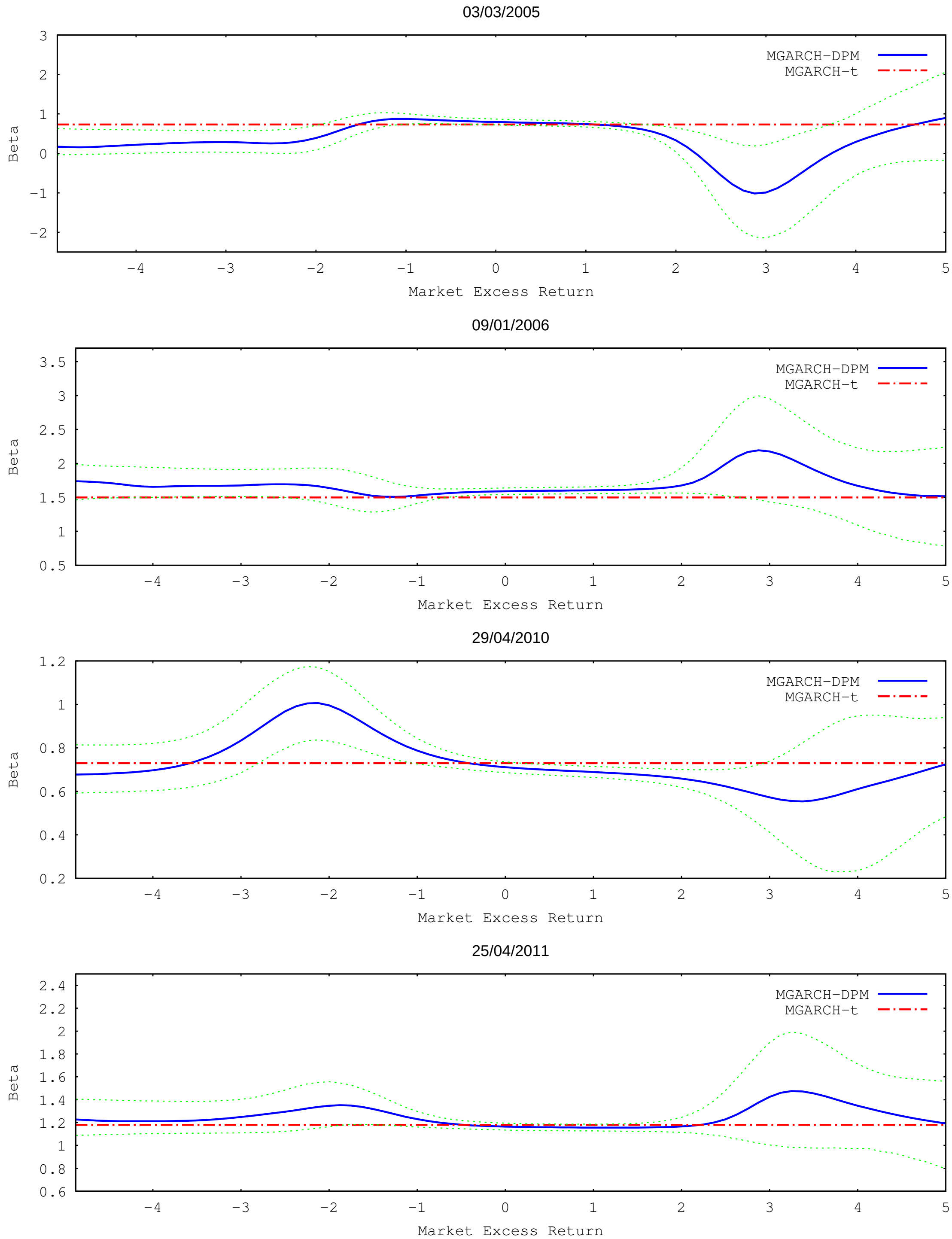

Figure 3.15: The posterior mean and 0.90 density intervals of XOM's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

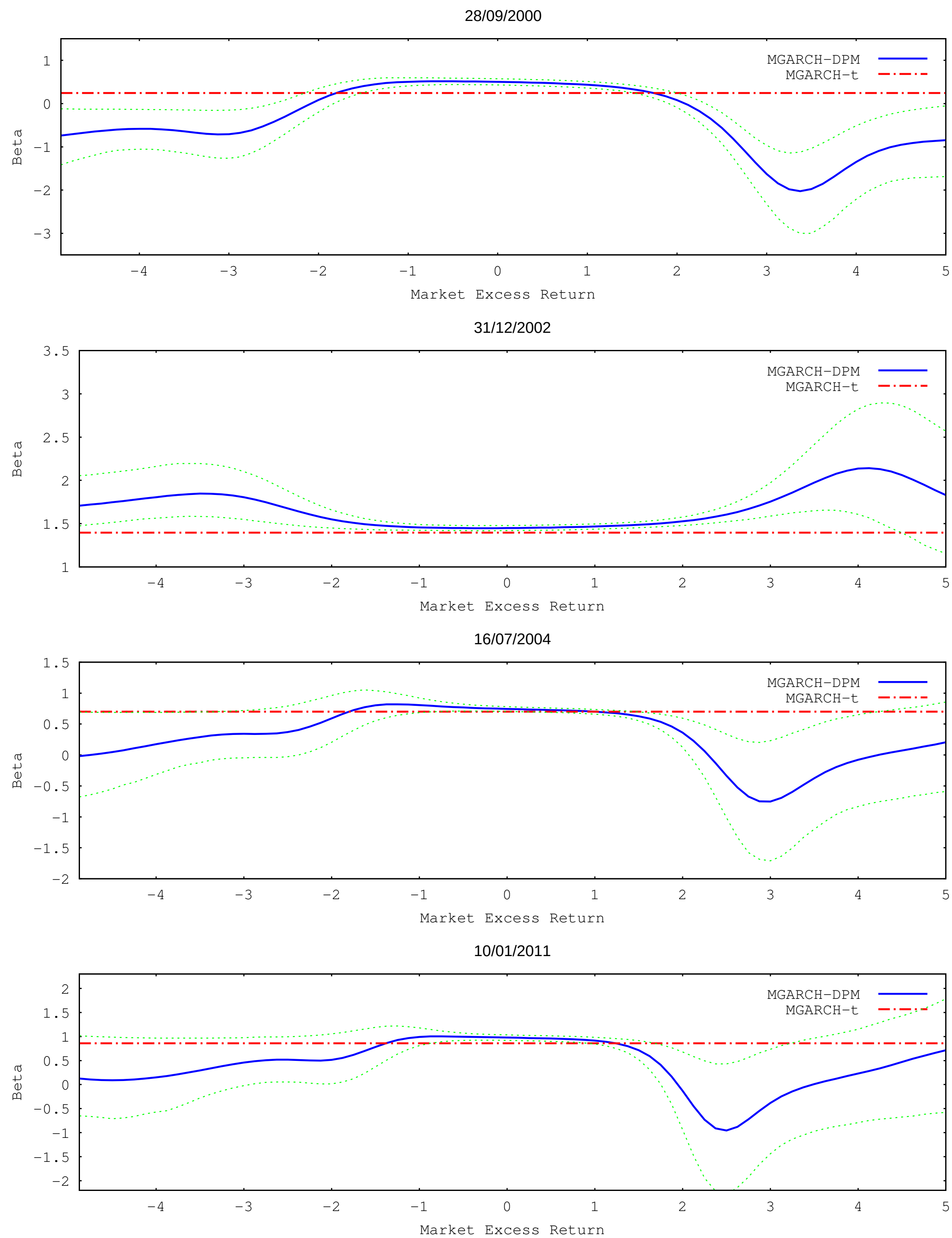

Figure 3.16: The posterior mean and 0.90 density intervals of GE's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

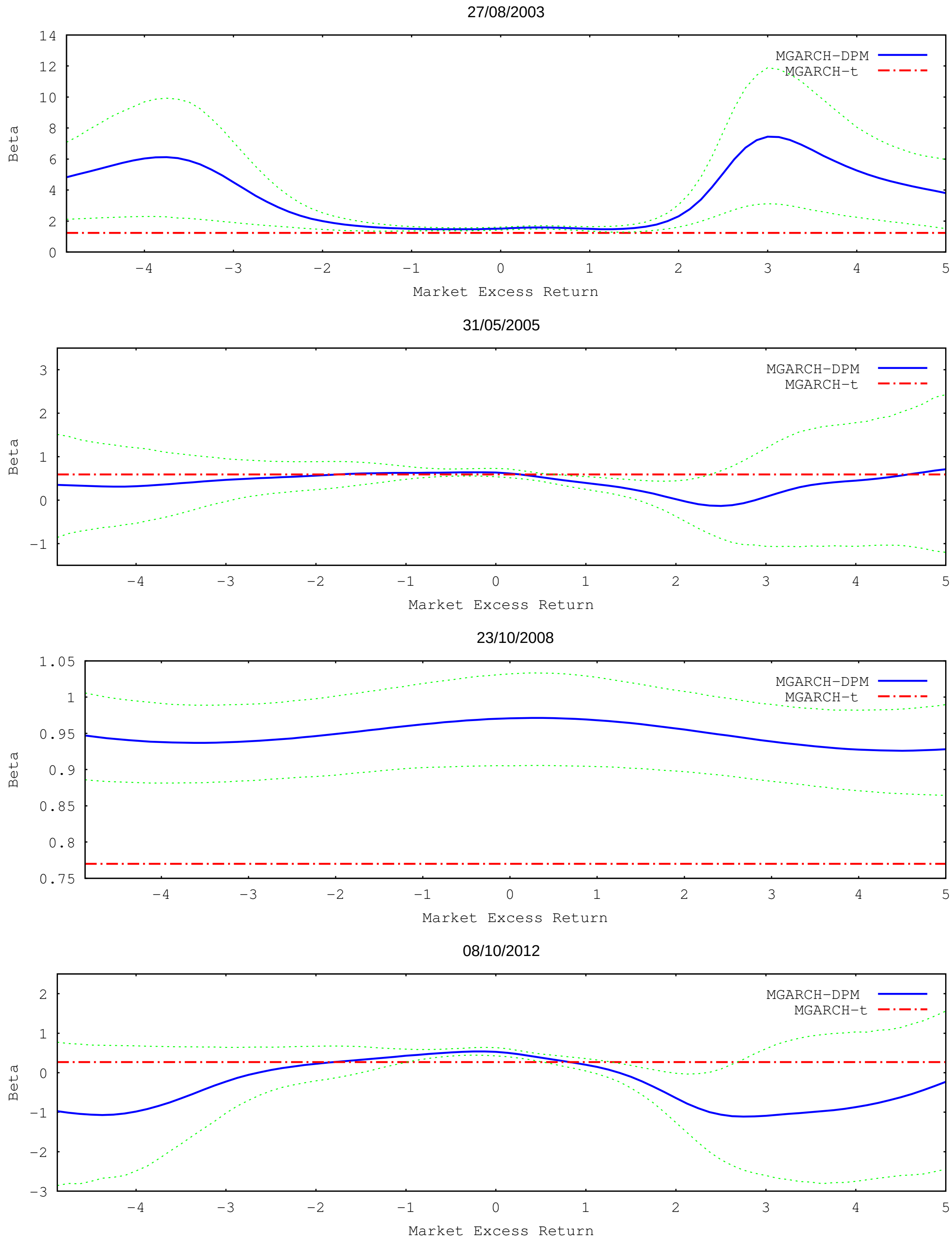

Figure 3.17: The posterior mean and 0.90 density intervals of AMGN's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

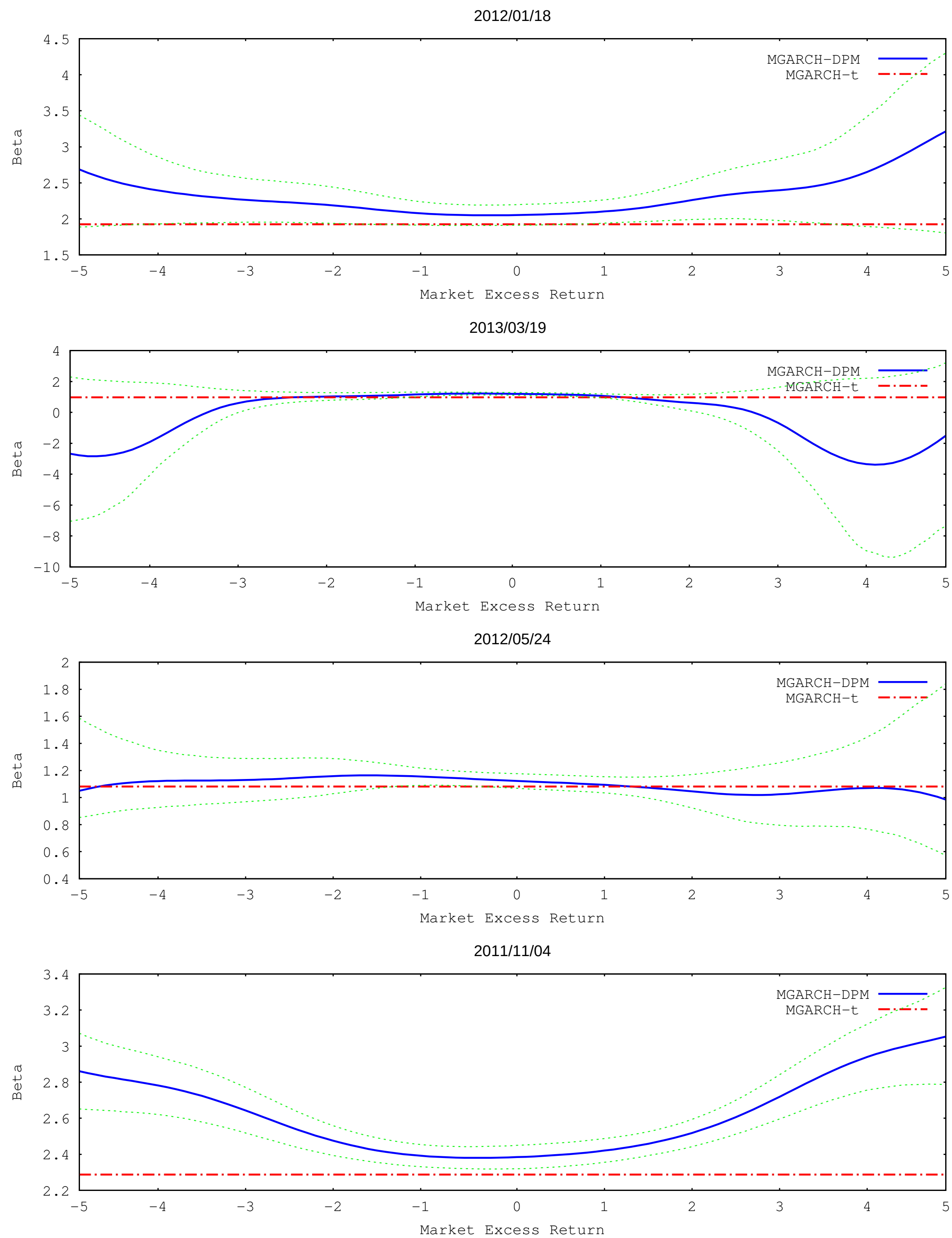

Figure 3.18: The posterior mean and 0.90 density intervals of BAC's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

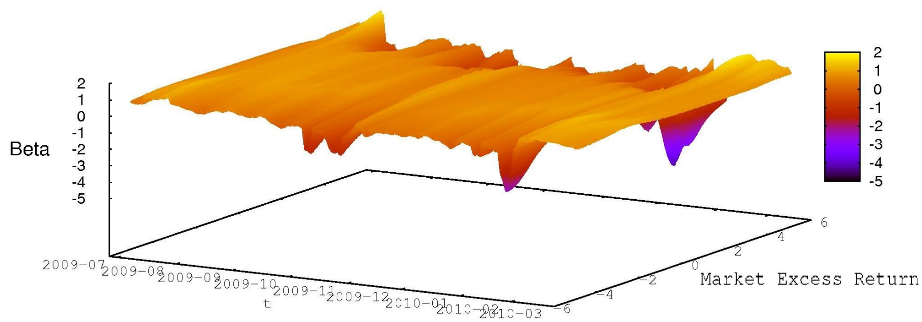

Figure 3.19: The posterior mean of IBM's nonparametric conditional beta as a function of excess market return and time from 2009-07 to 2010-03 estimated with MGARCH-DPM model.

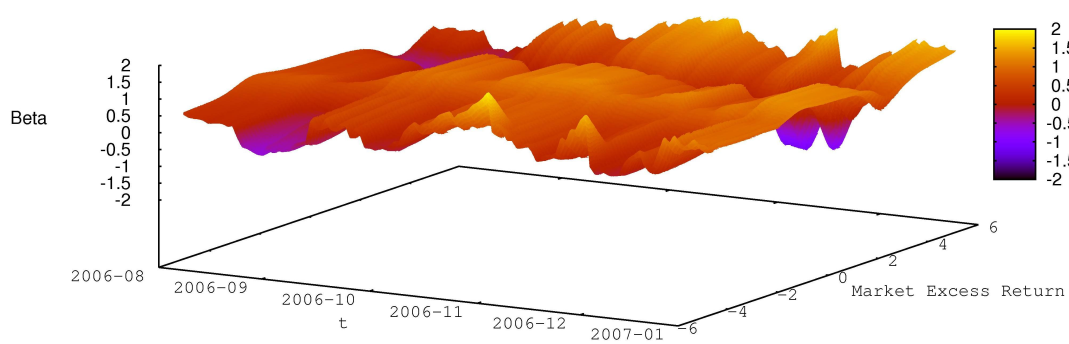

Figure 3.20: The posterior mean of XOM's nonparametric conditional beta as a function of excess market return and time from 2006-08 to 2007-01 estimated with MGARCH-DPM model.

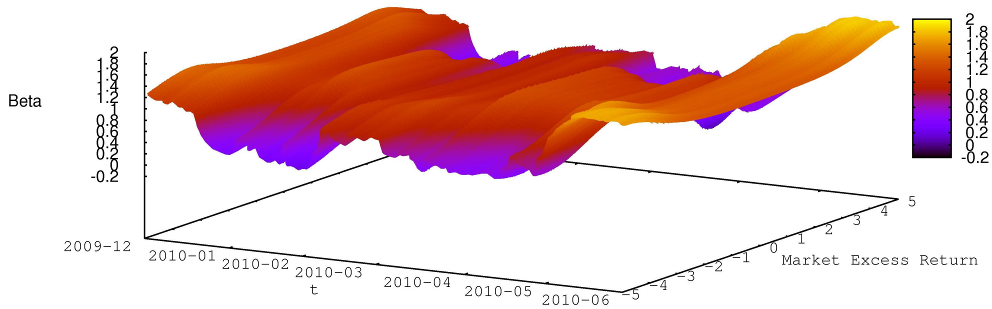

Figure 3.21: The posterior mean of GE's nonparametric conditional beta as a function of excess market return and time from 2009-12 to 2010-06 estimated with MGARCH-DPM model.

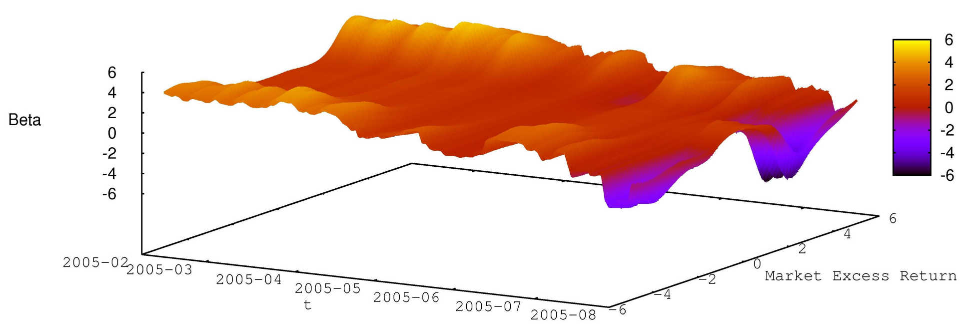

Figure 3.22: The posterior mean of AMGN's nonparametric conditional beta as a function of excess market return and time from 2005-02 to 2005-08 estimated with MGARCH-DPM model.

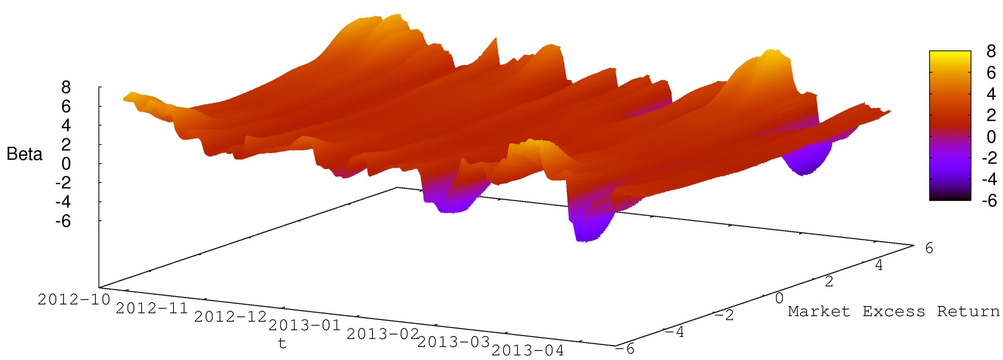

Figure 3.23: The posterior mean of BAC's nonparametric conditional beta as a function of excess market return and time from 2012-10 to 2013-04 estimated with MGARCH-DPM model.

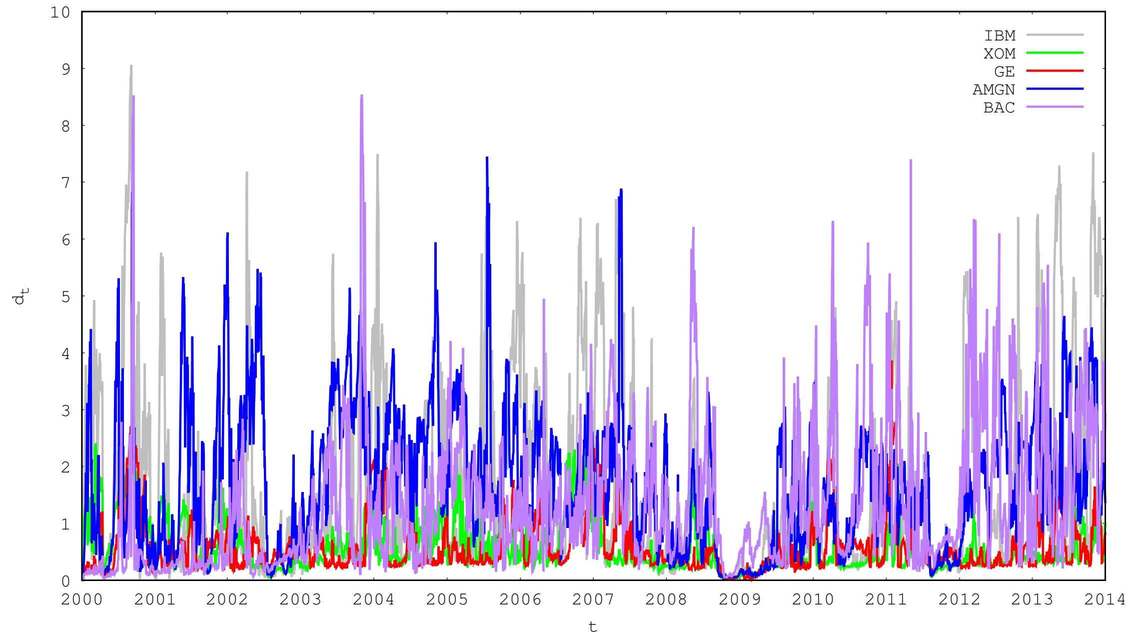

Figure 3.24: Variability of conditional beta with respect to the contemporaneous value of market excess returns over time for different stocks.

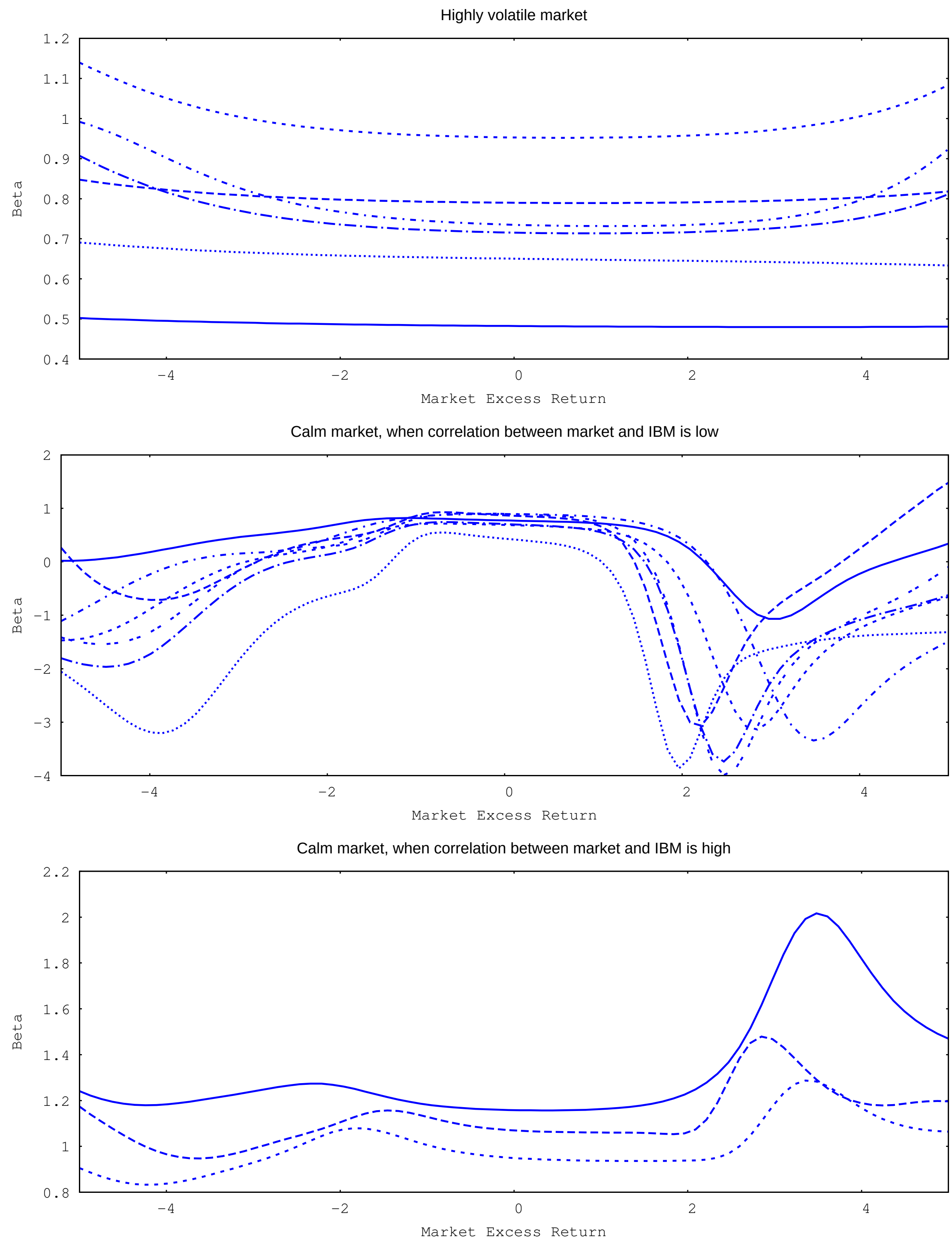

Figure 3.25: IBM: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

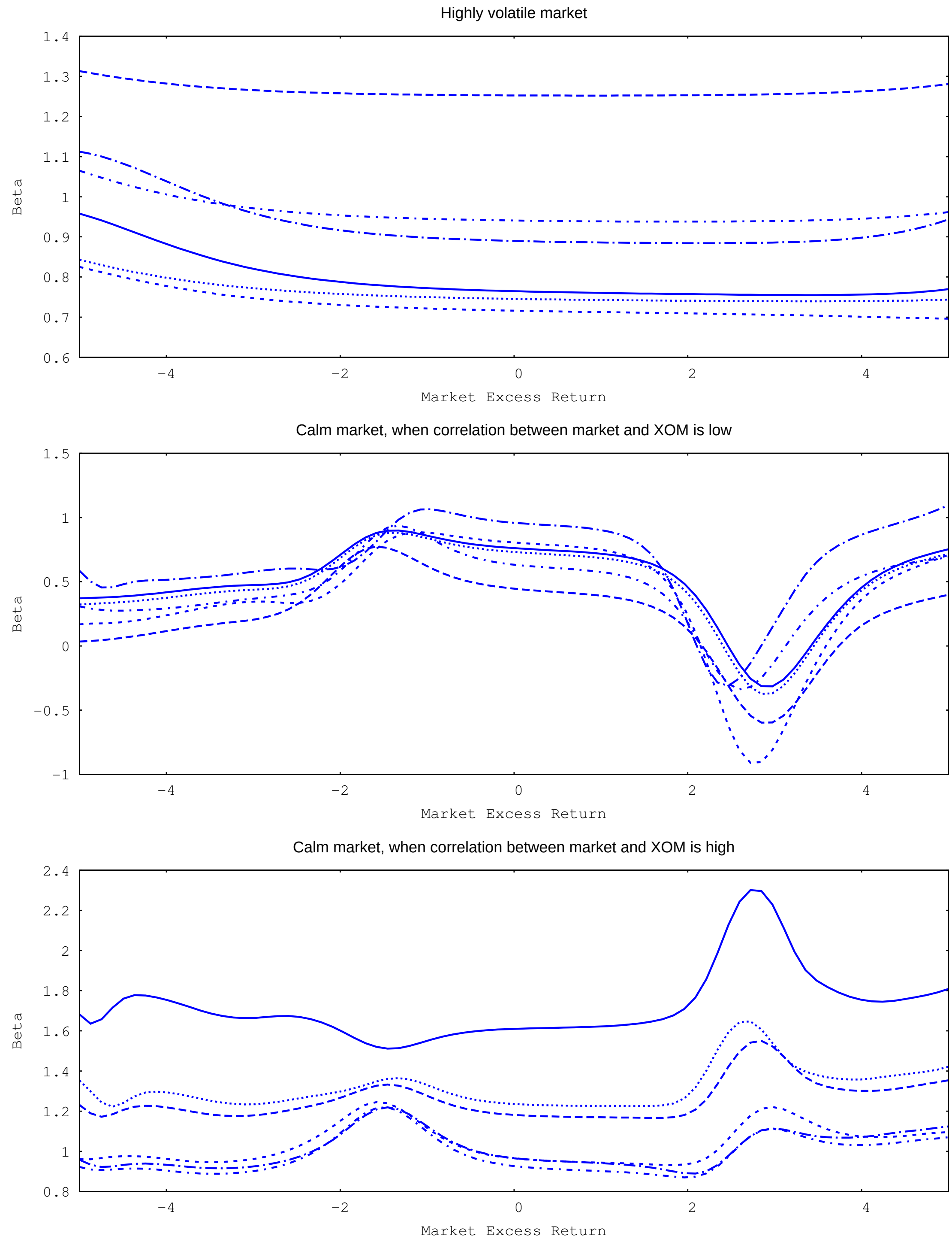

Figure 3.26: XOM: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

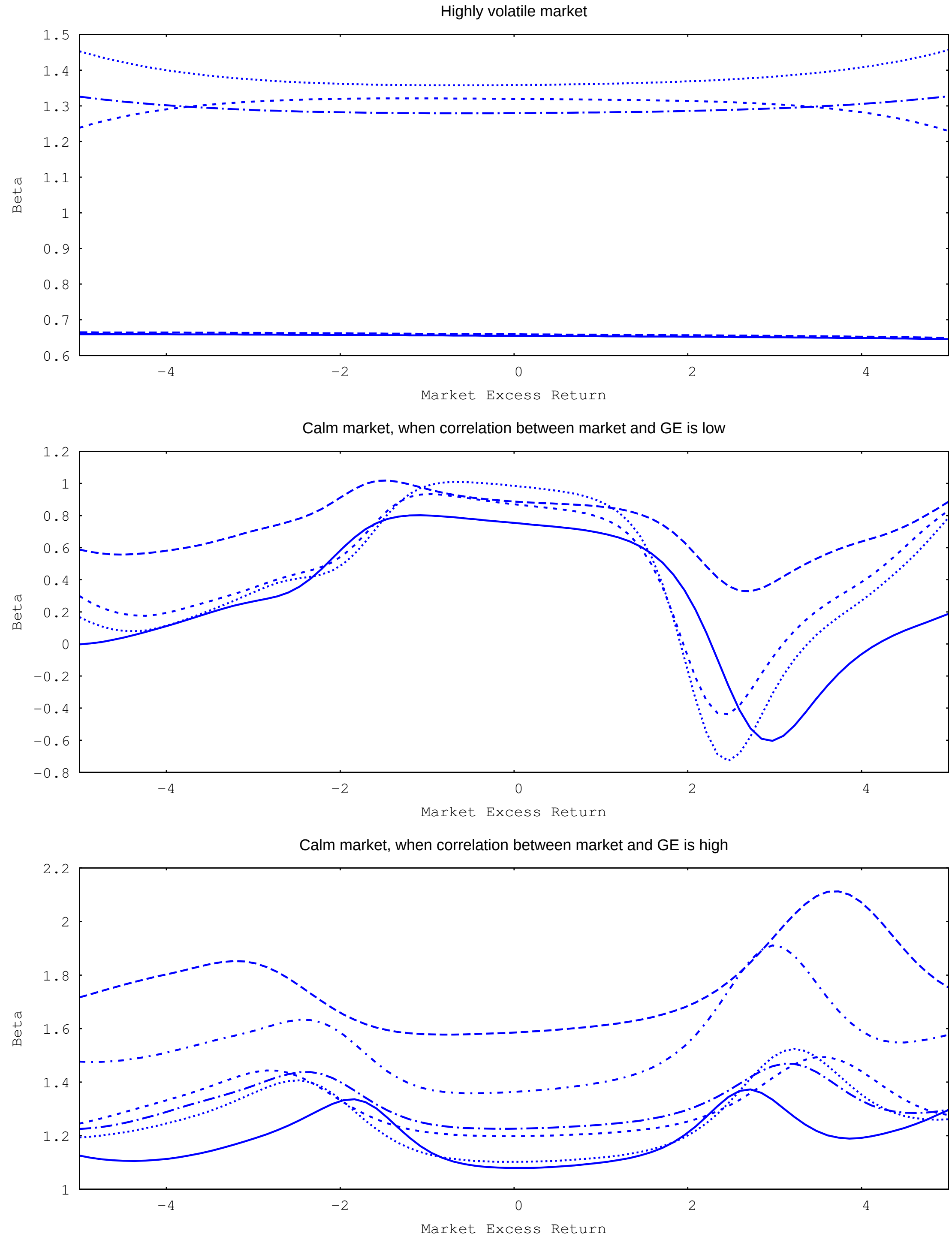

Figure 3.27: GE: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

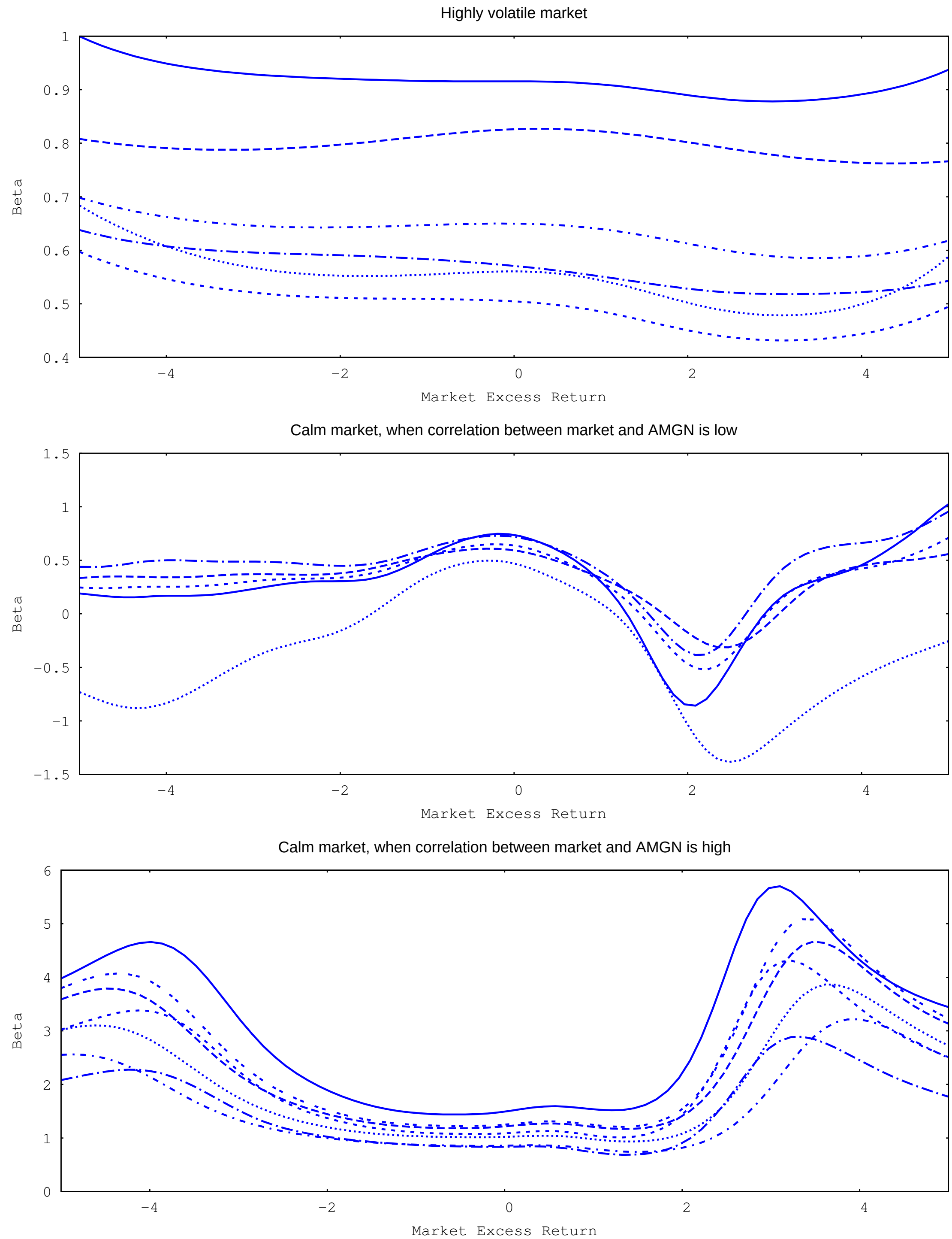

Figure 3.28: AMGN: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

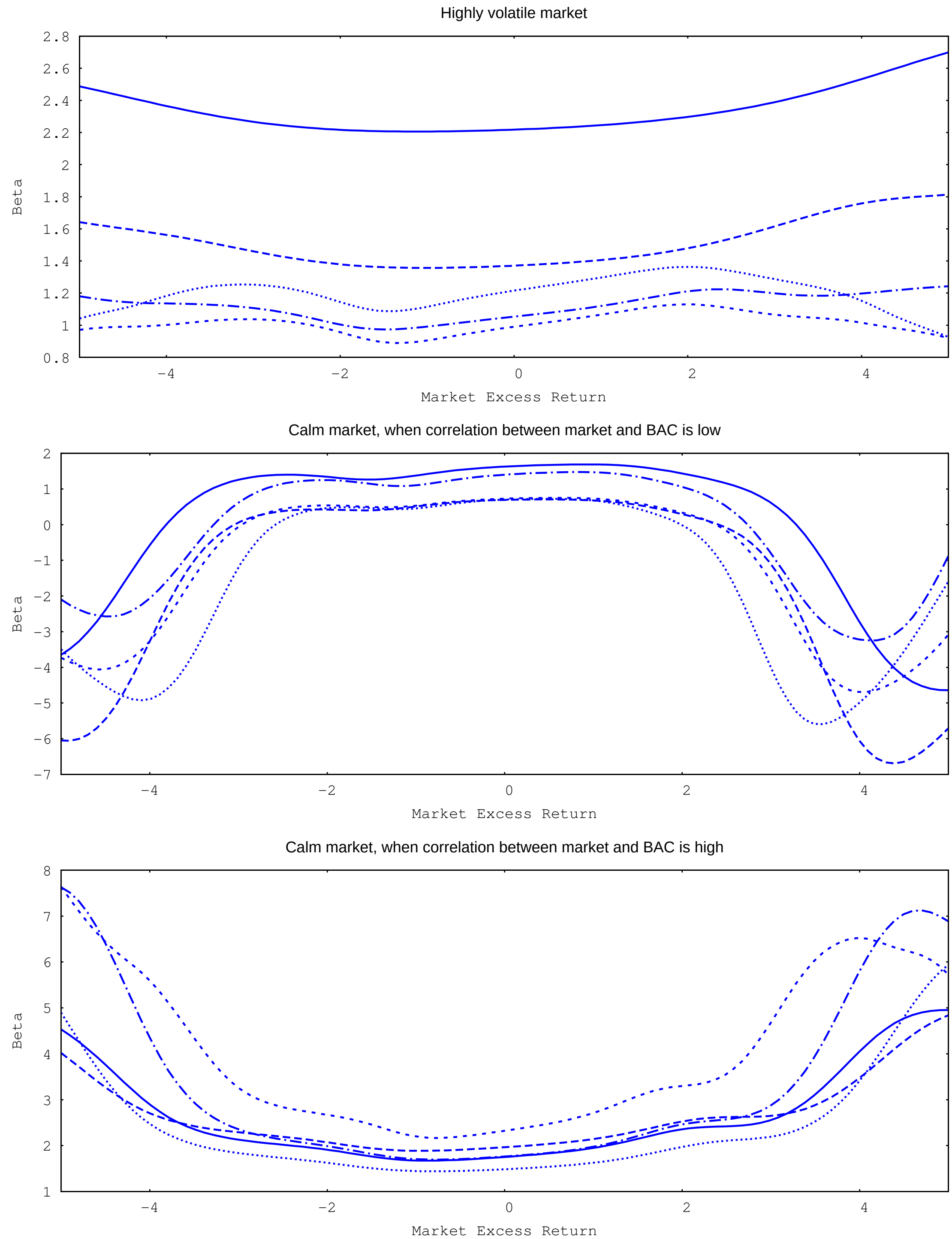

Figure 3.29: BAC: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

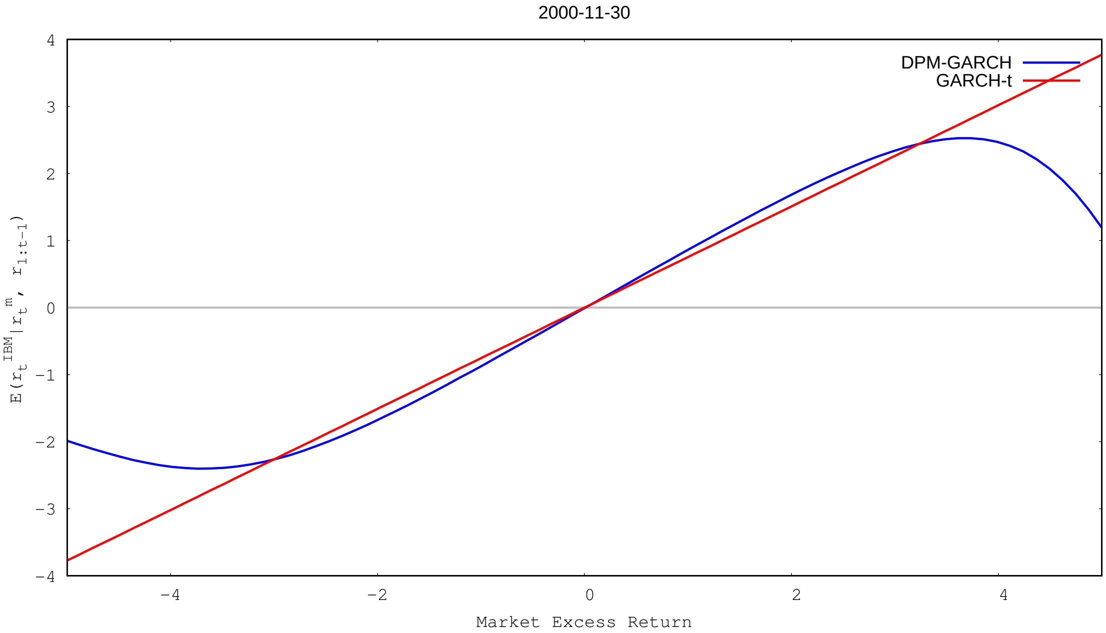

Figure 3.30: IBM: Predictive conditional expected return of IBM derived from MGARCH-DPM and MGARCH-t model.

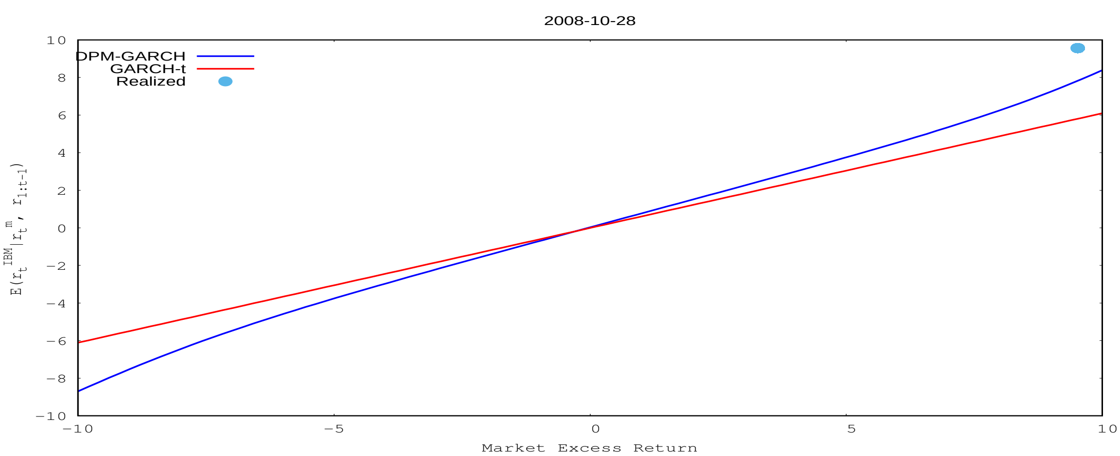

Figure 3.31: IBM: Predictive expected return from MGARCH-t and MGARCH-DPM models compared with the realized excess return of IBM when we expect a big shock to the market.

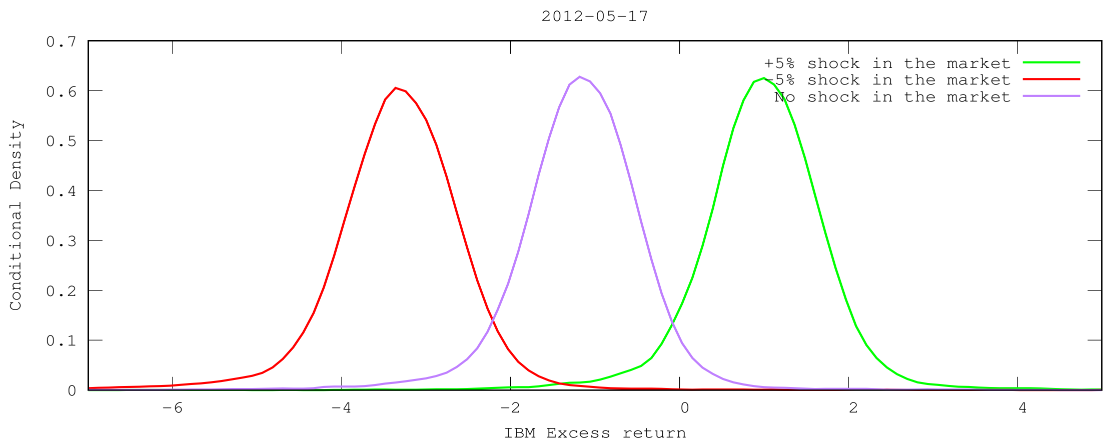

Figure 3.32: IBM: Effect of big shocks in the market return on IBM's predictive conditional density.

Chapter 4

Essay 2: How Does Risk Propagate? Contemporaneous and Lagged effects

4.1 Introduction

With the speed of information transmission between different markets these days, it is beneficial for investors, policymakers and regulators to be able to determine the contemporaneous shocks spillover from one market to another and how it affects various features (such as moments, tails, and value-at-risk) of the conditional density of the other markets. In this chapter, we study the information transmission among different markets. We develop a joint model of returns on two separate markets governed by an infinite mixture model from which the conditional density of one market given the returns of the other market can be derived. This makes it possible to study how a shock in one market influences the contemporaneous and future (one-day-ahead, one-week-ahead, and one-month ahead) conditional density of the other market in a general setting. The proposed model extends the literature on "spillover effects" or "contagion effects" that focus on transmission through conditional moments to spillover effects on the entire density.

More importantly, we provide a way to assess different risk scenarios, contemporaneously. In the efficient markets where information is transmitted almost simultaneously, it is beneficial for policymakers and regulators to be able to determine for example how the Canadian market will be contemporaneously affected if a big shock happens in the oil market. There is a vast literature studying the lagged effects on the volatility and some ad hoc procedures to estimate the contemporaneous spillover effects on the expected returns. To the best of our knowledge, this is the first work that provides us with insights into the contemporaneous spillover effects and market interactions in a general setting and enables us to carry out risk scenario analyses which are particularly useful in stress testing.

Understanding the cross-market relationships is important to global investors, regulators and policymakers, hedge fund managers, and other participants in the financial markets. There is a large body of literature on the topic of return spillover and volatility spillover among different markets. We can categorize the related articles into three classes: The lagged spillover effect of the returns of one market on the expected returns of the other market, the lagged spillover effect of the shocks to one market on the variance of the other market, and the contemporaneous spillover effect of the shocks to one market on the expected returns of the other market.

The first category of the studies, which is the most natural one, is done through a time series regression model or a vector autoregressive (VAR) model by using the lagged returns of one market as regressors on the right hand side of the regression equation of the returns of the other market.

The second group of studies estimates the lagged volatility spillover among markets. Many articles in the literature find significant volatility spillovers between developed and emerging markets (Wei et al., 1995), within individual equity markets (Hamao et al., 1990; Reyes, 2001), between commodities (Apergis and Rezitis, 2003), between derivatives (Abhyankar, 1995; Pan and Hsueh, 1998; Eom et al., 2002), and between financial markets in the same country (Darbar and Deb, 1997). Estimating the dynamic volatility spillover effects is mostly done by working in the multivariate GARCH-BEKK[1] framework which flexibly allows volatility interactions among markets. Caporale et al. (2006) study the transmission of South East Asia financial crisis into other developed stock markets by estimating a bivariate GARCH-BEKK model using the likelihood ratio tests. They find that during crises volatility spillover effects are unidirectional from the markets in turmoil to the others as opposed to the bidirectional spillovers in the normal periods. Gardebroek and Hernandez (2013) examine volatility transmission in oil, ethanol and corn prices, and show how a shock in one market affects other markets. Using GARCH-BEKK and DCC-GARCH models, Mensi et al. (2014) study how OPEC news announcements influence the spillover effects between oil and cereal markets.

The third strand in the literature studies the contemporaneous spillover effect of shocks in one market on the expected return of the other market. In the literature, this is usually done by an ad hoc multi-step procedure; first, a model is estimated for the returns of the first market, and then the residuals (or squared residuals) are used on the right-hand side of the regression equation for the returns of the other market. Applying this procedure, Ng (2000) studies the spillovers from a regional shock in Tokyo market and a global shock in the US market into the expected return of a particular PacificBasin market. He first assumes a bivariate asymmetric dynamic GARCH(1,1) model for the weekly returns on the Tokyo stock market and S&P 500 to estimate the shocks to the Tokyo market, e_t^{ja}, and to the US market, e_t^{us}. Then the estimated shocks are used as regressors in the return equation of each of the PacificBasin markets in a univariate GARCH framework.

$$r_t = \mu_0 + \beta r_{t-1} + \gamma_{t-1} r_{t-1}^{ja} + \delta_{t-1} r_{t-1}^{us} + \phi e_t^{us} + \psi e_t^{ja} + e_t, \quad e_t \sim N(0, \sigma_t^2)$$

In this model, the coefficients ϕ and ψ denote the contemporaneous spillover effects on the

[1]GARCH-BEKK representation proposed by Engle and Kroner (1995).

market under analysis from the world and the regional market, respectively. Note that there is a strong assumption here; S&P 500 returns and Tokyo stock returns are assumed to be external drivers for the local market returns. Another restrictive assumption here is the linearity; the shocks to the world and to the region affect the local market returns linearly.

Christiansen (2007) follows the same logic in three steps; she first estimates a model for the US returns as a proxy for the world market, then estimates a model for European market returns, using the estimated residuals of the first step on the right hand side of the regression equation. In the last step, she estimates a model for the local market returns, assuming that the estimated residuals from the first and second steps are exogenous regressors for the local market returns, and the corresponding coefficients are the spillover effects. Basing on the multi-step procedure, Balli et al. (2015) finds evidence of significant contemporaneous spillover effects from developed markets to emerging markets. Baele (2005) considers a Markov switching model in the second step and estimates the time-varying contemporaneous spillovers from the US and European markets into the local markets.

Apergis and Miller (2008) follow the same multi-step process and recover the oil-supply shocks, global aggregate-demand shocks, and global oil-demand shocks from the first step. Then they determine the contemporaneous effect of these structural shocks on the stock market returns in eight countries in a vector autoregressive framework.

The multi-step procedure for estimating the contemporaneous spillover effect assumes that the shocks to the first market are external drivers for the other market, which is not necessarily true. This approach also assumes that the contemporaneous shocks to the first market affect the expected return of the other market linearly. We propose a model for capturing the contemporaneous spillover effects in a general setting without imposing any assumption on the linearity of spillover effects or the externality of the shocks. In our approach, we model the return density of both markets jointly, allowing an unconstrained interaction. Moreover, by estimating the joint density and then deriving the conditional density, we are able to study the effects of the contemporaneous shocks in either of the markets on the entire shape of the conditional density of the other market. This enables us to explore the contemporaneous spillover effects of big positive or negative surprises in one market on different features of the other market's conditional probability density function such as the conditional expected return, volatility, kurtosis, and value-at-risk. The proposed model can be used to assess risk transmissions among any class of assets. For example, it can estimate how a stock market crash in the US market will affect the Canadian markets, contemporaneously. More generally, it can measure financial stress. For instance, conditional on a 3% drop in the oil market what is the probability that the US market will drop by more than 3%? Density intervals and other probability statements can be directly derived from this model.

In this chapter, we consider a semiparametric model for the multivariate joint density of the assets without imposing any restriction on the form of the joint density. Volatility clustering is taken into account by the flexible BEKK multivariate GARCH model. BEKK model discloses any significant lagged information transmission among asset volatilities. Moreover, we define a second component for the covariance matrix, a state-specific constituent for the part of the covariance matrix that is not captured by the GARCH structure. In this model, the mean vector of the assets is assumed to be state-specific as well. We estimate the joint density in a Bayesian framework. The full conditional density can be backed out to study how a shock in one market will affect the conditional density of the other market. Note that this is a general case that nests all the cases under the heading of the conditional moments' spillover effect.

We apply the proposed model to a dataset on S&P 500 and oil returns (difference in log-prices), and also a dataset on S&P TSX returns and oil returns, exploring the contemporaneous and lagged spillover effects of a surprise in oil market on Canada market and the US market. The results show that the contemporaneous transmission of shocks from oil market to both Canada market and the US market is more significant than the lagged spillovers. I provide empirical evidence that a big positive or negative shock in oil market can affect the shape of the entire conditional density of the Canada market returns and the US market returns, increasing the value-at-risk of an investment in these markets in the event of both positive and negative shocks.

The remainder of the chapter is structured as follows. Section 4.2 provides a general theoretical setting of the model used in this study, key features of the semiparametric MGARCH model, and the use of the Dirichlet process prior followed by details of the posterior sampling. The benchmark model is introduced in Section 4.3. Section 4.4 presents an outline of the data, followed by the empirical results. Section 4.5 concludes.

4.2 Semiparametric Model

In this section, we introduce in detail the model framework that we use to estimate the spillover effects. Let $\mathbf{r_t} = (r_t^1, ..., r_t^q)$ be the vector of returns at time t. In case we are interested in studying the spillover effects between two markets, q equals 2. We model the joint distribution of $\mathbf{r_t}$ nonparametrically by an infinite mixture of normal distributions which can approximate any continuous multivariate distribution (Ghosal, 1999). This semiparametric model, referred to as MGARCH-DPM, can be written as the following equation in a Dirichlet process mixture (DPM) model framework

$$\mathbf{r_t} = \mu_\mathbf{t} + \Gamma_t \mathbf{r_{t-1}} + \epsilon_\mathbf{t}, \quad \epsilon_\mathbf{t} \sim N(0, H_t^{\frac{1}{2}} B_t (H_t^{\frac{1}{2}})'), \; t = 1, ..., T \tag{4.1}$$

$$\theta_t \equiv \{\mu_\mathbf{t}, \Gamma_t, B_t\} | G \sim G, \tag{4.2}$$

$$G|\alpha, G_0 \sim DP(\alpha, G_0), \tag{4.3}$$

$$H_t = C_0' C_0 + A' \epsilon_\mathbf{t-1} \epsilon_\mathbf{t-1}' A + F' H_{t-1} F. \tag{4.4}$$

In this model, $\mu_{\mathbf{t}}$ and the residuals $\epsilon_{\mathbf{t}} = \mathbf{r_t} - \mu_{\mathbf{t}} - \Gamma_t \mathbf{r_{t-1}}$ are q-vectors, and B_t and H_t are $q \times q$ matrices. $H_t^{1/2}$ is the Cholesky decomposition of H_t. G is the mixing measure over a continuous kernel. The mixing is over the mean vector and the component B of the covariance matrix of a normal distribution. The second component, H_t, of the covariance matrix captures volatility clustering over time.

Equation (4.4) represents the MGARCH structure used in this chapter which is based on the BEKK multivariate GARCH model of Engle and Kroner (1995). The sufficient condition for positive definiteness of H_t is that at least one of the C_0 or F be of full rank (Engle and Kroner, 1995). C_0 is a q by q lower triangular matrix, and A and F are matrices of dimmension $q \times q$. A is a matrix of coefficients a_{ij} that capture the lagged shock interactions among markets, and F is a matrix of coefficients f_{ij} that capture the lagged volatility spillover effects between markets i and j. MGARCH-BEKK structure in Equation (4.4) has been used in the literature mostly to derive the lagged spillover effects of variances among markets. The expanded formulas for the covariance with this structure for $q = 2$ are given by

$$\begin{aligned} h_{11,t} = c_{11}^2 + a_{11}^2 \epsilon_{1,t-1}^2 + 2a_{11}a_{21}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{21}^2\epsilon_{2,t-1}^2 \\ + f_{11}^2 h_{11,t-1} + 2f_{11}f_{21}h_{12,t-1} + f_{21}^2 h_{22,t-1}, \end{aligned} \tag{4.5}$$

$$\begin{aligned} h_{12,t} = c_{11}c_{12} + a_{11}a_{12}\epsilon_{1,t-1}^2 + (a_{11}a_{22} + a_{21}a_{12})\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{22}a_{21}\epsilon_{2,t-1}^2 \\ + f_{11}f_{12}h_{11,t-1} + (f_{11}f_{22} + f_{21}f_{12})h_{12,t-1} + f_{21}f_{22}h_{22,t-1}, \end{aligned} \tag{4.6}$$

$$\begin{aligned} h_{22,t} = c_{12}^2 + c_{22}^2 + a_{12}^2\epsilon_{1,t-1}^2 + 2a_{12}a_{22}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{22}^2\epsilon_{2,t-1}^2 \\ + f_{12}^2 h_{11,t-1} + 2f_{12}f_{22}h_{12,t-1} + f_{22}^2 h_{22,t-1}. \end{aligned} \tag{4.7}$$

In Equation (4.3), G is the unknown distribution that governs the mixing over μ_t, the

VAR coefficients, Γ_t, and the first part of the covariance matrix of the normal kernel in the mixture model, B_t. B_t is a symmetric positive definite matrix that accounts for the part of the covariance matrix that is not captured by GARCH dynamics. A draw from a DP, $G \sim DP(\alpha, G_0)$, is almost surely a discrete distribution and is governed by two parameters. The concentration parameter α, a positive scalar and a base distribution G_0. The nonparametric distribution G is centered on the base distribution G_0, which can be considered as the prior guess; $E(G(\Delta)) = G_0(\Delta)$ for the measurable set Δ. The concentration parameter measures the strength of belief in G_0. The larger α, the stronger belief in G_0 and the more distinct elements we have with non-negligible mass.

Sethuraman (1994) characterizes a stick-breaking representation of the DP. Combining this with the normal kernel gives the associated stick breaking representation of the MGARCH-DPM density as

$$p(r_t|\mu, \Gamma, B, W, H_t) = \sum_{j=1}^{\infty} \omega_j N(r_t|\mu_j + \Gamma_j r_{t-1}, H_t^{1/2} B_j (H_t^{1/2})'), \tag{4.8}$$

$$\omega_1 = v_1, \ \omega_j = v_j \prod_{l=1}^{j-1} (1 - v_l), \ j > 1, \tag{4.9}$$

$$v_j \overset{iid}{\sim} \text{Beta}(1, \alpha) \tag{4.10}$$

$$(\mu_j, \Gamma_j) \overset{iid}{\sim} N(\beta_0, D), \quad B_j \overset{iid}{\sim} \mathcal{W}^{-1}(B_0, \nu_0), \tag{4.11}$$

where $N(r_t|\mu_j + \Gamma_j r_{t-1}, H_t^{1/2} B_j (H_t^{1/2})')$ denotes the multivariate normal density evaluated at r_t. The model nests several special cases. First, the Gaussian model is obtained when $\alpha \to 0$ and $B_1 = I$. The Student-t model results from μ and Γ being constant for all mixture components and $\alpha \to \infty$, since $G \to G_0$, the inverse Wishart distribution.

To estimate the unknown parameters in Equations (4.8)-(4.11), we apply an MCMC sampler along with the slice sampler of Walker (2007). Slice sampling introduces a latent variable, $u_t \in (0, 1)$, to elegantly convert an infinite sum to a finite mixture model which makes the sampling feasible. Estimating the joint posterior density of u_t and other model parameters and then integrating out the slice variable u_t recovers the desired posterior density. In practice, this means jointly sampling all parameters including the slice variable but then discarding u_t. Define u_t such that the joint density of (r_t, u_t) given $(W = \{\omega_1, \omega_2, ...\}, \Theta \equiv (\mu, \Gamma, B))$ is given by

$$f(r_t, u_t|W, \Theta) = \sum_{j=1}^{\infty} \mathbf{1}(u_t < \omega_j) N(r_t|\mu_j + \Gamma_j r_{t-1}, (H_t^{1/2})' B_j H_t^{1/2}). \tag{4.12}$$

Let $s_{1:T} = \{s_1, ..., \ s_T\}$ be the configuration set that partitions the data $r_{1:T}$ into c distinct clusters such that observation r_t is assigned parameter $\theta_{s_t} = (\mu_{s_t}, \Gamma_{s_t}, B_{s_t})$. Let $n_j = \{\#t|s_t = j\}$ be the number of observations allocated to state j. The full likelihood is

$$p(r_{1:T}, u_{1:T}, s_{1:T}|W, \theta) = \Pi_{t=1}^{T} \mathbf{1}(u_t < \omega_{s_t}) N(r_t|\mu_{s_t} + +\Gamma_{s_t} r_{t-1}, (H_t^{1/2}) B_{s_t} (H_t^{1/2})'), \tag{4.13}$$

and the joint posterior is proportional to

$$p(W_{1:K}) \Pi_{j=1}^{K} p(\mu_j, B_j) \Pi_{t=1}^{T} \mathbf{1}(u_t < \omega_{s_t}) N(r_t|\mu_{s_t} + \Gamma_{s_t} r_{t-1}, (H_t^{1/2}) B_{s_t} (H_t^{1/2})') \tag{4.14}$$

where K is the smallest natural number that satisfies the condition $\sum_{j=1}^{K} \omega_j > 1 - \min\{u_t\}_{t=1}^{T}$ and $W_{1:K}$ denotes the finite set of W and similarly for other parameters $\mu_{1:K}$, $\Gamma_{1:K}$ and $B_{1:K}$.

Note that in the proposed model we have path dependency in the GARCH recursions (Equation 4.4). At each iteration g of the MCMC, we compute $K^{(g)}$ different GARCH processes to take into consideration the path dependency through $\epsilon_{\mathbf{t-1}}$ (Haas et al., 2004). At each iteration of the MCMC, we have $K^{(g)}$ distinct clusters which is analogous to having $K^{(g)}$ different states in a Markov switching framework. Therefore, we need to keep track of

$K^{(g)}$ different GARCH processes, $H_t^j,\ j=1,..K^{(g)},\ t=1,...,T$:

$$H_t^j = C_0'C_0 + A'\epsilon_{t-1}^j{\epsilon^j}'_{t-1}A + F'H_{t-1}^jF.$$

Having defined the notation, the steps of the MCMC algorithm are discussed next.

4.2.1 Sampling Steps

Step 1 The posterior distribution of $\theta_j = (\mu_j, \Gamma_j, B_j),\ j=1,...,K$: Equation (4.28) can be rewritten as

$$\mathbf{r_t} = X_t\beta + \epsilon_\mathbf{t}, \quad \epsilon_\mathbf{t} \sim N(0, H_t^{\frac{1}{2}}B_t(H_t^{\frac{1}{2}})'),\ t=1,...,T \tag{4.15}$$

where X_t is a vector of 1 (for the intercept) and the lagged returns, and $\beta = Vec(\mu,\Gamma)$. Using the transformation $\mathbf{z}_t = H_t^{-1/2}\mathbf{r}_t$, $\bar{X}_t = H_t^{-1/2}X_t$, and $\mathbf{e}_t = H_t^{-1/2}\epsilon_t$, we have

$$\mathbf{z_t} = \bar{X}_t\beta + \mathbf{e_t}, \quad \mathbf{e_t} \sim N(0, B_t),\ t=1,...,T \tag{4.16}$$

And we can apply multivariate regression results for conditionally conjugate priors (Karlsson, 2013) to obtain the posterior distributions of the parameters:

$$B_j|r_{1:T}, S, \mu_j, \Gamma_j, H_{1:T} \sim iW\left(n_j + \nu_0, B_0 + \Sigma_{t:s_t=j}(z_t - \bar{X}_t\beta_j)(z_t - \bar{X}_t\beta_j)'\right) \tag{4.17}$$

$$\beta_j|r_{1:T}, s_{1:T}, B_j, H_{1:T} \sim N(\bar{\beta}, \bar{D}) \tag{4.18}$$

in which

$$\bar{D} = (\Sigma_{s_t=j}\bar{X}_t'B_j^{-1}\bar{X}_t + D^{-1})^{-1} \tag{4.19}$$

$$\bar{\beta} = \bar{D}(\Sigma_{s_t=j}\bar{X}_t'B_j^{-1}z_t + D^{-1}\beta_0) \tag{4.20}$$

Step 2 Updating $v_j,\ j = 1, ..., K$ (Ishwaran and James, 2001):

$$v_j|S \sim Beta(1 + \Sigma_{t=1}^{T} 1(s_t = j), \alpha + \Sigma_{t=1}^{T} 1(s_t > j)). \tag{4.21}$$

Then we update W based on recently generated v_js: $\omega_1 = v_1,\ \omega_j = v_j \prod_{i=1}^{j-1}(1 - v_i),\ j > 1$.

Step 3 Updating $u_t,\ t = 1, ..., T$ (Walker, 2007):

$$u_t|S \sim Uni(0, \omega_{s_t})$$

Then we update K such that $\Sigma_{j=1}^{K}\omega_j > 1 - min\{u_t\}_{t=1}^{T}$. Additional ω_j and θ_j will need to be generated from the priors if K is incremented.

Step 4 Updating $s_{1:T}$ (Walker, 2007): For each $t = 1, ..., T$,

$$p(s_t = j|r_{1:T}) \propto 1(\omega_j > u_t)N(r_t|\mu_j, H_t^{1/2} B_j (H_t^{1/2})'), j = 1, ..., K \tag{4.22}$$

Step 5 Updating α: Assuming $Gamma(a_0, b_0)$ as the prior distribution of α, we can update α following the two steps below (See Escobar and West (1995))

(a) Given the most recent values of α and d, sample η from $(\eta|\alpha, d) \sim Beta(\alpha+1, T)$.

(b) Given d and just generated η, sample the new α from

$$\alpha|\eta, d \sim \Pi_\eta \Gamma(a_0 + d, b_0 - log(\eta)) + (1 - \Pi_\eta)\Gamma(a_0 + d - 1, b_0 - log(\eta)), \tag{4.23}$$

where d is the number of alive clusters to which at least one observation has been assigned, and Π_η is defined by $\frac{\Pi_\eta}{1-\Pi_\eta} = \frac{a_0+k-1}{n(b_0-log(\eta))}$.

Step 6 Updating GARCH parameters $\Psi = \{c_{0,ij}, a_{ij}, b_{ij}\}_{i,j\leq q}$:

$$p(\Psi|\beta, B, s_{1:T}, r_{1:T}, H_{1:T}) \propto p(\Psi) \times \prod_{t=1}^{T} N(r_t|\beta_{s_t}, H_t^{1/2} B_{s_t} (H_t^{1/2})'), \tag{4.24}$$

which is not of standard form, and we apply a Metropolis Hastings sampler. Given the current value Ψ of the chain, the proposal Ψ' is computed by

$$\Psi' = \Psi + N(0, \widehat{V}). \tag{4.25}$$

Ψ' is accepted with probability $\min\{p(\Psi'|\beta, B, S, r_{1:T}, H_{1:T})/p(\Psi|\beta, B, S, r_{1:T}, H_{1:T}), 1\}$, otherwise, Ψ' is rejected and Ψ is selected as the draw. In Equation (4.25), $\widehat{V}$ is the inverse Hessian matrix of $\ell = \log[p(\Psi|\beta, B, S, r_{1:T}, H_{1:T})]$ at its posterior mode, $\widehat{\Psi}$.[2] We can obtain $\widehat{V}$ in Equation (4.25) by numerically optimizing ℓ.[3]

4.2.2 Conditional Density

Applying the posterior sampling algorithm, we sample model parameters for many iterations and after dropping a set of burn-in draws we have the following set of sampled parameters:

$$\{(\mu_j^{(g)}, \Gamma_j^{(g)}, B_j^{(g)}), v_j^{(g)}, j = 1, ..., K^{(g)}\}, \{s_t^{(g)}, u_t^{(g)}, t = 1, ..., T\}, H_{1:T}^{(g)} = \{H_1^{(g)}, ..., H_T^{(g)}\}, \tag{4.26}$$

[2] To achieve an acceptance rate of parameters between 0.2 and 0.5, we can scale $\widehat{V}$.
[3] We apply Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to find the posterior mode of ℓ.

for $g = 1, ..., M$ where M is the number of MCMC iterations. At each iteration $g = 1, ..., M$ of the algorithm, a draw of $G|r_{1:T}$, can be written as

$$G^{(g)} = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)} \delta_{\theta_j^{(g)}} + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) G_0(\theta), \tag{4.27}$$

where $\theta_j^{(g)} = (\mu_j^{(g)}, \Gamma_j^{(g)}, B_j^{(g)})$ and $\delta_{\theta_j^{(g)}}$ is a mass point at $\theta_j^{(g)}$.

We can estimate the joint density of $\mathbf{r_t}$. For the sake of simplicity, we consider a special case of our model, $\mathbf{r_t} = (r_t^o, r_t^{us})$ where r_t^o and r_t^{us} represent the oil market return and the US market return, respectively. Combining Equation (4.27) with the kernel density $f(.)$ gives the predictive joint density of oil returns and the US market returns conditional on $G^{(g)}$ as

$$p(r_t^o, r_t^{us}|r_{1:T}, G^{(g)}) = \Sigma_{j=1}^{K^{(g)}} \omega_j^{(g)} f(r_t^o, r_t^{us}|\theta_j^{(g)}) + \left(1 - \Sigma_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f(r_t^o, r_t^{us}|\theta) G_0(\theta) d\theta \tag{4.28}$$

The kernel density here is a normal density. Applying the properties of the conditional and marginal distributions of the jointly normal random variables (Bauwens et al., 2000), we are able to simplify the conditional distribution of r^{us} given r^o. To minimize notational clutter, we drop the iteration superscript, g, but note that at each iteration we use the most recently updated parameters.

$$\begin{aligned} p_{r^{us}|r^o}(r_t^{us}|r_t^o, r_{1:T}, G) = {} & \Sigma_{j=1}^K q_j(r_t^o) N(r_t^{us}|\mu_{j,us} + \Gamma_{j,us}\mathbf{r_{t-1}} + \beta_{jt}(r_t^o - \mu_{j,o} - \Gamma_{j,o}\mathbf{r_{t-1}}), c_{jt}) \\ & + \left(1 - \Sigma_{j=1}^K q_j(r_t^o)\right) \frac{A_1}{A_2} \end{aligned} \tag{4.29}$$

where

$$\beta_{jt} = \frac{(H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{us,o}}{(H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{o,o}} \tag{4.30}$$

$$c_{jt} = (H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{us,us} - (H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{us,o} (H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{o,o}^{-1} (H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{us,o} \tag{4.31}$$

$$q_j(r_t^o) = \frac{\omega_j N(r_t^o|\mu_{j,o} + \Gamma_{j,o}\mathbf{r_{t-1}}, (H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{o,o})}{\Sigma_{j=1}^K \omega_j N(r_t^o|\mu_{j,o} + \Gamma_{j,o}\mathbf{r_{t-1}}, (H_t^{\frac{1}{2}} B_j H_t^{\frac{1}{2}'})_{o,o}) + \left(1 - \Sigma_{j=1}^K \omega_j\right) A_2} \tag{4.32}$$

And

$$A_1 = \int N(r_t^o, r_t^{us}|\mu + \Gamma\mathbf{r_{t-1}}, H_t^{\frac{1}{2}} B H_t^{\frac{1}{2}'}) N(\beta|\beta_{\mathbf{0}}, D) \times iW(B|B_0, \nu_0) d\beta dB, \tag{4.33}$$

$$A_2 = \int N(r_t^o|\mu_o + \Gamma_o\mathbf{r_{t-1}}, (H_t^{\frac{1}{2}} B H_t^{\frac{1}{2}'})_{o,o}) N(\beta|\beta_{\mathbf{0}}, D) \times iW(B|B_0, \nu_0) d\beta dB. \tag{4.34}$$

where the subscript (o,o) in $()_{o,o}$ denotes element (1,1) of the matrix, and other subscripts are defined similarly. β includes elements of vector μ and matrix Γ, and A_1 and A_2 can be easily approximated by Monte Carlo simulation. Therefore, the formula for the conditional distribution will be written as

$$\begin{aligned} p(r_t^{us}|r_t^o, r_{1:T}, G) &= \Sigma_{j=1}^K q_j(r_t^o) N(r_t^{us}|\mu_{j,us} + \Gamma_{j,us}\mathbf{r_{t-1}} + \beta_{jt}(r_t^o - \mu_{j,o} - \Gamma_{j,o}\mathbf{r_{t-1}}), c_{jt}) \\ &+ \left(1 - \Sigma_{j=1}^K q_j(r_t^o)\right) \frac{\Sigma_n N(r_t^o, r_t^{us}|\mu_{\mathbf{n}} + \Gamma_n\mathbf{r_{t-1}}, H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})}{\Sigma_n N(r_t^o|\mu_{n,o} + \Gamma_{n,o}\mathbf{r_{t-1}}, (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{o,o})} \end{aligned} \tag{4.35}$$

where

$$\beta_{nt} = \frac{(H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{us,o}}{(H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{o,o}} \tag{4.36}$$

and

$$c_{nt} = (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{us,us} - (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{us,o} (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{o,o}^{-1} (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{us,o}. \tag{4.37}$$

μ_n, Γ_n, and $B_n,\ n = 1, ..., N$ are i.i.d draws from the priors. Equation (4.35) can be rewritten as the following equation so the contemporaneous shock spillover from oil returns into the

US return density is more obvious.

$$p(r_t^{us}|r_t^o, r_{1:T}) = \Sigma_{j=1}^K q_j(\epsilon_t^o) N(r_t^{us}|\mu_{j,us} + \Gamma_{j,us}\mathbf{r_{t-1}} + \beta_{jt}(\overbrace{r_t^o - \mu_{j,o} - \Gamma_{j,o}\mathbf{r_{t-1}}}^{\epsilon_t^o}), c_{jt})$$
$$+ \left(1 - \Sigma_{j=1}^K q_j(\epsilon_t^o)\right) \mathcal{A}(\epsilon_t^o) \tag{4.38}$$

where ϵ_t^o denotes the shock in the oil returns (if state j occurs), and $\mathcal{A}(\epsilon_t^o)$ is a nonlinear function of ϵ_t^o. Unlike the literature, we pursue the joint estimation first and then derive the density of the US market conditional on the oil market return. By estimating the joint density, we do not lean on the externality assumption; that the shocks to the oil market are external drivers for the US market. Moreover, the contemporaneous effect of a shock in oil returns on the expected return of the US market is potentially nonlinear while the ad hoc approaches in the literature for studying the contemporaneous spillover effect assume linearity.

4.2.3 Transmission of shocks from time t to $t + h$:

Thus far, we have looked at the contemporaneous spillover effect of shocks in one market on the other market. Now, we study the propagation of these shocks into the other market's conditional density over time. If we have a shock in the oil market today, how does this shock transmit to the US market over time? To investigate the effect of a shock (for example 5% shock) in the oil return at time t on the conditional density of the US market at time $t + 1$, we take the following steps at each iterations of the MCMC:

1. Set $r_{t,shock}^{oil}$.

   The sample mean of the oil return is zero with 4 decimal point precision. We define shocks as $r_{t,shock}^{oil} = 5\% = 0.05$.

2. Take a draw for $r^{us}_{t,shock} \sim p(r^{us}_t | r^{oil}_{t,shock})$ which is a mixture model:

$$p(r^{us}_t | r^o_t, r_{1:T}, G) = \sum_{j=1}^{K} q_j(r^o_t) N(r^{us}_t | \mu_{j,us} + \Gamma_{j,us}\mathbf{r_{t-1}} + \beta_{jt}(r^o_t - \mu_{j,o} - \Gamma_{j,o}\mathbf{r_{t-1}}), c_{jt})$$
$$+ \left(1 - \Sigma_{j=1}^K q_j(r^o_t)\right) \frac{\Sigma_n N(r^o_t, r^{us}_t | \mu_{\mathbf{n}} + \Gamma_n \mathbf{r_{t-1}}, H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})}{\sum_n N(r^o_t | \mu_{n,o} + \Gamma_{n,o}\mathbf{r_{t-1}}, (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{o,o})} \tag{4.39}$$

(4.40)

To do so, we take a draw of the component $j = 1, 2, ..., K+1$ with probabilities $q_1(r^o_t), q_2(r^o_t), ..., q_K(r^o_t), 1 - \sum_{j=1}^K q_j(r^o_t)$. If we choose the last component, then take a draw from a mixture model with M components and weights $w_n = \frac{N(r^o_t | \mu_{n,o} + \Gamma_{n,o}\mathbf{r_{t-1}}, (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{o,o})}{\sum_n N(r^o_t | \mu_{n,o} + \Gamma_{n,o}\mathbf{r_{t-1}}, (H_t^{\frac{1}{2}} B_n H_t^{\frac{1}{2}'})_{o,o})}$.

3. Set $\mathbf{r}_{t,shock} = (r^{us}_{t,shock}, r^{oil}_{t,shock})$.

4. Using $\mathbf{r}_{t,shock}$, update the GARCH components $\{H_{t+1}\}$.

5. $\mathbf{r}_{t+1,shock}$ has a mixture density:

$$p(r^{us}_{t+1,shock}, r^o_{t+1,shock} | r_{1:T}, G) = \Sigma_{j=1}^K w_j N(r^o_{t+1,shock}, r^{us}_{t+1,shock} | \mu_j + \Gamma_j \mathbf{r}_{t,shock}, H_{t+1}^{\frac{1}{2}} B_j H_{t+1}^{\frac{1}{2}'})$$
$$+ \left(1 - \Sigma_{j=1}^K w_j\right) \int f(r^{us}_{t+1,shock}, r^o_{t+1,shock} | \theta) G_0(\theta) d\theta$$

6. Take a draw from the joint density: $(r^{us}_{t+1,shock}, r^o_{t+1,shock})$.

7. Obtain the conditional density of r^{us}_{t+1} given $r^o_{t+1,shock}$ from

$$p(r^{us}_{t+1} | r^o_{t+1,shock}, r_{1:T}, G) = \tag{4.41}$$
$$\Sigma_{j=1}^K q_j(r^o_{t+1,shock}) N(r^{us}_{t+1} | \mu_{j,us} + \Gamma_{j,us}\mathbf{r}_{t,shock} + \beta_{j,t+1}(r^o_{t+1,shock} - \mu_{j,o} - \Gamma_{j,o}\mathbf{r}_{t,shock}), c_{j,t+1})$$
$$+ \left(1 - \Sigma_{j=1}^K q_j(r^o_{t+1,shock})\right) \frac{\Sigma_n N(r^o_{t+1,shock}, r^{us}_{t+1} | \mu_{\mathbf{n}} + \Gamma_n \mathbf{r}_{t,shock}, H_{t+1}^{\frac{1}{2}} B_n H_{t+1}^{\frac{1}{2}'})}{\Sigma_n N(r^o_{t+1,shock} | \mu_{n,o} + \Gamma_{n,o}\mathbf{r}_{t,shock}, (H_{t+1}^{\frac{1}{2}} B_n H_{t+1}^{\frac{1}{2}'})_{o,o})}$$

8. Repeat steps 6 and 7 (e.g., 1000 times) and take the average of conditional density in Equation (4.41) over this 1000 iterations as $p(r^{us}_{t+1}|r^{o}_{t,shock}, r_{1:T})$ in the current iteration of MCMC.

9. Go to the next iteration of MCMC.

We can generalize this process to get the effect of a 5% shock in oil market at time t on the conditional density of the US market at time $t+5$ by adding the following steps after step 6:

At each iterations of the MCMC, after drawing unknowns from the posteriors we do the following steps:

1. Set $r^{oil}_{t,shock} = 5\%$ or $r^{oil}_{t,shock} = 5\%$.

2. Take a draw for $r^{us}_{t,shock} \sim p(r^{us}_{t}|r^{oil}_{t,shock})$ which is a mixture model.

3. Set $\mathbf{r}_{t,shock} = (r^{us}_{t,shock}, r^{oil}_{t,shock})$.

4. Using $\mathbf{r}_{t,shock}$, update the GARCH components $\{H_{t+1}\}$.

5. $\mathbf{r}_{t+1,shock}$ has a mixture density:

$$
\begin{aligned}
p(r^{us}_{t+1,shock}, r^{o}_{t+1,shock}|r_{1:T}, G) = {} & \Sigma^{K}_{j=1} w_j N(r^{o}_{t+1,shock}, r^{us}_{t+1,shock}|\mu_j + \Gamma_j \mathbf{r}_{t,shock}, H^{\frac{1}{2}}_{t+1} B_j H^{\frac{1}{2}'}_{t+1}) \\
& + \left(1 - \Sigma^{K}_{j=1} w_j\right) \int f(r^{us}_{t+1,shock}, r^{o}_{t+1,shock}|\theta) G_0(\theta) d\theta
\end{aligned}
$$

Take a draw from this joint density: $(r^{us}_{t+1,shock}, r^{o}_{t+1,shock}) \sim p(r^{us}_{t+1,shock}, r^{o}_{t+1,shock}|r_{1:T}, G)$.

6. for $j = 1 : 4$ do the following steps:

   - Using $\mathbf{r}_{t+j,shock}$, update the GARCH components $\{H_{t+j+1}\}$.

- $\mathbf{r}_{t+j+1,shock}$ has a mixture density:

$$\begin{aligned} & p(r^{us}_{t+j+1,shock}, r^{o}_{t+j+1,shock}|r_{1:T}, G) = \\ & \Sigma_{j=1}^{K} w_j N(r^{o}_{t+j+1,shock}, r^{us}_{t+j+1,shock}|\mu_j + \Gamma_j \mathbf{r}_{t+j,shock}, H^{\frac{1}{2}}_{t+j+1} B_j H^{\frac{1}{2}'}_{t+j+1}) \\ & + \left(1 - \Sigma_{j=1}^{K} w_j\right) \int f(r^{us}_{t+j+1,shock}, r^{o}_{t+j+1,shock}|\theta) G_0(\theta) d\theta \end{aligned}$$

- Take a draw from the joint density: $(r^{us}_{t+j+1,shock}, r^{o}_{t+j+1,shock})$.

7. Obtain the conditional density of r^{us}_{t+5} given $r^{o}_{t+5,shock}$ from

$$\begin{aligned} & p(r^{us}_{t+5}|r^{o}_{t+5,shock}, r_{1:T}, G) = \\ & \Sigma_{j=1}^{K} q_j(r^{o}_{t+5,shock}) N(r^{us}_{t+5}|\mu_{j,us} + \Gamma_{j,us}\mathbf{r}_{t+4,shock} + \beta_{j,t+5}(r^{o}_{t+5,shock} - \mu_{j,o} - \Gamma_{j,o}\mathbf{r}_{t+4,shock}), c_{j,t+5}) \\ & + \left(1 - \Sigma_{j=1}^{K} q_j(r^{o}_{t+5,shock})\right) \frac{\Sigma_n N(r^{o}_{t+5,shock}, r^{us}_{t+5}|\mu_{\mathbf{n}} + \Gamma_n \mathbf{r}_{t+4,shock}, H^{\frac{1}{2}}_{t+5} B_n H^{\frac{1}{2}'}_{t+5})}{\Sigma_n N(r^{o}_{t+5,shock}|\mu_{n,o} + \Gamma_{n,o}\mathbf{r}_{t+4,shock}, (H^{\frac{1}{2}}_{t+5} B_n H^{\frac{1}{2}'}_{t+5})_{o,o})} \end{aligned} \tag{4.42}$$

8. Repeat steps 7 and 8 (e.g., 1000 times) and take the average of conditional density in Equation (4.42) over this 1000 iterations as $p(r^{us}_{t+5}|r^{o}_{t,shock}, r_{1:T})$ in the current iteration of MCMC.

9. Go to the next iteration of MCMC.

In step 7, if we had new draws, we add the new component with $w_{K+1} = v_{K+1}\Pi_{j<K+1}(1-v_j)$ where $v_{K+1} \sim Beta(1, \alpha)$.

Following the steps mentioned above, we are able to obtain the conditional density $p(r^{us}_{t+5}|r^{o}_{t,shock}, r_{1:T})$. Similarly, we simulate the conditional density of the US market h-period-ahead (h = 1, 5, 22 for one day, one week, and one month, respectively) given a shock in the oil market return today, $r^{oil}_{t,shock}$. By setting $r^{oil}_{t,shock}$, we can perform different scenario analyses and examine how for example a big positive shock in oil market today would transmit to the US market in the future. It is very important for policy makers and also for investors to understand these relationships.

4.2.4 Lagged Shocks Spillover Effect on Volatility

In the proposed semiparametric model, estimating the lagged spillover of a shock in one market to the volatility of the other market is not as straightforward as the models in the literature. In the semiparametric model, the volatility at each time t has two components; H_t that follows a GARCH-BEKK process, and a second component B_{s_t} which is component-specific. Therefore, the lagged spillover effect of a shock in oil market on the volatility of the US market can be numerically approximated by

$$\frac{\partial \mathrm{vol}_{us,t}}{\partial \epsilon_{o,t-1}} = \frac{(H_t^{\frac{1}{2}} B_{s_t} H_t^{\frac{1}{2}'} |_{\epsilon_{t-1}})_{11} - (H_t^{\frac{1}{2}} B_{s_t} H_t^{\frac{1}{2}'} |_{\epsilon_{t-1}+(0,\delta)'})_{11}}{\delta} \tag{4.43}$$

where $\mathrm{vol}_{us,t}$ is the volatility of the US market at time t.

4.3 Benchmark: MGARCH-BEKK Model with Student-t Innovations

As our benchmark model, we consider a multivariate GARCH-BEKK-t model where the innovations follow a multivariate Student-t distribution. In estimating the spillover effects in the benchmark model, we follow the same steps as we do in the semiparametric model. We, first, estimate the joint density of the returns from two separate markets and then derive the density of one market conditional on the other market's returns. In the benchmark model, we assume that this joint density is a multivariate Student-t density. This is a restrictive assumption compared to the proposed model, but it still is more extensive than what has been considered in the literature. In the following formulation, $\mathbf{r_t}$ is the vector of returns, μ

is the mean vector, Γ is the autoregressive coefficient matrix, and H_t is the dynamic scale matrix of the innovations, ϵ_t.

$$\mathbf{r_t} = \mu + \Gamma \mathbf{r_{t-1}} + \epsilon_\mathbf{t}, \quad \epsilon_\mathbf{t} \sim t(0, H_t, \nu), \; t = 1, ..., T, \tag{4.44}$$

$$H_t = C_0' C_0 + A' \epsilon_\mathbf{t-1} \epsilon_\mathbf{t-1}' A + F' H_{t-1} F. \tag{4.45}$$

The on-diagonal elements of Γ provide the measures of own-mean lagged spillovers, whereas the off-diagonal elements measure the cross-mean lagged spillovers between markets. The scale matrix, H_t, follows an MGARCH structure with the set of parameters A, F, and C_0. C_0 is a lower triangular matrix, and A is a q by q matrix of coefficients a_{ij} that capture the lagged shock interactions among markets. F is a q by q matrix of coefficients f_{ij} that capture the lagged volatility spillover effects between markets. The attractive property of GARCH-BEKK model is that the conditional covariance matrices are positive definite (Silvennoinen, Annastiina Teräsvirta, 2009). This model, with normal density instead of student-t density, is commonly used in the literature to study the lagged volatility spillover among different markets. This is a special case of the proposed semiparametric model where μ_j is constant for all js and $\alpha \longrightarrow \infty$, since $G \longrightarrow G_0$. We use this model to estimate the lagged volatility spillovers and the contemporaneous shock spillovers between markets in a parametric setting.

4.3.1 Sampling Steps

We have

$$\mathbf{r_t} = \mu + \Gamma \mathbf{r_{t-1}} + \epsilon_\mathbf{t}, \quad \epsilon_\mathbf{t} \sim t(0, H_t, \nu), \; t = 1, ..., T \tag{4.46}$$

Step 1 Updating parameters $\Psi = (c_{0,11}, c_{0,21}, c_{0,22}, a_{11}, a_{12}, a_{21}, a_{22}, b_{11}, b_{12}, b_{21}, b_{22})$ (with identification restrictions $c_{0,11} > 0, c_{0,22} > 0, a_{11} > 0$ and $b_{11} > 0$), ν, and elements of vector μ and matrix Γ ($\beta =$ elemnts of μ and Γ). Assuming the prior $p(\beta) \sim N(\beta_0, D)$

and $p(\Psi) \sim N(\Psi_0, E)$ (truncated), then the posterior density is

$$p(\Psi, \beta, \nu | r_{1:T}, H_{1:T}) \propto p(\Psi)p(\beta)p(\nu) \times \prod_{t=1}^{T} t(r_t|\mu, \Gamma, H_t, \nu) \tag{4.47}$$

which is not of standard form, and we apply a Metropolis-Hastings sampler. Given the current value (Ψ, β, ν) of the chain, the proposal $(\Psi, \beta, \nu)'$ is computed by

$$(\Psi, \beta, \nu)' = (\Psi, \beta, \nu) + N(0, \widehat{V}) \tag{4.48}$$

$(\Psi, \beta, \nu)'$ is accepted with probability $\min\{p(\Psi, \beta, \nu)'|r_{1:T}, H_{1:T})/p((\Psi, \beta, \nu)|r_{1:T}, H_{1:T}), 1\}$, otherwise, $(\Psi, \beta, \nu)'$ is rejected and (Ψ, β, ν) is selected as the draw. In Equation (4.48), $\widehat{V}$ is the inverse Hessian matrix of $\ell = \log[p(\Psi, \beta, \nu|r_{1:T}, H_{1:T})]$ at its posterior mode, $(\widehat{\Psi, \beta, \nu})$.[4] We can obtain $\widehat{V}$ in Equation (4.48) by numerically optimizing ℓ.

[4]To achieve an acceptance rate of parameters between 0.2 and 0.5, we can scale $\widehat{V}$.

4.4 Empirical Results

4.4.1 Spillover Effects of a Shock in Oil Market on Canada Market

In this section, we study shock spillovers from oil market to Canada market. Data are from Toronto Stock Exchange returns, and West Texas Oil returns, from Bloomberg website, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations). We estimate the model parameters for both the benchmark model and the semiparametric model. We go on to illustrate the contemporaneous and lagged spillover effects of big surprises in oil market on the conditional density of the Canada market returns. Returns are calculated as log returns. The time series of daily returns are illustrated in Figure 4.1, and Table 4.1 displays the summary statistics of the datasets. The skewness coefficients are different from zero, and the kurtosis coefficients are above three for both series. These findings indicate that the probability distributions of the oil and S&P TSX returns are skewed and leptokurtic.

Model Performance

The criteria that we use to compare different models is the value of the log-predictive likelihood and the log-Bayes factor. Bayesian model comparisons favours parsimonious model specifications. Complex models are only chosen if they provide an improved description of the data. For each particular model $\mathcal{M}$ (i.e., MGARCH-BEKK-t or MGARCH-DPM), the predictive likelihood for $r_{L:T}, L < T$ is expressed in terms of the one-step-ahead predictive likelihoods,

$$m(r_{L:T}|r_{1:L-1}, \mathcal{M}) = \Pi_{t=L}^{T} f(r_t|r_{1:t-1}, \mathcal{M}). \tag{4.49}$$

We compare the models on a subset of data (starting from t=L) which tends to minimize the effect the prior distributions have. Each predictive likelihood is estimated as,

$$f(r_t|r_{1:t-1},\mathcal{M}) = \int f(r_t|\theta, r_{1:t-1})p(\theta|r_{1:t-1})d\theta, \tag{4.50}$$

$$\approx \frac{1}{M}\sum_{g=1}^{M} f(r_t|\theta^{(g)}, r_{1:t-1}). \tag{4.51}$$

where $\theta^{(g)},\ g = 1, ..., M$ is the posterior draws of the model parameters given $r_{1:t-1}$.

Based on Equation (4.51), the predictive likelihoods for the two models can be written as

$$f(r_t|r_{1:t-1}, \text{MGARCH-BEKK-t}) \approx \frac{1}{M}\sum_{g=1}^{M} t(r_t|\mu^{(g)}, \Gamma^{(g)}, H_t^{(g)}, \nu^{(g)}) \tag{4.52}$$

and

$$f(r_t|r_{1:t-1}, \text{MGARCH-DPM}) \approx \frac{1}{M}\sum_{g=1}^{M} N(r_t|\mu^{(g)}_{s_t^{(g)}}, \Gamma^{(g)}_{s_t^{(g)}}, H_t^{(g)^{1/2}} B^{(g)}_{s_t^{(g)}} H_t^{(g)^{1/2'}}) \tag{4.53}$$

In MGARCH-DPM model $s_t^{(g)}$ at each iteration g is one of the $K^{(g)}$ components, say component j, with probability $\omega_j^{(g)}$, or is a new component with probability $1-\sum_{j=1}^{K^{(g)}}\omega_j^{(g)}$.

The model with the larger predictive likelihood value is the one most consistent with the data. Usually in computing these quantities, which tend to be small, it is best to work with the log-predictive likelihood. In comparing two models ($\mathcal{M}_1$ versus $\mathcal{M}_2$) a rough guide is that a log-Bayes factor bigger than 3 is a strong evidence for outperformance of $\mathcal{M}_1$ and a log-Bayes factor bigger than 5 is a very strong evidence for outperformance of $\mathcal{M}_1$ against $\mathcal{M}_2$.

The following priors are used in estimation. In the MGARCH-BEKK-t model, $\nu \sim$

$\mathcal{U}(2, 100)$, and $\beta \sim N(0, 0.1I)$ for both models where $\beta = (\mu, \Gamma)$. For each of the GARCH parameters in both models, we set $\Psi \sim N(0, 100)$ truncated to impose identification. In MGARCH-DPM, for the concentration parameter we set $\alpha \sim \mathcal{G}(0.1, 0.3)$. The prior on α controls the number of the distinct components in the mixture model, although with a large number of observations the effect of the prior is diminished. For the hyper-parameters of the base measure G_0, we set $B_0 = (\nu_0 - q - 1)I$ which makes $E(B) = I$ and centers the conditional covariance of r_t at H_t. $\nu_0 = 8$, but other values for ν_0 do not change our conclusions.

Table 4.2 reports the log-predictive likelihoods and the log-Bayes factor over 2012/10/03 to 2015/12/31 with $L = T - 800$. The results strongly support our semi-parametric model relative to the benchmark model.

Figure 4.2 displays the time-series of the oil returns and S&P TSX returns as well as the difference in the log-predictive likelihood of the two models at each time using

$$\log p(r_t | r_{1:t-1}, \text{MGARCH-DPM}) - \log p(r_t | r_{1:t-1}, \text{MGARCH-BEKK-t}). \tag{4.54}$$

Positive values favour the MGARCH-DPM specification. This figure shows that the MGARCH-DPM model almost always outperforms MGARCH-BEKK-t model. There are large differences when the returns are extreme.

Parameter Estimation for S&P TSX and WTO

Table 4.3 reports the posterior mean of the constant parameters estimated using the benchmark model and the proposed model along with the %95 density intervals. Both models show strong positive values for f_{11} and f_{22}. The semiparametric model uses about 4.3 mixture components in average to estimate the joint density of the S&P TSX and the oil market returns.

Figure 4.3 shows the estimated time-varying posterior mean of the correlation between S&P TSX and oil returns obtained from the benchmark and the semiparametric model. It shows that the benchmark model overestimates the correlation.

Density Estimation

Figure 4.4 illustrates the posterior density of S&P TSX at time t conditional on the realized value of oil return at time t for three different days selected from a highly volatile period, a low volatile period, and a normal period. This figure supports our semiparametric model since the shape of the estimated posterior conditional density of each day is compatible with the period it lies in, in terms of the volatility level.

Figure 4.5 shows the posterior density of S&P TSX at time t conditional on the realized value of oil return at time t from 2007-01 to 2011-01. This figure illustrates how the estimated density evolves over time.

Scenario Analysis

In this section, we carry out a set of comparative risk scenario analyses. How a big shock in the oil market would contemporaneously affect the shape of conditional density of the Canada market return. Risk scenario analysis is useful in stress testing, evaluating risks to the financial system. One goal in stress testing might be assessing the impact of various potential risks that can be transmitted to a market. As an example, we would like to answer the question of 'How the mean or volatility of a market would be affected by a severe but plausible shock in other markets?' With the provided model in this chapter, we are able to answer a more comprehensive question: *How the conditional density of a market would be affected by a contemporaneous shock in another market?* The answer to these questions gives valuable insights into the portfolio optimization problems, the diversification opportunities, and the future regulations.

Figures 4.6 and 4.7 illustrate the contemporaneous spillover effect from a shock in the oil market to the S&P TSX conditional density and compare the results with the benchmark.

As the second panel illustrates, with the benchmark model, the contemporaneous effect of a big positive or negative shock in oil market on the TSX market is symmetric. When the market is volatile, a positive shock in oil market shifts the contemporaneous conditional density of TSX to the right, and a negative shock in oil market shifts the contemporaneous conditional density of TSX to the left. When the market is calm, the effect is the opposite; a positive shock in oil market shifts the contemporaneous conditional density of TSX to the left, and a negative shock in oil market shifts the contemporaneous conditional density of TSX to the right.

The contemporaneous spillover effect of a shock in the oil market on the TSX market derived from the benchmark is symmetric; the shifts in the conditional density in these two cases are almost the same amount but in opposite directions. This is not the case with the semiparametric model. Using the proposed model, we find that a positive 10% and a negative 10% shock in the oil market spill over into the TSX market asymmetrically. When the market is calm, a positive shock in the oil market shifts the contemporaneous conditional density of TSX to the right, and a negative shock in the oil market shifts the contemporaneous conditional density of TSX to the left. These shifts are not the same amount for positive and negative shocks, and the resulting conditional densities are skewed and leptokurtic. When the market is volatile, a positive shock in the oil market has almost no effect on the TSX conditional density contemporaneously while a negative shock in oil market shifts the contemporaneous conditional density of TSX to the left. Both positive and negative shocks result in a fatter lower tail than the case with no shock to the oil market.

Figures 4.8 and 4.10 show how a ±10% shock in the oil market propagates to the conditional density of the S&P TSX over time in the semiparametric model and the benchmark model for volatile and calm periods. Figures 4.9 and 4.11 illustrate the corresponding log of the conditional densities in order to make the tail comparison easier. The first plot on top left shows the transmission of a ±10% in oil market into the TSX market contemporaneously. The remaining three plots illustrate the lagged spillover of these shocks into the TSX market after one day (h=1), one week (h=5), and one month (h=22). Assuming a shock in oil market at time t, the returns of the TSX market at time t+1 (h=1), t+5 (h=5) and t+22 (h=22) are simulated using the semiparametric model, and the conditional densities are estimated. These plots show how the effect of a big shock in oil market on the TSX market disappears as we go forward over time. The contemporaneous effect is the most significant one.

Spillover Effects on Value-at-Risk

In this section, we assess downside risks faced by an investor who invests in the market portfolio. We look at the one-period-ahead value-at-risk for $10,000 investment in S&P TSX at 5% level. We need to find the value $\mathrm{VaR}_{t+1}^{5\%}$ that with probability 0.05 the loss on our investment is bigger than $\mathrm{VaR}_{t+1}^{5\%}$. In another words,

$$p(R_{t+1} < -\mathrm{VaR}_{t+1}^{5\%}) = 0.05$$

where R_{t+1} is the return on our investment after one period, $R_{t+1} = 10,000 \times e^{r_{t+1}^{TSX}}$. Table 4.4 reports the effect of a shock in oil market on the one-period-ahead value-at-risk for $10,000 investment in S&P TSX at 5% level. In the semiparametric model, positive and negative shocks both increase the predictive value-at-risk. The benchmark model underestimates the predictive value-at-risk of the investment in S&P TSX when we have a shock in oil market.

Lagged Spillover Effect

We are also able to estimate how a shock in the oil market today affects the volatility of the TSX market tomorrow. This lies under the heading of volatility spillover in the literature.

To compute the lagged spillover effect, we use the posterior mean of the following

$$\frac{\partial h_{11,t}}{\partial \epsilon_{2,t-1}} = \frac{(H_t^{\frac{1}{2}} B_{s_t} H_t^{\frac{1}{2}'} |_{\epsilon_{t-1}})_{11} - (H_t^{\frac{1}{2}} B_{s_t} H_t^{\frac{1}{2}'} |_{\epsilon_{t-1}+(0,\delta)'})_{11}}{\delta} \tag{4.55}$$

Figure 4.12 illustrates an oil market shock spillover into the volatility of TSX market over time (Using the posterior mean of Equation (4.55)). Unlike similar work in the literature, the estimated spillover effect from both models is not constant over time and increases when market volatility goes up. Moreover, this figure shows that the lagged volatility spillover effects are underestimated in the benchmark model.

4.4.2 Spillover Effects of a Shock in the Oil Market on the US Market

In this section we study shock spillovers from oil market to the US market. Data are from S&P 500 returns and West Texas Oil returns, from Bloomberg website, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations). We estimate the model parameters for both the benchmark model and the semiparametric model. We go on to illustrate the contemporaneous and lagged spillover effects of big surprises in oil market on the conditional density of the US market returns. Returns are calculated as log returns. The time series of daily returns are illustrated in Figure 4.13, and Table 4.5 displays summary statistics of the datasets.

The skewness coefficients are different than zero and the kurtosis coefficients are above three for both series. These findings indicate that the probability distributions of the oil and S&P returns are skewed and leptokurtotic.

Model Performance

Table 4.2 reports the log-predictive likelihoods and the log-Bayes factor with $L = T - 500$. The results strongly support our semi-parametric model relative to the benchmark model.

Figure 4.14 displays the time-series of the oil returns and S&P TSX returns as well as the difference in the log-predictive likelihood of the two models at each time using

$$\log p(r_t|r_{1:t-1}, \text{MGARCH-DPM}) - \log p(r_t|r_{1:t-1}, \text{MGARCH-BEKK-t}). \tag{4.56}$$

Positive values favour the MGARCH-DPM specification. This figure shows that the MGARCH-DPM model almost always outperforms MGARCH-BEKK-t model. There are large differences when the returns are extreme.

Parameter Estimation

Table 4.7 reports the posterior mean of the constant parameters estimated using the benchmark model and the proposed model along with the %95 density intervals. Both models show strong positive values for f_{11} and f_{22}. The semiparametric model uses about 3.1 mixture components in average to estimate the joint density of the S&P 500 and the oil market returns. Figure 4.15 shows the estimated time-varying posterior mean of the correlation between S&P 500 and oil returns obtained from the benchmark and the nonparametric model.

Density Estimation

Figure 4.16 shows the posterior density of S&P500 at time t conditional on the realized value of oil return at time t. This figure illustrates how the estimated density evolves over time.

Scenario Analysis

Investors, fund managers, and regulators are interested in knowing how various potential risks in other markets can be transmitted to the markets of their interest. For example, one might want to know how a big shock in the oil market would contemporaneously affect the conditional density of the US market returns. In this section, we use the benchmark model and the proposed model to examine how the conditional density of the US market would be affected by a contemporaneous shock in the oil market.

Figures 4.17 and 4.18 illustrate the contemporaneous spillover from a shock in the oil market to the conditional density of the S&P 500 returns and compare the results with the benchmark. In the benchmark model, the shifts in the conditional density in the cases of positive and negative shocks in the oil market are almost the same amount but in opposite directions. While in the proposed model, a positive 10% and a negative 10% shock in the oil market spill over into the S&P 500 market asymmetrically. When the market is calm, a positive shock in oil market shifts the contemporaneous conditional density of S&P 500 to the right, and a negative shock in oil market shifts the contemporaneous conditional density of S&P 500 to the left. These shifts are not the same amount for positive and negative shocks, and the resulting conditional densities are skewed and leptokurtic. When the market is volatile, a positive shock in oil market has almost no effect on the S&P 500 conditional density contemporaneously while a negative shock in oil market shifts the contemporaneous conditional density of S&P 500 to the left. Both positive and negative shocks result in a fatter lower tail than the case with no shock to the oil market.

This also can be of interest to investors and policy makers to examine how a ±10% shock in oil market propagates to the conditional density of the S&P TSX over time in the DPM-GARCH model and benchmark model for volatile and calm periods. Figures 4.19 and 4.21 show risk transmission into the entire conditional density of the US market, and Figures 4.20 and 4.22 illustrate the corresponding log of the conditional densities in order to make tail comparison easier. The first plot on top left shows the transmission of a ±10% in oil market into the S&P 500 market contemporaneously. The remaining three plots illustrate the lagged spillover of these shocks into the S&P 500 market after one day (h=1), one week (h=5), and one month (h=22). These plots show how the effect of a big shock in oil market on the S&P 500 market disappears as we go forward over time. The contemporaneous effect is the most significant one.

We can use the proposed model to measure financial stress. For instance, conditional on a 10% drop in oil market, what is the probability that U.S. market will realize a loss more than 5%? To this end, we examine the contemporaneous and lagged effects of shocks in oil market on the left tail of the conditional densities where the losses occur. Table 4.8 reports the effect of a drop in oil market on the probability of 5% loss in the U.S. market contemporaneously and after one day, one week and one month at three different dates. In general, the contemporaneous effect of a 10% drop in the oil market on the probability of 5% loss in the US market is more significant than the lagged effects, particularly during a volatile period.

Spillover Effects on Value-at-risk

Table 4.9 reports the effect of a shock in oil market on the one-period-ahead value-at-risk for $10,000 investment in S&P 500 at 5% level. In the semiparametric model, positive and negative shocks both increase the predictive value-at-risk. The benchmark model underestimates the predictive value-at-risk of the investment in S&P 500 when we have a shock in oil market.

4.5 Conclusion

We propose a model to allow a shock in one market to influence the contemporaneous and future (one-day-ahead, one-week-ahead, and one-month ahead) conditional density of the other market. This model extends the literature on spillover effects or contagion effects that focus on the transmission of shocks through moments to spillover effects on the conditional density. We provide a general approach to assess different risk scenarios, contemporaneously. With the speed of information transmission nowadays, it is beneficial for policymakers and regulators to be able to determine the contemporaneous spillover from shocks in one market to different aspects of the conditional density of the other market such as the conditional moments, tails, and value-at-risk.

We apply the proposed model to study how a shock in oil market affects the conditional density of S&P 500 returns and S&P TSX returns. The contemporaneous spillover effect of a shock in oil market on the TSX and S&P 500 markets derived from the benchmark is symmetric; the shifts in the conditional density in these two cases are almost the same amount but in opposite directions. This is not the case with the semiparametric model. Using the proposed model, we find that a positive 10% and a negative 10% shock in the oil market spill over into the TSX and S&P 500 asymmetrically. When the market is calm, a positive shock in oil market shifts the contemporaneous conditional density of TSX and S&P 500 to the right, and a negative shock in oil market shifts the contemporaneous conditional density of TSX and S&P 500 to the left. These shifts are not the same amount for positive and negative shocks, and the resulting conditional densities are skewed and leptokurtic. When the market is volatile, a positive shock in oil market has almost no effect on the TSX and S&P 500 conditional density contemporaneously while a negative shock in oil market shifts the contemporaneous conditional density of TSX and S&P 500 to the left. Both positive and negative shocks result in a fatter lower tail than the case with no shock to the oil market.

We also study the effect of the shocks in oil market on the value-at-risk of an investment in S&P 500 and S&P TSX. Positive and negative shocks in oil market both increase the predictive value-at-risk. The benchmark model underestimates the predictive value-at-risk of the investment in the market portfolio when we have a shock in the oil market.

	Mean	Variance	Skewness	Kurtosis	Min	Max
TSX	0.0047	0.250	-0.646	9.010	-4.251	4.069
WT Oil Return	0.002	1.019	-0.030	5.479	-7.228	9.539

Table 4.1: Summary statistics of the daily returns on oil returns and S&P TSX, from 2000/01/03 to 2015/12/31 (4021 observations).

Model	log(predictive likelihood)
MGARCH-DPM	-976.20
MGARCH-BEKK-t	-1033.10
log Bayes factor	56.90

Table 4.2: log The Predictive Likelihood of the last 800 observations. Data are daily returns on oil returns and S&P TSX, from 2000/01/03 to 2015/12/31 (4021 observations).

TSX	MGARCH-DPM		MGARCH-BEKK-t	
Parameter	Post. Mean	%95 DI	Post. Mean	%95 DI
$c_{0,11}$	0.025	(0.012, 0.031)	0.038	(0.029 , 0.048)
$c_{0,12}$	0.002	(-0.010, 0.010)	-0.004	(-0.036, 0.031)
$c_{0,22}$	0.005	(0.000, 0.011)	0.060	(0.043, 0.082)
$a_{0,11}$	0.149	(0.113, 0.189)	0.234	(0.205, 0.269)
a_{12}	-0.036	(-0.064, -0.012)	-0.028	(-0.089, 0.036)
a_{21}	0.017	(0.009, 0.025)	0.004	(-0.008, 0.016)
a_{22}	0.021	(0.008, 0.034)	0.165	(0.142, 0.189)
$f_{0,11}$	0.928	(0.918, 0.938)	0.960	(0.947, 0.969)
f_{12}	0.001	(-0.009, 0.015)	0.008	(-0.011, 0.029)
f_{21}	-0.054	(-0.063, -0.044)	-0.001	(-0.003, 0.002)
f_{22}	0.965	(0.957, 0.974)	0.980	(0.972, 0.984)
μ_1	–	–	0.018	(0.010, 0.028)
μ_2	–	–	0.005	(-0.015, 0.029)
Γ_{11}	–	–	0.033	(0.007, 0.064)
Γ_{12}	–	–	0.005	(-0.006, 0.016)
Γ_{21}	–	–	0.034	(-0.022, 0.099)
Γ_{22}	–	–	-0.035	(-0.068, 0.00)
ν	–	–	8.56	(7.26, 10.1)
k	4.3	(3.24,6.02)	–	–

Table 4.3: Posterior mean and %95 density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-BEKK-t models. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

Date	Model	One-period-ahead value-at-risk		
		-10 % Shock	No Shock	+10% Shock
15/01/2015	MGARCH-DPM	90.983	75.314	113.550
	MGARCH-BEKK	74.025	70.450	73.826
30/11/2011	MGARCH-DPM	108.111	77.795	107.913
	MGARCH-BEKK	89.596	80.573	81.262
14/10/2008	MGARCH-DPM	271.253	246.900	252.946
	MGARCH-BEKK	275.046	263.175	258.500
06/09/2013	MGARCH-DPM	93.856	47.078	80.772
	MGARCH-BEKK	78.788	48.788	79.280

Table 4.4: Effect of a shock in oil market on the one-period-ahead value-at-risk for $10,000 investment in S&P TSX at 5% level. Data used for estimation are the daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

	Mean	Variance	Skewness	Kurtosis	Min	Max
S&P	0.015	1.572	-0.181	6.892	-9.035	10.79
WT Oil	0.002	1.019	-0.030	5.479	-7.228	9.539

Table 4.5: Summary statistics of the daily returns on oil returns and S&P 500, from 2000/01/03 to 2015/12/31 (3965 observations).

Model	log(marginal likelihood)
MGARCH-DPM	-1217.693
MGARCH-BEKK	-1261.147
log Bayes factor	44.058

Table 4.6: log Marginal Likelihood of the last 500 observations. Data are daily returns on oil returns and S&P 500, from 2000/01/03 to 2015/12/31 (3965 observations).

S&P 500	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	%95 DI	Post. Mean	%95 DI
$c_{0,11}$	0.21	(0.011,0.033)	0.104	(0.089,0.124)
$c_{0,12}$	-0.004	(-0.012,0.005)	-0.011	(-0.034,0.014)
$c_{0,22}$	0.011	(0.006,0.016)	0.054	(0.035,0.071)
$a_{0,11}$	0.113	(0.100,0.140)	0.255	(0.231,0.282)
a_{12}	-0.009	(-0.019,0.000)	-0.020	(-0.040,0.003)
a_{21}	0.029	(0.020,0.038)	0.021	(-0.003,0.050)
a_{22}	0.030	(0.021,0.038)	0.159	(0.140,0.181)
$f_{0,11}$	0.943	(0.934,0.949)	0.952	(0.942,0.960)
f_{12}	0.006	(0.002,0.010)	0.007	(0.000,0.014)
f_{21}	-0.054	(-0.058,-0.051)	-0.004	(-0.010,0.001)
f_{22}	0.971	(0.967,0.973)	0.981	(0.976,0.985)
μ_1	–	–	0.022	(0.005,0.048)
μ_1	–	–	0.016	(-0.008,0.037)
Γ_{11}	–	–	-0.050	(-0.079,-0.026)
Γ_{12}	–	–	0.001	(-0.018,0.021)
Γ_{21}	–	–	0.010	(-0.015,0.037)
Γ_{22}	–	–	-0.035	(-0.070,-0.004)
k	3.1	(2.65,4.12)	–	–

Table 4.7: Posterior mean and %95 density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data are daily return on oil from http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm, and S&P 500 from CRSP: https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/index.cfm, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).

Probability of 5% loss in the US market when we have a 10% drop in the oil market:

Date	Shock in oil market	Contemporaneously	After 1 day	After 1 week	After 1 month
15/01/2015	No shock	0.0000524	0.0000185	0.0000192	0.000548
	-10% shock	0.00222	0.000824	0.00084	0.00221
10/12/2008	No shock	0.10863	0.11549	0.10687	0.09031
	-10% shock	0.21112	0.13696	0.12384	0.10564
14/10/2008	No shock	0.08332	0.08865	0.08559	0.07846
	-10% shock	0.18354	0.10802	0.10059	0.09135

Table 4.8: Probability of 5% loss in the US market when we have a 10% drop in the oil market. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).

Date	Model	One-period-ahead value-at-risk		
		-10 % Shock	No Shock	+10% Shock
15/01/2015	MGARCH-DPM	186.636	181.532	293.603
	MGARCH-BEKK	174.951	158.038	162.662
30/11/2011	MGARCH-DPM	263.661	237.435	338.730
	MGARCH-BEKK	262.201	238.509	233.432
14/10/2008	MGARCH-DPM	609.062	569.726	589.790
	MGARCH-BEKK	645.990	572.552	598.257
06/09/2013	MGARCH-DPM	179.371	102.075	233.465
	MGARCH-BEKK	131.033	100.590	93.460

Table 4.9: Effect of a shock in oil market on the one-period-ahead value-at-risk for $10,000 investment in S&P 500 at 5% level. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).

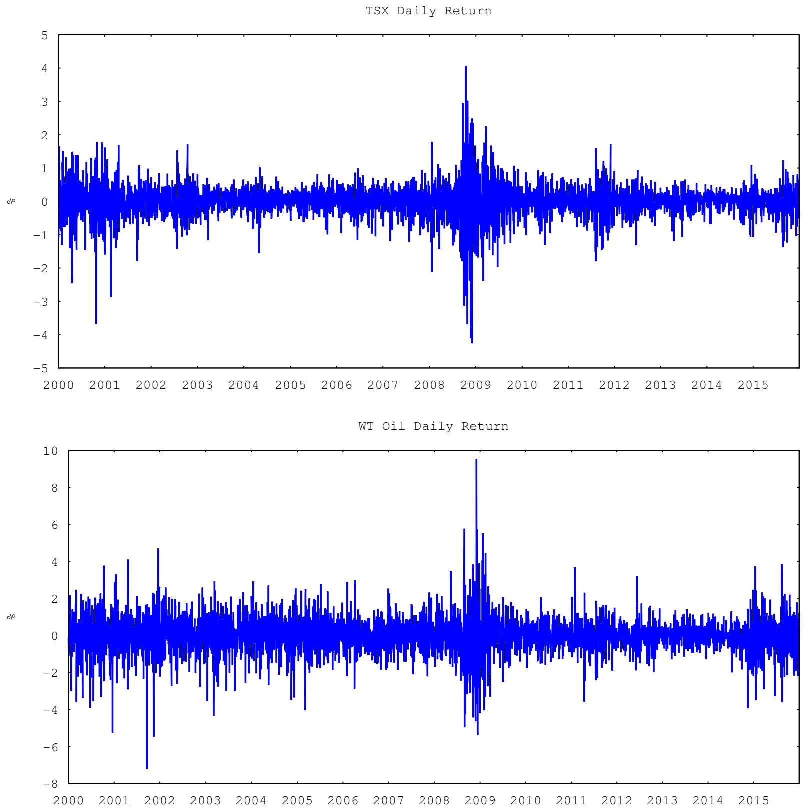

Figure 4.1: Time series of the daily returns on oil returns and S&P TSX, from 2000/01/03 to 2015/12/31 (4021 observations)

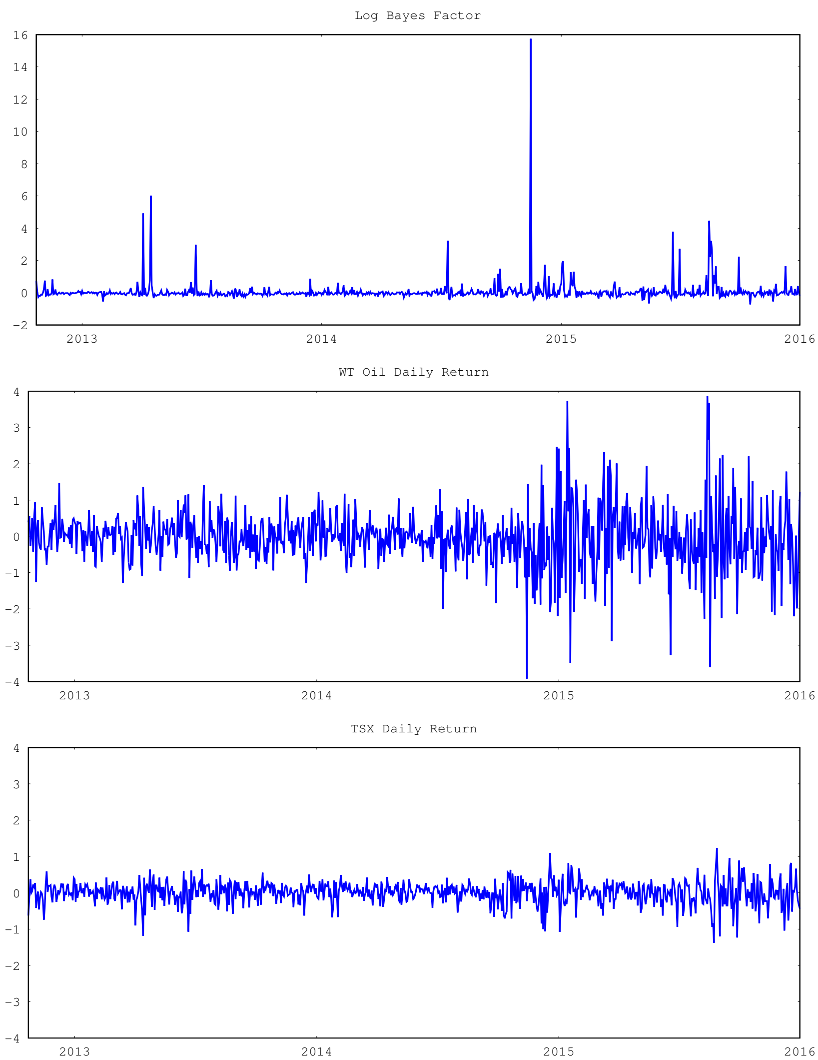

Figure 4.2: First panel: Log Bayes facor for the last 800 days. Second and third panel: Time series of the daily returns of oil returns and S&P TSX, from 2010/10/23 to 2015/12/31 (800 observations)

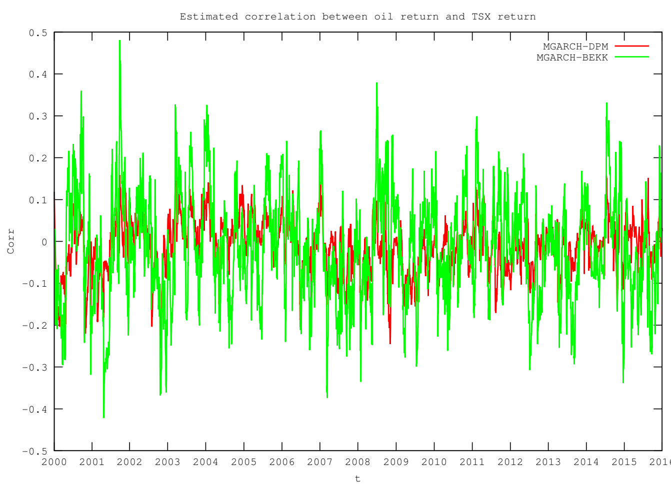

Figure 4.3: Estimated correlation between oil return and S&P TSX. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

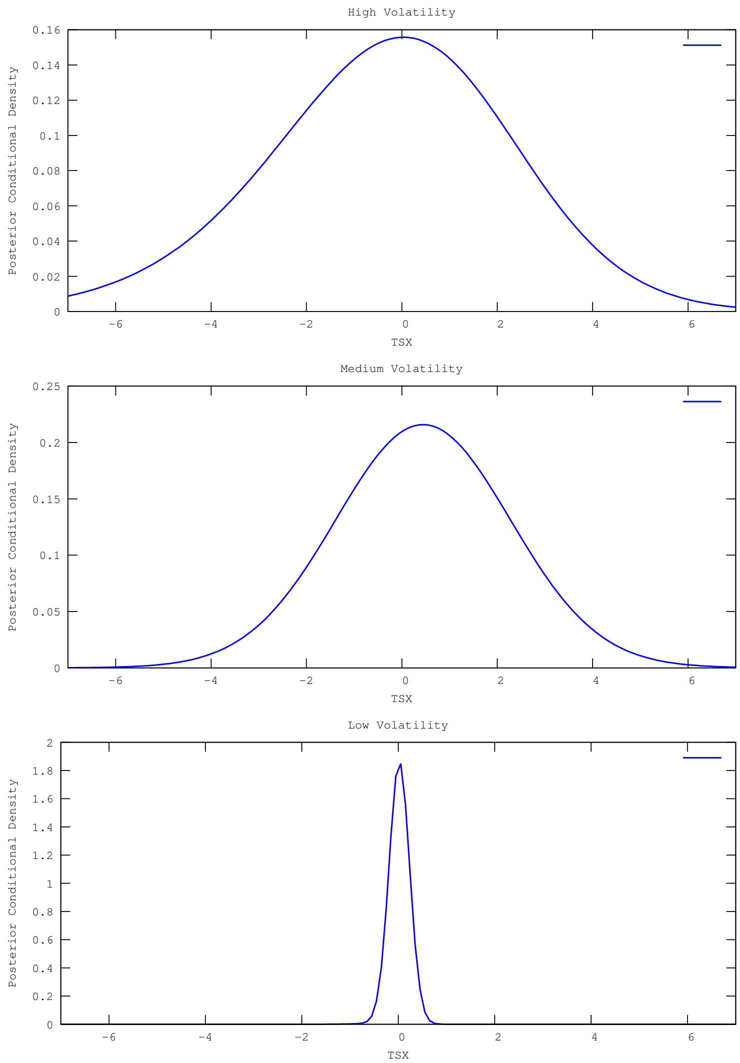

Figure 4.4: Posterior conditional density estimated using the semiparametric model. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

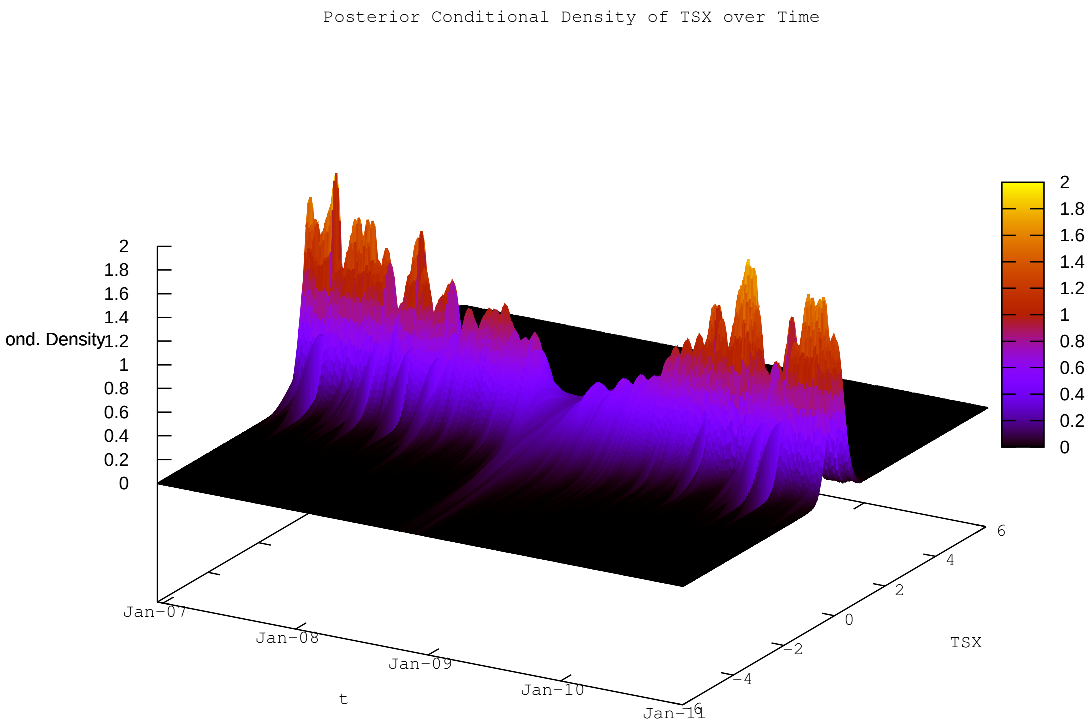

Figure 4.5: Posterior conditional density in three dimension estimated using the semiparametric model. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

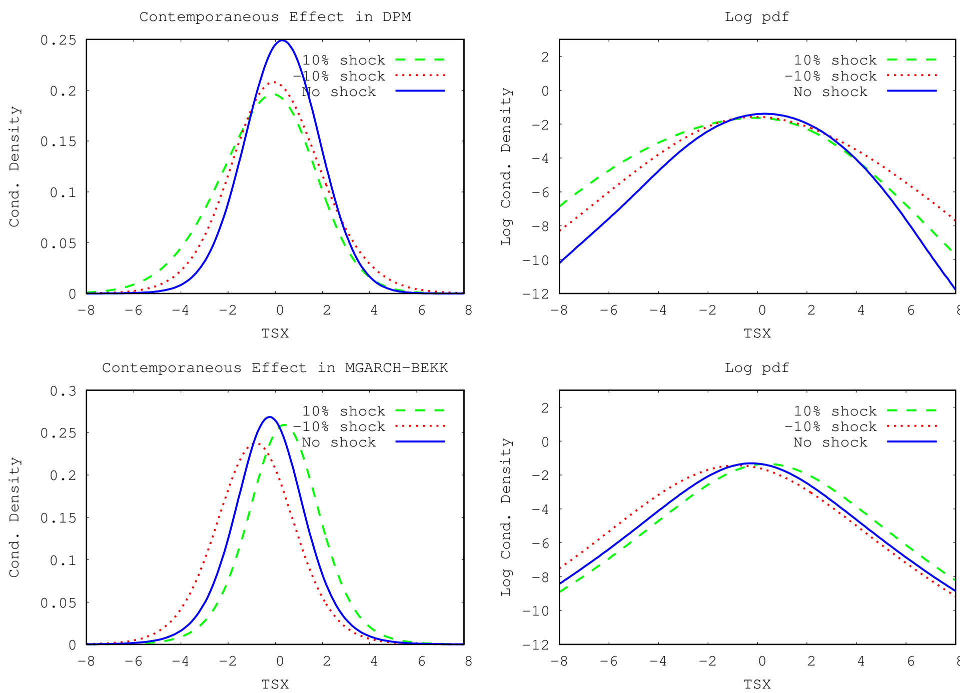

Figure 4.6: Risk Scenario Analysis when the market is highly volatile (14/10/2008). Contemporaneous spillover effect from a ±10% shock in oil market return to the conditional density of the S&P TSX. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

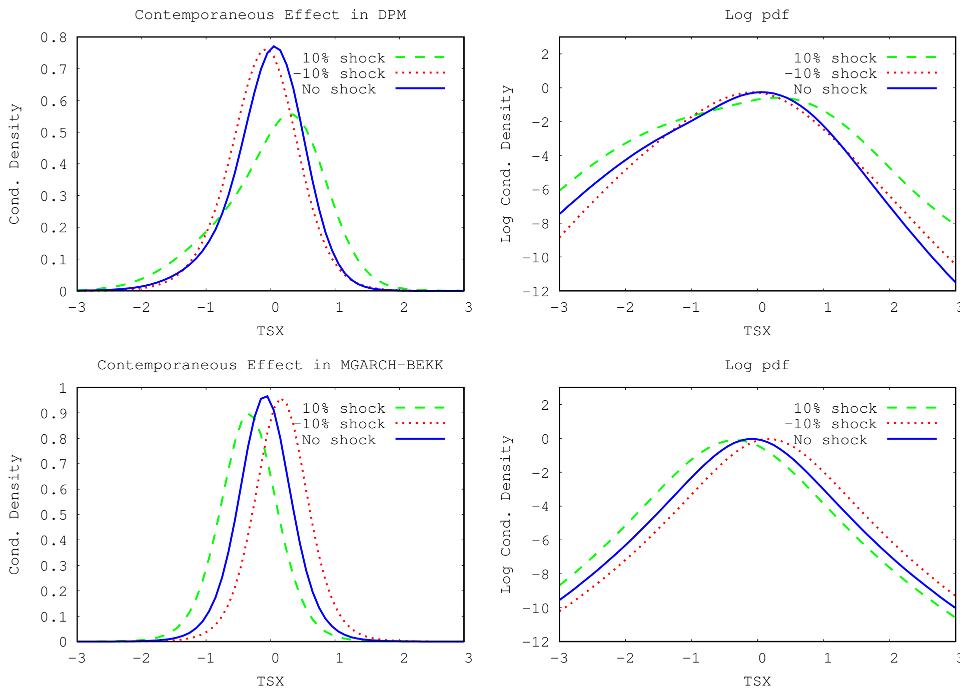

Figure 4.7: Risk Scenario Analysis when the market is calm (15/01/2015). Comparing the contemporaneous spillover effect from a ±10% shock in oil market return to the conditional density of the S&P TSX derived from the DPM model and the benchmark model. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

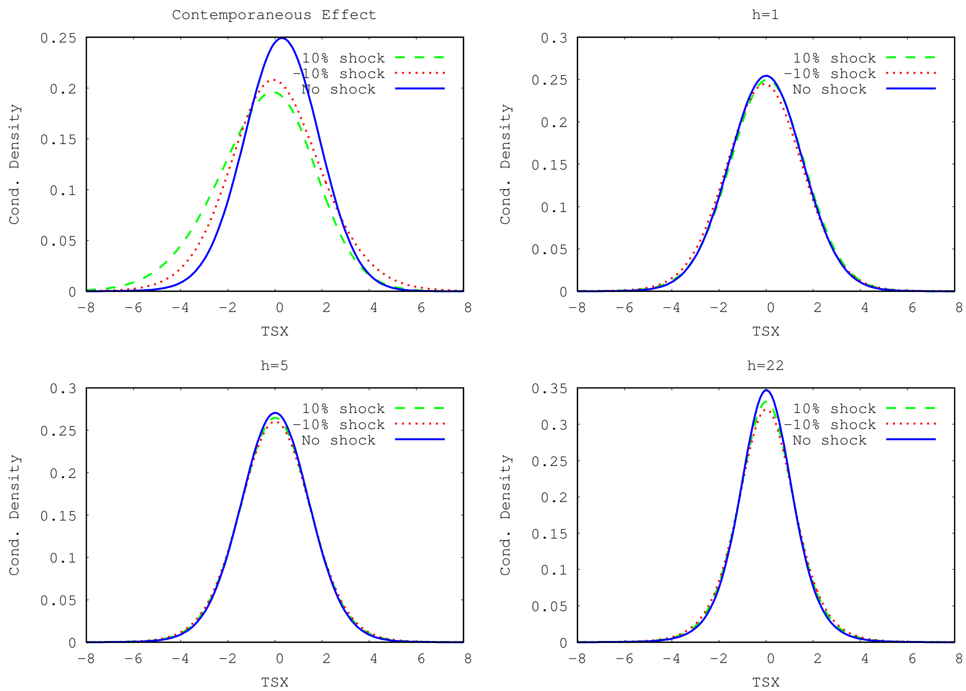

Figure 4.8: Risk Scenario Analysis when the market is highly volatile (14/10/2008). Propagation of a ±10% shock in oil market return to the conditional density of the S&P TSX in the semiparametric model. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations). The first plot on top left shows the contemporaneous spillover effect of ±10% in oil market on the TSX market. The second plot on top right shows the lagged spillover effect of ±10% in oil market on the TSX market after one day (h=1). The third and fourth plots show the lagged spillover effect of ±10% in oil market on the TSX market after one week (h=5), and one month (h=22), respectively.

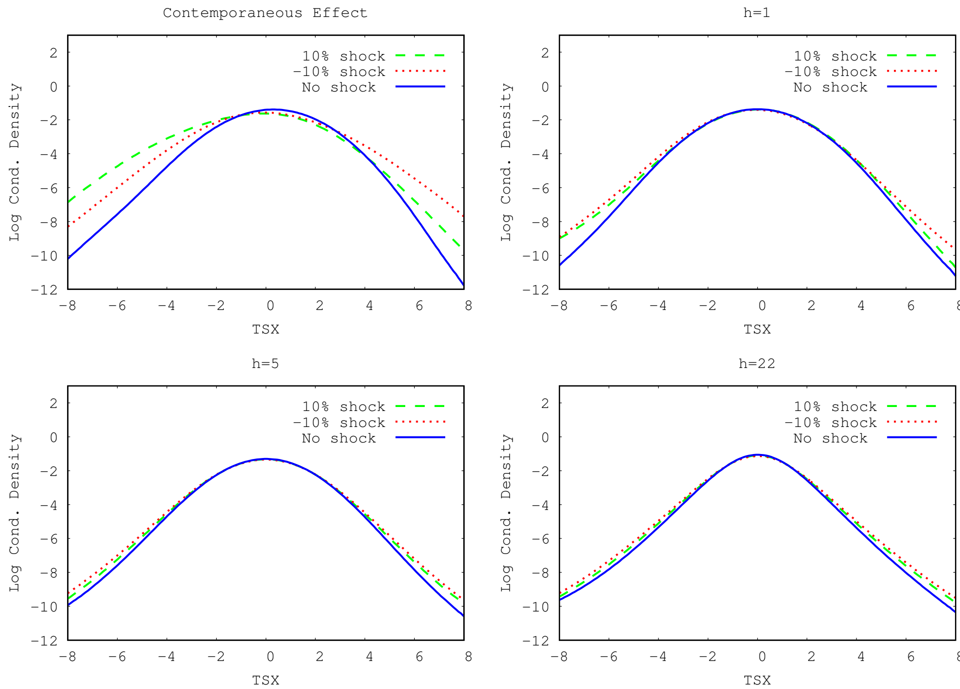

Figure 4.9: Risk Scenario Analysis when the market is highly volatile (14/10/2008). Log of the conditional density of S&P TSX if we have a $\pm 10\%$ shock in oil market return in the semiparametric model. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).
The first plot on top left shows the contemporaneous spillover effect of $\pm 10\%$ in oil market on the TSX market. The second plot on top right shows the lagged spillover effect of $\pm 10\%$ in oil market on the TSX market after one day (h=1). The third and fourth plots show the lagged spillover effect of $\pm 10\%$ in oil market on the TSX market after one week (h=5), and one month (h=22), respectively.

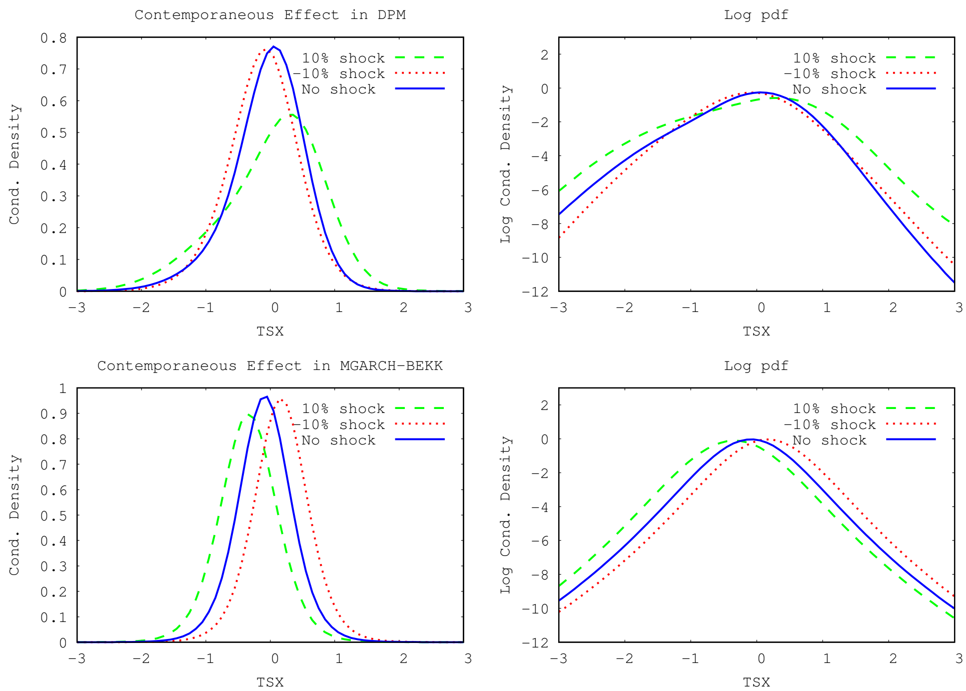

Figure 4.10: Risk Scenario Analysis when the market is calm (15/01/2015). Propagation of a ±10% shock in oil market return to the conditional density of the S&P TSX in the semiparametric model. Data are daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).
The first plot on top left shows the contemporaneous spillover effect of ±10% in oil market on the TSX market. The second plot on top right shows the lagged spillover effect of ±10% in oil market on the TSX market after one day (h=1). The third and fourth plots show the lagged spillover effect of ±10% in oil market on the TSX market after one week (h=5), and one month (h=22), respectively.

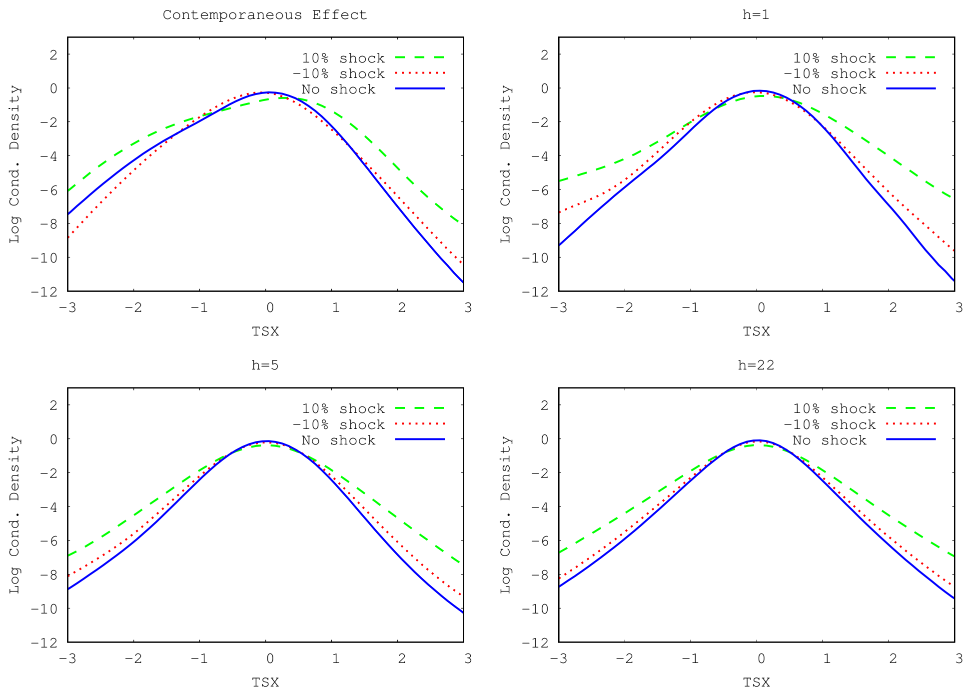

Figure 4.11: Risk Scenario Analysis when the market is calm (15/01/2015). Log of the conditional density of S&P TSX if we have a $\pm10\%$ shock in oil market return in the semiparametric model. Data are daily returns on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).
The first plot on top left shows the contemporaneous spillover effect of $\pm10\%$ in oil market on the TSX market. The second plot on top right shows the lagged spillover effect of $\pm10\%$ in oil market on the TSX market after one day (h=1). The third and fourth plots show the lagged spillover effect of $\pm10\%$ in oil market on the TSX market after one week (h=5), and one month (h=22), respectively.

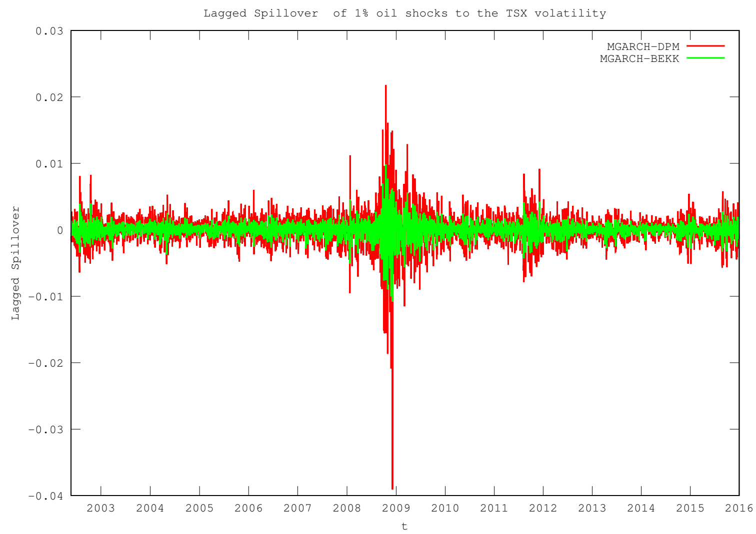

Figure 4.12: Lagged spillover of one 1% of oil shocks into the TSX volatility. Data used for estimation are the daily return on oil and S&P TSX, ranging from Jan 3, 2000 to Dec 31, 2015 (4021 observations).

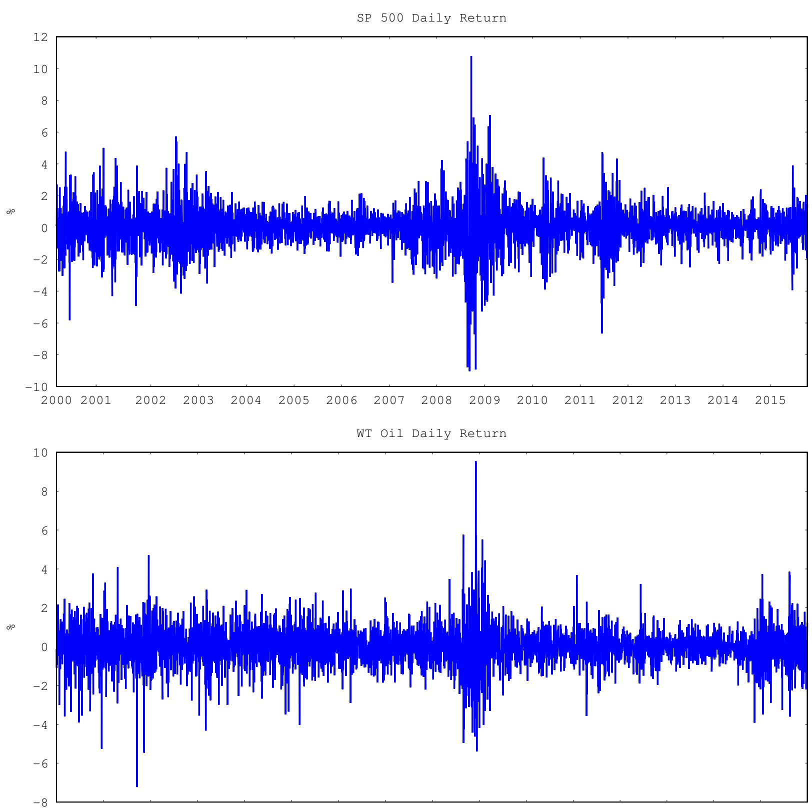

Figure 4.13: Time series of the daily returns on oil returns and S&P 500, from 2000/01/03 to 2015/12/31 (3965 observations)

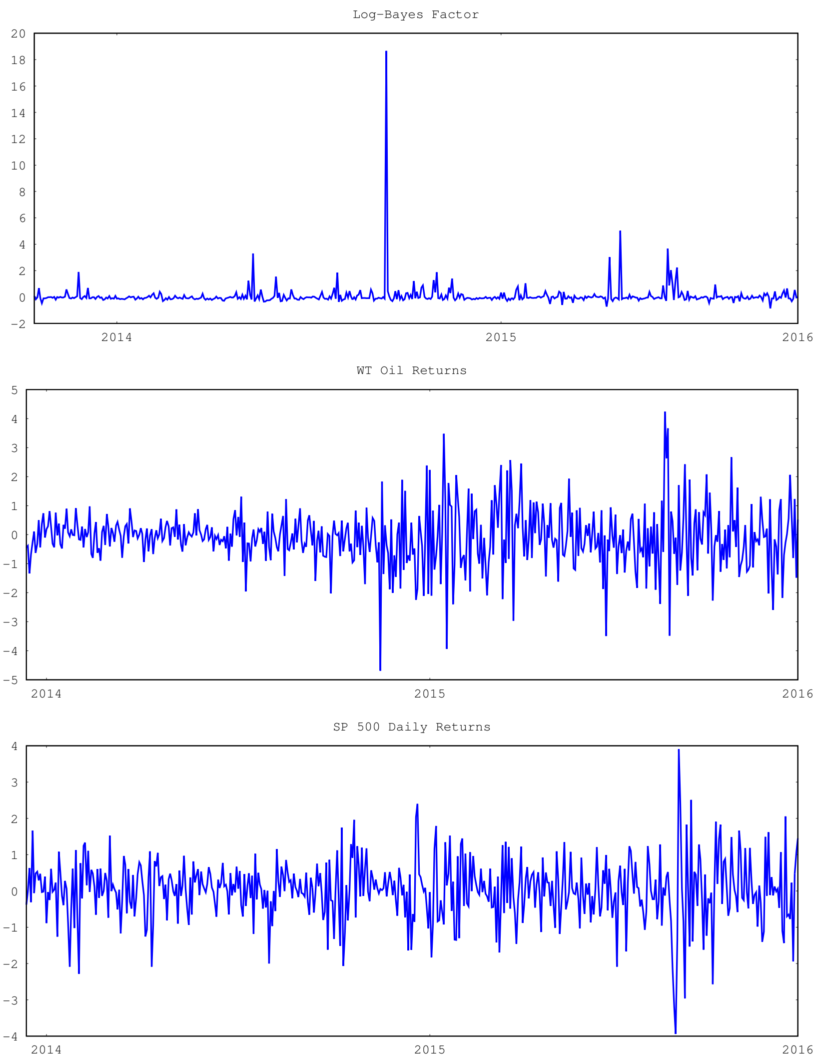

Figure 4.14: First panel: Log Bayes facor for the last 500 days. Second and third panel: Time series of the daily prices of oil returns and S&P 500, from 2013/10/23 to 2015/12/31 (500 observations)

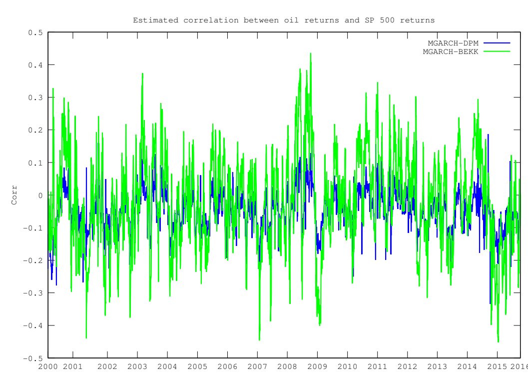

Figure 4.15: Estimated correlation between oil return and S&P 500. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).

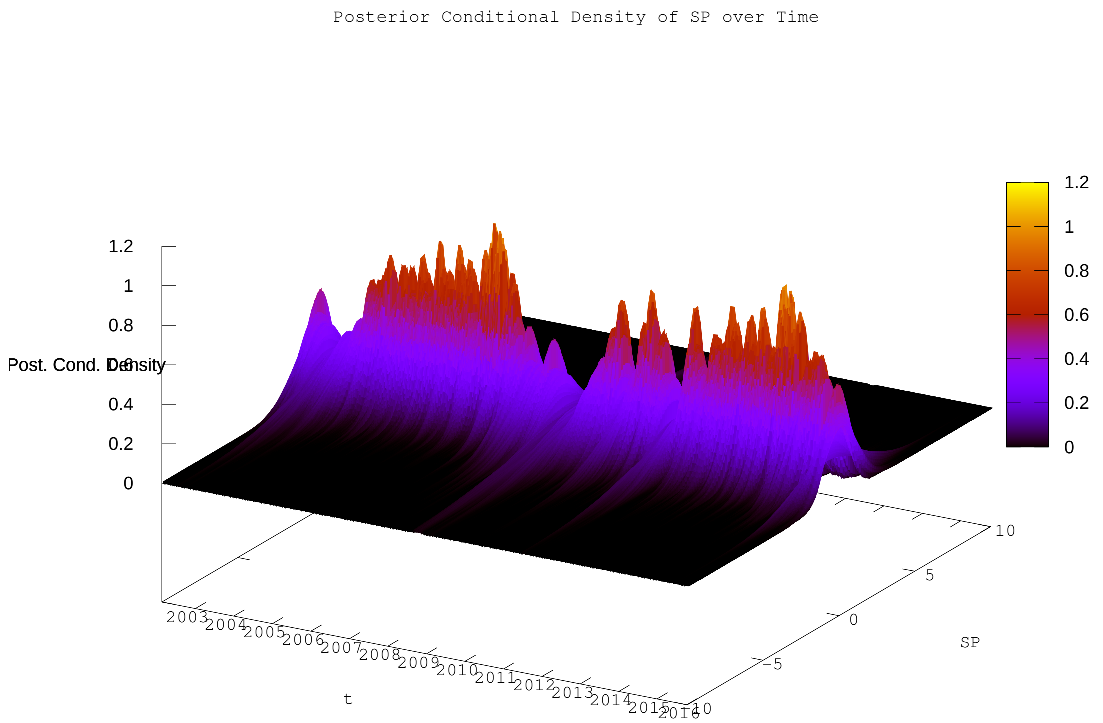

Figure 4.16: Realized conditional density. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).

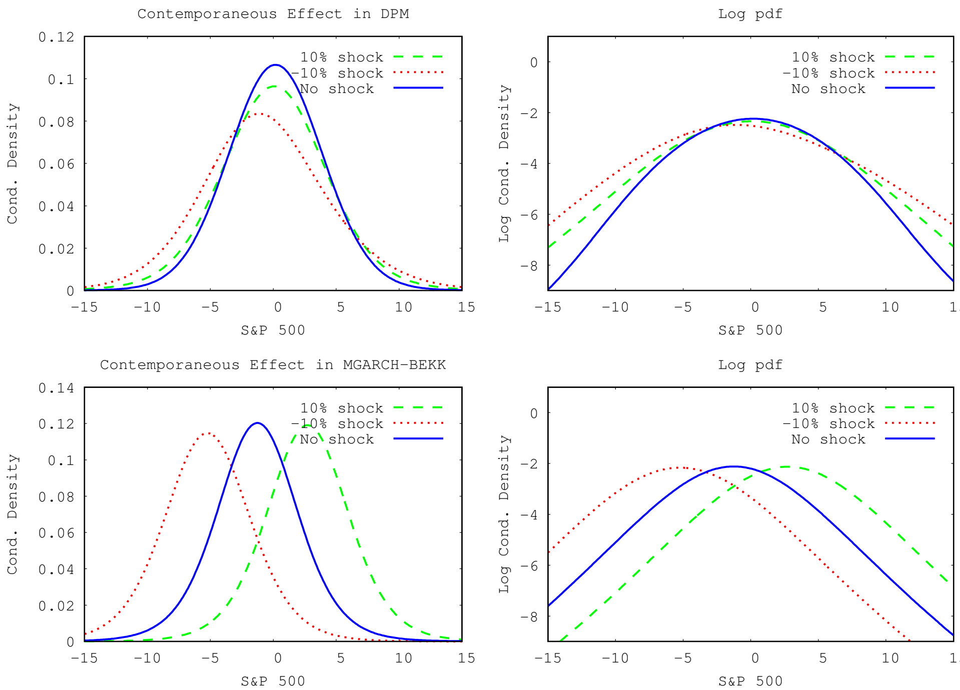

Figure 4.17: Risk Scenario Analysis when the market is highly volatile (15/10/2008). Contemporaneous spillover effect from a ±10% shock in oil market return to the conditional density of the S&P 500. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).

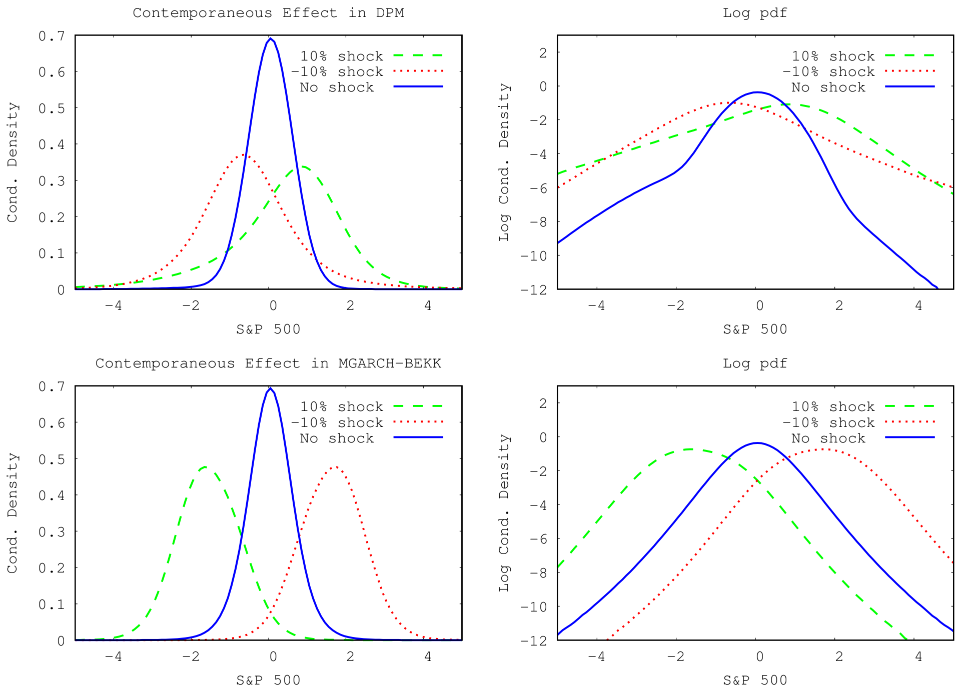

Figure 4.18: Risk Scenario Analysis when the market is calm (31/12/2013). Comparing the contemporaneous spillover effect from a ±10% shock in oil market return to the conditional density of the S&P 500 derived from the DPM model and the benchmark model. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).

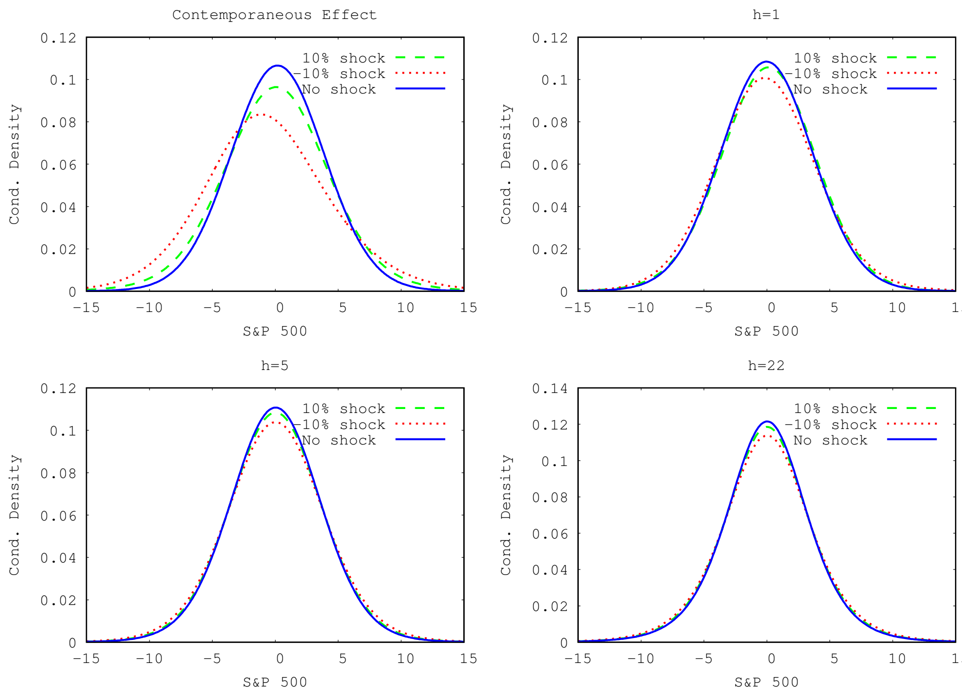

Figure 4.19: Risk Scenario Analysis when the market is highly volatile (15/10/2008). Propagation of a $\pm10\%$ shock in oil market return to the conditional density of the S&P 500 in the semiparametric model. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations). The first plot on top left shows the contemporaneous spillover effect of $\pm10\%$ in oil market on the S&P 500 market. The second plot on top right shows the lagged spillover effect of $\pm10\%$ in oil market on the S&P 500 market after one day (h=1). The third and fourth plots show the lagged spillover effect of $\pm10\%$ in oil market on the S&P 500 market after one week (h=5), and one month (h=22), respectively.

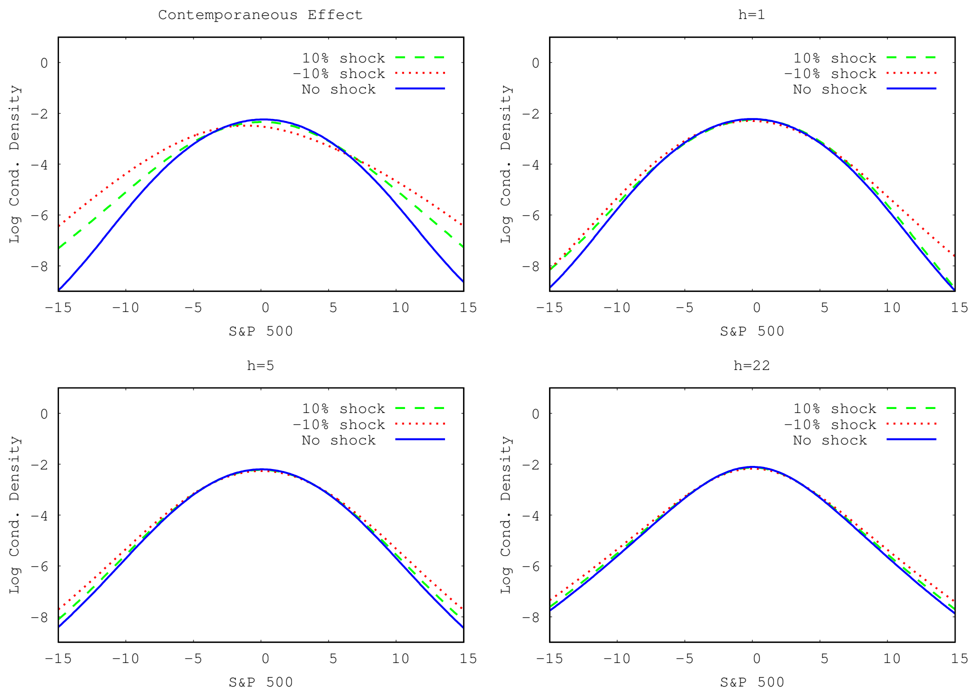

Figure 4.20: Risk Scenario Analysis when the market is highly volatile (15/10/2008). Log of the conditional density of S&P TSX if we have a ±10% shock in oil market return in the semiparametric model. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).
The first plot on top left shows the contemporaneous spillover effect of ±10% in oil market on the S&P 500 market. The second plot on top right shows the lagged spillover effect of ±10% in oil market on the S&P 500 market after one day (h=1). The third and fourth plots show the lagged spillover effect of ±10% in oil market on the S&P 500 market after one week (h=5), and one month (h=22), respectively.

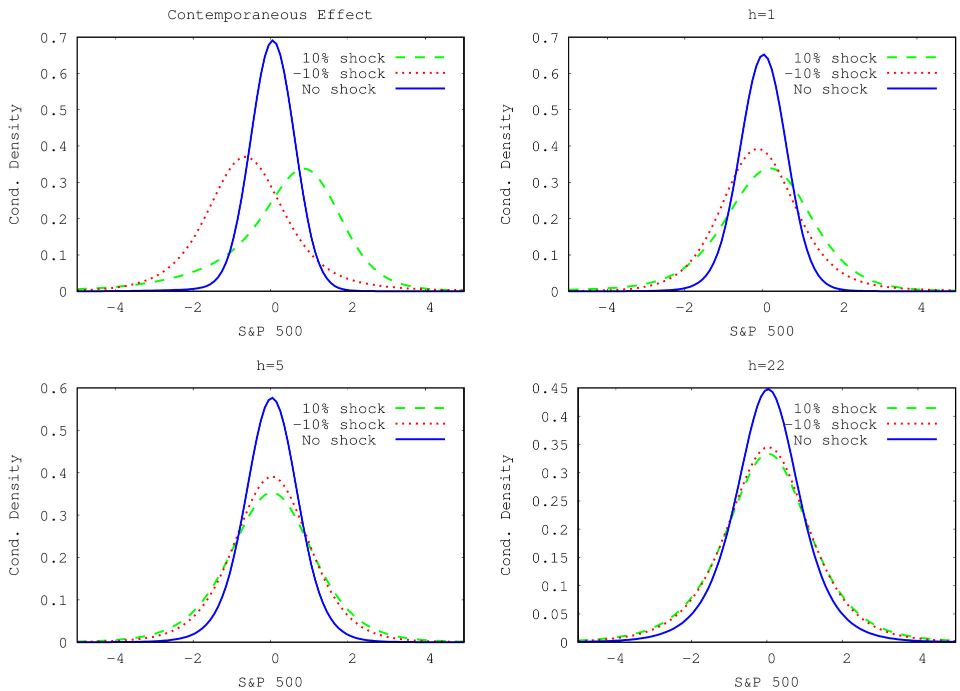

Figure 4.21: Risk Scenario Analysis when the market is calm (31/12/2013). Propagation of a ±10% shock in oil market return to the conditional density of the S&P 500 in the semiparametric model. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).
The first plot on top left shows the contemporaneous spillover effect of ±10% in oil market on the S&P 500 market. The second plot on top right shows the lagged spillover effect of ±10% in oil market on the S&P 500 market after one day (h=1). The third and fourth plots show the lagged spillover effect of ±10% in oil market on the S&P 500 market after one week (h=5), and one month (h=22), respectively.

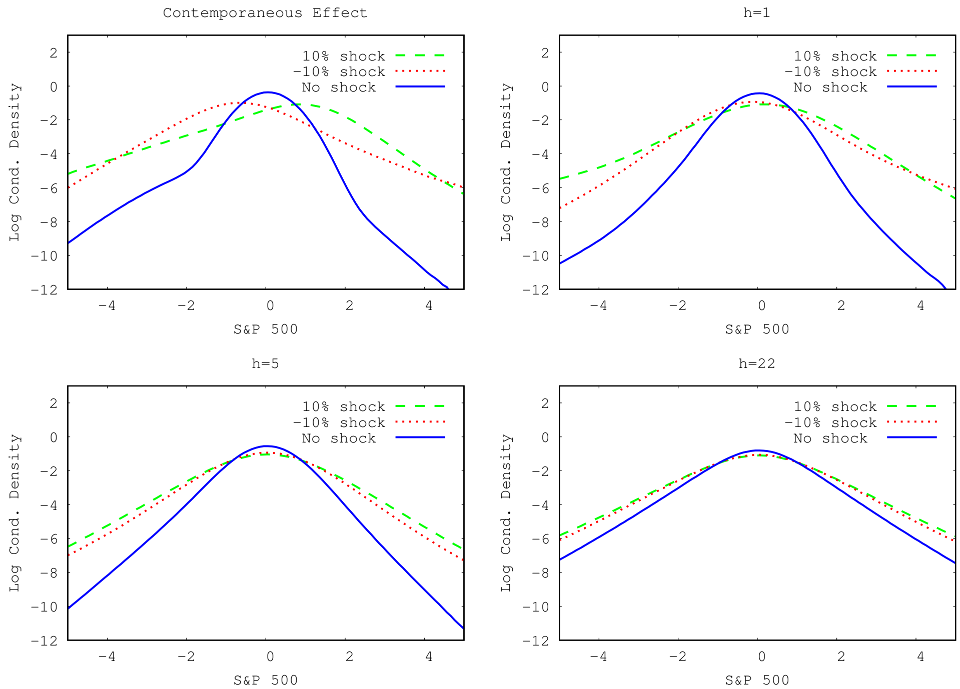

Figure 4.22: Risk Scenario Analysis when the market is calm (31/12/2013). Log of the conditional density of S&P 500 if we have a ±10% shock in oil market return in the semiparametric model. Data are daily return on oil and S&P 500, ranging from Jan 3, 2000 to Dec 31, 2015 (3965 observations).
The first plot on top left shows the contemporaneous spillover effect of ±10% in oil market on the S&P 500 market. The second plot on top right shows the lagged spillover effect of ±10% in oil market on the S&P 500 market after one day (h=1). The third and fourth plots show the lagged spillover effect of ±10% in oil market on the S&P 500 market after one week (h=5), and one month (h=22), respectively.

Chapter 5

Essay 3: Do financial variables help predict the conditional distribution of the market portfolio?

The extent of predictability in stock market returns has been a very active area of research. Predictability of market returns has important implications for asset allocation and risk management. Predictability is usually examined by regressing the stock market returns on a lagged variable such as dividend yield as the predictor. There are contradicting conclusions on whether available information on the financial and macroeconomic variables can improve out-of-sample forecasts of market excess returns. In theory, the predictability of stock returns can be consistent with an efficient stock market if the predictable component reflects a time-varying risk premium.

Some articles provide empirical evidence that lends support to the predictability of market returns. Rozeff (1986); Fama and French (1988); Campbell and Shiller (1988); Cutler et al. (1991); Hodrick (1992); Lettau and Ludvigson (2001); Balvers et al. (1990) and Rapach and Wohar (2006) report that certain financial variables such as dividend yield, price-earnings ratio, and the ratio of the dividend yield to the short-term interest rate can explain a significant fraction of movements in monthly, quarterly, or annual stock returns. Pesaran and Timmermann (1995) find that during calm periods predictability of stock returns is low while during more volatile periods predictability is increased and could be exploited by investors. Keijsers (2017) studies the predictive power of dividend payout ratio in a mixture model with parameter instability and shows that assuming stable parameters can lead to enormous losses for the long-term investor.

On the other hand, many studies find evidence against the forecasting power of the financial and macroeconomic variables (see for example Goetzmann and Jorion (1993); Nelson and Kim (1993); Bossaerts and Hillion (1999); Ferson et al. (2003); Engstrom (2003); Goyal and Welch (2003); Lettau and Ludvigson (2005); Ang and Bekaert (2006) and Welch and Goyal (2008)). These studies either conclude that there is no strong evidence supporting the out-of-sample forecasting power of these variables or show that the statistics for many of these predictors are spuriously significant.

The studies mentioned above along with many others only investigate the predictability of mean returns. For an extensive review see Rapach and Zhou (2013). However, out-of-sample point forecasts of the conditional mean do not convey any information about the uncertainty of the realized returns nor the spread and shape of the return distribution, while this information may be valuable for investors. The economic value of a return forecast depends not only on the point forecasts but the entire return distribution.

Although the value of predictors for the predictive mean is unclear, there is evidence that predictors help explain other features of the predictive distribution. Schwert (1989), Engle et al. (2012), Christiansen et al. (2012), Paye (2012), and Asgharian et al. (2013) use various financial and macroeconomic factors to predict the volatility of stock returns. Allowing for predictability of the conditional mean and the conditional variance of stock returns, Cenesizoglu and Timmermann (2012) examine different forecasting models with different predictors using Welch and Goyal (2008) data. They show that using these predictors can improve the economic performance of the prediction models upon constant mean and variance, although this is not the case when the focus is only on statistical measures.

The literature on the predictability of the entire density of market returns using financial and economic predictors is scarce. Some studies investigate this matter through predictability of the conditional mean and conditional variance, as mentioned above. Cenesizoglu and Timmermann (2008) propose a quantile approach to capturing predictability in the distribution of stock returns and find that many of the economic variables studied in the literature are useful in predicting the tails of the return distribution but not necessarily its center.

In this chapter, instead of focusing only on the conditional mean, conditional variance, or conditional quantiles of market excess returns, we investigate whether these financial and macroeconomic variables are useful in predicting the one-period-ahead full density of monthly US stock returns. We attempt to predict features of the market return density in addition to what is captured by the first and second moments. This is particularly important in financial and econometric applications where we usually have heavy tails, asymmetry, or multi-modality. We provide evidence that these features can be important in applications such as portfolio selection.

We consider a mixture model for the distribution of market excess returns over time where the predictors are used not only in forecasting the conditional mean but in predicting the shape of the conditional density of market returns. This is done by allowing the weights of the mixture to be determined using available information on the predictors.

To implement this, we use a Bayesian nonparametric mixture model that allows the mixing distribution to change with time. In general, we can allow time-variation in the weights of the mixture, the atoms, or both. See for example MacEachern (1999); De Iorio et al. (2004); Gelfand et al. (2005); Teh et al. (2006); Griffin and Steel (2006); Duan et al. (2007); Rodriguez and ter Horst (2008). In this chapter, we assume that atoms are constant, and variation of the mixing distribution over time comes from the time-varying weights. To determine the weights at each time, we use available information on financial and macroeconomic variables. This is a novel and flexible approach by which we examine whether the financial/macroeconomic variables carry any useful information to predict the time-varying density of the market excess returns.

The time-varying weights in the model allow the mixture density to change over time, and this variation in density is in addition to the changes caused by a time-varying mean and variance. The infinite mixture model with time-varying weights can reveal changes to the density not captured by the first and second moments. This makes a considerable difference in the statistical and economic performance of the model as we demonstrate in our empirical results.

To give a picture of this difference, we look at an example where we use the dividend payout ratio as the only predictor for market excess returns in the proposed model and in the linear regression model where the predictor is only used in forecasting the conditional mean. Figure 5.1 illustrates the conditional density estimated by the two models in two different periods. This reveals a notable difference in the level of uncertainty around the conditional mean. The conditional means, however, are almost the same for both models. This figure also shows how using available information on dividend payout ratio changes the density over time in the proposed model while using this information only in the conditional mean does not affect the shape of the density over time considerably.

We use the predictors studied in Welch and Goyal (2008), updated until December 2015. Besides studying the out-of-sample point forecasts of the conditional mean which show little or no improvement over the benchmark models, we evaluate the improvement in the predictive density of market excess return by log-predictive Bayes factors. Moreover, the economic gains of the proposed model is illustrated by a market timing exercise as well as a portfolio selection problem for an investor who maximizes her expected utility. Our results show that the statistical and economic performance of the proposed model is superior over different benchmark models. The greatest predictive ability is found from the kitchen sink model which includes all predictors under study. We also illustrate how these financial and macroeconomic variables are useful in predicting the conditional density, the conditional moments and the conditional quantiles.

The remainder of the chapter is structured as follows. We begin by theoretical settings of the proposed model and the posterior sampling steps in Section 5.1. Section 5.2 presents simulation results, and applications of the proposed model are found in Section 5.3. Section 5.4 concludes.

5.1 Model Specification

The question under study is whether stock returns can be predicted using a set of financial and macroeconomic variables. To examine the predictive power of a variable, researchers mainly focus on the conventional statistical criteria such as root mean squared forecasting errors of the predicted mean. By far the most popular model in studying the predictability of market returns using financial and economic variables is the following linear regression model

$$y_t = x_{t-1}\mu + e_t \tag{5.1}$$

where y_t is the market return in excess of the risk-free rate, x_{t-1} denotes a row vector of the lagged predictors, including 1 for the intercept term, used to forecast the conditional mean of market returns, and e_t is the error vector (innovation) usually assumed to be i.i.d with mean equal to zero. μ is the vector of regression coefficients including the intercept. This model is estimated, and the in-sample and out-of-sample forecasting power of the prediction model is usually compared with the constant mean model. The constant mean model is the primary benchmark in the literature and corresponds to the case where no predictor is used in Equation (5.1) except for an intercept.

In contrast to Equation (5.1), the volatility could be time-varying (e.g., following a GARCH process). Some articles study predictability in the conditional volatility through entering the predictors in the volatility of the market returns (see for example Schwert (1989), Christiansen et al. (2012), Paye (2012), and Asgharian et al. (2013)). One example is given in the following specification

$$\begin{aligned} y_t &= x_{t-1}\mu + e_t \quad e_t \sim \mathcal{N}(0, \sigma_t^2) \\ \sigma_t^2 &= \alpha_0 + x_{t-1}\alpha_1 + \delta \frac{e_t}{\sigma_t} + \gamma \sigma_{t-1}^2 \end{aligned} \tag{5.2}$$

The models in Equations (5.1) and (5.2) focus on the predictability of the conditional mean and volatility but maintain the same innovation density for e_t. However, the density of e_t may change over time as well. In this chapter, instead of focusing only on the predictive mean or variance of market excess returns, we investigate the predictability of the entire density which provides a complete measure of the uncertainty around the predictive mean; the whole distribution for all possible outcomes.

To this end, we assume an infinite mixture model for the distribution of market excess returns over time. The predictors are used not only in forecasting the conditional mean but in predicting the general shape of the conditional density of market returns. This is done by allowing the weights of the mixture to be determined using available information on the predictors.

5.1.1 An Infinite Mixture Model with Time-varying Weights

We propose a novel approach to use different financial and macroeconomic variables to estimate the time-varying predictive density of market excess returns. The density of excess returns is assumed to follow

$$f_t(y) = \int f(y|\theta)dG_t(\theta), \quad t = 1, ..., T, \tag{5.3}$$

where y_t is the market excess return at time t, $f(.|\theta)$ is the kernel density with parameters θ, and G_t is the time-varying mixing distribution. G_t can be continuous or discrete. Here, we focus on the discrete case. To allow the mixing distribution, G_t, to change over time, we can introduce dependence in the weights, the atoms, or both. Define,

$$\begin{aligned} G_t(.) &= \sum_{j=1}^{\infty} \omega_j(t)\delta_{\theta_j(t)}(.) \\ \omega_j(t) &= v_j(t)\prod_{l<j}(1 - v_l(t)) \end{aligned} \tag{5.4}$$

where $0 \leq \omega_j(t) \leq 1$ and $\sum_{j=1}^{\infty} \omega_j(t) = 1$ for all t, and $\delta_{\theta_j(t)}(.)$ denotes a point mass at $\theta_j(t)$. This is the general stick-breaking model (MacEachern, 1999; Rodriguez and ter Horst, 2008; Griffin and Steel, 2006). $\theta_j(t)$s are independent and identically distributed sample paths from a stochastic process, and $v_j(t)$s are i.i.d draws from a distribution with support over $(0, 1)$[1].

The general stick-breaking prior can be simplified by introducing dependence only in the weights (constant atoms), or in the atoms (constant weights). Constant weight models are usually computationally simpler but lack flexibility (MacEachern, 2000). In this chapter, we assume that the atoms are constant over time, and introduce time-variation in the weights in a model specification called Dependent Probit Stick-Breaking Prior put forward by Rodriguez and Dunson (2011). The dependent probit stick-breaking prior is similar to the model in Equation (5.4) when atoms are constant and $v_j(t)$s are determined by probit transformations of normal random variables:

[1]In the literature, one common way to define the time-varying weights is to let the sticks, $v_j(t)$s, be i.i.d. realizations from a stochastic process $\mathcal{B}(a_j(t), b_j(t))$ where $\mathcal{B}(a, b)$ denotes the beta distribution.

$$G_t = \sum_{j=1}^{\infty} \omega_j(t)\delta_{\theta_j}, \tag{5.5}$$

$$\omega_j(t) = v_j(t)\prod_{l<j}(1 - v_l(t)), \tag{5.6}$$

$$v_j(t) = \Phi(\alpha_j(t)), \quad j = 1, 2, ... \tag{5.7}$$

where $\Phi(.)$ denotes the cumulative distribution function for the standard normal distribution, $\{\alpha_j(t)\}_{j=1}^{\infty}$ have Gaussian densities, and θ_js are i.i.d draws from the base measure G_0; $\theta_j \overset{i.i.d}{\sim} G_0,\ j = 1, 2, ...$ The probit transformations ensure that the weights are well-defined and satisfy $0 \leq \omega_j(t) \leq 1$ and $\sum_{j=1}^{\infty} \omega_j(t) = 1$ for all t.

We propose a nonparametric mixture model based on the dependent stick-braking prior of Rodriguez and Dunson (2011) where available information on the predictors, x_{t-1}, is used in determining the time-varying sticks, $v_j(t)$s, and hence in estimating the time-varying density of the market excess return at time t.

The proposed model is specified as follows

$$y_t \sim f_t(y) = \sum_{j=1}^{\infty} \omega_j(t)\mathcal{N}(y_t|x_{t-1}\mu_j, \sigma_t^2\nu_j^2), \quad t = 1, ..., T \tag{5.8}$$

$$\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \quad \text{where } \epsilon_t = y_t - \eta \tag{5.9}$$

$$(\mu_j, \nu_j^2) \overset{i.i.d}{\sim} G_0, \quad j = 1, 2, ... \tag{5.10}$$

$$\omega_j(t) = \Phi(x_{t-1}\beta_j)\prod_{j'<j}(1 - \Phi(x_{t-1}\beta_{j'})),\ j = 1, 2, ... \tag{5.11}$$

$$G_0 \equiv \mathcal{N}(\mu|\mu_0, V_0) \times \mathcal{IG}(\nu^2|\frac{s_0}{2}, \frac{v_0}{2}) \tag{5.12}$$

$$\beta_j \overset{i.i.d}{\sim} \mathcal{N}(\beta_0, B_0),\ j = 1, 2, ... \tag{5.13}$$

where T is the number of observations. The vector of the predictors at time $t-1$ including 1 for the intercept term, x_{t-1}, the vector of the unknown coefficients in the conditional mean, μ, and the vector of the unknown coefficients in probit transformations, β_j, are $(p+1)$-dimensional, where p is the number of predictors. The vector of the predictors at each time, x_t, includes 1 and the predictors of interest. For example, when examining the predicting power of the dividend payout ratio, $\frac{d_t}{p_t}$, we set $x_t = (1 \;\; \frac{d_t}{p_t})$. β_js are the unknown parameter vectors describing how the weights of the mixture density respond to different values of the predictors.

Here, the kernel density is Gaussian. The mixing is over μ_j and the component ν_j of the variance, which are i.i.d draws from the base measure G_0 and constant over time. The second component of the variance, σ_t, captures volatility clustering over time. Equation (5.9) denotes the GARCH(1,1) process for σ_t^2 where the parameter η permits a nonlinear asymmetric response to the shocks, ϵ_t. The component-specific parameter ν_j is a positive number which scales σ_t to yield a better estimate of the density of y_t. Variation of the mixing distribution over time comes from the time-varying weights. The time-varying weights have stick-breaking structure where the time-varying sticks are determined by probit transformations of a linear combination of the lagged predictors, $\Phi(x_{t-1}\beta_j)$.

Equation (5.12) shows the base measure that we use for the component-specific parameters, μ_j and ν_j. $\mathcal{IG}(s_0, v_0)$ represents an inverse gamma distribution with shape parameter s_0 and scale parameter v_0[2]. The vector of coefficients β_j in the probit transformations for each mixture component j has a normal prior (Equation (5.13)).

Available information on the predictors, x_{t-1}, in addition to being used in predicting the conditional mean of market returns (where the majority of the literature has focused), are

[2]The inverse gamma distribution's probability density function is defined over the support $x > 0$

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} exp(-\frac{\beta}{x}).$$

Here, $\Gamma(.)$ denotes the gamma function.

employed to predict the time-varying mixture weights. This translates into non-constant weights if the predictors are statistically and practically useful in forecasting the entire predictive density. If a predictor conveys no information about the predictive density of market returns, the corresponding β_js will have estimated values near zero. If none of the predictors are significant in forecasting the weights, we will have an infinite mixture model with weights that do not change significantly over time, which in practice is equivalent to a mixture model with constant weights.

The proposed model (Equations (5.8)-(5.13)) nests several special cases that we use as our benchmarks. The first case is when x_{t-1} has only one element equal to 1 (i.e., when we do not use any predictor; p=0). This case is equivalent to an infinite mixture model with constant weights. The second case is when the mixture has only one component ($\omega_1(t) = 1$, $\omega_j(t) = 0, \forall j > 1$ and $\nu_1 = 1.$). This case is equivalent to a linear regression model with time-varying mean and time-varying variance. The third case is achieved from the second case where variance does not change over time (in GARCH specification we have $\delta = 0$, and $\gamma = 0$). This case is equivalent to a Gaussian linear regression model with time-varying mean and constant variance. The fourth case is obtained from the third case when we use no predictor in the conditional mean. This case is equivalent to linear regression model with constant mean and constant variance which is the main benchmark considered in the literature.

In Table 5.1, we introduce the benchmarks against which we compare the statistical and economic performance of the proposed model later in our empirical experiments. The first three models are popular linear regression models with constant/time-varying mean and volatility. $\mathcal{M}_0$ is the constant mean and constant variance model and is analogous to the prevailing mean model of Welch and Goyal (2008). No predictor is used in this model. In $\mathcal{M}_1$, we use the predictors in the conditional mean. $\mathcal{M}_2$ extends $\mathcal{M}_1$ to allow for time-varying volatility through a GARCH(1,1) process.

The next benchmark model we consider, $\mathcal{M}_3$, is an infinite mixture model in which, unlike the proposed model in this chapter, weights are fixed; the vector of predictors includes only 1. Our proposed model is an infinite mixture model where the predictors are used in predicting the conditional mean and also in forecasting the time-varying weights of the mixture. We could only use the predictors to forecast the weights of the mixture since it is possible that adding new parameters to the model in the conditional mean might only increase the dimension of the problem without adding any statistical or economic value. Therefore, in our empirical studies, we consider two versions of the proposed model, $\mathcal{M}_4$ and $\mathcal{M}_5$.

Model	Description	
	Benchmark Models	
$\mathcal{M}_0$	Constant mean and constant variance	$y_t = \mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_1$	Time-varying mean and constant variance	$y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_2$	Time-varying mean and variance	$y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$
$\mathcal{M}_3$	Infinite mixture with constant weights	$y_t = \mu_t + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \quad (\mu_t, \nu_t^2) \sim G$
	Proposed Models	
$\mathcal{M}_4$	Infinite mixture with time-varying weights	$y_t = \mu_t + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \quad (\mu_t, \nu_t^2) \sim G_t$
$\mathcal{M}_5$	Infinite mixture with time-varying weights and predictors in the conditional mean	$y_t = x_{t-1}\mu_t + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \quad (\mu_t, \nu_t^2) \sim G_t$

Table 5.1: List of the benchmark models and the proposed models. In all models, y_t is the market excess return and x_t is a row vector of the predictors (financial or macroeconomic variables) including 1 for the intercept.

5.1.2 Sampling Steps

To estimate the unknown parameters in Equation (5.8)-(5.13), we apply a Gibbs sampler along with the slice sampler of Walker (2007), the collapsed sampler of Rodriguez and Dunson (2011), and Metropolis-Hastings sampler. In the following and other sections of the chapter $y_{j:t}$ denotes $\{y_j, y_{j+1}, ..., y_t\}$.

Step 1 The posterior distribution of (μ_j, v_j^2), $j = 1, ..., K$: Assuming the conjugate priors $G_0 \equiv \mathcal{N}(\mu|\mu_0, V_0) \times \mathcal{IG}(v^2|, \frac{s_0}{2}, \frac{v_0}{2})$, sampling from the posterior distribution of the component-specific parameters is straightforward. First, we introduce indicator variables $\{\xi_t\}_{t=1}^T$ such that $\xi_t = j$ if and only if observation y_t is sampled from component j. Then, we make the following transformation:

$$\sigma_t^{-1} y_t = \sigma_t^{-1} x_{t-1} \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, v_t^2).$$

Then we can use the linear regression results to take posterior draws for (μ_j, v_j^2).

$$\nu_j^2 | y_{1:T}, \xi, \mu_j \sim \mathcal{IG}\left(\frac{n_j + v_0}{2}, \frac{s_0 + \sum_{\xi_t = j}(\sigma_t^{-1} y_t - \sigma_t^{-1} x_{t-1} \mu_j)^2}{2}\right),$$

$$\mu_j | y_{1:T}, \xi, \nu_j^2 \sim \mathcal{N}(\bar{\mu}, \bar{V}), \tag{5.14}$$

in which

$$\bar{V} = \left(V_0^{-1} + \sum_{\xi_t = j} \frac{x_{t-1} x'_{t-1}}{\sigma_t^2 \nu_j^2}\right)^{-1}, \tag{5.15}$$

$$\bar{\mu} = \bar{V}\left(\sum_{\xi_t = j} \frac{x_{t-1} y_t}{\sigma_t^2 \nu_j^2} + V_0^{-1} \mu_0\right). \tag{5.16}$$

Step 2 To overcome the infinite-dimensionality of the problem, we apply the slice sampler introduced in Walker (2007) by defining auxiliary variables u_t such that

$$f(y_t, u_t|W) = \sum_{j=1}^{\infty} \mathbf{1}(u_t < \omega_j(t))\mathcal{N}(y_t|x_{t-1}\mu_j, \sigma_t^2\nu_j^2). \tag{5.17}$$

then the joint posterior is proportional to

$$\Pi_{t=1}^{T} p(\omega_1(t), ..., \omega_K(t))\Pi_{j=1}^{K} p(\mu_j, \nu_j)\Pi_{t=1}^{T}\mathbf{1}(u_t < \omega_{\xi_t}(t))\mathcal{N}(y_t|x_{t-1}\mu_{\xi_t}, \sigma_t^2\nu_{\xi_t}^2)$$

K is the smallest number that satisfies the condition $1 - \sum_{j=1}^{K}\omega_j(t) < u_t$, for all t. Details are not discussed here. We update the auxiliary variables $\{u_t\}_{t=1}^{T}$ as in Equation 5.18, and K is the smallest number that satisfies Equation 5.19.

$$u_t|\xi \sim \mathcal{U}(0, \omega_{\xi_t}(t)) \tag{5.18}$$

$$\max_t\{1 - \sum_{j=1}^{K}\omega_j(t)\} < \min\{u_1, ..., u_T\} \tag{5.19}$$

Additional $\omega_j(t)$ and (μ_j, ν_j) will need to be generated from the priors if K is incremented.

Step 3 Updating $\xi_t,\ t = 1, ..., T,$

$$p(\xi_t = j|y_{1:T}) \propto 1(\omega_j(t) > u_t)\mathcal{N}(y_t|x_{t-1}\mu_j, \sigma_t^2 v_j^2), j = 1, ..., K \tag{5.20}$$

In order to sample the value of β_js, we use a data augmentation scheme developed in Rodriguez and Dunson (2011) which allows us to implement another Gibbs sampling scheme. We first define the set of conditionally independent auxiliary variables $z_j(t) \sim \mathcal{N}(x_{t-1}\beta_j, 1)$. If we define $\xi_t = j$ if and only if $z_j(t) > 0$ and $z_r(t) < 0, \forall r < j$, then we have

$$p(\xi_t = j) = p(z_j(t) > 0, z_r(t) < 0, \forall r < j) = \Phi(x_{t-1}\beta_j) \prod_{j'<j} (1 - \Phi(x_{t-1}\beta_{j'})) = \omega_j(t)$$

Then the posterior for $\{z_j(t)\}_{t,j}$ is written as

$$z_j(t)|\xi_t \sim \begin{cases} \mathcal{N}(x_{t-1}\beta_j, 1)1_{\mathbb{R}^-} & \text{for } j < \xi_t \\ \mathcal{N}(x_{t-1}\beta_j, 1)1_{\mathbb{R}^+} & \text{for } j = \xi_t \end{cases} \tag{5.21}$$

where $\mathcal{N}(\mu, v)1_A$ denotes the truncated normal distribution over A.

Then conditional on the augmented variables, we update $\{\beta_j\}_{j=1}^K$:

$$\begin{aligned} \beta_j|z_{j,1:T} &\sim \mathcal{N}(\bar{\beta}_j, B_j^{-1}) \\ \bar{\beta}_j &= B_j^{-1}\left(\sum_{t|\xi_t \geq j} x_{t-1}z_j(t) + B_0^{-1}\beta_0\right) \\ B_j &= \left(\sum_{t|\xi_t \geq j} x_{t-1}x'_{t-1} + B_0^{-1}\right) \end{aligned} \tag{5.22}$$

Step 4 Updating GARCH parameters: Assuming $p(\alpha, \delta, \gamma, \eta)$ as the prior for GARCH parameters, the posterior will be written as

$$p(\alpha, \delta, \gamma, \eta|y_{1:T}) \propto p(\alpha, \delta, \gamma, \eta) \prod_{t=1}^{T} N(y_t|x_{t-1}\mu_{\xi_t}, \sigma_t^2 \upsilon_{\xi_t}^2) \tag{5.23}$$

which is not of standard form, and we apply a Metropolis-Hastings sampler.

5.1.3 Predictive Density Estimation

The predictive density is defined as the conditional distribution of y_{T+1} given $y_{1:T}$ where all parameter uncertainty has been integrated out. This is the key quantity used in forecasting and model comparison.

At each iteration g of the MCMC algorithm, we have the following set of sampled parameters

$$\{(\mu_j^{(g)}, \nu_j^{(g)}), \beta_j^{(g)}, \{z_j^{(g)}(t)\}_{t=1}^{T}\}_{j=1}^{K^{(g)}}, \quad \{\xi_t^{(g)}, u_t^{(g)}, \sigma_t^{(g)}\}_{t=1}^{T}, \qquad g = 1, ..., M$$

where M is the number of the MCMC iterations. At each iteration g of the algorithm, a draw of $G_{T+1}|y_{1:T}$ can be written as

$$G_{T+1}^{(g)} = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}(T+1)\delta_{(\mu_j^{(g)}, \nu_j^{(g)})} + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}(T+1)\right) G_0(\mu, \nu^2),\ \ g = 1, ..., M \quad (5.24)$$

Combining this with the normal kernel gives us the predictive density of y_{T+1} at each iteration g as

$$\begin{aligned} p(y_{T+1}|y_{1:T}, G_{T+1}^{(g)}) = & \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}(T+1)\mathcal{N}(y_{T+1}|x_T\mu_j^{(g)}, (\sigma_{T+1}^2)^{(g)}\nu_j^{(g)}) \\ & + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}(T+1)\right) \int \mathcal{N}(y_{T+1}|x_T\mu, (\sigma_{T+1}^2)^{(g)}\nu^2)G_0(\mu, \nu^2)d\mu d\nu \end{aligned} \quad (5.25)$$

At each iteration, y_{T+1} has a normal density with parameters specific to the component j with weight $\omega_j^{(g)}(T+1)$, $j = 1, \ldots, K^{(g)}$ or follows a Normal density with new parameters (μ, ν^2) drawn from the base measure, G_0, with weight $1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}(T+1)$. The predictive density of y_{T+1} is estimated by averaging Equation (5.25) over all iterations of the MCMC algorithm.

$$p(y_{T+1}|y_{1:T}) = \frac{1}{M}\sum_{g=1}^{M} p(y_{T+1}|y_{1:T}, G_{T+1}^{(g)}) \tag{5.26}$$

5.2 Simulation Results

In this section, the proposed model is tested on simulated data. The primary goal is to assess how well the model can estimate the time-varying weights of the mixture density over time.

First, we simulate σ_t^2 for $t = 1, 2, ..., 500$ from a univariate GARCH(1,1) process, $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$, with $\sigma_0^2 = 0.01$ and parameters $\alpha = 0$, $\delta = 0.005$, $\gamma = 0.90$, and $\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0, 0.0003)$. Then, the data ($y_t$, $t = 1, ...500$) is simulated according to a mixture of two Normal densities.

$$\mathcal{N}(-0.4, 0.2\sigma_t^2),\ \mathcal{N}(0.5, 0.3\sigma_t^2) \tag{5.27}$$

To draw the data, y_ts, from the mixture density, we need to have $\omega(t) = (\omega_1(t), \omega_2(t))$ for all t. To this end, we first generate the vector $x_{t-1} = (1, x_{t-1}^*)$ for all t where x_t^*s are from a Guassian density. Then, the 2-vectors β_1 and β_2 are generated from Guassian densities, and the weights are calculated using the stick-breaking structure given in Equation (5.11). After normalizing the weights so that $\omega_1(t) + \omega_2(t) = 1$ for all ts, we generate our sample y_t, $t = 1, ...500$ from a mixture of two normal densities given in Equation (5.27). This process produces data observations (x_t, y_t), $t = 1, ..., 500$ to which the proposed model ($\mathcal{M}_5$) is applied, where x_{t-1}s are used as values of the predictors over time to forecast the time-varying weights.

The analysis reported here is based on 15000 iterations of the MCMC algorithm. The first 5000 draws were dropped as burn-in and the following 10000 used for inference. The average acceptance rate of the GARCH parameters is about 30%. Table 5.3 displays the posterior mean and 95% density intervals (DI) for the GARCH parameters in proposed model, $\mathcal{M}_5$. The posterior mean of the number of components in the mixture used to estimate the unknown density is 2.65.

The proposed infinite mixture with time-varying weights, $\mathcal{M}_5$, can potentially have an infinite number of mixture components. But as the estimated weights in Figure 5.2 show, only two components with weights significantly different than zero are recognized. The left (right) plots illustrate the estimated and true weights of the first three components with the highest weights for times that have been generated from the first (second) mixture component in Equation (5.27). Note that the true weights of the third components are zero. As we see in Figure 5.2, the weight forecasts using the proposed model capture the truth well.

If instead we apply the infinite mixture model with constant weights, $\mathcal{M}_3$, to the simulated data, the weights are constant over time. Figure 5.3 illustrates the estimated and true weights of the first three components with the highest weights for times that have been generated from the first (left plots) and second (right plots) mixture component. As we see, the estimated weights are constant over time and consequently the same in the left and right plots. The estimated weights are far from the true weights.

This study shows how the infinite mixture model with time-varying weights, $\mathcal{M}_5$, is able to forecast the true weights of the mixture density over time using the predictor variables if they are useful in determining the weights.

5.3 Empirical Results

5.3.1 Data

In our empirical analysis, we use monthly data on stock returns and a set of predictors studied in Welch and Goyal (2008) from 1927 Jan to 2015 Dec (1067 observations). We use monthly data for S&P 500 returns (continuously compounded including dividends) minus

3-month T-bill as the market excess return. The annual mean and variance of the excess market returns are 6.02% and 42.26%, respectively, for the full sample.

- Stock Variance (**svar**): Sum of squared daily returns on the S&P 500.
- Earnings Price Ratio (**e/p**): The difference between the log of earnings and the log of prices.
- Dividend Payout Ratio (**d/e**): The difference between the log of dividends and the log of earnings.
- Dividend Price Ratio (**d/p**): The difference between the log of dividends and the log of prices.
- Dividend Yield (**d/y**): The difference between the log of dividends and the log of lagged prices.
- Book-to-Market Ratio (**b/m**): The ratio of book value to market value for the Dow Jones Industrial Average.
- Treasury Bill (**tbl**)
- Long Term Yield (**lty**)
- Long Term Spread (**tms**): The difference between the long term yield and treasury bill.
- Default Yield Spread (**dfy**): The difference between BAA yield and AAA yield.
- Inflation (**infl**)
- Net Equity Expansion (**ntis**): The ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks.

- A Kitchen Sink Regression (all): This includes all the aforementioned variables.
- Changes in industrial production (IP)

5.3.2 Estimates

The following priors are used in estimation. In linear regression models, $\mu \sim \mathcal{N}(0, 0.1I_p)$. For GARCH parameters, we set $(\alpha, \delta, \gamma, \eta) \sim \mathcal{N}(0, 100I_4)$. In the probit stick-breaking models, for the hyper-parameters of the base measure G_0, we set $\mu_0 = 0_p$, $V_0 = 10I_p$, $v_0 = 2$, and $s_0 = 8$. We set $\mathcal{N}(0, 0.01I_p)$ as the prior for the vector of coefficients in the probit transformation, β_js, reducing the effect of less informative predictors.

Table 5.4 reports posterior mean and 95% density intervals for the GARCH parameters in the proposed model, $\mathcal{M}_5$, when we use different predictors. For example, the first row correspond to the case where we use the stock variance (svar) as the only predictor; $x_t = (1 \;\; x_t^*) = (1 \;\; \text{svar}_t)$. There is clear evidence of GARCH-type heteroskedasticity in the data. This table also reports the posterior mean of K, the number of alive components in the mixture used to estimate the unknown predictive density. On average, the density of market returns is estimated using about 5-9 distinct components. Estimates of η are consistently positive indicating a larger response to the conditional variance from negative shocks.

5.3.3 Out-of-sample Point Forecasts of the Conditional Mean

In this section, we examine the out-of-sample performance of the proposed model (ability to predict the conditional mean) and compare it with the out-of-sample performance of the benchmark models. The out-of-sample point forecasts are generated using a 30-year rolling window of the observations.

Table 5.5 displays the out-of-sample root mean squared forecasting error of the infinite mixture model $\mathcal{M}_5$ with time-varying weights and predictors in the conditional mean divided by that obtained by the candidates. The candidates are the benchmark models ($\mathcal{M}_0$ to $\mathcal{M}_3$) and the proposed model when the predictors do not enter in the conditional mean ($\mathcal{M}_4$). The results cover the out-of-sample period 1970-Jan to 2015-Dec (540 observations). Values less than one indicate that $\mathcal{M}_5$ outperforms the candidate, producing a lower root mean squared forecasting error. For example, the second column compares the out-of-sample performance of the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, with that of time-varying mean and constant variance model, $\mathcal{M}_1$. In the second column, each row compares $\mathcal{M}_5$ with $\mathcal{M}_1$ when we use a specific predictor in both models. For example, the second row corresponds to the case where we use dividend payout ratio (d/e) as the only predictor; $x_t = (1 \;\; x_t^*) = (1 \;\; d_t/e_t)$. As we see, there is little or no difference in forecasting power of the two models when we focus on the predictability of the conditional mean using the same predictor in both models. Comparing $\mathcal{M}_5$ with other candidates leads to the same conclusion.

Figures 5.4 and 5.5 demonstrate the time series of the cumulative out-of-sample root mean squared forecasting errors of the probit stick-breaking model with time-varying weights, $\mathcal{M}_5$, divided by that of the constant mean and variance model, $\mathcal{M}_0$, for different predictors. Looking at these figures, we are able to diagnose months with better (decreasing graph) or worse (increasing graph) relative performance of $\mathcal{M}_5$ over the out-of-sample forecasting period. In general, all figures show increasing pattern, approaching one after 1990 (except for the inflation). This translates into no strong preference to any of the models, $\mathcal{M}_5$ or $\mathcal{M}_0$. For the inflation case, the time series pattern illustrates superior performance of $\mathcal{M}_5$ relative to $\mathcal{M}_0$ most of the times throughout the out-of-sample period.

Comparison of the models based on the root mean squared forecasting errors focuses on the accuracy of point forecasts and the predictability of the conditional mean. In the next section, instead of studying the predictive mean, we compare different models based on their ability to forecast the entire density of market returns.

5.3.4 Predictive Likelihood

To compare the performance of the models, we calculate each model's predictive likelihood. The predictive likelihood for $y_{L:T}$, $L < T$ is expressed in terms of the one-step-ahead predictive likelihoods,

$$m(y_{L:T}|y_{1:L-1}, \mathcal{M}) = \Pi_{t=L}^{T} p(y_t|y_{1:t-1}, \mathcal{M}) \tag{5.28}$$

where $\mathcal{M}$ denotes a particular model, and $L > 1$ is chosen to eliminate the influence of the priors on model comparison. The model with the larger value of predictive likelihood is the one most consistent with the data. We can approximate the one-step-ahead predictive likelihoods, $p(y_t|y_{1:t-1}, \mathcal{M})$, by averaging the data density over draws of the unknown parameters conditional on the data history $y_{1:t-1}$. This integrates out parameter and distributional uncertainty as

$$\begin{aligned} p(y_t|y_{1:t-1}, \mathcal{M}) &= \int p(y_t|\Theta, y_{1:t-1}, \mathcal{M}) p(\Theta|y_{1:t-1}, \mathcal{M}) d\Theta \\ &\approx \frac{1}{M} \sum_{g=1}^{M} p(y_t|\Theta^{(g)}, y_{1:t-1}, \mathcal{M}) \end{aligned} \tag{5.29}$$

where Θ denotes the model parameters, $\Theta^{(g)}$ is a posterior draw from $p(\Theta|y_{1:t-1}, \mathcal{M})$, and $p(y_t|\Theta^{(g)}, y_{1:t-1}, \mathcal{M})$ is the data density given $\Theta^{(g)}$ and $y_{1:t-1}$ for model $\mathcal{M}$. Note that we are able to compute σ_t at each iteration of the MCMC since we have σ_{t-1} and the GARCH parameters: $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$.

Based on Equation (5.29), the predictive likelihood for the proposed model is estimated as

$$p(y_t|y_{1:t-1}, \mathcal{M}_5) \approx \frac{1}{M} \sum_{g=1}^{M} \mathcal{N}(y_t|\mu_{\xi_t^{(g)}}^{(g)}, (\sigma_t^{(g)} v_{\xi_t^{(g)}}^{(g)})^2). \tag{5.30}$$

At each iteration g, $\xi_t^{(g)}$ is drawn from one of the $K^{(g)}+1$ components with weights $\omega_j^{(g)}(t)$ $j=1,\ldots,K^{(g)}$ and $1-\sum_{j=1}^{K^{(g)}}\omega_j^{(g)}(t)$. When $\xi_t^{(g)}=K^{(g)}+1$, a new pair of parameters (μ,ν^2) is drawn from the base measure G_0.

Table 5.6 reports the values of the log-predictive likelihoods for different models. Among all models studied here, the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, when we use all predictors (the kitchen sink case), has the largest value of the predictive likelihood and hence is the most consistent with the data.

To compare the proposed model with any of the candidates when we use the same predictors in both models, we compute the log-Bayes factor for the proposed model against the candidate as the ratio of the predictive likelihoods (Equation 5.31). The candidates are the benchmark models ($\mathcal{M}_0$ to $\mathcal{M}_3$) and the proposed model when the predictors are not entered in the conditional mean ($\mathcal{M}_4$).

$$\text{log Bayes factor of } \mathcal{M}_5 \text{ vs } \mathcal{M} = \log\frac{m(y_{L:T}|y_{1:L-1},\mathcal{M}_5)}{m(y_{L:T}|y_{1:L-1},\mathcal{M})} \tag{5.31}$$

Positive values indicate that $\mathcal{M}_5$ produces better density forecasts relative to the candidate model, and a log-Bayes factor bigger than 5 is a very strong evidence for outperformance of $\mathcal{M}_5$ against the candidate model. Table 5.7 reports the values of the log-Bayes factor of the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, versus the candidate model when both models use the same predictor. Each row compares $\mathcal{M}_5$ with the candidate when we use a specific predictor in both models. For example, the second row corresponds to the case where we use dividend payout ratio (d/e) as the only predictor; $x_t=(1\ \ x_t^*)=(1\ \ d_t/e_t)$. This row shows that when we use dividend payout ratio as the only predictor, $\mathcal{M}_5$ dominates all candidate models.

Compared to the constant mean and variance model, $\mathcal{M}_0$, and the time-varying mean model, $\mathcal{M}_1$, the results strongly support the proposed model when we use any of the predictors. Column 4 shows that even after the GARCH effects are accounted for ($\mathcal{M}_2$), the time-varying mixture model, $\mathcal{M}_5$, improves density forecasts.

The log-Bayes factor of $\mathcal{M}_5$ versus $\mathcal{M}_3$ (column 5) shows how allowing the weights of the infinite mixture model to be determined by available information on predictors can improve the performance of the model relative to when we have an infinite mixture model with constant weights over time. Almost in all individual predictive models (except net equity expansion), using the predictor in predicting the weights of the mixture density improves the performance of the model. Our results reveal the stronger predictive ability for the kitchen sink case, the stock variance, the dividend payout ratio, the earning price ratio, and the dividend price ratio.

The log-Bayes factor of $\mathcal{M}_5$ versus $\mathcal{M}_4$ (the last column) shows whether using the predictive variable in predicting the conditional mean of the density in addition to using it in predicting the weights of the mixture density adds any value to the model. In general, the absolute values are smaller than 1 (except for the kitchen sink case), showing no significant preference of the two models. For the kitchen sink case, the difference in log-predictive likelihood is 3.862. This means that using all predictors together in predicting the conditional mean of the density in addition to using them in predicting the weights of the mixture density improves the performance of the model in spite of increasing the model dimension.

Table 5.7 compares different models based on the entire log-predictive likelihood criteria where all predictions of the sample contribute equally to the criteria. In practical applications such as risk management, the ability of the models to predict the extremes might be more interesting. Here, we examine the accuracy of the proposed model in describing the left tail of the return distribution. Concentrating on the extreme events, we define the tail predictive likelihood criteria to compare the proposed model with the benchmarks.

$$m^{\alpha}(y_{L:T}|y_{1:L-1},\mathcal{M}) = \Pi_{t=L}^{T} \frac{p(y_t|y_{1:t-1},\mathcal{M})1_{\{y_t<z_\alpha\}}}{\int_{-\infty}^{z_\alpha} f(v|y_{1:t-1,\mathcal{M}})dv} \tag{5.32}$$

$$\simeq \Pi_{t=L}^{T} \frac{\frac{1}{M}\sum_{g=1}^{M}\mathcal{N}(y_t|\mu_{\xi_t^{(g)}}^{(g)},(\sigma_t^{(g)}v_{\xi_t^{(g)}}^{(g)})^2)1_{\{y_t<z_\alpha\}}}{\frac{1}{M}\sum_{g=1}^{M}\Phi(\frac{z_\alpha-\mu_{\xi_t^{(g)}}^{(g)}}{\sigma_t^{(g)}v_{\xi_t^{(g)}}^{(g)}})} \tag{5.33}$$

where z_α is the lower $\alpha\%$ quantile of the sample. $1_{\{y_t<z_\alpha\}}$ equals 1 if $y_t < z_\alpha$ and equals 0 otherwise. The division by the integration constant $\int_{-\infty}^{z_\alpha} f(v|y_{1:t-1,\mathcal{M}})dv$ normalizes the density on the tail region (Diks et al., 2011), and $\Phi(.)$ denotes the cumulative distribution function for the univariate standard normal.

The second column of Table 5.8 reports the log-Bayes factor of the proposed model, $\mathcal{M}_5$, versus the time-varying mean and constant variance model, $\mathcal{M}_1$. The results are for $L = 100$ and $\alpha = 0.05$. This table illustrates outperformance of the proposed model in predicting the extreme events. To examine whether this ourperformance is related to the time-varying weights or is only because of the GARCH structure of the conditional variance, we also report the log-Bayes factor of the proposed model, $\mathcal{M}_5$, versus the time-varying mean and variance model, $\mathcal{M}_2$, in the third column. The results strongly support $\mathcal{M}_5$ against $\mathcal{M}_2$ which shows using these financial variables in predicting the weights of the mixture improves the predictive ability in the lower tail.

Weights over time

The difference between the infinite mixture model with constant weights and the infinite mixture model proposed in this chapter is that in the latter the weights of the mixture density at each period are potentially predicted by the available information on different financial and macroeconomic variables. Since values of these predictors change over time, we expect the predicted weights to change over time as well provided that these predictors are useful in forecasting the weights of the mixture density.

For the infinite mixture model with constant weights, $\mathcal{M}_3$, the weights do not change over time. Figure 5.6 illustrates the posterior mean of the weights of the first five components of the mixture density with the highest weights over time predicted by $\mathcal{M}_3$. When we estimate $\mathcal{M}_5$ using a predictor x_t^* ($x_t = (1 \;\; x_t^*)$), if the predictor is useful in forecasting the weights, the weights will not be constant over time anymore. For example, Figure 5.7 illustrates the weights over time estimated by $\mathcal{M}_5$ when we use the dividend payout ratio as the predictor, $x_t = (1 \;\; d_t/e_t)$. The predicted variation of weights over time results in a more accurate predictive density.

We illustrate the weights predicted by $\mathcal{M}_5$ for different predictors (Figures 5.7-5.22). These figures show that some of these predictors such as the kitchen sink case, the stock variance, the dividend payout ratio, the earning price ratio, and the dividend price ratio have strong predictive ability while the remaining variables seem not useful in forecasting the time-varying weights, rendering almost constant weights over time. The predicted variation of weights over time results in more accurate predictive densities as we illustrated in Table 5.7.

Density over time

In this section, we illustrate estimates of the predictive density of the market excess return obtained by the proposed model and the benchmarks. The time-varying weights in the proposed model make the mixture density change over time and this variation in density is in addition to the changes caused by the time-varying conditional mean and conditional variance. The infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, is able to reveal changes to the density not captured by the first and second moments which are overlooked by the linear regression models. This is important in empirical distributions in finance and economics where we might have features such as heavy tails and asymmetry. Exploiting these features can make a big difference in applications such as risk management and portfolio selections.

Figure 5.8 illustrates the out-of-sample predictive density of the market excess return at different dates for four different months[3] for the kitchen sink case. The predictive density obtained from three models are illustrated; the proposed model, $\mathcal{M}_5$, the linear regression model with time-varying mean and constant variance, $\mathcal{M}_1$, and the linear regression model with time-varying mean and variance, $\mathcal{M}_2$.

This figure shows how market return density changes over time. When we apply the linear regression model with constant variance, $\mathcal{M}_1$, the density is almost the same over time while allowing time-variation in the volatility, $\mathcal{M}_2$, permits the volatility of the density to change over time. However, the density remains symmetric and bell-shaped. Applying the proposed model, $\mathcal{M}_5$, enables us to capture the asymmetry and fat tails. To facilitate the tail comparisons, we illustrate the log of the probability density functions in the right panels.

This figure also shows how the conditional density derived by different models changes over time. Particularly, the blue density shows how available information on all predictors helps estimate the conditional density. Although the conditional mean is almost the same over time (implying that the predictors do not help forecast the conditional mean), the shape of the density changes.

If we only focus on the point forecasts of the conditional mean, there is not much difference between different models. For example, consider a month in oil shock period (1974-09). Figure 5.9 compares the conditional mean and the conditional density of the market returns derived from the linear regression model, $\mathcal{M}_1$, and the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, when we use all predictors in both models (the kitchen sink case). As we see in the figure, the estimated mean derived from the two models are not significantly different. However, the shape of the estimated

[3]The dates are 1932-09 in the great depression period, 1974-09 in the oil shock period, 2000-08 in the bubble period, and 2009-02.

density is different; the density derived from the linear regression model is symmetric and does not illustrate any fat tails. The density obtained by applying the proposed model shows skewness and positive kurtosis implying that using available information on all predictors in forecasting the weights of the mixture density enables us to capture the skewness and the fatter tails in the predictive density for this month. This shows that these predictive variables are useful in forecasting the entire density, carrying valuable information on the riskiness of the market returns.

Figure 5.10 displays the 5%, 25%, 75%, and 95% conditional quantiles over the full sample period using the estimates of the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, when we use all predictors (the kitchen sink case). The variation of the conditional quantiles over time, particularly when we see asymmetry in the upper and lower quantiles, supports the usefulness of the predictive variables in forecasting the conditional density of market returns. This includes the period following 1941, 1973-1975, 1979-1980, mid-1994, and 1996 where the decrease in the lower quantiles is more than the increase in the upper quantiles, and 1931-1933, 2008-2009, and 1987 when the increase in the upper quantiles is more than the decrease in the lower quantiles. The variation in the quantiles during these periods is on top of what can be captured by time-varying mean and volatility alone. This figure also shows the conditional mean predicted using this model (the solid red line). As we see, the conditional mean is almost constant and equal to zero over time, implying that the predictors are not useful in predicting the mean of the conditional density.

We also examine the higher order moments of the stock return distribution (conditional skewness and kurtosis). Figure 5.11 and 5.12 plot the time series of the conditional skewness and the conditional excess kurtosis of the return distribution, respectively, over the full sample period using the estimates of the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, when we use all predictors (the kitchen sink case). The return distribution is negatively skewed most of the time, and the conditional excess kurtosis is mainly positive. Illustrating the periods with different levels of risk, these figures are informative for investors.

In assessing whether a model helps investors obtain economic value, in most cases, measures of forecasting performance based on the first and second moments contain little information, and, instead, the whole predictive distribution will be more important. In Section 5.3.5, we use the predictive density of models to make portfolio decisions and then compare performance of the proposed and benchmark models based on ex post utility outcomes. The results show the superiority of the proposed model, when used as the basis for investment decisions.

5.3.5 Economic Gains

In the previous sections, we demonstrated how using available information on certain financial variables can improve the predictive density of market returns statistically. Now, we investigate the economic improvement achieved by this predictability. We evaluate the economic significance of the predictability of each model with two practical applications. First, we use each model to find the optimal weights of a portfolio of two assets, the market portfolio and the risk-free asset, and investigate the utility improvement produced by each model for an investor who maximizes the expected utility to make her portfolio decisions. Second, we carry out an experiment to illustrate the market timing ability of each model. The results show the significant market timing ability of the proposed model as well as its considerable utility improvement power when employed in portfolio selection decisions.

Utility Improvement

Consider an asset allocation problem where the investor allocates his money to a portfolio of two assets: the market portfolio and the risk-free asset. At each time t, the investor obtains the optimal weight assigned to the market portfolio, w_t^*, by maximizing her expected utility. w_t^* is derived from

$$w_t^* = arg \max_{w_t} E[U(w_t, y_{t+1})|I_t, \mathcal{M}] \tag{5.34}$$

where I_t is the available information at time t, and $\mathcal{M}$ denotes a particular model. $1 - w_t^*$ is the optimal weight for risk-free asset.

When we have the full predictive density of y_{t+1}, w_t^* can be easily approximated as follows assuming the right hand side of Equation (5.34) exists

$$w_t^* \approx arg \max_{w_t} \frac{1}{M} \sum_{m=1}^{M} U(w_t, y_{t+1}^m) \tag{5.35}$$

where $y_{t+1}^m,\ m = 1, ..., M$ are M draws from the predictive density estimated by model $\mathcal{M}$. This is particularly useful when the investor has utility functions such as power utility for which there is no closed form expression for $E[U(w_t, y_{t+1})|I_t, \mathcal{M}]$. In order to limit the leverage more than 100% and avoid short-selling, we constrain the weight on the market portfolio by imposing $0 \leq w_t \leq 2$. The optimal weight at each time, w_t^*, can be found numerically.

To evaluate the economic gains obtained by the proposed model, $\mathcal{M}_5$, over the constant mean model, $\mathcal{M}_0$, we calculate Δ; the maximum monthly return the investor would be willing to give up for the economic benefit obtained by switching from $\mathcal{M}_0$ to $\mathcal{M}_5$.

$$\sum_{t=t_s}^{t_e} U(W_{t+1}^{re,\mathcal{M}_0}) = \sum_{t=t_s}^{t_e} U(\exp^{-\Delta} W_{t+1}^{re,\mathcal{M}_5}). \tag{5.36}$$

where t_s to t_e is the out-of-sample period. $W_{t+1}^{re,\mathcal{M}_0}$ and $W_{t+1}^{re,\mathcal{M}_5}$ are the realized wealth of the investor at time $t+1$ using the optimal weights determined at time t using $\mathcal{M}_0$ and $\mathcal{M}_5$, respectively.

$$W_{t+1}^{re} = w_t^* \exp(r_{ft} + y_{t+1}^{re}) + (1 - w_t^*) \exp(r_{ft}) \tag{5.37}$$

where y^{re}_{t+1} is the realized value of market excess return at time $t+1$. If $\Delta > 0$, then $\mathcal{M}_5$ is preferred to $\mathcal{M}_0$.

We consider three types of utility functions:

- Power Utility (CRRA)

$$U(W_{t+1}) = \frac{W_{t+1}^{1-\tau}}{1-\tau}$$

  In the case of the power utility function, we need to impose the restriction $0 \leq w_t \leq 1$ to avoid unbounded values for the expected utility (Geweke 2001).

- Exponential Utility(CARA)

$$U(W_{t+1}) = \frac{-1}{\tau} \exp^{-\tau(W_{t+1})}$$

- Quadratic Utility

$$U(W_{t+1}) = W_{t+1} - \frac{\tau}{2(1+\tau)} W_{t+1}^2$$

where $W_{t+1} = w_t \exp(r_{ft} + y_{t+1}) + (1 - w_t) \exp(r_{ft})$, and τ denotes the coefficient of investor's risk aversion.

Table 5.9 displays the annualized basis point fee an investor would be willing to pay to switch from $\mathcal{M}_0$ to $\mathcal{M}_5$. The out-of-sample period is from Jan 1985 to Dec 2015 (360 observations). Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations. Positive numbers show that $\mathcal{M}_5$ outperforms the constant mean model in terms of generating economic benefits. In most cases, an investor is willing to pay to switch from the constant mean model to the proposed model. This table also shows that a more risk-averse investor (higher τ) is ready to pay more to switch to the proposed model; exploiting available information on the financial variables to extract information about the riskiness of the market returns and shape of the market returns predictive density is more valuable for an investor with higher degree of risk aversion. The more risk averse investors care more about characteristics of the predictive densities other than the first and second moments. They value more the models that take into account the entire density and render better estimations of the riskiness of the assets.

Table 5.10 and 5.11 compare the economic gains of the proposed model, $\mathcal{M}_5$, to the time-varying mean and constant variance model, $\mathcal{M}_1$, and the time-varying mean and variance model, $\mathcal{M}_2$, respectively. Almost in all cases, the investor is willing to pay to switch to the proposed model.

Market Timing Portfolio

In another attempt to evaluate the economic improvement achieved by the proposed model, we implement a simple investment strategy (market timing). The one-month-ahead forecast of excess market returns (r^m_{t+1}) is compared to a target return (for example $r_{\text{target}} = 10\%$). If $\hat{P}(r^m_{t+1} > r_{\text{target}}) > p_{\text{target}}$, then the investor will invest only in the market in the following month ($t+1$), obtaining the actual market return in that period. However, if $\hat{P}(r^m_{t+1} > r_{\text{target}})$ is less than p_{target}, then the investor will invest only in the risk-free asset in the following month, obtaining the actual risk-free return. This produces a time series of realized investment returns (a portfolio of returns). We carry out this experiment for each forecasting model and examine the ability of the models to time the market by studying the performance of the portfolios obtained by each model.

The evaluation of portfolios obtained from different models is based on risk-return trade-off approach and utility comparison. The utility function used here is the quadratic utility with the risk aversion coefficient $\tau = 5$.

$$U(W_{t+1}) = W_{t+1} - \frac{\tau}{2(1+\tau)} W^2_{t+1}$$

To compare the performance of two models, $\mathcal{M}_a$ and $\mathcal{M}_b$, we calculate the performance fee (Δ) that an investor would pay to switch from the portfolio obtained by model a to the portfolio obtained by model b.

$$\sum_{t=t_s}^{t_e} U(W_{t+1}^{\mathcal{M}_a}) = \sum_{t=t_s}^{t_e} U(\exp^{-\Delta} W_{t+1}^{\mathcal{M}_b}). \tag{5.38}$$

where t_s to t_e is the out-of-sample period. $W_{t+1}^{\mathcal{M}} = e^{r_{t+1}}$ are the realized wealth of the investor at time $t+1$ using the forecasts obtained by Model $\mathcal{M}$. We also calculate the Sharp ratios for each model.

Table 5.12 reports the performance of market timing portfolios with $p_{\text{target}} = 0.5$ and $r_{\text{target}} = 0$. The out-of-sample market timing is conducted from Jan 1960-Dec 2015. The values reported in this table are annualized monthly values. Column 4 shows the values of the Sharp ratios, and the values in the last column are annualized basis point performance fees that an investor is willing to pay to switch from the model in the corresponding row to $\mathcal{M}_5$. From this table, it can be seen that the portfolio obtained by the proposed model yields higher average return, lower variance, and higher Sharpe ratio.

In order to truly examine the superiority of the proposed model in market timing, transaction costs must be taken into account. However, the emergence of ETFs has allowed for low transaction costs when switching between a market proxy portfolio (ETF) and a bond (possibly an ETF as well). Table 5.12 also shows the Sharpe ratios and Δ obtained by different models after accounting for transaction costs. Transaction costs are subtracted from monthly returns depending on whether stocks (market) or the risk free asset (T-bills) are chosen for that particular month. For example, if stocks are chosen in month t, then the adjusted return for t after accounting for transaction costs is $r_t \times (1 - c)$, where c is the appropriate cost for stocks. Similar analysis is done for the risk-free asset. Following Pesaran and Timmermann (1994), we use a cost of 0.5% for transaction costs of stocks, and a cost of 0.1% for the risk free asset. From Table 5.12, it can be seen that the Sharpe ratios and the annual fees are not significantly different after accounting for transaction costs.

5.4 Conclusion

In this chapter, we propose a model to study the one-period-ahead out-of-sample predictability of the US stock return density using financial and macroeconomic variables. In contrast to the extant literature that focuses on the point forecasts of the conditional mean, we examine the predictability of the entire density which can be valuable for asset allocation and risk management. We consider a Bayesian nonparametric mixture model that allows the mixing distribution to change with time. In the proposed model, the weights of the mixture are constructed as probit transformations of a linear combination of the predictors.

We compare statistical and economic measures of forecasting performance of the proposed model with a set of benchmark models. Despite little or no improvement in point forecasts, certain variables display significant out-of-sample predictive ability with respect to the predictive density of market returns and increase economic value for investors when employed in portfolio decisions. A risk-averse investor is willing to pay a performance fee to switch from the benchmark models to the infinite mixture model proposed in this chapter. We also illustrate how these financial and macroeconomic variables are useful in predicting the conditional density, the conditional moments and the conditional quantiles.

Model	Description	
	Benchmark Models	
$\mathcal{M}_0$	Constant mean and constant variance	$y_t = \mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_1$	Time-varying mean and constant variance	$y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_2$	Time-varying mean and variance	$y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$
$\mathcal{M}_3$	Infinite mixture with constant weights	$y_t = \mu_t + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \quad (\mu_t, \nu_t^2) \sim G$
	Proposed Models	
$\mathcal{M}_4$	Infinite mixture with time-varying weights	$y_t = \mu_t + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \quad (\mu_t, \nu_t^2) \sim G_t$
$\mathcal{M}_5$	Infinite mixture with time-varying weights and predictors in the conditional mean	$y_t = x_{t-1}\mu_t + e_t, \quad e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2)$ $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \quad (\mu_t, \nu_t^2) \sim G_t$

Table 5.2: List of the benchmark models and the proposed models. In all models, y_t is the market excess return and x_t is a row vector of the predictors (financial or macroeconomic variables) including 1 for the intercept.

	GARCH parameters: probit stick-breaking model with time-varying weights, $\mathcal{M}_5$.	
Parameter	Post. Mean	95% DI
α	0.0033	$(0.00, 0.009)$
δ	0.005	$(0.00, 0.010)$
γ	0.940	$(0.835, 0.947)$
η	0.518	$(0.111, 0.945)$

Table 5.3: This table displays posterior mean and 95% density intervals (DI) for the parameters of probit stick-breaking model with time-varying weights, $\mathcal{M}_5$, for simulated data.

Predictor (x_t^*)	α	δ	γ	η	K
Stock Variance	0.000	0.026	0.801	0.127	8.859
	$(0.000, 0.001)$	$(0.008, 0.057)$	$(0.725, 0.841)$	$(0.088, 0.176)$	$(6.542, 10.145)$
Dividend Payout Ratio	0.001	0.052	0.792	0.091	7.635
	$(0.000, 0.002)$	$(0.024, 0.094)$	$(0.718, 0.844)$	$(0.035, 0.141)$	$(5.895, 9.745)$
Long Term Spread	0.001	0.038	0.801	0.101	8.709
	$(0.000, 0.002)$	$(0.012, 0.068)$	$(0.721, 0.865)$	$(0.043, 0.162)$	$(6.312, 11.012)$
Default Yield Spread	0.001	0.048	0.794	0.107	5.046
	$(0.000, 0.002)$	$(0.021, 0.096)$	$(0.746, 0.856)$	$(0.049, 0.189)$	$(4.175, 6.953)$
Dividend Price Ratio	0.01	0.054	0.799	0.089	5.849
	$(0.000, 0.002)$	$(0.023, 0.092)$	$(0.699, 0.868)$	$(0.041, 0.138)$	$(3.859, 7.425)$
Book-to-Market Ratio	0.000	0.046	0.775	0.089	7.699
	$(0.000, 0.001)$	$(0.013, 0.115)$	$(0.699, 0.861)$	$(0.042, 0.134)$	$(7.425, 9.478)$
Earnings Price Ratio	0.000	0.053	0.835	0.102	6.497
	$(0.000, 0.001)$	$(0.012, 0.134)$	$(0.787, 0.874)$	$(0.037, 0.198)$	$(4.845, 8.024)$
Dividend Yield	0.001	0.057	0.807	0.083	5.861
	$(0.000, 0.002)$	$(0.023, 0.123)$	$(0.726, 0.857)$	$(0.024, 0.146)$	$(4.124, 7.451)$
Inflation	0.000	0.035	0.768	0.077	9.38
	$(0.000, 0.001)$	$(0.012, 0.133)$	$(0.714, 0.823)$	$(0.026, 0.155)$	$(7.152, 11.415)$
Net Equity Expansion	0.000	0.032	0.815	0.112	9.071
	$(0.000, 0.001)$	$(0.011, 0.086)$	$(0.743, 0.877)$	$(0.051, 0.192)$	$(7.182, 10.856)$
IP	0.001	0.051	0.749	0.094	7.307
	$(0.000, 0.002)$	$(0.017, 0.089)$	$(0.689, 0.824)$	$(0.042, 0.151)$	$(5.124, 9.047)$
Kitchen sink case	0.000	0.024	0.838	0.098	7.164
	$(0.000, 0.0000)$	$(0.014, 0.040)$	$(0.818, 0.858)$	$(0.071, 0.126)$	$(5.658, 9.175)$

Table 5.4: This table displays posterior mean and 95% density intervals for the GARCH parameters and the number of distinct mixture components in the probit stick-breaking model with time-varying weights and the predictors in the conditional mean, $\mathcal{M}_5$, for different predictors. For example, the first row correspond to the case where we use the stock variance (svar) as the only predictor; $x_t = (1 \;\; x_t^*) = (1 \;\; \text{svar}_t)$. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

Predictor (x_t^*)	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
Stock Variance	1.002	1.001	1.002	1.002	1.002
Dividend Payout Ratio	1.001	1.001	1.001	1.001	0.999
Long Term Spread	1.002	1.001	1.001	1.002	1.002
Default Yield Spread	1.000	0.999	1.000	0.999	0.999
Dividend Price Ratio	0.999	1.000	1.000	0.999	0.999
Book-to-Market Ratio	1.003	1.001	1.001	1.003	1.002
Earnings Price Ratio	0.999	1.000	1.000	0.999	0.998
Dividend Yield	1.001	1.001	1.001	1.000	0.999
Inflation	0.994	1.001	1.002	0.994	0.994
Net Equity Expansion	1.003	1.000	1.001	1.002	1.001
Changes in Industrial Production	0.989	1.001	0.997	0.999	1.000
Kitchen sink case	1.002	1.004	1.004	1.002	1.001

Table 5.5: This table displays the out-of-sample root mean squared forecasting error of the infinite mixture with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, divided by that of the candidate. Numbers less than 1 show that $\mathcal{M}_5$ outperforms the candidate. Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec and the out-of-sample period covers 1970-Jan to 2015-Dec.

$\mathcal{M}_0$: Constant mean and variance: $y_t = \mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_2$: Time-varying mean and variance: $y_t = x_{t-1}\mu + e_t, \ e_t \sim \mathcal{N}(0, \sigma_t^2), \ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

$\mathcal{M}_3$: Infinite mixture with constant weights: $y_t = \mu_t + e_t, \ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2), \ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \ (\mu_t, \nu_t^2) \sim G$

$\mathcal{M}_4$: Infinite mixture with time-varying weights: $y_t = \mu_t + e_t, \ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2), \ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \ (\mu_t, \nu_t^2) \sim G_t$

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t, \ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2), \ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2, \ (\mu_t, \nu_t^2) \sim G_t$

Predictor (x_t^*)	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
Stock Variance	−49.701	−50.175	86.419	100.038	108.308	107.311
Dividend Payout Ratio	−49.701	−51.321	85.460	100.038	106.348	107.259
Long Term Spread	−49.701	−50.310	83.189	100.038	102.986	101.301
Default Yield Spread	−49.701	−50.403	84.681	100.038	101.053	101.619
Dividend Price Ratio	−49.701	−50.588	85.286	100.038	105.409	105.863
Book-to-Market Ratio	−49.701	−51.239	83.396	100.038	104.516	104.012
Earnings Price Ratio	−49.701	−50.126	85.438	100.038	106.045	105.524
Dividend Yield	−49.701	−50.551	85.031	100.038	104.403	104.444
Inflation	−49.701	−48.961	92.665	100.038	100.084	100.184
Net Equity Expansion	−49.701	−48.907	86.946	100.038	98.655	97.848
Changes in Industrial Production	−49.701	−52.132	84.452	100.038	102.104	103.025
Kitchen sink case	−49.701	−50.798	93.246	100.038	107.384	111.246

Table 5.6: This table displays the log predictive likelihood of the differnt models (with L=300). Positive values show that $\mathcal{M}_5$ outperforms the candidate. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_0$: Constant mean and variance: $y_t = \mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_2$: Time-varying mean and variance: $y_t = x_{t-1}\mu + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$
$\mathcal{M}_3$: Infinite mixture with constant weights: $y_t = \mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G$
$\mathcal{M}_4$: Infinite mixture with time-varying weights: $y_t = \mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

Predictor (x_t^*)	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
Stock Variance	157.012	157.486	20.892	7.273	−0.997
Dividend Payout Ratio	156.960	158.580	21.799	7.221	0.911
Long Term Spread	151.002	151.611	18.112	1.263	−1.685
Default Yield Spread	151.320	152.022	16.938	1.581	0.566
Dividend Price Ratio	155.564	156.451	20.576	5.825	0.453
Book-to-Market Ratio	153.714	155.252	20.616	3.974	−0.503
Earnings Price Ratio	155.226	155.651	20.086	5.486	−0.520
Dividend Yield	154.145	154.995	19.412	4.406	0.040
Inflation	149.886	149.146	7.519	0.146	0.101
Net Equity Expansion	147.550	146.756	10.902	−2.189	−0.806
Changes in Industrial Production	152.720	155.157	18.573	2.987	0.921
Kitchen sink case	160.947	162.044	18.000	11.208	3.862

Table 5.7: This table displays the log predictive likelihood of the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, minus log predictive likelihood of the candidate (with L=300). Positive values show that $\mathcal{M}_5$ outperforms the candidate. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_0$: Constant mean and variance: $y_t = \mu + e_t,\ \ e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t,\ \ e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_2$: Time-varying mean and variance: $y_t = x_{t-1}\mu + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$
$\mathcal{M}_3$: Infinite mixture with constant weights: $y_t = \mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G$
$\mathcal{M}_4$: Infinite mixture with time-varying weights: $y_t = \mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

Predictors	$\mathcal{M}_1$	$\mathcal{M}_2$
Stock Variance	40.396	34.925
Dividend Payout Ratio	42.174	37.419
Long Term Spread	41.788	37.195
Default Yield Spread	41.353	37.129
Dividend Price Ratio	42.704	37.127
Book-to-Market Ratio	41.364	36.993
Earnings Price Ratio	41.286	35.950
Dividend Yield	42.404	37.129
Inflation	43.436	35.509
Net Equity Expansion	42.050	36.672
Changes in Industrial Production	40.351	35.221
Kitchen sink case	41.187	36.361

Table 5.8: This table displays the log of tail predictive likelihood of the infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, minus the log of tail predictive likelihood of the candidate, with L=100 and $\alpha = 0.05$. Positive values show that $\mathcal{M}_5$ outperforms the candidate. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t,\ \ e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_2$: Time-varying mean and variance: $y_t = x_{t-1}\mu + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

	CRRA			CARA			Quadratic		
Predictor	$\tau = 2$	$\tau = 5$	$\tau = 10$	$\tau = 2$	$\tau = 5$	$\tau = 10$	$\tau = 2$	$\tau = 5$	$\tau = 10$
Stock Variance	−28.116	−5.752	**53.915**	−33.206	−5.810	**64.728**	−31.605	**1.593**	**66.021**
Dividend Payout Ratio	−9.493	**31.029**	**60.871**	**7.497**	**33.146**	**86.383**	**5.958**	**42.342**	**103.711**
Long Term Spread	−32.232	**0.083**	**59.043**	−31.992	**0.484**	**217.134**	−31.235	**5.879**	**77.516**
Default Yield Spread	**0.771**	**26.295**	**78.749**	**0.937**	**26.794**	**93.609**	**1.216**	**29.722**	**90.900**
Dividend Yield	−8.070	**12.716**	**69.832**	**0.618**	**38.481**	**253.709**	**6.482**	**61.446**	**128.024**
Dividend Price Ratio	−7.439	**16.727**	**73.647**	**0.544**	**39.354**	**275.707**	**6.449**	**62.519**	**129.722**
Earnings Price Ratio	−5.218	**20.945**	**69.279**	**3.334**	**39.985**	**261.037**	**7.905**	**61.405**	**128.564**
Book-to-Market Ratio	−3.838	**31.847**	**75.432**	**4.146**	**34.310**	**166.117**	**4.756**	**45.885**	**111.569**
Inflation	−0.738	**16.148**	**50.475**	−0.774	**16.519**	**50.314**	−0.587	**17.881**	**59.521**
Net Equity Expansion	−18.541	**4.244**	**46.251**	−18.482	**4.068**	**57.519**	−18.251	**6.396**	**55.652**
Δ in Industrial Production	−25.213	**0.258**	**61.124**	−26.195	**1.241**	**229.352**	−28.723	**6.123**	**79.257**
Kitchen sink case	−26.123	**15.764**	**61.544**	−19.889	**19.136**	**393.716**	−18.153	**36.484**	**110.919**

Table 5.9: This table displays the annualized basis point fee an investor would be willing to pay to switch from the constant mean and variance model, $\mathcal{M}_0$, to the proposed model, $\mathcal{M}_5$. Positive numbers show that $\mathcal{M}_5$ outperforms the constant mean model in terms of generating economic benefits. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec and the out-of-sample period is from Jan-1985 to Dec-2015. Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations.

$\mathcal{M}_0$: Constant mean and variance: $y_t = \mu + e_t,\ \ e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

	CRRA			CARA			Quadratic		
Predictor	$\tau = 2$	$\tau = 5$	$\tau = 10$	$\tau = 2$	$\tau = 5$	$\tau = 10$	$\tau = 2$	$\tau = 5$	$\tau = 10$
Stock Variance	**1.412**	**28.168**	**105.928**	−3.585	**28.257**	**43.191**	−2.042	**32.660**	**121.657**
Dividend Payout Ratio	−3.325	**37.264**	**72.407**	**13.685**	**39.344**	**107.882**	**12.088**	**48.478**	**116.322**
Long Term Spread	**27.321**	**34.132**	**59.083**	**17.599**	**33.774**	**109.121**	**15.435**	**33.741**	**63.416**
Default Yield Spread	**13.179**	**36.907**	**84.501**	**13.357**	**37.833**	**64.884**	**13.562**	**39.880**	**94.944**
Dividend Yield	**11.756**	**29.280**	**80.019**	**20.478**	**55.182**	**260.879**	**26.203**	**77.413**	**136.878**
Dividend Price Ratio	**10.988**	**31.966**	**82.881**	**18.998**	**54.313**	**282.878**	**24.777**	**77.211**	**137.747**
Earnings Price Ratio	**7.684**	**30.710**	**72.204**	**16.241**	**49.619**	**268.208**	**20.709**	**70.531**	**129.730**
Book-to-Market Ratio	−7.007	**27.839**	**72.385**	**0.963**	**30.435**	**108.598**	**1.505**	**41.568**	**108.522**
Inflation	**7.639**	**26.051**	**70.488**	**7.618**	**26.492**	**50.314**	**7.804**	**28.038**	**83.151**
Net Equity Expansion	−7.475	**6.906**	**37.381**	−7.510	**6.718**	**4.117**	−7.377	**7.928**	**43.645**
Δ in Industrial Production	−2.123	**39.122**	**73.184**	**14.125**	**40.254**	**108.214**	**14.014**	**49.978**	**117.925**
Kitchen sink case	**20.775**	**70.412**	**120.943**	**36.950**	**75.679**	**314.556**	**38.184**	**89.889**	**166.288**

Table 5.10: This table displays the annualized basis point fee an investor would be willing to pay to switch from the time-varying mean and constant variance model, $\mathcal{M}_1$, to the proposed model, $\mathcal{M}_5$. Positive numbers show that $\mathcal{M}_5$ outperforms $\mathcal{M}_1$ in terms of generating economic benefits. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec and the out-of-sample period is from Jan-1985 to Dec-2015. Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations.

$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t,\ \ e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

	CRRA			CARA			Quadratic		
Predictor	$\tau = 2$	$\tau = 5$	$\tau = 10$	$\tau = 2$	$\tau = 5$	$\tau = 10$	$\tau = 2$	$\tau = 5$	$\tau = 10$
Stock Variance	**14.573**	**17.798**	**34.199**	**9.511**	**17.374**	−14.391	**10.825**	**20.037**	**38.684**
Dividend Payout Ratio	**0.246**	**25.827**	**24.368**	**16.735**	**27.673**	**28.263**	**15.020**	**35.697**	**61.085**
Long Term Spread	−9.933	**3.726**	**9.372**	−9.913	**3.392**	**51.017**	−9.895	**2.805**	**12.795**
Default Yield Spread	−6.983	**1.421**	**10.174**	−6.901	**0.436**	**28.950**	−6.838	**0.072**	**15.493**
Dividend Yield	**10.574**	**11.779**	**26.545**	**17.223**	**37.173**	**181.766**	**24.824**	**58.483**	**77.953**
Dividend Price Ratio	**8.987**	**12.845**	**25.069**	**16.944**	**34.480**	**210.978**	**22.553**	**56.549**	**74.048**
Earnings Price Ratio	−1.281	**5.657**	**12.148**	**7.218**	**24.532**	**174.654**	**11.563**	**44.148**	**64.435**
Book-to-Market Ratio	−7.711	**10.757**	**16.465**	**0.221**	**12.750**	**86.956**	**0.663**	**23.093**	**46.395**
Inflation	−3.008	**1.731**	**8.992**	−3.051	**2.422**	−21.628	−2.959	**2.577**	**15.203**
Net Equity Expansion	**0.763**	**7.520**	**15.289**	**0.848**	**7.552**	**21.602**	**0.883**	**8.361**	**19.487**
Δ in Industrial Production	**1.421**	**26.714**	**24.983**	**17.155**	**28.153**	**29.124**	**15.992**	**36.166**	**63.117**
Kitchen sink case	**14.146**	**45.846**	**73.116**	**20.227**	**48.512**	**213.045**	**21.392**	**62.619**	**108.449**

Table 5.11: This table displays the annualized basis point fee an investor would be willing to pay to switch from the time-varying mean and variance model, $\mathcal{M}_2$, to the proposed model, $\mathcal{M}_5$. Positive numbers show that $\mathcal{M}_5$ outperforms $\mathcal{M}_2$ in terms of generating economic benefits. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec and the out-of-sample period is from Jan-1985 to Dec-2015. Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations.

$\mathcal{M}_2$: Time-varying mean and variance: $y_t = x_{t-1}\mu + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

Model	Mean	St. Dev.	Sharpe Ratio	Δ
	Without Transaction Costs			
$\mathcal{M}_0$	0.388%	6.170%	0.062	41.640
$\mathcal{M}_1$	1.27%	22.342%	0.057	48.306
$\mathcal{M}_2$	1.09%	20.249%	0.054	33.271
$\mathcal{M}_3$	1.71%	21.105%	0.081	42.589
$\mathcal{M}_5$	5.26%	23.531%	0.223	−−
Buy-Hold	4.69%	26.592%	0.179	−−
	With Transaction Costs			
$\mathcal{M}_0$	0.385%	6.144%	0.061	41.293
$\mathcal{M}_1$	1.267%	22.326%	0.056	47.944
$\mathcal{M}_2$	1.094%	20.225%	0.054	32.987
$\mathcal{M}_3$	1.674%	21.076%	0.079	42.242
$\mathcal{M}_5$	5.182%	23.425%	0.221	−−
Buy-Hold	4.69%	26.592%	0.179	−−

Table 5.12: Performance of market timing portfolios with $p_{\text{target}} = 0.5$ and $r_{\text{target}} = 0$. The results are for the kitchen sink case when we use all predictors under study. The first and second columns reports the annualized mean and standard deviation of the portfolios. The values in the last column are annualized basis point performance fees that an investor is willing to pay to switch from the model in the corresponding row to $\mathcal{M}_5$. The risk aversion coefficient is $\tau = 5$. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec and the out-of-sample period is from Jan-1960 to Dec-2015. Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations.

$\mathcal{M}_0$: Constant mean and variance: $y_t = \mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_2$: Time-varying mean and variance: $y_t = x_{t-1}\mu + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

$\mathcal{M}_3$: Infinite mixture with constant weights: $y_t = \mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G$

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

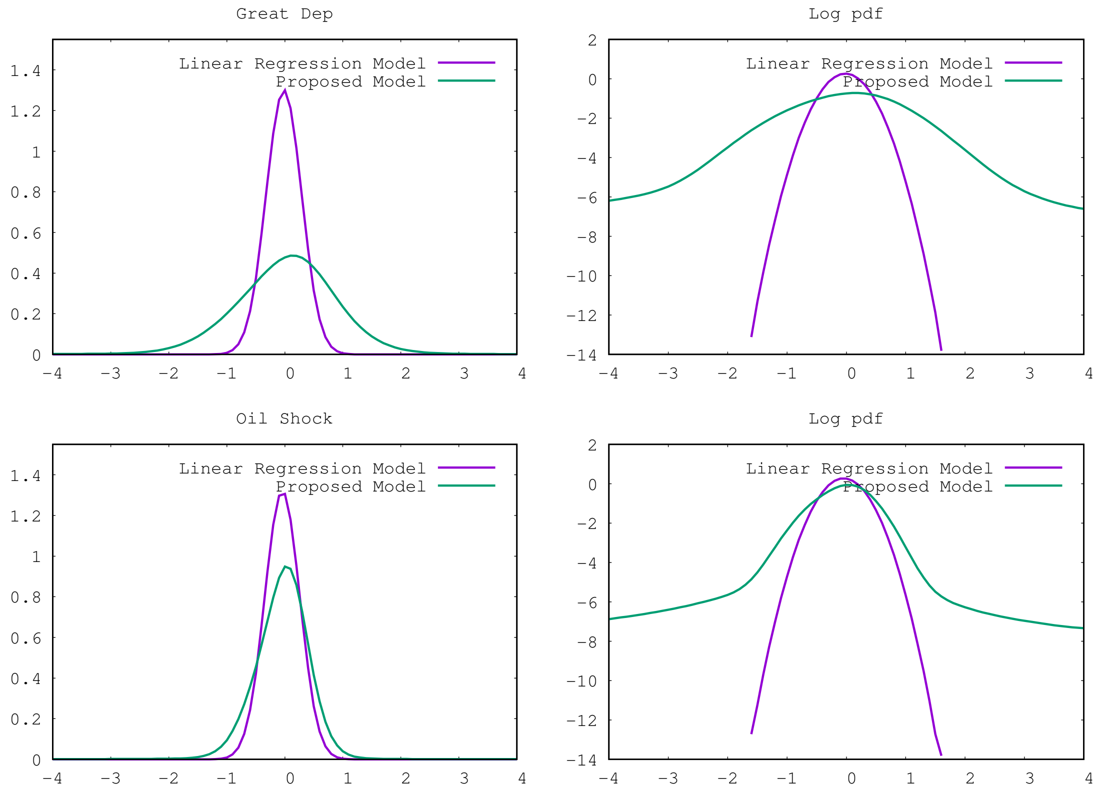

Figure 5.1: Density estimation obtained by the proposed model and the linear regression model at two different dates when we use dividend payout ratio as the predictor. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.

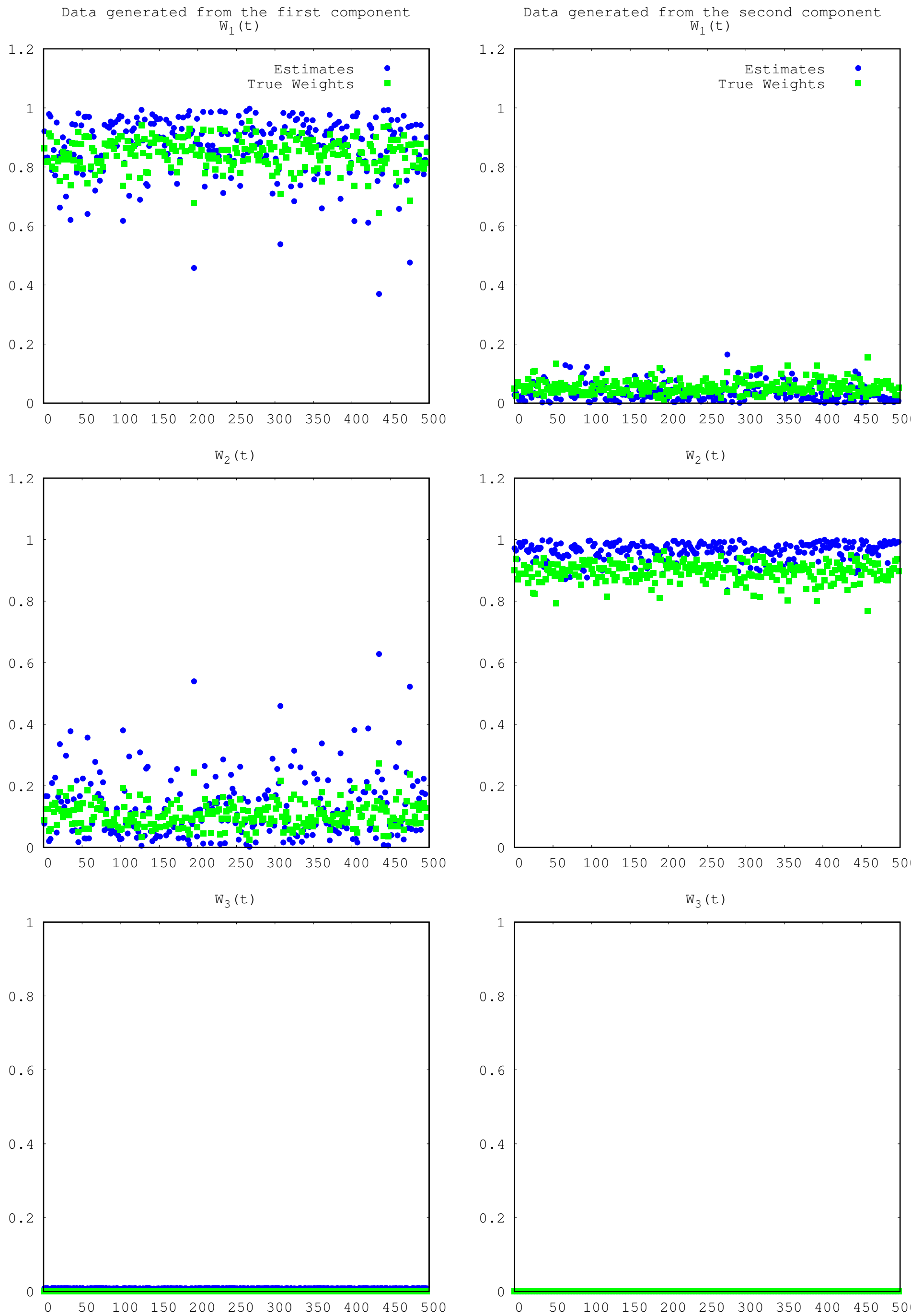

Figure 5.2: Simulated data. Comparison of the true weights with the estimated weights from the infinite mixture model with time-varying weights, $\mathcal{M}_5$.

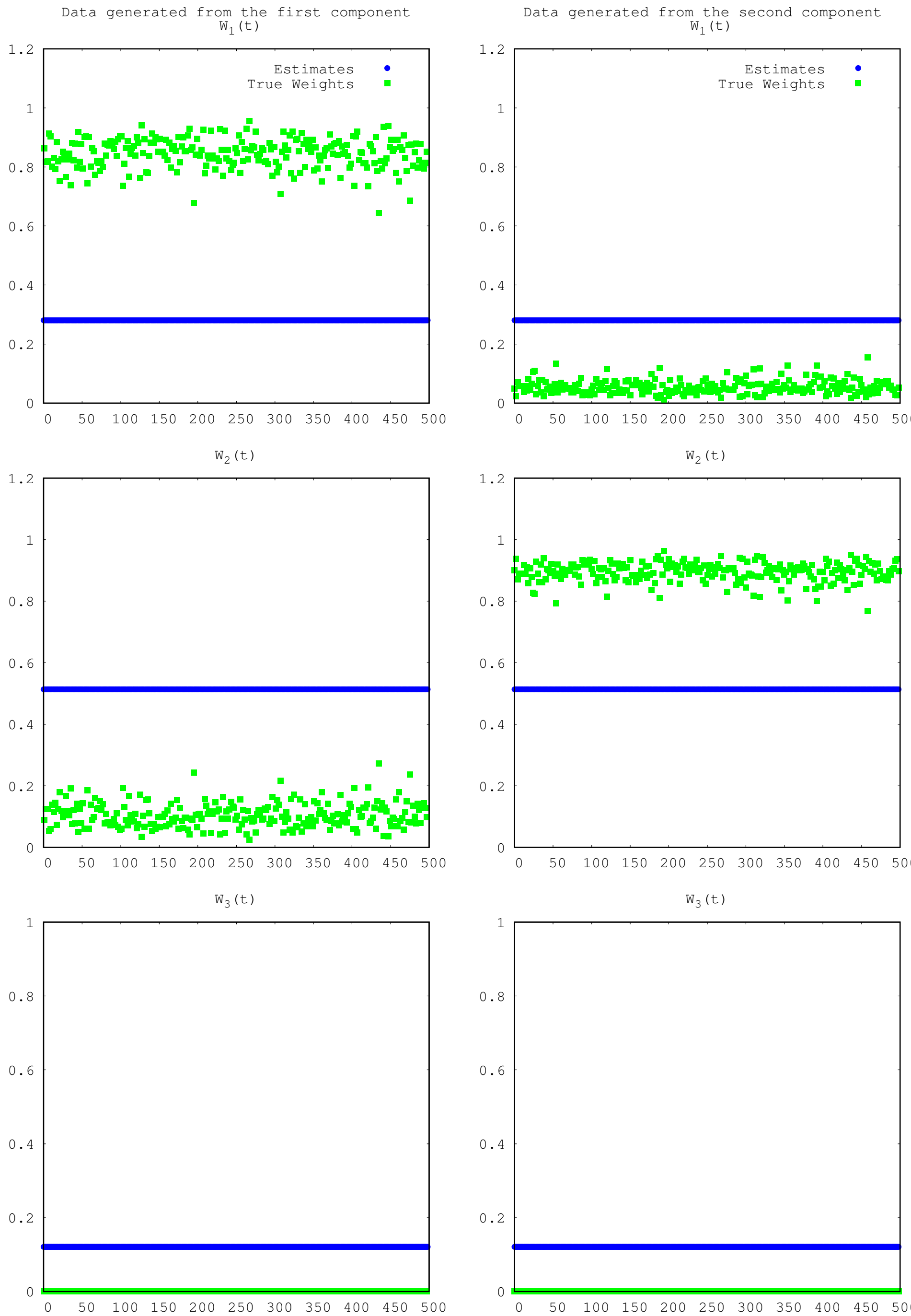

Figure 5.3: Simulated data. Comparison of the true weights with the estimated weights fromthe infinite mixture model with constant weights, $\mathcal{M}_3$,.

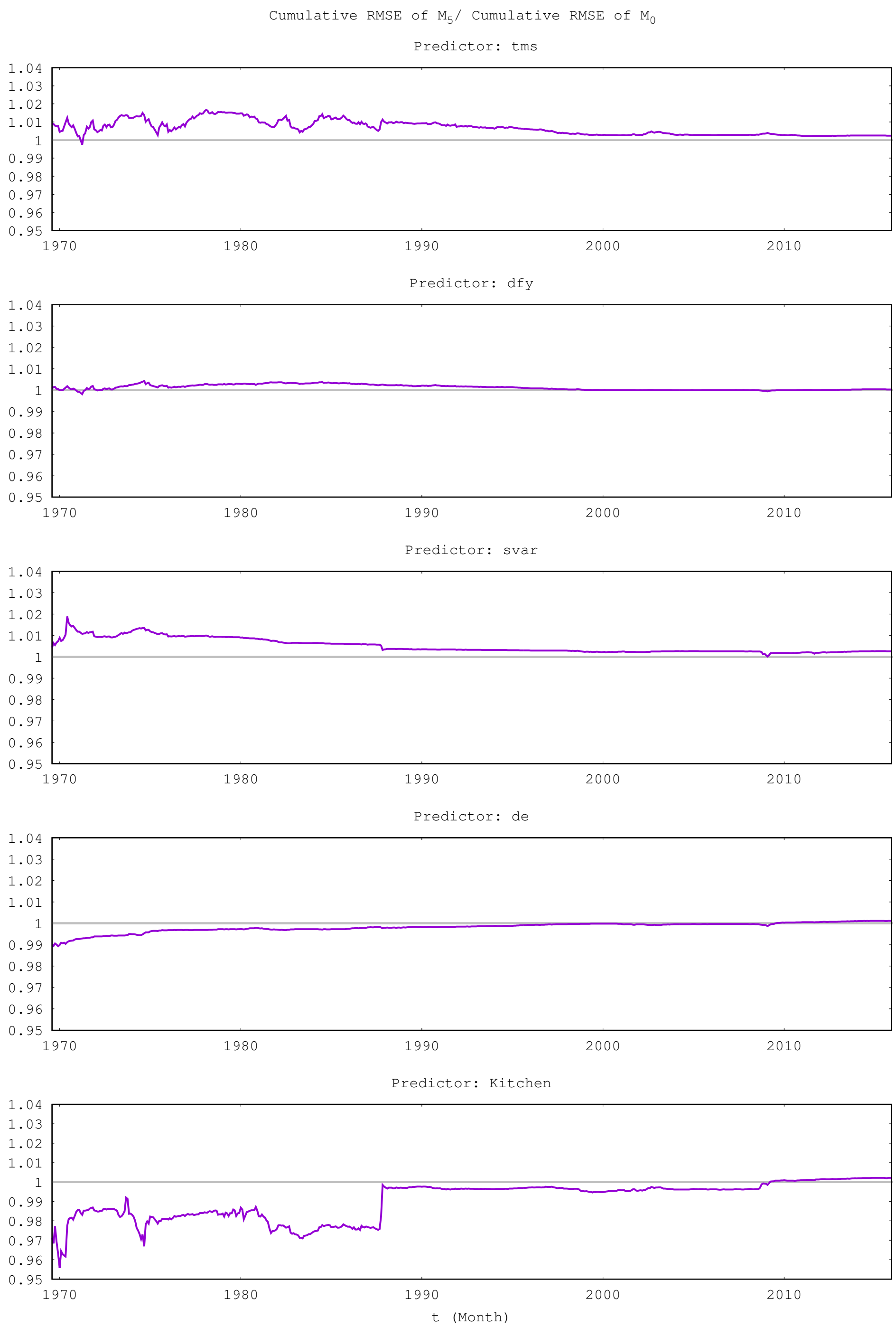

Figure 5.4: Cumulative out-of-sample root mean squared forecasting error of infinite mixture model with time-varying weights and predictors in mean, $\mathcal{M}_5$, divided by that of constant mean and variance model, $\mathcal{M}_0$. Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec and the out-of-sample period covers 1970-Jan to 2015-Dec.

$\mathcal{M}_0$: Constant mean and variance: $y_t = \mu + e_t, \;\; e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\; e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\; \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\; (\mu_t, \nu_t^2) \sim G_t$

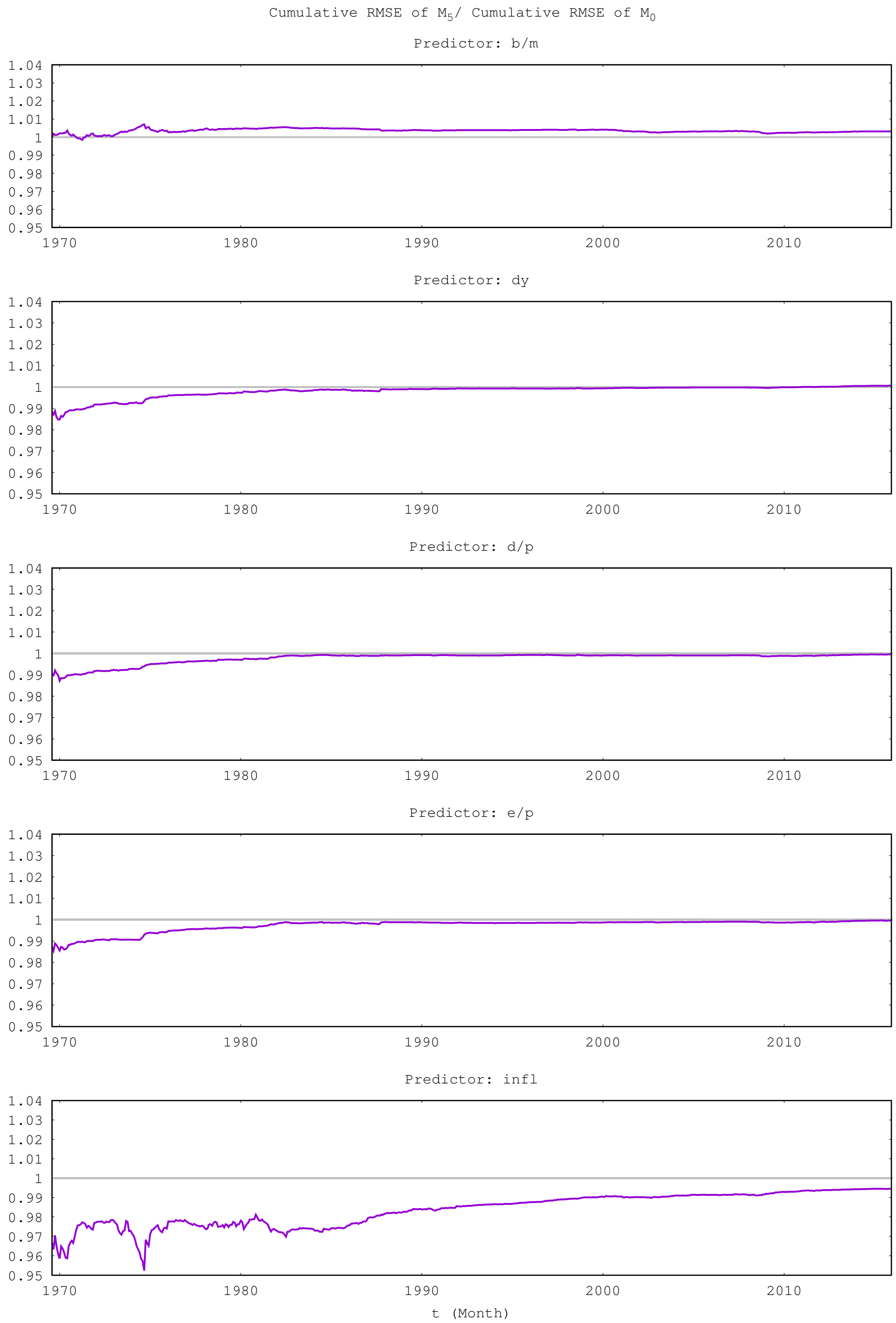

Figure 5.5: Cumulative out-of-sample root mean squared forecasting error of infinite mixture model with time-varying weights and predictors in mean, $\mathcal{M}_5$, divided by that of constant mean and variance model, $\mathcal{M}_0$. Forecasts are obtained by estimating the models applying a rolling window with the most recent 30 years of observations. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec and the out-of-sample period covers 1970-Jan to 2015-Dec.
$\mathcal{M}_0$: Constant mean and variance: $y_t = \mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t$, $e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2)$, $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$, $(\mu_t, \nu_t^2) \sim G_t$

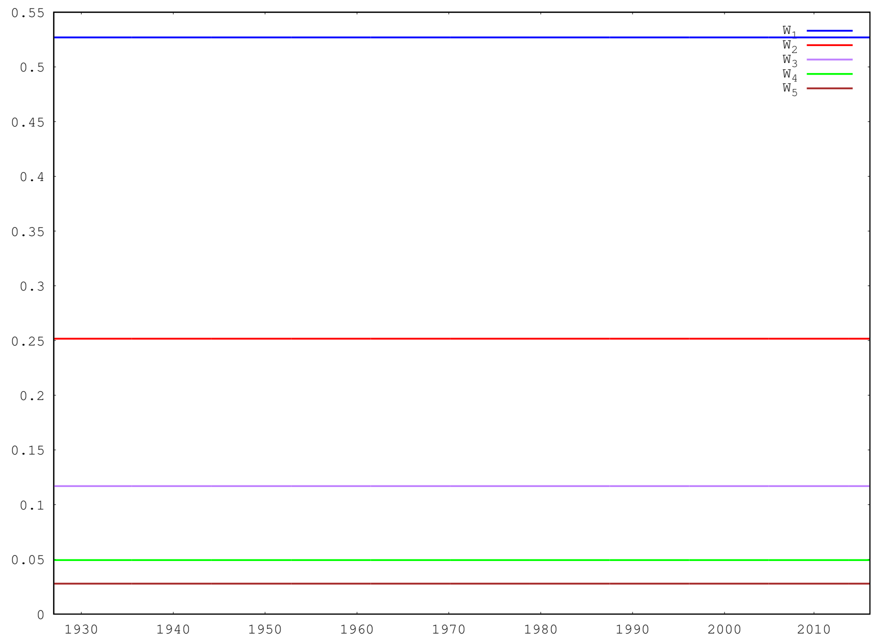

Figure 5.6: Weights estimated by infinite mixture model with constant weights, $\mathcal{M}_3$. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_3$: Infinite mixture with constant weights: $y_t = \mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G$

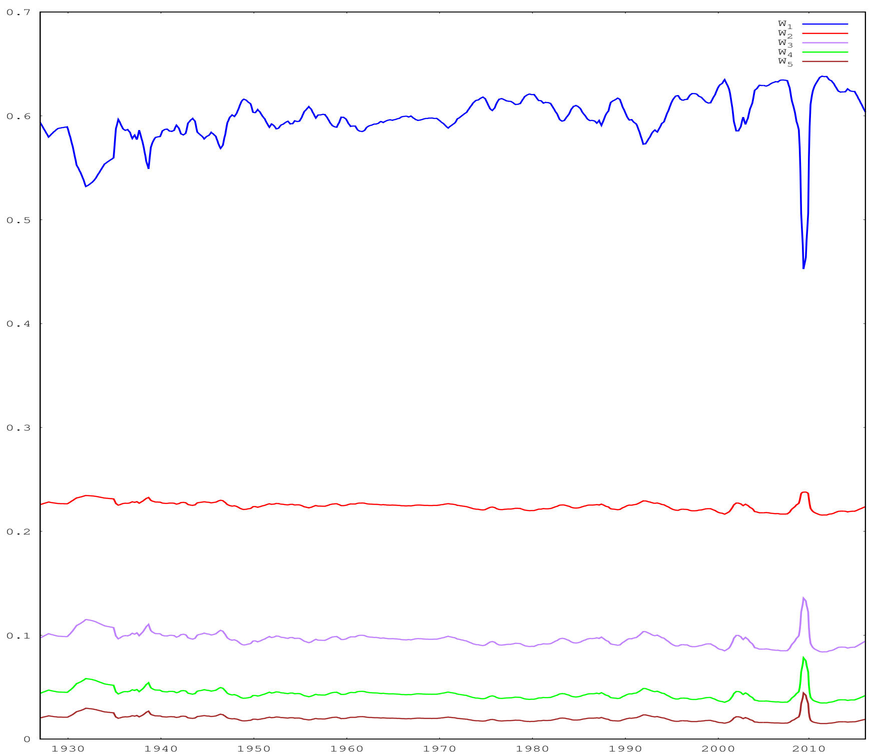

Figure 5.7: Weights estimated by infinite mixture model with time-varying weights and predictors in the conditional mean, $\mathcal{M}_5$, when we have dividend payout ratio as the predictor in the weights and the conditional mean, $x_t = (1 \;\; d_t/e_t)$. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

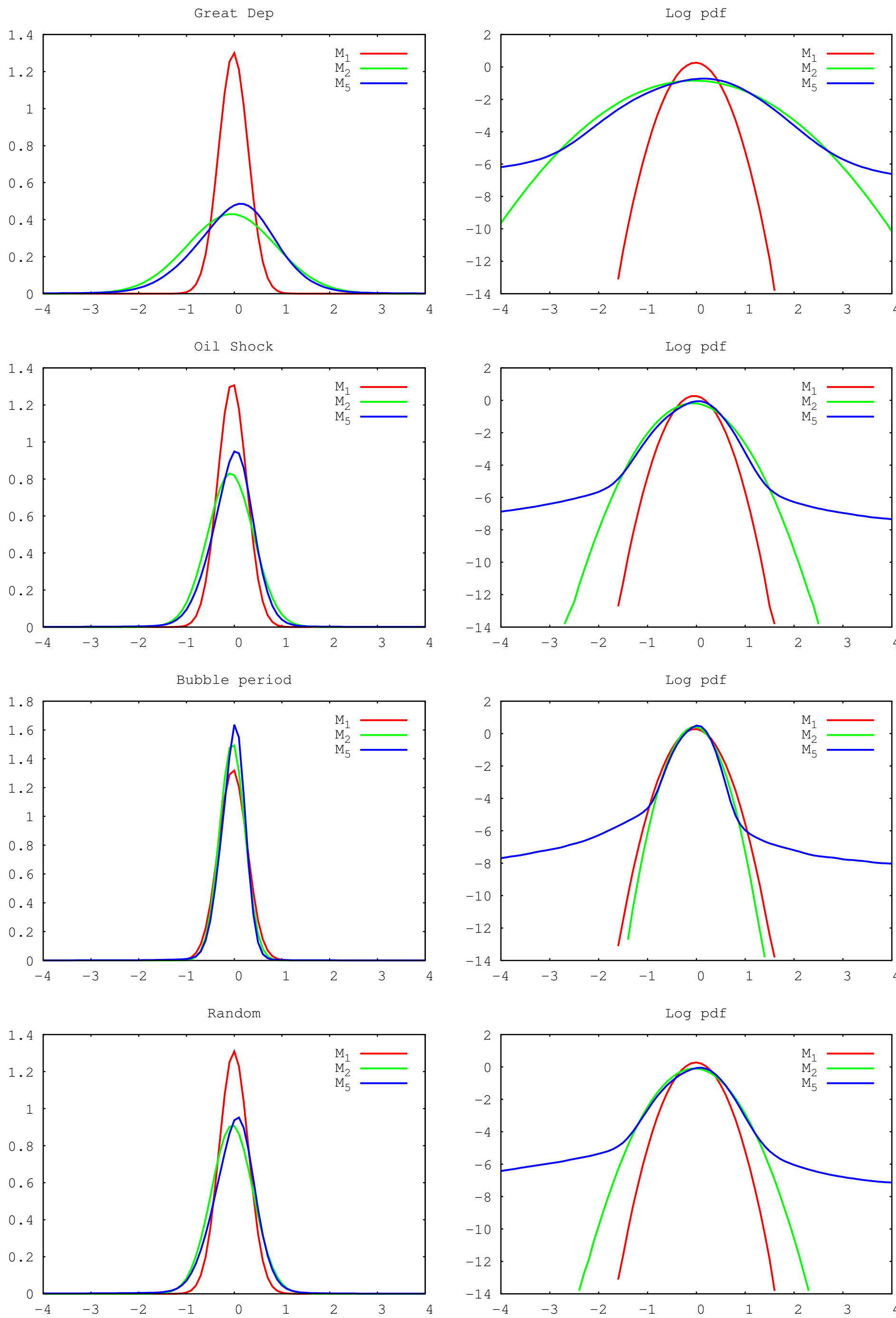

Figure 5.8: Density estimation obtained by different models at four different dates when we use all predictors (the kitchen sink case) as the predictors. The dates are 1932-09 in the great depression period, 1974-09 in the oil shock period, 2000-08 in the bubble period, and 2009-02. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t,\ \ e_t \sim \mathcal{N}(0, \sigma^2)$
$\mathcal{M}_2$: Time-varying mean and variance: $y_t = x_{t-1}\mu + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

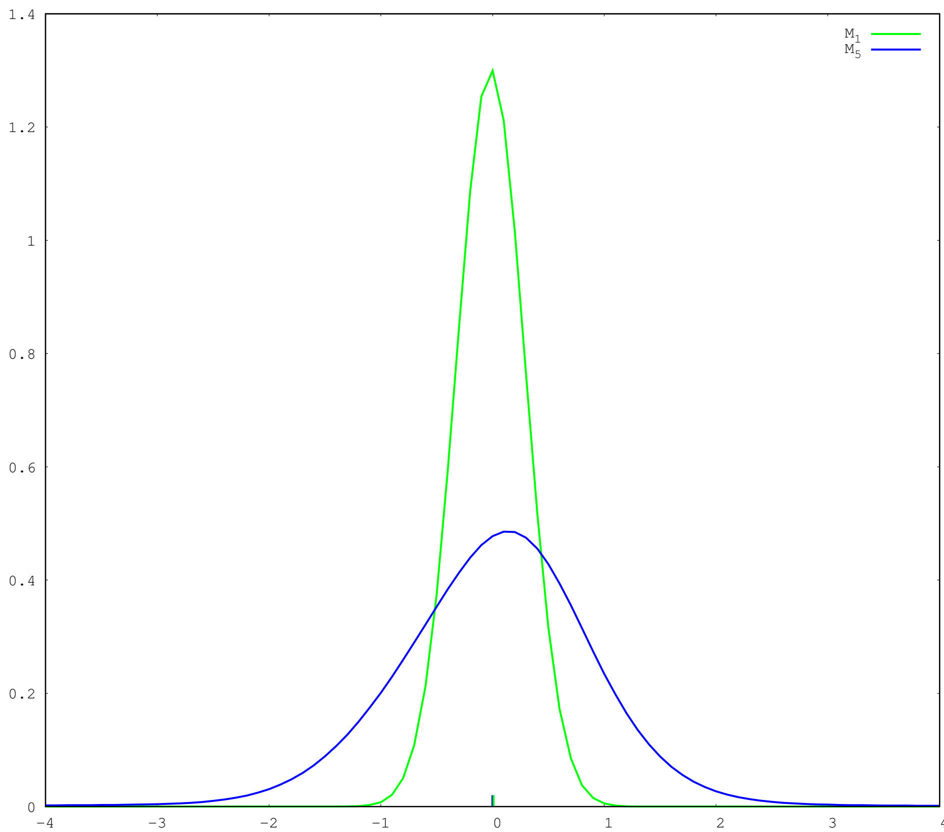

Figure 5.9: Comparison of the conditional mean and density of market excess returns estimated by the time-varying mean and constant variance model, $\mathcal{M}_1$, and the infinite mixture model with time-varying weights and predictor in the conditional mean, $\mathcal{M}_5$, when we use all predictors (the kitchen sink case) as the predictors. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.

$\mathcal{M}_1$: Time-varying mean and constant variance: $y_t = x_{t-1}\mu + e_t, \quad e_t \sim \mathcal{N}(0, \sigma^2)$

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t$, $e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2)$, $\sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$, $(\mu_t, \nu_t^2) \sim G_t$

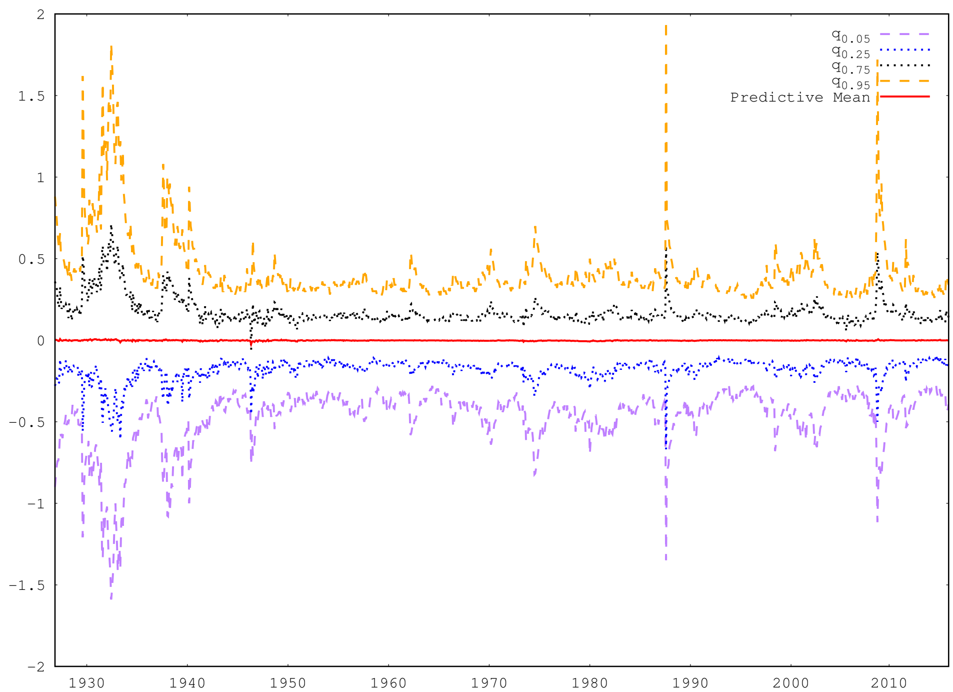

Figure 5.10: This figure displays the 5%, 25%, 75%, and 95% conditional quantiles over the full sample period using estimates of the infinite mixture model with time-varying weights and predictor in the conditional mean, $\mathcal{M}_5$, when we use all predictors (the kitchen sink case) as the predictors. This figure also shows the conditional mean predicted using this model (the solid red line). Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.

$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2\nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

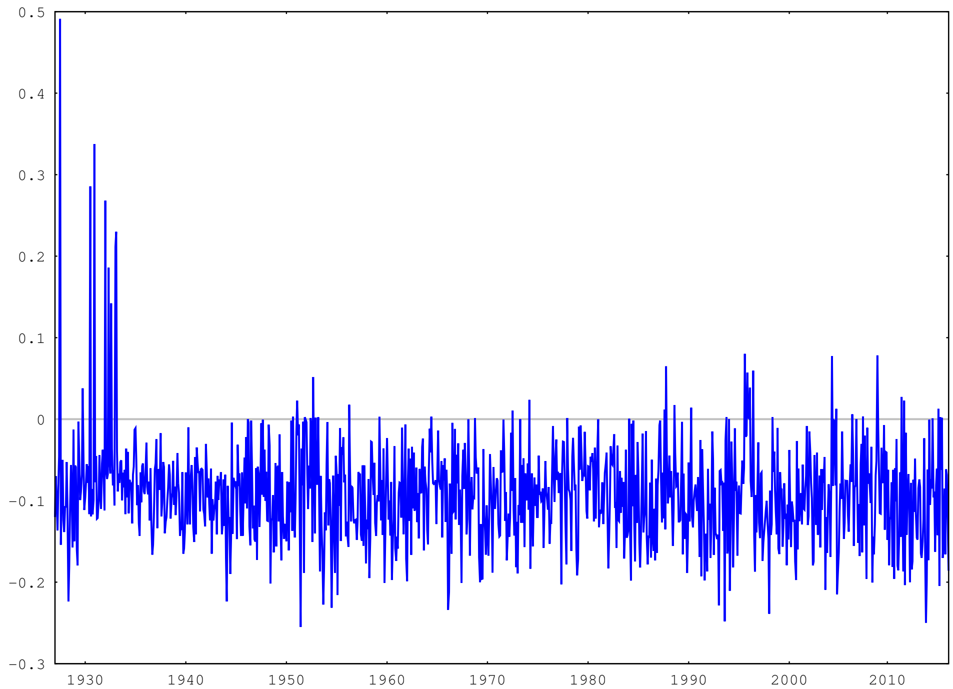

Figure 5.11: Conditional skewness estimates using the infinite mixture model with time-varying weights and predictor in the conditional mean, $\mathcal{M}_5$, when we use all predictors (the kitchen sink case) as the predictors. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

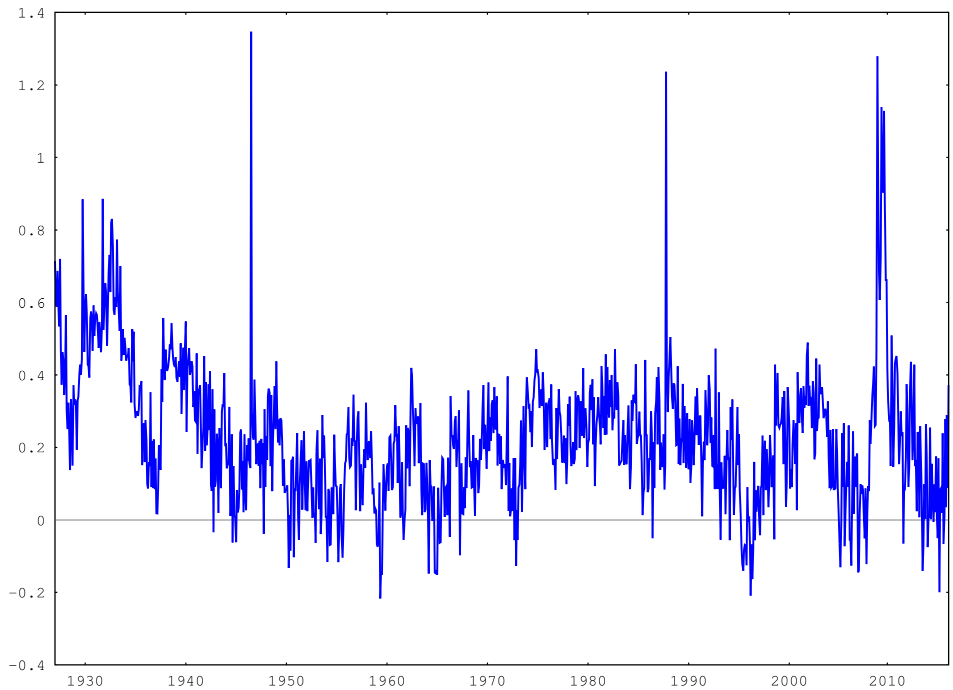

Figure 5.12: Conditional kurtosis estimates using the infinite mixture model with time-varying weights and predictor in the conditional mean, $\mathcal{M}_5$, when we use all predictors (the kitchen sink case) as the predictors. Data are monthly market returns and predictor values from 1927-Jan to 2015-Dec.
$\mathcal{M}_5$: Infinite mixture with time-varying weights and predictors in the conditional mean: $y_t = x_{t-1}\mu_t + e_t,\ e_t \sim \mathcal{N}(0, \sigma_t^2 \nu_t^2),\ \sigma_t^2 = \alpha + \delta\epsilon_{t-1}^2 + \gamma\sigma_{t-1}^2,\ (\mu_t, \nu_t^2) \sim G_t$

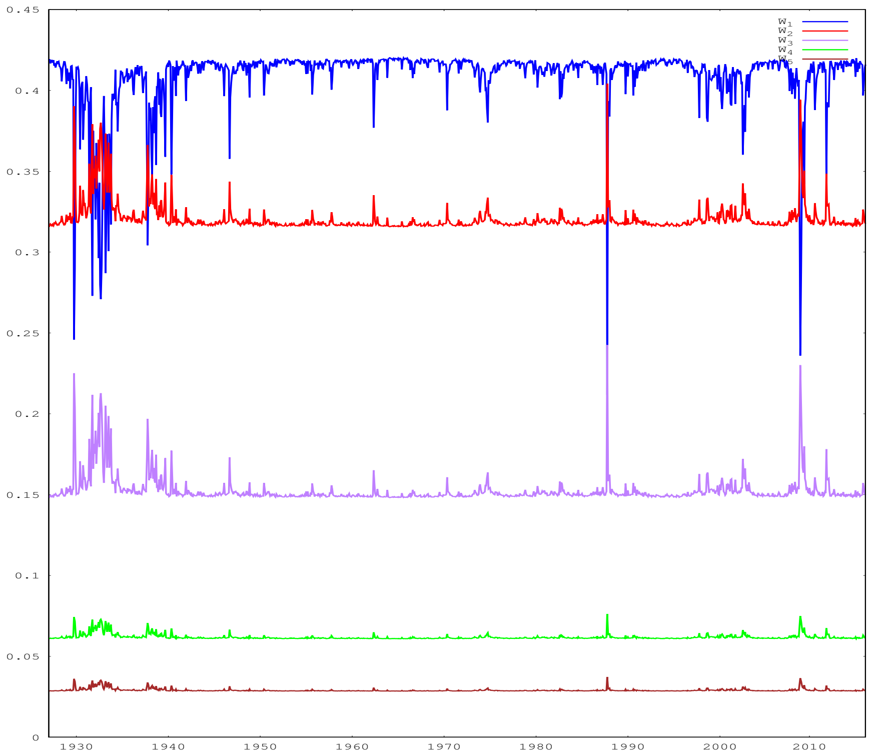

Figure 5.13: Weights when we have **Stock Variance** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

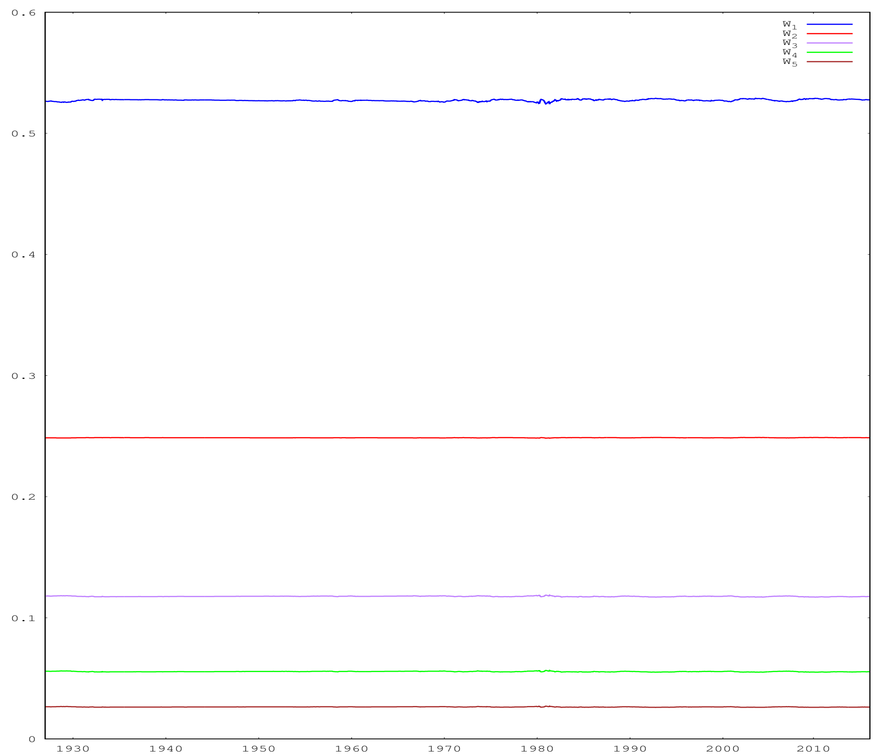

Figure 5.14: Weights when we have **Long Term Spread** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

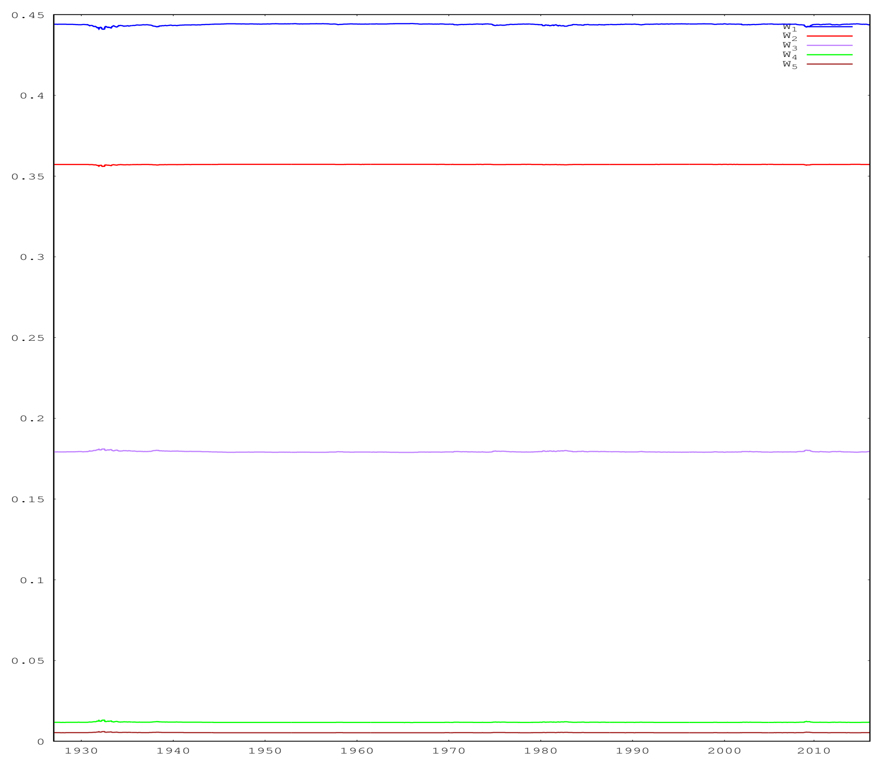

Figure 5.15: Weights when we have **Default Yield Spread** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

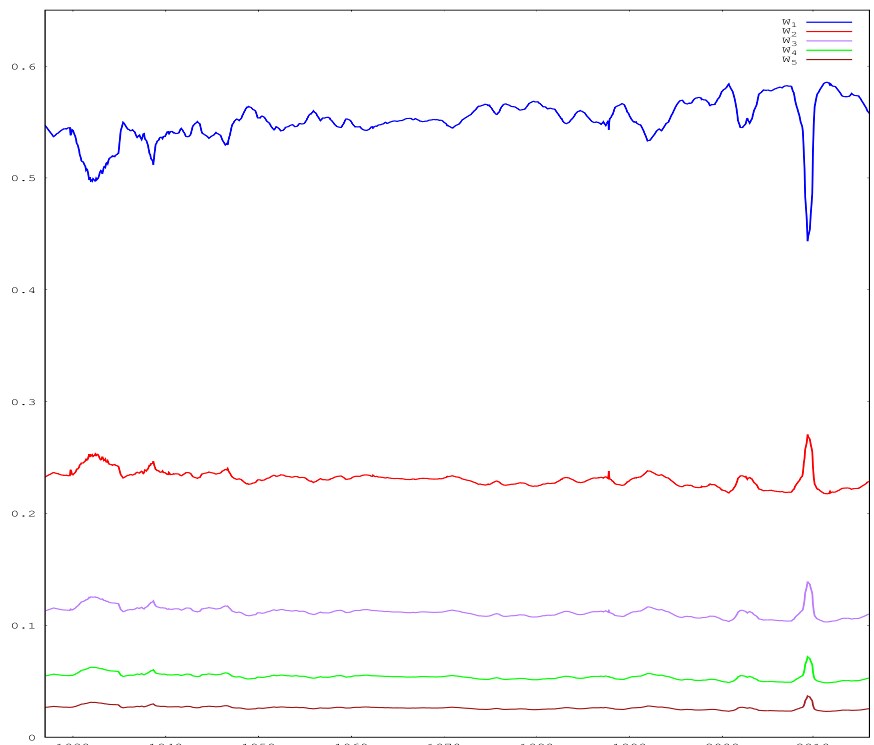

Figure 5.16: Weights when we have all predictors (**kitchen sink case**) as the predictors in the weights and the predictive mean in $\mathcal{M}_5$.

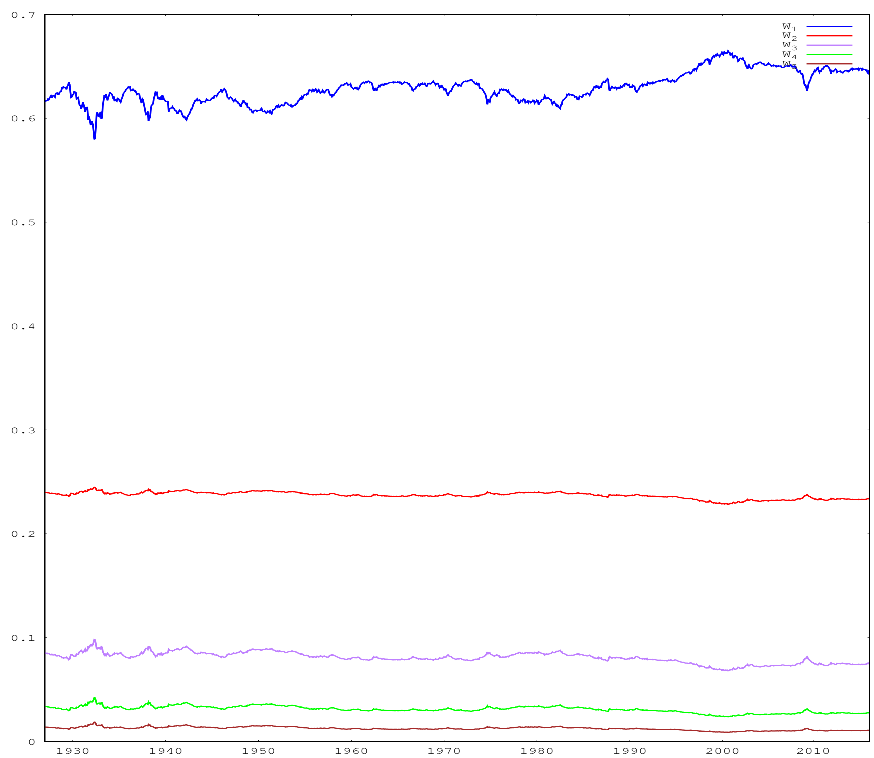

Figure 5.17: Weights when we have **Dividend Price Ratio** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

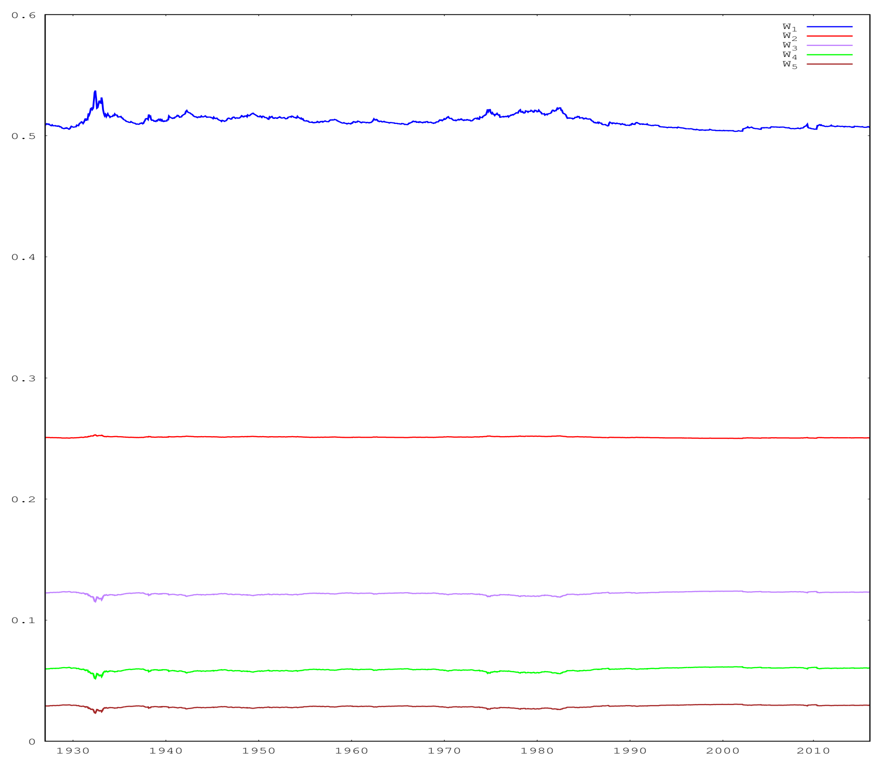

Figure 5.18: Weights when we have **Book-to-Market Ratio** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

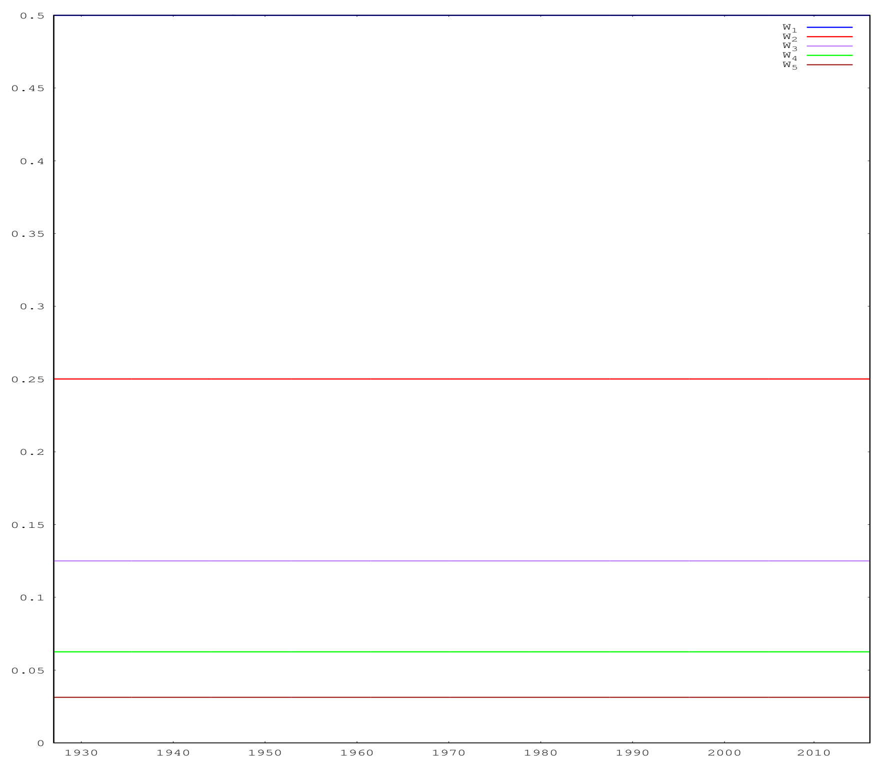

Figure 5.19: Weights when we have **Inflation** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

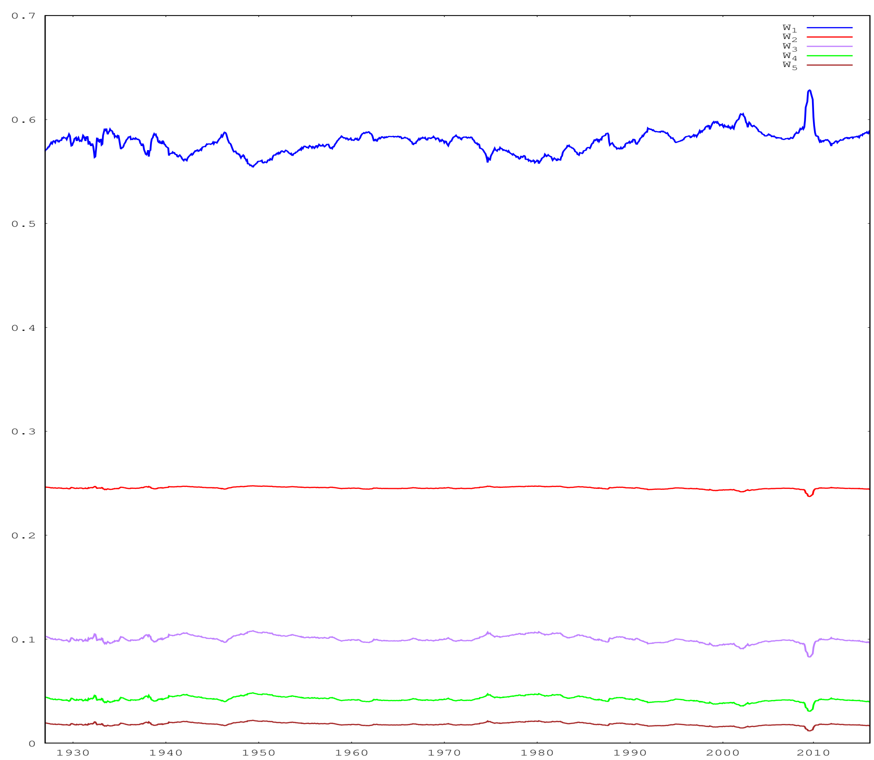

Figure 5.20: Weights when we have **Earnings Price Ratio** as the predictor in weights and mean in $\mathcal{M}_5$.

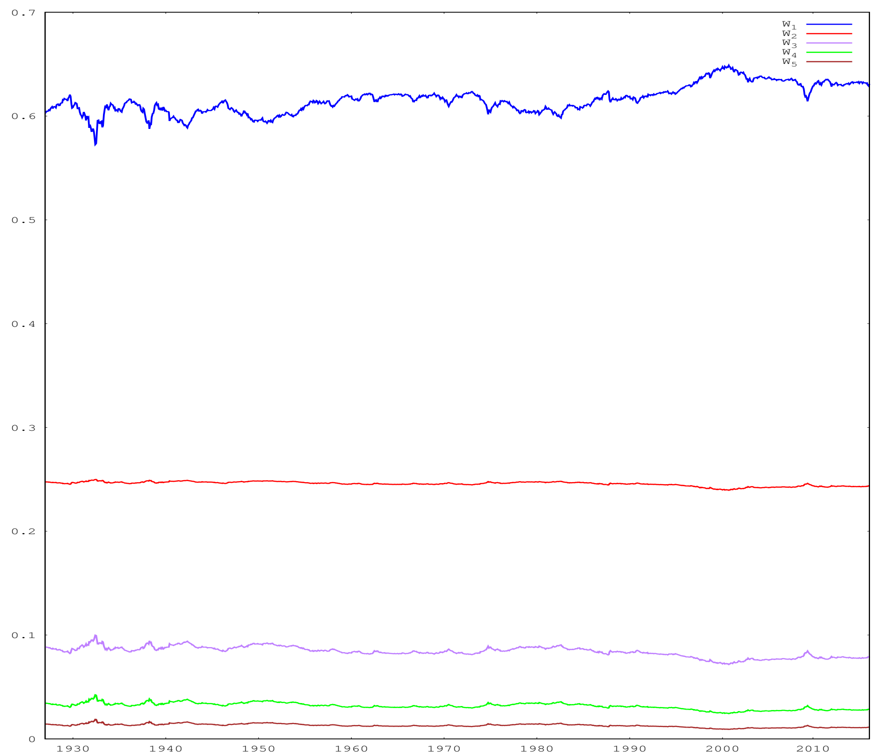

Figure 5.21: Weights when we have **Dividend Yield** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

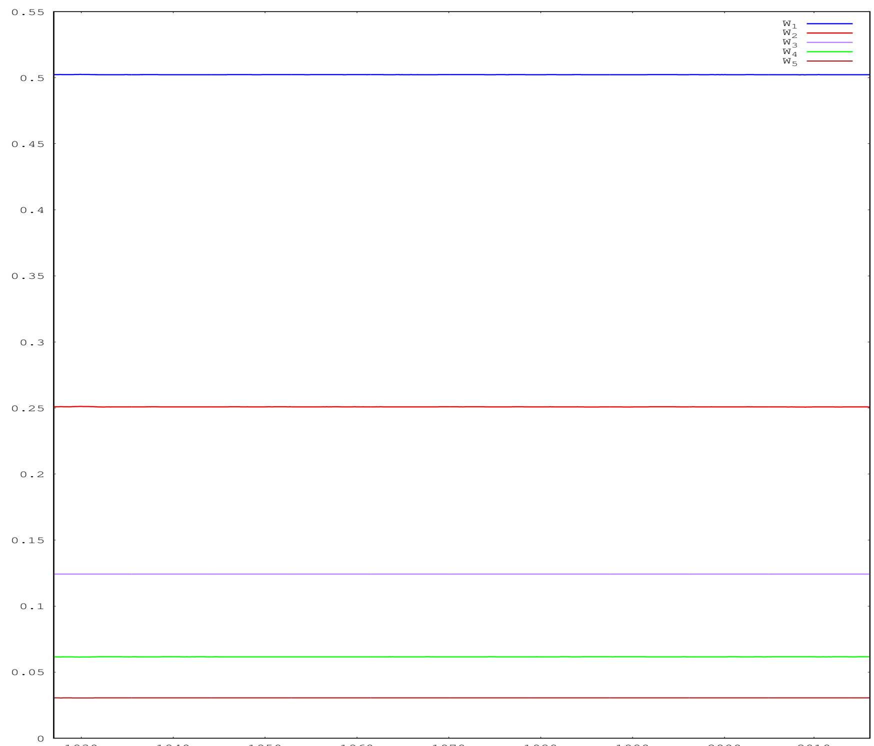

Figure 5.22: Weights when we have **Net Equity Expansion** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

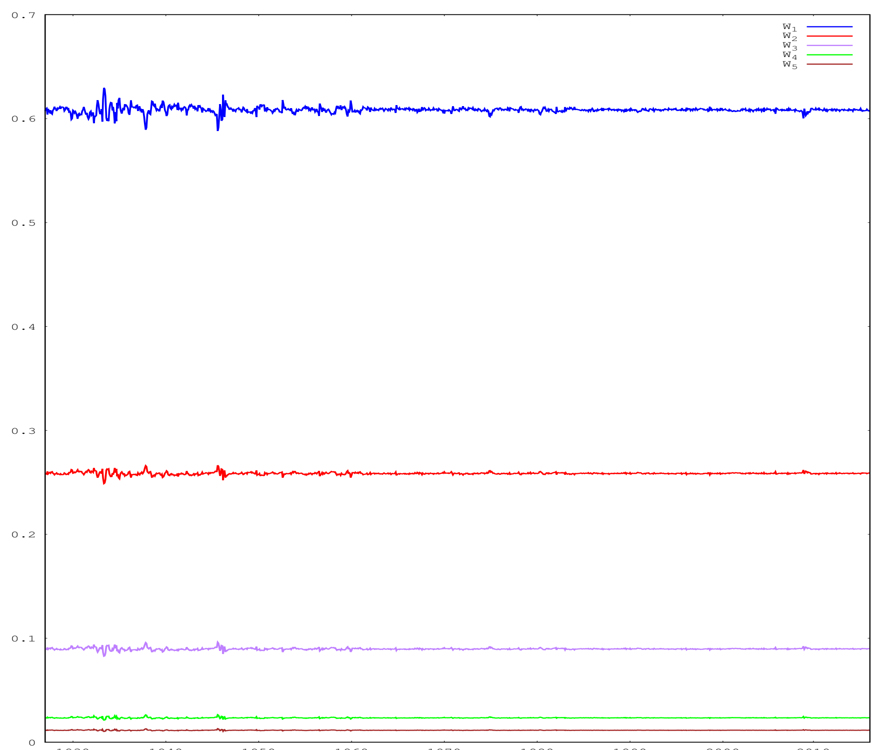

Figure 5.23: Weights when we have **Industrial Production** as the predictor in the weights and the predictive mean in $\mathcal{M}_5$.

Chapter 6

Conclusion

This thesis makes contributions in the financial econometrics area with a particular focus on estimating an accurate model for financial asset returns using Bayesian nonparametric methods. Most of the literature has concentrated on modeling the conditional (co)variance of returns in forms of stochastic volatility and (multivariate) GARCH models with a fixed parametric innovation density. This means that once volatility dynamics are removed, the innovations distribution is constant over time. Empirical research that allows the innovation distribution to be unknown and to change over time is scarce. The attempt in this thesis is to fill this gap and advance the methodology of modeling asset returns by estimating an unknown density for the return innovations rather than making a specific assumption. This can significantly enhance our understanding of the tails and higher moments of the returns density which are relatively unexplored.

The first essay derives a dynamic conditional beta representation using a Bayesian semiparametric multivariate GARCH model. I show how to select the number of factors and that the predictive Bayes factors strongly support this semiparametric model over a multivariate GARCH with Student-t innovations. Empirically, I find the time-varying beta from the proposed model nonlinearly depends on the contemporaneous value of the market excess returns. In highly volatile markets, beta is almost constant, while in stable markets, the beta coefficient can depend asymmetrically on the contemporaneous value of the market excess returns.

In the second essay, I propose a model that allows a shock in one market to influence the contemporaneous and future (one-day-ahead, one-week-ahead, and one-month ahead) conditional density of the other market. This model extends the literature on spillover effects or contagion effects that focus on the transmission of shocks through moments to spillover effects on the conditional density. I provide a general approach to assess different risk scenarios, contemporaneously. With the speed of information transmission nowadays, it is beneficial for policymakers and regulators to be able to determine the contemporaneous shocks spillover from one market to different aspects of the conditional density of the other market such as the conditional moments, tails, and value at risk. I apply the proposed model to study how shocks in oil market affect the conditional density of S&P 500 returns and S&P TSX returns. The contemporaneous spillover effect of a shock in oil market on the TSX and S&P 500 markets derived from the benchmark is symmetric; the shifts in the conditional density in these two cases are almost the same amount but in opposite directions. This is not the case with the semiparametric model. Using the proposed model, I find that a positive 10% and a negative 10% shock in the oil market spill over into the TSX and S&P 500 asymmetrically. When the market is calm, a positive shock in oil market shifts the contemporaneous conditional density of TSX and S&P 500 to the right, and a negative shock in oil market shifts the contemporaneous conditional density of TSX and S&P 500 to the left. These shifts are not the same amount for positive and negative shocks, and the resulting conditional densities are skewed and leptokurtic. When the market is volatile, a positive shock in oil market has almost no effect on the TSX and S&P 500 conditional density contemporaneously while a negative shock in oil market shifts the contemporaneous conditional density of TSX and S&P 500 to the left. Both positive and negative shocks result in a fatter lower tail than the case with no shock to oil market. I also study the effect of the shocks in oil market on the value-at-risk of an investment in S&P 500 and S&P TSX. Positive and negative shocks in oil market both increase the predictive value-at-risk. The benchmark model underestimates the predictive value-at-risk of the investment in the market portfolio when we have a shock in oil market.

In the third essay, I propose a model to study the one-period-ahead out-of-sample predictability of the US stock return density using financial and macroeconomic variables. In contrast to the extant literature that focuses on the point forecasts of the conditional mean, I examine the predictability of the entire density which can be valuable for asset allocations and risk management. I consider a Bayesian nonparametric mixture model that allows the mixing distribution to change with time. In the proposed model, the weights of the mixture are constructed as probit transformations of a linear combination of the predictors. I compare statistical and economic measures of forecasting performance of the proposed model with a set of benchmark models. Despite little or no improvement in the point forecasts, certain variables display significant out-of-sample predictive ability with respect to the predictive density of market returns and increase economic value for investors when employed in portfolio decisions. A risk-averse investor is willing to pay a performance fee to switch from the benchmark models to the infinite mixture model proposed in this chapter. I also illustrate how these financial and macroeconomic variables are useful in predicting the conditional density, the conditional moments and the conditional quantiles.

Overall, this research emphasizes the importance of accurate density estimation and return predictability. Most work in the literature can explain the first and second moments, but what matters in practice is a good estimation of the entire density so we can study various features such as value-at-risk. This dissertation explores Bayesian nonparametric approaches to estimating the unknown density of returns lifting the restrictive parametric assumptions. This research should be of interest not only to the academic community but to the investors, practitioner, managers, and policymakers as well. In this light, potential applications are performed in each essay that could be implemented in real time.

Bibliography

Abhyankar, A. (1995). Trading round-the-clock: Return, volatility and volume spillovers in the Eurodollar futures markets. *Pacific-basin finance journal*, 3(1):75–92.

Ang, A. and Bekaert, G. (2006). Stock Return Predictability: Is it There? *Review of Financial Studies*, 20(3):651–707.

Apergis, N. and Miller, S. M. (2008). Department of Economics Working Paper Series Do Structural Oil-Market Shocks Affect Stock Prices? *Energy Economics*, 31(4):569–575.

Apergis, N. and Rezitis, A. (2003). Agricultural price volatility spillover effects : the case of Greece. 30(3):389–406.

Asgharian, H., Hou, A. I. J. U. N., and Javed, F. (2013). The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance : A GARCH-MIDAS Approach. *Journal of Forecasting*, 31(July):600–612.

Ausín, M. C., Galeano, P., and Ghosh, P. (2014). A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. *European Journal of Operational Research*, 232(2):350–358.

Baele, L. (2005). Volatility Spillover Effects in European Equity Markets All use subject to JSTOR Terms and Conditions Volatility Spillover Effects Markets in European Equity. *The Journal of Financial and Quantitative Analysis*, 40(2):373–401.

Bali, T. G. and Engle, R. F. (2010). The intertemporal capital asset pricing model with dynamic conditional correlations. *Journal of Monetary Economics*, 57(4):377–390.

Bali, T. G., Engle, R. F., and Tang, Y. (2016). Dynamic Conditional Beta is Alive and Well in the Cross-Section of Daily Stock Returns. *Working Paper*, (646):1–71.

Balli, F., HR, H., Basher, S. A., and Ghassan, H. B. (2015). An Analysis of Returns and Volatility Spillovers and their Determinants in Emerging Asian and Middle Eastern Countries. *International Review of Economics {&} Finance*, 39:311–325.

Balvers, R. J., Cosimano, T. F., and McDonald, B. (1990). Predicting Stock Returns in an Efficient Market. *The Journal of Finance*.

Bauwens, L., Hafner, C. M., and Rombouts, J. V. K. (2007). Multivariate Mixed Normal Conditional Heteroskedasticity. *Computational Statistics & Data Analysis*, 51(7):3551–3566.

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Bauwens, L., Lubrano, M., and Richard, J.-F. (2000). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson Distributions via Pólya Urn Schemes. *The annals of statistics*, 1(2):353–355.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of econometrics*, 31:307–327.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A Capital Asset Pricing Model with Time-varying Covariances. *Journal of Political Economy.*, 96(1):116–131.

Bossaerts, P. and Hillion, P. (1999). Implementing Statistical Criteria to Select Return Forecasting Models : What Do We Learn? *Review of Financial Studies*, 12(2):405–428.

Campbell, J. Y. and Shiller, R. J. . (1988). Stock Prices, Earnings, and Expected Dividends. *Journal of Finance*, 43(3):661–676.

Caporale, G. M., Pittis, N., and Spagnolo, N. (2006). Volatility transmission and financial crises. *Journal of Economics and Finance*, 30(3):376–390.

Cenesizoglu, T. and Timmermann, A. (2008). Is the Distribution of Stock Returns Predictable?

Cenesizoglu, T. and Timmermann, A. (2012). Do return prediction models add economic value? *Journal of Banking and Finance*, 36(11):2974–2987.

Choudhry, T. (2002). The Stochastic Structure of the Time-Varying Beta: Evidence from UK Companies. *The Manchester School*, 70(6):768–791.

Christiansen, C. (2007). Volatility-spillover effects in European bond markets. *European Financial Management*, 13(5):923–948.

Christiansen, C., Schmeling, M., and Schrimpf, A. (2012). A Comprehensive Look at Financial Volatility Prediction by Economic Variables. *Journal of Applied Econometrics*, 977(August):956–977.

Clark, P. K. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*, 41(1):135–155.

Cutler, D. M., Poterba, J. M., and Summers, L. H. (1991). Speculative Dynamics. *Review of Economic Studies*, 58(3):529–546.

Darbar, S. and Deb, P. (1997). Co-Mouvents in International Equity Markets. *Journal of Financial Research*, 20(3):305–322.

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). An ANOVA Model for Dependent Random Measures. *Journal of the American Statistical Association*, 99(465):205–215.

Delatola, E. and Griffin, J. E. (2013). A Bayesian semiparametric model for volatility with a leverage effect. *Computational Statistics & Data Analysis*, 60:97–110.

Diks, C., Panchenko, V., and Dijk, D. V. (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics*, (163):215–230.

Ding, Z. and Engle, R. F. (2001). Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing.

Duan, J. A., Guindani, M., and Gelfand, A. E. (2007). Generalized Spatial Dirichlet Process Models. *Biometrika*, 94(4):809–825.

Engle, R. F. (2016). Dynamic conditional beta. *Journal of Financial Econometrics*, 14(4):643–667.

Engle, R. F., Ghysels, E., and Sohn, B. (2012). On the Economic Sources of Stock Market Volatility. *Review of Economics and Statistics*.

Engle, R. F. and Kroner, K. F. (1995). Multivariate Simultaneous Generalized ARCH. *Econometric Theory*, 11(1):122–150.

Engle, R. F. and Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, 48(5):1749–1778.

Engstrom, E. (2003). The conditional relationship between the equity risk premium and the dividend price ratio. Working paper, Columbia University.

Eom, Y., Subrahmanyam, M., and Uno, J. (2002). The transmission of swap spreads and volatilities in the international swap markets. *Journal of Fixed Income,*, 12(1):6–28.

Escobar, M. D. (1994). Estimating Normal Means With a Dirichiet Process Prior. *Journal of the American Statistical Association*, 89(425):268–277.

Escobar, M. D. and West, M. (1995). Bayesian Density Estimation and Inference using Mixtures. *Journal of the american statistical*, 90:577–588.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34–105.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F. and French, K. R. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3–56.

Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The annals of statistics*, 1(2):209–230.

Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003). Spurious Regressions in Financial Economics? *Journal of Finance*, 58(4):1393–1413.

Fisher, M., Jensen, M. J., and Tkac, P. (2015). Cross-sectional mutual fund performance. Federal Reserve Bank of Atlanta.

Galeano, P. and Ausin, M. C. (2010). The Gaussian Mixture Dynamic Conditional Correlation Model: Parameter Estimation, Value at Risk Calculation, and Portfolio Selection. *Journal of Business & Economic Statistics*, 28(4):559–571.

Gardebroek, C. and Hernandez, M. a. (2013). Do energy prices stimulate food price volatility? Examining volatility transmission between US oil, ethanol and corn markets. *Energy Economics*, 40:119–129.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian Nonparametric Spatial Modeling with Dirichlet Process Mixing. *Journal of the American Statistical Association*, 100(471):1021–1035.

Ghosal, S., G. J. R. R. (1999). Posterior Consistency of Dirichlet Mixtures in Density Estimation. *Annals of Statistics*, 27(1):143–158.

Giannopoulos, K. (1995). Estimating the Time-varying Components of International Stock Markets Risk. *European Journal of Finance*, 1:129–164.

Goetzmann, W. N. . and Jorion, P. (1993). Testing the Predictive Power of Dividend Yields. *Journal of Finance*, 48(2):663–679.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios: a matter of balance. *Management Science*, 49(5):639–654.

Granger, C. W. (1980). Long Memory Relationships and theAggregation of Dynamic Models. *Journal of Econometrics*, 14:227–238.

Griffin, J. E. and Steel, M. F. J. (2006). Order-Based Dependent Dirichlet Processes. *Journal of the American Statistical Association*, 101(473):179–194.

Haas, M., Mittnik, S., and Paolella, M. (2004). A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530.

Hamao, Y., Masulis, R. W., and Ng, V. (1990). Correlations in Price Changes and Volatility across International Stock Markets. *Review of Financial Studies*, 3(2):281–307.

Hansen, L. P. and Richard, S. F. (1987). The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica*, 55(3):587–613.

Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61(2):247–264.

Hodrick, R. J. (1992). Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. *Review of Financial Studies*, 5(3):357–386.

Ishwaran, H. and James, L. F. (2001). Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, 96(453):161–174.

Ishwaran, H. and Zarepour, M. (2000). Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models. *Journal of Computational and Graphical Statistics*, 87(2):371–390.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185–212.

Jensen, M. and Maheu, J. M. (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157(2):306–316.

Jensen, M. and Maheu, J. M. (2013). Bayesian Semiparametric Multivariate GARCH Modeling. *Journal of Econometrics*, 176:3–17.

Jensen, M. and Maheu, J. M. (2014). Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics*, 178:523–538.

Kalli, M. and Griffin, J. E. (2015). Flexible Modeling of Dependence in Volatility Processes. *Journal of Business & Economic Statistics*, 33(1):102–113.

Kalli, M., Griffin, J. E., and Walker, S. G. (2009). Slice sampling mixture models. *Statistics and Computing*, 21(1):93–105.

Kalli, M., Walker, S. G., and Damien, P. (2013). Modeling the Conditional Distribution of Daily Stock Index Returns: An Alternative Bayesian Semiparametric Model. *Journal of Business & Economic Statistics*, 31(4):371–383.

Karlsson, S. (2013). Forecasting with Bayesian Vector Autoregressions. *Handbook of Economic Forecasting*, 2:791–897.

Keijsers, B. (2017). Long-Term Investing Under Possible Parameter Instability.

Kim, S. and Shephard, N. (1998). Stochastic Volatility : Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65:361–393.

Lau, J. W. and Cripps, E. (2012). Bayesian non-parametric mixtures of GARCH(1,1) models. *Journal of Probability and Statistics*, 2012:16.

Lettau, M. and Ludvigson, S. (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance*, 56(3):815–849.

Lettau, M. and Ludvigson, S. C. (2005). Expected returns and expected dividend growth. *Journal of Financial Economics*, 76(3):583–626.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47(1):13–37.

Lo, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The annals of statistics*, 12(1):351–357.

MacEachern, S. N. (1999). Dependent nonparametric processes. In *SA proceedings of the section on Bayesian statistical science*, pages 50–55.

MacEachern, S. N. (2000). Dependent Dirichlet Processes. Unpublished manuscript, Department of Statistics, The Ohio State University. Technical report.

Mandelbrot, B. B. (1997). *The Variation of Certain Speculative Prices*. Springer, New York.

McCurdy, T. H. and Morgan, I. (1992). Evidence of Risk Premiums in Foreign Currency Futures Markets. *Review of Financial Studies*, 5(1):65–83.

Mensi, W., Hammoudeh, S., Nguyen, D. K., and Yoon, S.-M. (2014). Dynamic spillovers among major energy and cereal commodity prices. *Energy Economics*, 43(January 2002):225–243.

Nakajima, J. and Omori, Y. (2009). Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics & Data Analysis*, 53(6):2335–2353.

Neal, R. M. (2000). Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9(2):249–265.

Nelson, C. R. . and Kim, M. J. . (1993). The Role of Small Sample Bias. *Journal of Finance*, 48(2):641–661.

Ng, A. (2000). Volatility spillover effects from Japan and the US to the PacificBasin. *Journal of International Money and Finance*, 19(2):207–233.

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140(2):425–449.

Pan, M.-s. and Hsueh, L. P. (1998). Transmission of Stock Returns and Volatility Between the U . S . and Japan : Evidence from the Stock Index Futures Markets. *Asia-Pacic Financial Markets*, 5:211–225.

Paye, B. S. (2012). Predictive Regressions for Aggregate Stock Market Volatility using Macroeconomic Variables. *Journal of Financial Economics*, 106(3):527–546.

Pesaran, H. and Timmermann, A. (1994). Forecasting stock returns An examination of stock market trading in the presence of transaction costs. *Jornal of Forecasting*, 13(August 1992):335–367.

Pesaran, H. M. and Timmermann, A. (1995). American Finance Association Predictability of Stock Returns: Robustness and Economic Significance. *Journal of Finance*, 50(4):1201–1228.

Rapach, D. E. and Wohar, M. E. (2006). In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance*, 13(2):231–247.

Rapach, D. E. and Zhou, G. (2013). Forecasting Stock Returns. In *Handbook of economic forecasting 2*, pages 328–383.

Reyes, M. G. (2001). Asymmetric volatility spillover in the Tokyo stock exchange. *Journal of Economics and Finance*, 25(2):206–213.

Robinson, P. M. (1978). Statistical Inference for a Random Coefficient Autoregressive Model. *Scandinavian Journal of Statistics*, 5(3):163–168.

Rodriguez, A. and Dunson, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian analysis (Online)*, 6(1):145–178.

Rodriguez, A. and ter Horst, E. (2008). Bayesian dynamic density estimation. *Bayesian Analysis*, 3(2):339–366.

Roth, M. (2013). On the Multivariate t Distribution. Technical Report 3059.

Rozeff, M. (1986). Dividend yields are equity risk premiums.

Schwert, G. W. (1989). Why Does Stock Market Volatility Change over Time? *Journal of Finance*, 44(5):1115–1153.

Sethuraman, J. (1994). A Constructive Definition of Dirichlet Priors. *Statistica Sinica*, 4:639–650.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, (6):425–431.

Silvennoinen, Annastiina Teräsvirta, T. (2009). Multivariate GARCH models. In *Handbook of financial time series*, pages 201–229.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4(2):183–204.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006). Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, 101(476):1566–1581.

Virbickaite, A., Ausin, M. C., and Galeano, P. (2015). Bayesian Inference Methods for Univariate and Multivariate Garch Models: a Survey. *Journal of Economic Surveys*, 29(1):76–96.

Walker, S. G. (2007). Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics Simulation and Computation*, (36):46–57.

Wei, K., Liu, Y., Yang, C., and Chaung, G. (1995). Volatility and price change spillover effects across the developed and emerging markets. *Pacific-Basin Finance Journal*, 3(1):113–36.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

West, M., Muller, P., and Escobar, M. D. (1994). *Hierarchical priors and mixture models , with application in regression and density estimation, in Aspects of Uncertainfy, eds. P. R. Freeman and A. F. M. Smith, New York: Wiley, pp. 363-386.* .