Three Essays on Forecasting Return Distributions with Mixture Modelling

Three Essays on Forecasting Return Distributions with Mixture Modelling

By Chenxing Li, M.Fin., B.Econ.

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University

Doctor of Philosophy (2020)

Hamilton, Ontario (DeGroote School of Business)

TITLE: Three Essays on Forecasting Return Distributions with Mixture Modelling

AUTHOR: Chenxing Li M.Fin., B.Econ. (McMaster University)

SUPERVISOR: Dr. John M. Maheu

COMMITTEE: Dr. Narat Charupat

COMMITTEE: Dr. Dean Mountain

COMMITTEE: Dr. Shui Feng

NUMBER OF PAGES: xvi, 186

To My Dear Parents

Abstract

This thesis examines three important topics in finance: co-jumps among risky assets and portfolios, proposing a new Bayesian semiparametric stochastic volatility model with Markovian mixtures, and whether better return density forecasts lead to economic gains in portfolio allocation practice.

First, jump models are useful in capturing skewed and/or leptokurtic financial returns. So far, research has focused on jumps in a single asset, including co-jumps between return and volatility. On the other hand, co-jumps among assets are also important especially in practices such as beta dynamics and portfolio allocation. I propose a parsimonious yet flexible multivariate GARCH-jump mixture model (MGARCH-jump model) with multivariate jumps that allows both jump sizes and jump arrivals to be correlated. The model identities co-jumps well and shows that both jump arrivals and jump sizes are highly correlated. The model also provides better prediction and better investment outcomes as opposed to the benchmark multivariate GARCH model with normal innovations (MGARCH-N model).

Second, I extend the Bayesian semiparametric stochastic volatility (SV-DPM) model of Jensen and Maheu (2010). Instead of using a Dirichlet process mixture (DPM) to capture return innovations, we use an infinite hidden Markov model (IHMM). This allows for time variation in the return density beyond that attributed to latent volatility dynamics. The new model (SV-IHMM) also nests the SV-DPM as a special

case and improves the density forecast from the SV model with Student-t innovations, SV-DPM and other benchmark models. Our model is applied to equity stock returns, foreign exchange rates, commodity prices and industrial growth rates. The model demonstrates robust and consistent gains in out-of-sample performance against other benchmarks. Furthermore, predictive densities from the SV-IHMM exhibit clear distributional change over time.

Third, I investigate the relationship between statistical improvements in density forecasts of returns and actual economic gains in portfolio allocation for a risk-averse investor. To aid this investigation, this chapter proposes a new multivariate Bayesian semiparametric model that has better out-of-sample density forecasts than benchmark models. Results show that this more sophisticated econometric model does provide positive economic gains whether the investor's utility is CRRA, CARA or quadratic. The economic gain diminishes when the investor is more risk-averse because she is moving away from risky investment positions.

Acknowledgements

I would have never been able to complete this thesis without the guidance of my great Ph.D. supervisor and other committee members, help from all my friends and unconditional support from my beloved family.

First, I want to thank my Ph.D. supervisor, Dr. John Maheu, for his patient guidance throughout the whole process of writing this thesis. He led me into the fabulous world of Bayesian econometrics. He has been helping me polishing my research, programming and writing skills. I also want to thank him for his completely unselfish mentoring and support for my academic and non-academic life. He supported me financially to numerous conferences, shared his experience on teaching and, more importantly, helped me get through a tough period in my personal life. Without him, this thesis wouldn't be possible from the very beginning.

Second, I want to thank my thesis committee members, Dr. Narat Charupat, Dr. Dean Mountain and Dr. Shui Feng, for their active help and constructive suggestions. They have been extremely kind and helpful in every stage of writing this thesis. Their sharp insights significantly helped this thesis to move forward. I also wish to thank all the audience to my BBL seminar presentations and the audience to RCEA Bayesian Workshop, SBIES, ESOBE and CFE conferences for their valuable comments.

Last but not least, I want to thank my parents for supporting my pursuit of a Ph.D. in the other side of the world, to thank the community of DeGroote School of Business for providing a supportive environment for my research, and to thank my

friends, Qiao Yang, Jia Liu, Azam Shamsi, Efthimios Nikolakoupolos, Javier Mella Barahona, Zehua Zhang, Punit Anand and many others, for creating a friendly and fun atmosphere through this stressful career.

Chapter 3 is joint work with Dr. John M. Maheu and Dr. Qiao Yang.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Chenxing Li, declare that this thesis titled, "Three Essays on Forecasting Return Distributions with Mixture Modelling" and the work presented in it are my own.

Chapter 1

Introduction

This thesis focuses on three important topics in forecasting return distributions, especially using mixture models. It differs from the traditional point forecast methods in that it can provide complete distributional information for the future instead of merely an expected value. This feature is of great importance to risk-related financial decision-making processes, for example, portfolio management, derivative investment, hedging strategies design, etc. This thesis enhances decisions by improving return density forecasts with new time-series econometric models.

The models proposed in this thesis are all in the rich class of discrete mixtures of distributions. This type of model usually consists of at least two components or kernels and mixes them with some particular weights. A common choice of kernel is Gaussian distribution for its desirable statistical properties (eg. stable) while other kernels are also possible. In this thesis, Gaussian kernels are implemented for all proposed models. In Bayesian econometrics, a mixture model is parametric if it has

a finite number of kernels and nonparametric if it has an infinite number of kernels. Both types are covered by the proposed models in this thesis.

The other crucial aspect of mixture models is the set of mixing weights. Mixing weights can be either static or dynamic over time. A static mixture is usually easier to estimate, requires fewer observations but the structure is more restrictive, while a dynamics mixture is opposite. Empirically, this is purely a trade-off problem. Therefore, there is no definitive answer to which mixture is better and it solely depends on the properties of data. This thesis provides examples of both static and dynamic mixtures.

Chapter 2 of this thesis proposes a new fully parametric mixture model to capture co-jumps among equity returns in addition to diffusion dynamics. Past researches have extensively investigated univariate jumps. Examples include Press (1967), Jorion (1988), Bates (1996), Chan and Maheu (2002), Bandi and Renò (2016) and many others. While co-jumps among different assets are not yet comprehensively researched due to its complexity. Current research is mainly restricted on either ex-post statistical tests that treat jump size and jump arrival as a whole or some primitive bivariate econometric models. For example, Bollerslev et al. (2008), At-Sahalia and Xiu (2016), Laurini and Mauad (2015) and Chua and Tsiaplias (2019). This chapter proposes a new multivariate mixture model that allows both jump sizes and jump arrivals to be correlated among assets separately at the same time.

This model consists of two components: a multivariate generalized autoregressive conditional heteroskedasticity (GARCH, Bollerslev 1986) component and a compound jump component. The multivariate GARCH (MGARCH) component incorporates a vector-diagonal representation of the BEKK-GARCH model (Engle and Kroner 1995). The compound jump component is the Hadamard product of a latent jump size variable and a jump arrival variable. The jump size variable is multivariate normal so that the jump sizes can be correlated. The jump arrival variable is multinomially distributed, and each outcome of trial represents a particular jump/co-jump combination so the jump arrivals can also be correlated and all possible jump/co-jump possibilities are allowed.[1]

The construction of this jump component has some nice features to return conditional moments. When further condition on jump arrivals, the conditional means and variances of returns are only affected when a jump occurs, and the conditional covariances are only affected when a co-jump occurs to the corresponding two asset returns. This allows the jump effects on conditional means vector and conditional covariance matrix to be activated/deactivated element-by-element or block-by-block. This feature helps to isolate jump/co-jump effects from normal periods.

This new model is tested by five trivariate datasets and one 5-dimensional dataset. The trivariate cases are General Electric (GE), Exxon (XOM), Wal-Mart (WMT), Microsoft (MSFT) and American Express (AXP) with their corresponding industry

[1]See Section 2.2 for details.

and the market. The 5-dimensional case is the combination of these five individual firms. I find that from January 1, 1990 to December 31, 2016, in the trivariate examples, most (but not all) of the market jumps coincide with an industrial jump and/or firm jump, and a considerable portion of industrial jumps is also accompanied with market and/or firm jumps. In the 5-dimensional example, the jumps are either mutual co-jumps or individual jumps, which shows low contagion effect across industries unless it's market-wide. As for jump sizes, all the jumps found are positively correlated. The new MGARCH-jump model forecasts significantly better than the benchmark MGARCH model with normal innovations (MGARCH-N model). Results show that this improvement is only restricted by some isolated periods. And in other periods, the two models perform approximately the same.

Two investment-related applications are also provided in the chapter. The first one extracts beta dynamics from those two models. Beta dynamics extracted from the MGARCH-N model is very different from that from the new model because it essentially performs excess smoothing when ignoring jumps. The second application shows the value at risk estimations by the two models over time for an equally weighted portfolio. The new model provides a similar estimation at a 10% level and a significantly more conservative estimation at a 1% level. This shows that the new model does not estimate the potential loss drastically when an investor cares about more regularly happened events, and this loss is properly reflected when an investor only cares about truly rare events.

Chapter 3 is joint work with John Maheu and Qiao Yang. This chapter extends the Bayesian semiparametric stochastic volatility model proposed by Jensen and Maheu (2010). Stochastic volatility (SV) models, first proposed by Taylor (1982), are very popular in modelling dynamic heteroskedasticity in financial time series. They can naturally generate fatter tails in unconditional distributions than homoskedastic models and GARCH models. However, the conditional distribution is usually assumed to be normal while much empirical evidence has been found against it. Examples can be found in Dieobold (1986), Barndorff-Nielsen (1997), Chib et al. (2002) and many others.

To resolve this issue, other than imposing a parametric assumption to innovations (eg. Student-t for a SV-t model), Jensen and Maheu (2010) propose a Bayesian semiparametric extension to an SV model by adding a Dirichlet process mixture (DPM) component. They assume the return innovation distribution is unknown and approximate it nonparametrically to generate potentially skewed and/or leptokurtic conditional returns. However, since the mixture in DPM is static, the innovation distribution, although approximated nonparametrically, is still constant over time.

Vo (2009), Shibata and Watanabe (2005) and others take another route and add a Markov switching (MS) scheme to SV models. The MS is also a mixture approach but the mixing weights depend on the state of the previous period. By adding an MS component, the innovation distribution is now allowed to change over time in a Markovian manner. The drawback is that one must predefine the number of states used in MS and it may be restrictive in many cases.

The model proposed in this chapter inherits the merits of both methods. We modify the model proposed by Jensen and Maheu (2010) and replace the DPM component by an infinite hidden Markov model (IHMM). The IHMM can be seen as a Bayesian nonparametric extension of an MS model, where the number of states is increased from a fixed finite number to infinity. The innovation distributions are unknown and approximated nonparametrically *and* they are allowed to be dynamic over time. The proposed SV-IHMM model is designed in such a way that the main level change in volatility dynamics is captured by the state-dependent component and the remaining smooth change is captured by the SV component.

We test four series to represent different data types: AAPL for equity returns, USD/CAD for foreign exchange returns, WTI crude oil for commodity returns and industrial production for macroeconomic growths. We find that SV parameters are estimated with reasonable values in every case with this extremely rich and flexible model. Many identified states suggest non-normality for innovation distributions. Our new model forecasts significantly better in terms of log predictive likelihood than all the benchmark models, including the already powerful SV-DPM and SV-t. Predictive density results show that innovation distributions are very likely to be dynamic over time. All the results are robust to different parameter settings of hyper-priors for precision parameters.

Chapter 4 investigates the relationship between a better return density forecasts and economic gains in portfolio allocation practice for a risk-averse investor. Traditionally, researchers mainly focus on improving volatility forecasts to achieve better portfolio

allocations. Examples are Fleming et al. (2001), Guidolin and Timmermann (2005, 2007), Kalotychou et al. (2014), Bollerslev et al. (2018) and many others. While other distributional information that is important to a risk-averse investor's decision making, such as skewness and tails, is conveniently ignored. In this chapter, I firstly propose a multivariate Bayesian semiparametric model that extends the SV-IHMM in Chapter 3 to a multivariate universe and employs a GARCH structure similar to the MGARCH-jump model in Chapter 2. The new model forecasts significantly better than the benchmark models. Then I prove it does provide positive ex-post economic gains in utility-based portfolio allocations.

The model proposed in this chapter has two components. The first component is a parametric MGARCH component. This component is slightly different from the one used in the MGARCH-jump model proposed in Chapter 2. An additional parameter is introduced here to capture potential asymmetric feedback in volatility dynamics. The second component is a Bayesian nonparametric component, IHMM. Similar to the SV-IHMM in Chapter 3, the introduction of the IHMM enables to approximate an unknown and time-varying innovation distribution nonparametrically. Unlike the SV-IHMM, the new MGARCH-IHMM captures the long-run volatility change through MGARCH dynamics. The additional state-dependent covariance parameter is parametrized centred at the identity matrix and serves as a multiplier to amplify or shrink the overall conditional covariances.

The proposed model is tested by monthly returns of the Fama-French 5 industry portfolios from July 1926 to August 2018. Empirical results show a strong regime-switching behaviour *in addition to* MGARCH dynamics, indicating that an MGARCH model alone is not sufficient in modelling equity returns. Further results reveal that the conditional covariance matrix is boosted up by the state-dependent component when the MGARCH covariance is increasing and vice versa. Out-of-sample log predictive likelihoods show very strong evidence supporting the MGARCH-IHMM against the benchmark MGARCH models and the IHMM.

To test the portfolio allocation performance, firstly, I solve a utility-based portfolio optimization problem using all the distributional information predicted by a model. The problem assumes a risk-averse investor allocates her wealth to a portfolio that contains multiple risky assets and a risk-free asset. This portfolio is optimized and rebalanced at the end of each period, and the beginning wealth of each period is normalized to 1. To evaluate the optimized portfolio, the investor is assumed to carry it over the next period when an ex-post return is realized. Then a break-even management fee that the investor is willing to pay to switch from one econometric model to another can be computed by the set of realized utilities generated by each model.

Results show that a risk-averse investor is always willing to pay a significant annual fee to switch from any benchmark model to my new model. This result is robust to three different utility functions (constant relative risk aversion, constant absolute risk aversion and quadratic utility) and robust to different risk aversion parameter levels.

The rest of this thesis is organized as follows. Chapter 2 proposes and tests a new multivariate GARCH-jump mixture model. Chapter 3 proposes a Bayesian semi-parametric SV model with Markovian mixture and tests it with a variety of asset returns. Chapter 4 investigates the economic gains in portfolio allocations of using a new MGARCH-IHMM model that forecasts better. Chapter 5 concludes.

Chapter 2

A Multivariate GARCH-Jump Mixture Model

2.1 Introduction

It is well-known that daily stock returns exhibit both continuous and occasional discontinued changes, also known as jumps. Popular volatility models include generalized autoregressive conditional heteroskedasticity (GARCH) (Bollerslev 1986) and stochastic volatility (SV). Both models work well, especially when volatility is persistent. As regarding jumps, many efforts have been made to model a univariate stock price process since Press (1967), who appends a simple geometric Brownian motion with a compounded Poisson counting process. However, how jump arrivals and jump sizes affect each other among assets remain unclear. This chapter proposes a parsimonious and yet flexible model that allows both jump arrivals and jump sizes to

be cross-sectionally correlated. We find that although jumps arrive infrequently in daily data, it's more likely to have multiple assets jump together rather than independently. Moreover, whenever they jump together, they are very likely to jump in the same direction, and the magnitudes are strongly correlated.

Introducing jumps can affect the conditional mean, conditional variance as well as higher-order unconditional moments such as skewness and kurtosis. This captures the empirical fact that the unconditional distribution of stock returns is skewed and leptokurtic relative to a normal distribution (for example, Corrado and Su 1996; Fama 1965; Kon 1984; Mandelbrot 1963; Mills 1995; Peiró 1994, 1999; Praetz 1972). Moreover, jumps are especially helpful in explaining large extreme return changes like market crushes.

Single-asset based jump models have been extensively investigated. The most commonly used model is the compounded Poisson process model introduced by Press (1967), where the jump arrival is a Poisson counting process and the jump size is normally independently distributed. Based on that, Ball and Torous (1983) provide a Bernoulli jump model through a discrete approximation. Jorion (1988) implements the Poisson jump model with an autoregressive conditional heteroskedasticity (ARCH) diffusion component on both the foreign exchange and the stock market. Vlaar and Palm (1993) test the Bernoulli jump model on the former European Monetary System for different drift (AR) and diffusion (ARCH/GARCH) representations, and Nieuwland et al. (1994) further test the Poisson model on the same topic. Bates (1996, 2000)

and Pan (2002) find a SV-jump (Poisson) model can best explain the price behaviour of foreign exchange and stock options than other available alternatives.

Aside from return jumps, Eraker (2004) and Caporin et al. (2014) confirm jumps also exist in volatility dynamics. Eraker et al. (2003) also study its impact on option pricing. Bandi and Renò (2016) and Jacod et al. (2017) take a further step and inspect the relation between return jumps and volatility jumps, and Chorro et al. (2017) further study how such return-volatility co-jumps affect density forecast. Another direction of extension from the basic jump-diffusion model is to consider time-varying jumps. Oldfield et al. (1977) expand the Poisson jump model into autoregressive jump sizes, while Chan and Maheu (2002) introduce an autoregressive conditional jump intensity (ARJI) model that allows jump intensities to be autocorrelated. Implementations of the ARJI model include Maheu and McCurdy (2004), who found jumps usually arrive in clusters, Chan and Feng (2012) and Maheu et al. (2013), who inspect the relation of jumps and risk premiums.

As mentioned above, most of the research on jumps have been focused on modelling a univariate asset return, with covariation among different assets ignored. Bollerslev et al. (2008) are among the first who identify the existence of co-jumps and provide a test for co-jumps in multiple assets. Gilder et al. (2014) confirm the empirical findings in Bollerslev et al. and provide another test. Mancini and Gobbi (2012) suggest a nonparametric estimator based on realized covariation. Similarly, At-Sahalia and Xiu (2016) decompose quadratic variation into continuous and discontinuous components to estimate co-jumps. Bibinger and Winkelmann (2015), Winkelmann et al. (2016)

also concentrate on extracting co-jump from quadratic covariation and introduce a truncated estimator. Caporin et al. (2017) further apply this estimator in a higher dimensional experiment. Other attempts are Gobbi and Mancini (2007) to derive a bivariate parametric co-jump estimator, and Novotný and Urga (2018) to introduce a new approach to test the existence of co-jumps.

All of the above co-jump estimators and tests rely on asymptotic assumptions and focus on identifying co-jumps but not the dynamics which are needed for prediction purpose. Moreover, exploiting co-jumps from quadratic covariation makes it impossible to study the cross-sectional relation of jump arrivals and jump sizes separately. This type of nonparametric estimator is designed to investigate the ex-post relation of jump sizes but ignores the information embedded in jump arrivals. Also, when extracting jumps from quadratic variation and bipower variation, one loses important information on the signs of jump sizes. One may argue that the nonparametric methods are superior as they don't rely on any distributional assumption, but to forecast jumps, an additional econometric model is still required. And the proposed MGARCH-jump model can flexibly model jump size distributions and covariances, and all jump arrival combinations are also allowed.

For parametric models, Laurini and Mauad (2015) propose a bivariate SV model with built-in co-jumps, but idiosyncratic jumps are not allowed in the model; and Zhang et al. (2019) provide a goodness-of-fit test for this type of models. Chua and Tsiaplias (2019) introduce another model with correlated jump sizes but independent jump arrivals and autocorrelated jump intensities.

In this chapter, I propose a new fully parametric model in which there's an embedded component that allows the returns to jump separately or jump together with correlated jump sizes. This model overcomes the drawbacks mentioned above. The proposed MGARCH-jump model is well identified with reasonable GARCH parameters and high no-jump probabilities. The results also show that when jumps occur, it is more likely that several stocks jump together, with strong and positive jump size correlations. This is especially important when extreme returns occur, which is confirmed by predictive likelihoods. The MGARCH-jump model performs similarly to the benchmark MGARCH-N model in normal times but outperforms it in high volatility times.

Some pure jump processes like Hawkes process (Hawkes 1971) are also popular in modelling jumps. Examples of Hawkes process models include Bacry et al. (2013), Rambaldi et al. (2014), At-Sahalia et al. (2015), At-Sahalia and Hurd (2015), Bormetti et al. (2015), etc. However, these intriguing works view all price changes as pure jumps with self-exciting intensities, hence unrelated to the discussion of this chapter.

In this chapter, Section 2.2 formally describes the layout and some properties of the proposed multivariate GARCH-jump mixture model (MGARCH-jump model hereafter) in detail. Section 2.3 illustrates the data sources along with the cleaning and transformation methods. Section 2.4 discusses the estimated and forecast results from the MGARCH-jump model for five trivariate examples of an individual asset, corresponding industry and the market. Section 2.5 tests the model in a higher dimension

with five stocks altogether. Section 2.6 shows several applications by the MGARCH-jump model, including the effect of jumps on beta dynamics and the predictive Value at Risk for an equally-weighted portfolio. And Section 2.7 is the conclusion.

2.2 Model

In this section, we present a discrete-time MGARCH-jump model for financial returns. The model has a multinomial jump arrival and a multivariate normal jump size component. Let $\boldsymbol{r}_t = (r_{t,1},\ r_{t,2},\ \cdots,\ r_{t,N})'$ be a $N \times 1$ vector of returns of N assets at time t. $\boldsymbol{r}_t$ is specified as

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t, \tag{2.1}$$

$$\boldsymbol{\epsilon}_t = \boldsymbol{\epsilon}_{1,t} + \boldsymbol{\epsilon}_{2,t}, \tag{2.2}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)'$ is a $N \times 1$ vector of constant drift, $\boldsymbol{\epsilon}_{1,t}$ is a $N \times 1$ vector of return innovations with $\mathrm{E}\left(\boldsymbol{\epsilon}_{1,t} | \boldsymbol{r}_{1:t-1}\right) = 0$, where $\boldsymbol{r}_{1:t-1} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_{t-1}\}$ denotes the information set at time $t-1$. In particular,

$$\boldsymbol{\epsilon}_{1,t} = \boldsymbol{H}_t^{1/2} \boldsymbol{z}_t, \tag{2.3}$$

$$\boldsymbol{z}_t \sim NID\left(\boldsymbol{0}, \boldsymbol{I}\right), \tag{2.4}$$

where $\boldsymbol{z}_t$ is a $N \times 1$ vector of multivariate NID shocks, $\boldsymbol{H}_t^{1/2}$ is the Cholesky decomposition of a $N \times N$ conditional covariance matrix regulated by a multivariate GARCH

structure. $\boldsymbol{\epsilon}_{2,t}$ is a $N \times 1$ vector of jump innovations also with mean of zero.

$$\boldsymbol{\epsilon}_{2,t} = \boldsymbol{J}_t - \mathrm{E}\left(\boldsymbol{J}_t | \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right), \tag{2.5}$$

where $\boldsymbol{J}_t = (J_{t,1}, J_{t,2}, \ldots, J_{t,N})'$ is a $N \times 1$ vector of jumps, $\boldsymbol{\Theta}$ is the union set of all parameters, and $\mathrm{E}\left(\boldsymbol{\epsilon}_{2,t} | \boldsymbol{r}_{1:t-1}\right) = 0$. Note that the conditional expectation of jumps is removed from the model, so $\mathrm{E}\left(\boldsymbol{r}_t | \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu}$ for all t. This feature provides a constant drift without jump effects as in Merton (1976). $\boldsymbol{\epsilon}_{1,t}$ and $\boldsymbol{\epsilon}_{2,t}$ are contemporaneously independent with each other.

2.2.1 Vector-Diagonal GARCH (VD-GARCH)

There are many approaches to extend the traditional version of the GARCH model (Bollerslev 1986) from a univariate to a multivariate world. We use a slightly modified version of the vector diagonal GARCH (VD-GARCH) model introduced by Ding and Engle (2001):

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}{}' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}, \tag{2.6}$$

where $\odot$ is the Hadamard product operator that performs element-by-element multiplication, $\boldsymbol{C}$ is a $N \times N$ lower triangular matrix, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are all $N \times 1$ vector of parameters. $\boldsymbol{\epsilon}_{t-1}$ includes both continuous shocks $\boldsymbol{\epsilon}_{1,t-1}$ and jump shocks $\boldsymbol{\epsilon}_{2,t-1}$. It is also natural to consider $\boldsymbol{\epsilon}_{t-1}$ incorporating only continuous shocks with $\boldsymbol{\epsilon}_{t-1} = \boldsymbol{r}_{t-1} - \boldsymbol{\mu} - \boldsymbol{J}_{t-1} + \mathrm{E}\left(\boldsymbol{J}_{t-1} | \boldsymbol{r}_{1:t-2}\right)$, but then the past jump series are propagated

into future $\boldsymbol{H}_t$ making the model likelihood path-dependent and thus the sampling for jump components extremely difficult.

The VD-GARCH specification is a simplified version of the BEKK model (Engle and Kroner 1995) and inherits the property that guarantees $\boldsymbol{H}_t$ to be positive definite if $\boldsymbol{H}_0$ is positive definite. Specifically, for each element $h_{t,ij}$ in matrix $\boldsymbol{H}_t$,

$$h_{t,ij} = \omega_{ij} + \alpha_i \alpha_j \epsilon_{t-1,i} \epsilon_{t-1,j} + \beta_i \beta_j h_{t-1,ij}, \tag{2.7}$$

where $\omega_{ij} = (\boldsymbol{C}\boldsymbol{C}')_{ij}$. No further restriction on parameters is required other than stationary conditions of $\alpha_i^2 + \beta_i^2 < 1 \ \forall i$ and $(\boldsymbol{C})_{ii} > 0$ for better identification. Note that ω_{ij} does not need to be restricted.

2.2.2 A Compound Multinomial Jump Structure

Most of the past univariate jump models parametrize jumps as a compound Poisson process follows Press (1967). Although a Poisson process fits well in univariate continuous-time models, it is not easily extended to a higher dimension with sufficient flexibility and dependence. While empirically, all the observed data is discrete in time, so as suggested in Ball and Torous (1983), Bernoulli jump is a good discrete approximation of a Poisson process over a small time interval and also more intuitive. One nice feature of Bernoulli jump is that it's much easier to generalize into the multivariate universe. A multinomially distribution with only one trial can perfectly fit

all possible jump/co-jump combination patterns. Therefore, for vector $\boldsymbol{J}_t$,

$$\boldsymbol{J}_t = \boldsymbol{Y}_t \odot \boldsymbol{B}_t, \tag{2.8}$$

$$\boldsymbol{Y}_t \sim N\left(\boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J\right), \tag{2.9}$$

where $\boldsymbol{Y}_t$ is a $N \times 1$ vector of jump sizes that are multivariate normally distributed with mean vector $\boldsymbol{\mu}_J$ and covariance matrix $\boldsymbol{\Sigma}_J$, and $\boldsymbol{B}_t = (B_{t,1}, B_{t,2}, \ldots, B_{t,N})'$ is a $N \times 1$ vector of jump indicators with each element $B_{t,i}$ being the result of a Bernoulli trial, $B_{t,i} \in \{0, 1\}$ for $i = 1, \ldots, N$. Let $L = 2^N$,

$$\boldsymbol{B}_t \sim \textit{multinomial}\left(1, p_1, \ldots, p_L\right), \tag{2.10}$$

where $\sum_{j=1}^{L} p_j = 1$, $B_{t,i} = 1$ indicates there is a jump for asset i at time t, and $B_{t,i} = 0$ otherwise. The parameter p_j is the jump/co-jump probability. Unlike univariate models, where the jump intensity parameter represents the probability of jump arrivals, in this specification, the jump/co-jump probability p_j is a separate probability assigned to each possible jump/co-jump outcome. Admittedly, it may look a little abstract and tedious, and the number of probabilities to be estimated increases exponentially as N grows, but it's necessary so to maintain the flexibility in jump arrivals, and it's relatively easy to sample. To be more specific, define a $2^N \times N$ matrix $\boldsymbol{\Omega}_B$ that contains all possible outcomes of $\boldsymbol{B}_t$, with each row being one exclusive possible value of $\boldsymbol{B}_t$, and $\boldsymbol{p} = (p_1, p_2, \ldots, p_L)'$ is a vector of corresponding jump probabilities. In a trivariate case, there are $2^3 = 8$ possible outcomes of $\boldsymbol{B}_t$: one trivariate co-jump

(1 1 1); three bivariate co-jumps (1 1 0), (1 0 1) and (0 1 1); three idiosyncratic jumps (1 0 0), (0 1 0) and (0 0 1); and one no jump outcome (0 0 0). This covers all possible jump patterns including all-asset co-jumps and subset co-jumps. Each outcome is associated with one probability element in $\boldsymbol{p}$. Note that $\boldsymbol{\Omega}_B$ is neither a parameter or a latent variable. It's a generated constant and solely depends on N. The order of combinations of $\boldsymbol{B}_t$ in $\boldsymbol{\Omega}_B$ (how rows are stacked in $\boldsymbol{\Omega}_B$) does not matter, but it needs to be consistent throughout the sampling procedure to avoid any unnecessary order switching.

One merit of this specification is that one can easily verify whether the jumps are cross-sectionally independent through these probabilities. Our empirical results show that the jump arrivals are strongly correlated cross-sectionally, so the multinomial assumption offers an accurate model of jump dependencies.

Besides jump arrivals, the multivariate normal structure naturally connects jump sizes among assets through the covariance matrix $\boldsymbol{\Sigma}_J$. As a result, in this model, one can easily extract the correlation of jump arrivals and that of jump sizes separately, so question like "whether and when do they jump together" and "how do they jump together" can be answered explicitly.

2.2.3 Conditional Moments

Most of the past research regarding cross-sectional co-jumps focus on estimating the compound jump process ($\boldsymbol{J}_t$) directly, and study its effect on the conditional

moments of return. Define the moments only conditional on the past information set as "ex-ante", and the moments further conditional on jump arrivals $\boldsymbol{B}_t$ as "ex-post". The first two ex-ante conditional moments of jump $\boldsymbol{J}_t$ are[1]

$$\mathrm{E}\left(\boldsymbol{J}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p} = \boldsymbol{\mu}_J \odot \left(\sum_{j=1}^{2^N} \boldsymbol{\Omega}_j p_j\right), \tag{2.11}$$

and

$$\mathrm{Cov}\left(\boldsymbol{J}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \left(\boldsymbol{\Sigma}_J + \boldsymbol{\mu}_J\boldsymbol{\mu}_J{}'\right) \odot \left(\sum_{j=1}^{2^N} p_j \boldsymbol{\Omega}_j \boldsymbol{\Omega}_j{}'\right) - \boldsymbol{\mu}_J\boldsymbol{\mu}_J{}' \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}\boldsymbol{p}'\boldsymbol{\Omega}_B, \tag{2.12}$$

where $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \boldsymbol{\theta}_H, \boldsymbol{p}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma_J})$, and $\boldsymbol{\Omega}_j$ is the jth row of $\boldsymbol{\Omega}_B$. Similarly, the first two conditional moments of return are

$$\mathrm{E}\left(\boldsymbol{r}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu}, \tag{2.13}$$

$$\mathrm{Cov}\left(\boldsymbol{r}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{H}_t + \mathrm{Cov}\left(\boldsymbol{J}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right). \tag{2.14}$$

The conditional mean of returns is simply $\boldsymbol{\mu}$ because both MGARCH volatility and jump innovation have a mean of zero. The conditional covariance of returns is the aggregation of conditional MGARCH volatility and conditional covariance of jumps.

[1]Proof can be found in Appendix 1.

Unique in this new model, the ex-post moments conditional on jump arrivals ($\boldsymbol{B}_t$) show more interesting properties. The first two conditional moments are:

$$\mathrm{E}\left(\boldsymbol{r}_t|\boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu} + \boldsymbol{\mu}_J \odot \left(\boldsymbol{B}_t - \boldsymbol{\Omega}_B{}'\boldsymbol{p}\right) \tag{2.15}$$

$$\mathrm{Cov}\left(\boldsymbol{r}_t|\boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{H}_t + \boldsymbol{B}_t\boldsymbol{B}_t{}' \odot \boldsymbol{\Sigma}_J \tag{2.16}$$

Because $\boldsymbol{B}_t\boldsymbol{B}_t{}'$ is positive semi-definite, and both $\boldsymbol{H}_t$ and $\boldsymbol{\Sigma}_J$ are positive definite, the conditional covariance of $\boldsymbol{r}_t$ is positive definite. To be more specific,

$$\mathrm{E}\left(\boldsymbol{r}_t|\boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu} + \begin{pmatrix} B_{t,1}\mu_{J,1} \\ B_{t,2}\mu_{J,2} \\ \vdots \\ B_{t,N}\mu_{J,N} \end{pmatrix} - \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p} \tag{2.17}$$

$$\mathrm{Cov}\left(\boldsymbol{r}_t|\boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{H}_t + \begin{pmatrix} B_{t,1}^2\sigma_{J,1}^2 & B_{t,1}B_{t,2}\sigma_{J,12} & \cdots & B_{t,1}B_{t,N}\sigma_{J,1N} \\ B_{t,2}B_{t,1}\sigma_{J,21} & B_{t,2}^2\sigma_{J,2}^2 & \cdots & B_{t,2}B_{t,N}\sigma_{J,2N} \\ \vdots & \vdots & \ddots & \vdots \\ B_{t,N}B_{t,1}\sigma_{J,N1} & B_{t,N}B_{t,2}\sigma_{J,N2} & \cdots & B_{t,N}^2\sigma_{J,N}^2 \end{pmatrix} \tag{2.18}$$

Clearly,

$$B_{t,i}B_{t,j} = \begin{cases} 1 & \text{if } B_{t,i} = B_{t,j} = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

This ensures which element(s) in $\boldsymbol{\mu}_J$ and $\boldsymbol{\Sigma}_J$ should be turned on and thus affect conditional means and covariances among asset returns. The corresponding element $\mu_{J,i}$ and $\sigma^2_{J,i}$ will be turned on if and only if asset i jumps, and $\sigma_{J,ij}$, where $i \neq j$, will be turned on if and only if asset i and asset j both jump at the same time. This property helps to capture the co-jump behaviour among assets and reflect it directly to return covariances. If there's no jump for all N assets, then $\boldsymbol{B}_t = (0, 0, \ldots, 0)'$, so $\mathrm{E}\left(\boldsymbol{r}_t | \boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu} - \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}$ and $\mathrm{Cov}\left(\boldsymbol{r}_t | \boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{H}_t$, which reduces to the results from basic dynamic volatility models such as MGARCH. If all N assets jump, then $\boldsymbol{B}_t = (1, 1, \ldots, 1)'$, so $\mathrm{E}\left(\boldsymbol{r}_t | \boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu} + \boldsymbol{\mu}_J - \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}$ and $\mathrm{Cov}\left(\boldsymbol{r}_t | \boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{H}_t + \boldsymbol{\Sigma}_J$. In other cases, only a sub-block of $\boldsymbol{\Sigma}_J$ is turned on. For example, in a trivariate case with a bivariate co-jump occurring, say $\boldsymbol{B}_t = (1, 1, 0)'$, two elements in $\boldsymbol{\mu}_J$ and four elements in $\boldsymbol{\Sigma}_J$ are turned on:

$$\mathrm{E}\left(\boldsymbol{r}_t | \boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu} + \begin{pmatrix} \mu_{J,1} \\ \mu_{J,2} \\ 0 \end{pmatrix} - \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}$$

$$\mathrm{Cov}\left(\boldsymbol{r}_t | \boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{H}_t + \begin{pmatrix} \sigma^2_{J,1} & \sigma_{J,12} & 0 \\ \sigma_{J,21} & \sigma^2_{J,2} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This is consistent with the intuition that conditional mean and variance can only be affected when the corresponding asset jumps and conditional covariance can only be

affected when the two corresponding assets jump together. Obviously, this model supports all jump/co-jump possibilities and channels the jump/co-jump effects into conditional moments as desired. As for correlations, computed by $\frac{h_{t,ij}+B_{t,i}B_{t,j}\sigma_{J,ij}}{\sqrt{\left(h_{t,ii}+B_{t,i}^2\sigma_{J,i}^2\right)\left(h_{t,jj}+B_{t,j}^2\sigma_{J,j}^2\right)}}$ ex-post and $\frac{h_{t,ij}+\sigma_{J,ij}}{\sqrt{\left(h_{t,ii}+\sigma_{J,i}^2\right)\left(h_{t,jj}+\sigma_{J,j}^2\right)}}$ if there's a co-jump, the jump effect depends on the scale of the correlation computed by $\boldsymbol{H}_t$ (MGARCH correlation hereafter). The co-jumps do not necessarily increase the overall return correlation as in the covariance case. For example, the jump impact on the overall ex-post correlation is determined by the comparison of the square of MGARCH correlation $\rho^2_{MGARCH} = \frac{h_{t,ij}^2}{h_{t,i}h_{t,j}}$ and $\tilde{\rho}^2 = \frac{\sigma_{J,ij}^2+2h_{t,ij}\sigma_{J,ij}}{\sigma_{J,i}^2\sigma_{J,j}^2+h_{t,i}\sigma_{J,j}^2+h_{t,j}\sigma_{J,i}^2}$. If ρ^2_{MGARCH} is greater than $\tilde{\rho}^2$, then the overall return correlation will decrease, and vice versa. This behaviour clearly can affect diversification benefits in a portfolio.

2.2.4 Sampling Algorithm

This model consists of two latent variables, $\boldsymbol{Y}_t$ and $\boldsymbol{B}_t$, so not easy to estimate by classical methods. Instead, we apply a typical Bayesian method, Markov chain Monte Carlo (MCMC). Bayesian methods are powerful since they allow one to treat latent variables as parameters to estimate and to break complex problems into relatively simple pieces. In each piece, one can estimate only a subset of parameters and treat others as given. The estimation of this model also takes advantage of these properties. A full MCMC run contains M_0+M iterations, where the first $M_0 = 10000$ are burn-in samples, and the rest $M = 10000$ are posterior draws. Each MCMC iteration is as follow:

1. $\boldsymbol{\mu}|\boldsymbol{r}_{1:T}, \boldsymbol{H}_{1:T}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}, \boldsymbol{p}$.

2. $\boldsymbol{\theta}_H|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}, \boldsymbol{p}$, where $\boldsymbol{\theta}_H = (\boldsymbol{C}, \boldsymbol{\alpha}, \boldsymbol{\beta})'$.

3. $\boldsymbol{B}_t|\boldsymbol{r}_t, \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{p}$.

4. $\boldsymbol{p}|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{H}_{1:T}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}$.

5. $\boldsymbol{Y}_t|\boldsymbol{r}_t, \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}$.

6. $\boldsymbol{\mu}_J|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{H}_{1:T}, \boldsymbol{\Sigma}_J, \boldsymbol{Y}_{1:T}, \boldsymbol{B}_{1:T}, \boldsymbol{p}$.

7. $\boldsymbol{\Sigma}_J|\boldsymbol{\mu}_J, \boldsymbol{Y}_{1:T}$.

Steps 1, 3, 5, 7 are simply Gibbs samplers, and steps 2, 4, 6 are Metropolis-Hastings (MH) due to unknown type of posterior distributions.[2] And the jump arrivals ($\boldsymbol{B}_t$) and jump sizes ($\boldsymbol{Y}_t$) can be estimated as:

$$\mathrm{E}\left(\boldsymbol{B}_t\right) \approx \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{B}_t^{(i)} \tag{2.20}$$

and

$$\mathrm{E}\left(\boldsymbol{Y}_t\right) \approx \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{Y}_t^{(i)} \tag{2.21}$$

[2]Details of each sampling step can be found in Appendix 2.

We apply uninformative priors for all parameters and let data determine the posteriors. The prior choices are:

$$\begin{aligned}
\boldsymbol{\mu} &\sim N\left(\mathbf{0}, 100\boldsymbol{I}\right) \\
\boldsymbol{\theta}_H &\sim N\left(\mathbf{0}, 100\boldsymbol{I}\right) \\
\boldsymbol{p} &\sim Dir\left(1, \ldots, 1\right) \\
\boldsymbol{\mu_J} &\sim N\left(\mathbf{0}, 100\boldsymbol{I}\right) \\
\boldsymbol{\Sigma_J} &\sim IW\left(N+2, \boldsymbol{I}\right)
\end{aligned}$$

2.2.5 Predictive Likelihood

Recall that $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \boldsymbol{\theta}_H, \boldsymbol{p}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma_J})$, and the predictive likelihood is computed by integrating out all parameters $\boldsymbol{\Theta}$. From equation (2.15) and (2.16), the conditional distribution of returns conditional on jump arrivals is simply a multivariate normal distribution. In order to compute the predictive likelihood for the whole model, one can first compute the likelihood of this conditional distribution $p\left(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}, \boldsymbol{B}_{t+1}\right)$ then integrate out jump arrivals:

$$\begin{aligned}
p\left(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}\right) &= \int\int p\left(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}, \boldsymbol{B}_{t+1}\right) p\left(\boldsymbol{B}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}\right) p\left(\boldsymbol{\Theta}|\boldsymbol{r}_{1:t}\right) \mathrm{d}\boldsymbol{\Theta}\mathrm{d}\boldsymbol{B}_{t+1} \\
&\approx \sum_{i=1}^{M}\sum_{j=1}^{L} p\left(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}^{(i)}, \boldsymbol{B}_{t+1}^{(j)}\right) p_j^{(i)}, \qquad (2.22)
\end{aligned}$$

$$\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}, \boldsymbol{B}_{t+1} \sim N\left(\boldsymbol{\mu} + \boldsymbol{\mu}_J \odot \left(\boldsymbol{B}_{t+1} - \boldsymbol{\Omega}_B{}'\boldsymbol{p}\right), \boldsymbol{H}_{t+1} + \boldsymbol{B}_{t+1}\boldsymbol{B}_{t+1}{}' \odot \boldsymbol{\Sigma}_J\right), \quad (2.23)$$

$$\boldsymbol{B}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta} \sim Multinomial\left(1, p_1, \ldots, p_L\right), \qquad (2.24)$$

where $p(\boldsymbol{\Theta}|\boldsymbol{r}_{1:t})$ is the posterior of all parameters conditional on sub-sample $\boldsymbol{r}_{1:t}$, $\boldsymbol{\Theta}^{(i)}$ and $p_j^{(i)}$ are the parameters drawn in the ith iteration, and $\boldsymbol{B}_{t+1}^{(j)}$ is the jth possible outcome from $\boldsymbol{\Omega}_B$. The out-of-sample likelihood is the product of the predictive likelihood evaluated in each period:

$$p(\boldsymbol{r}_{t+1:T}|\boldsymbol{r}_{1:t}) = \prod_{l=t}^{T} p(\boldsymbol{r}_{l+1}|\boldsymbol{r}_{1:l}). \tag{2.25}$$

In practice, one usually computes the log-predictive likelihood by summing the log-predictive likelihood evaluated in each out-of-sample period.

2.3 Data

The properties of this model enable an easy solution to questions such as "What's the jump/co-jump relation among different assets?" and "Does a particular asset more likely to jump idiosyncratically or with the corresponding industry or even with the market?" To answer these questions, we select General Electric (GE), Exxon (XOM), Wal-Mart (WMT), Microsoft (MSFT) and American Express (AXP), representing electrical equipment industry, petroleum and natural gas industry, retail industry, computer software industry and banking industry, respectively. All the return data is retrieved from the Center for Research in Security Prices (CRSP) database, specifically daily holding-period returns of each selected stock and the value-weighted market portfolio (MKT). Also, daily risk-free rates and Fama-French 49 industry portfolio are

obtained from Kenneth French's website. To match each stock with its corresponding industry portfolio, SIC codes of the above stocks are also acquired from CRSP. All the returns are collected from January 1, 1990 to December 31, 2016, with 6805 observations in total. Earning announcement dates are gathered from the I/B/E/S database.

Table 2.1 illustrates descriptive statistics of daily continuously compounded returns for the selected stocks as well as the value-weighted market portfolio. The statistics are computed from log returns in percentage value after dropping all the missing values and non-number observations.

2.4 Individual Stocks, Corresponding Industry and the Market Co-Jumps

2.4.1 Estimation

The first example is to estimate the trivariate model each for GE, XOM, WMT, MSFT and AXP coupled with their corresponding industry and the market respectively. Table 2.2 reports the results for these trivariate estimates. All the posteriors are in reasonable regions proved by a vast amount of previous researches, with small means (μ_i of 0.02 – 0.05), low shock parameters (α_i of 0.15 – 0.20), and high persistent parameter (β_i of 0.97 – 0.98) from the MGARCH specification.

All trivariate models indicate that "no jump" is the most likely outcome. No-jump probabilities ($p_{\overline{STK},\overline{IND},\overline{MKT}}$) ranges from 0.82 to 0.88. The jump size variances are considerably large, with jump size variance for individual stocks ($\sigma^2_{J,STK}$) ranging from 2.14 to 9.45, for industries ($\sigma^2_{J,IND}$) ranging from 1.61 to 3.66, for the market ($\sigma^2_{J,MKT}$) ranging from 1.00 to 1.52. Jump size covariances are all positive and also relatively large, with covariance for stocks and corresponding industry ($\sigma_{J,STK,IND}$) ranging from 1.83 to 4.25, for stocks and the market ($\sigma_{J,STK,MKT}$) ranging from 1.03 to 3.08, for industries and the market ($\sigma_{J,IND,MKT}$) ranging from 1.09 to 2.43. This confirms the fact that jumps are rare but extreme movements in stock returns.

From the basic probability rules, if jumps are cross-sectionally independent, a co-jump joint probability should be equal to the product of marginal jump probabilities for the corresponding assets. Panel A of Table 2.3 compares the co-jump joint probabilities with the product of its marginal probabilities. The co-jump probabilities range from 0.0595 to 0.0984, while the product of marginal probabilities ranges from 0.0007 to 0.0250. Jump arrivals are strongly correlated as the joint probabilities and product of marginal probabilities are very different from each other. The differences are even greater when the number of assets in a co-jump is greater. For example, the bivariate co-jump probabilities of GE and its industry, GE and the market, GE's industry and the market are 0.0984, 0.0980, 0.0986 respectively, while the products of marginal jump probabilities are 0.0181, 0.0159, 0.0121, respectively. They are very different but still the same decimals. In contrast, the joint probability of a trivariate co-jump with GE, its industry and the market jump all together is 0.0954, while the

product of marginal jump probabilities is 0.0019, 50 times less than the corresponding co-jump probability.

Panel B further computes the co-jump probabilities conditional on different univariate jumps, which indicate the proportion of co-jumps in each univariate jumps. The results show that if the market jumps, each selected stock and its industry will most likely jump as well. The electrical equipment industry, retail industry and banking industry are more likely to jump along with the market when an unusual condition occurs, more than half of jumps in the petroleum and natural gas industry and software industry coincide with the market jumps. For XOM, WMT and MSFT, all the firms in their industry are most likely to jump together with them, with probabilities of co-jump with their industries conditional on stock jumps being 0.9496, 0.9517 and 0.9783 respectively. When WMT jumps, the whole market is very likely to follow, with a probability of co-jump with the market conditional on stock jumps being 0.8102. GE and AXP also have a strong influence on their industry when they jump, with co-jump probability conditional on stock jumps of 0.6392 and 0.5457 respectively.

Figure 2.1 plots the posteriors of jump arrival probabilities for each of the five stocks with their corresponding industry and the market. Most of the jump arrivals are aligned together, which confirms the results in panel B of Table 2.3. Figure 2.2 plots jump size realizations over time. The figure shows jump size realizations are relatively large (up to 10% and -10%) and infrequent. The results are more clear if we focus on a small period. Take AXP from January 1, 2007 to December 31, 2009 as an example shown in Figure 2.3, jump probability is usually high around

quarterly earnings announcement dates. Beyond that, the progression of the sub-prime mortgage crisis plays an important role in jump dynamics. For instance: on March 13, 2007, reacted to the potential risk of sub-prime mortgages, causing a -2.93% jump on AXP, a -2.73% jump on the banking industry and a -1.86% jump on the market. On November 1, 2007, after a previous interest rate cut, the Federal Reserve injected 41 billion dollars into money supply with a response of -3.14% AXP jump, -3.49% industrial jump and -2.12% market jump. On September 29, 2008, the House of Representatives rejected the bailout plan, accompanying a -5.24% of the AXP jump, a -3.85% of the industrial jump and a -3.18% of market jump. All the above jumps have posterior jump probabilities greater than 0.9. Other major events can also be matched with these realizations. The proposed MGARCH-jump model correctly identifies the time and magnitude of a jump event.

Figure 2.4 scatters the jump probabilities and jump sizes across AXP, the industry and the market pairwise during the recent financial crisis. The top three graphs plot the jump probabilities with a 45-degree line. In most cases, the jump probabilities are low with points concentrating at the bottom left corner. Between AXP and the industry, it's a lot more likely to jump only in AXP than only in the industry, as most of the points lie below the 45-degree line with only a few exceptions. As for AXP and the market, nearly all the points are below the 45-degree line, indicating that any market jump is very likely an AXP-market co-jump but not conversely. In the top right graph, most of the points lie along the 45-degree line, so the financial industry and the market are more likely to jump together rather than jump separately. The

bottom three graphs plot the jump sizes with the linear regression line of the vertical axis variable against the horizontal axis variable. In all three cases, the points spread quite well along the regression line, indicating the jump sizes are highly correlated.

Table 2.4 outlines the jump size correlations for the five selected stocks with their industry and the market. The first observation is all the co-jumps have a positive correlation whenever it occurs. As for magnitude, all the five stocks are highly correlated with their corresponding industry, and each of the five industries is also highly correlated with the market when co-jump arrives. GE, XOM and AXP strongly follow the market in jump sizes, while WMT and MSFT are just moderately correlated with the market. The relatively low jump size correlation between WMT and the market is probably because of the defensive nature of WMT in the business cycle, while that between MSFT and the market is more likely due to the comparably lower stock-market co-jump probability. The high jump size correlations imply that when extreme events, such as crisis, occur, diversification benefits may be greatly affected as the overall correlation among asset returns could be significantly altered by jump effects. Details are further discussed in Section 2.5.

2.4.2 Prediction

This subsection compares the forecasts between the MGARCH-jump model and a benchmark MGARCH model (VD-GARCH) with normal renovations (MGARCH-N

model) by computing their predictive likelihood respectively. These predictive likelihoods are computed by comparing each of the five stocks along with their corresponding industry and the market. The last 100 observations are used for out-of-sample density forecast evaluation, and prediction is implemented one period ahead recursive forecasting, following equation (2.22) to (2.25).

Log-Bayes factor is computed by subtracting the log-predictive likelihoods of the MGARCH-N model from that of the MGARCH-jump model. A rule of thumb of this measure is that if a log-Bayes factor is greater than 5, then the evidence for MGARCH-jump is considered as "very strong". Table 2.8 lists the log-predictive likelihoods and log-Bayes factors from different cases. The MGARCH-jump model overwhelms the benchmark MGARCH-N model in all six predictions, with log-Bayes factors from around 12.70 to 61.81, That is, the MGARCH-jump model is approximately 3.2812×10^5 to 6.9782×10^{26} times better than the MGARCH-N model collectively in terms of predictive likelihood. This dominance is robust to a larger sample and/or longer prediction horizons.

Figure 2.5 plots the predictive likelihoods at each period. During normal days, both models perform very similarly due to the same VD-GARCH component; while in days with drastic return change, the predictive likelihood is significantly greater for the MGARCH-jump model. This is especially important because a risk-averse investor would like to be able to design some special strategies to diversify or hedge against these one-time, enormous risk events in advance. Comparing to the commonly

used MGARCH-N models, adding a jump component is more suitable to design those strategies.

2.5 Jumps/Co-jumps among Individual Stocks

The second application is to estimate a 5-dimensional model with GE, XOM, WMT, MSFT and AXP all together. Table 2.5 lists posterior results for jump component in this model. Again, the posterior estimates are in a reasonable region with a low mean ($\boldsymbol{\mu}$ of 0.03–0.06), low innovation parameter ($\boldsymbol{\alpha}$ of 0.11–0.15) and high persistent parameter ($\boldsymbol{\beta}$ greater than 0.98) as shown in Panel B of Table 2.5, and jump probabilities strongly favour "no jump" ($p_{\overline{GE},\overline{XOM},\overline{WMT},\overline{MSFT},\overline{AXP}} = 0.7102$) in Panel C, and jump size variances are large (all greater than 4.6). This shows that the proposed model is correctly specified and well-identified even with five stocks. Furthermore, the probability of only one stock jump while others don't is higher than that of any co-jumps, as the former are all above 0.024 and the latter are generally below 0.01 with the only exception of a 5-asset mutual jump probability of 0.017.

Panel A of Table 2.6 compares the joint co-jump probability and the product of corresponding marginal univariate jump probabilities. The jump arrivals across assets are strongly correlated. For instance, the XOM, MSFT, AXP co-jump has the lowest joint probability among all of 0.0010, but the product of marginal jump probabilities is less than 0.00005. This pattern is consistent among all co-jump cases. The joint co-jump probability is at least 0.0010, while the product of marginal probabilities is

at most 0.0001. Panel B exhibits that the majority of co-jumps among these five stocks are mutual co-jumps, which consist of 20.41% of the GE-jump, 18.74% of the XOM-jump, 16.22% of the WMT-jump, 14.90% of the MSFT-jump and 19.41% of the AXP-jump. And partial co-jumps with one or more stocks not jumping are a lot less frequent, each type of which consists of less than 10% of each asset jumps. This implies that for major stocks from different industries, they either all jump together, which is probably a market jump, or jump separately. One possible explanation is that when a jump arrives in some industry, it's not very likely to spread into other not directly related industries unless it's a market jump.

Table 2.7 lists jump size correlations among the five stocks. Most of the time, all five stocks jump in the same direction, and most of the stocks are also highly correlated in jump magnitude except for XOM in the oil industry. As mentioned before, these jump size correlations could severely change the overall return correlations. The final effect is rather complicated as shown in Figure 2.6, with plots the differences by subtracting the correlations of the GARCH component from those of ex-ante and ex-post covariances separately. These differences are usually around zero (no jump or very low probability of jump), but they can also go up to 0.4 and down to -0.4 as a result of jumps. As shown in the figure, jumps increase the return correlations when the MGARCH correlation is relatively low and thus decrease the diversification benefits, and vice versa. This is consistent with the theoretical implications based on the model structure.

The out-of-sample forecast comparison results are robust in high dimension. In the last row of Table 2.8, the log-Bayes factor for the MGARCH-jump model relative to the MGARCH-N model is 67.43, equivalent to 1.9206×10^{29} times better in terms of predictive likelihood. Similarly, in the bottom right plot in Figure 2.5, the MGARCH-jump model greatly outperforms the MGARCH-N model in a few particular periods and performs about the same for the rest time.

2.6 Applications

2.6.1 Impact on beta Dynamics

Consider a bivariate volatility model for excess returns of some individual stock and the market, then the beta of this stock can be computed simply, by definition, from its covariance matrix. Therefore, the dynamics of beta are equivalent to the dynamics of the conditional covariance matrix. Compared to the traditional approach that treats beta as a regression slope, this method naturally allows for dynamic beta as long as the covariances changing over time.

Past researches defining beta dynamics vary considerably. Bali et al. (2017) emphasize the role of dynamic beta in investment practice; Engle (2016) derives an estimator based on dynamic conditional correlation (DCC) model for continuous beta; from a completely different perspective, Todorov and Bollerslev (2010) on the other hand disentangle jumps into systematic jumps and idiosyncratic jumps, and then estimate continuous beta and jump beta accordingly. As for the MGARCH-jump model, we

follow a similar method to Engle (2016) but with Bayesian techniques. Unlike Todorov and Bollerslev (2010), we do not specifically separate jump beta, but rather focus on how beta changes with or without jumps.

One can either predict an ex-ante beta before knowing the exact jump arrivals or compute an ex-post beta after taking jump arrivals into account. Based on results from Section 2.2.3, if $\tilde{\boldsymbol{r}}_t = (\tilde{r}_{t,i}, \tilde{r}_{t,m})'$, where $\tilde{r}_{t,i}$ is the excess return of an arbitrary asset i, and $\tilde{r}_{t,m}$ is the excess return of the market. Then,

$$\mathrm{Cov}\left(\tilde{\boldsymbol{r}}_t | \boldsymbol{B}_t, \boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \begin{pmatrix} h_{t,ii} + B_{t,i}^2 \sigma_{J,i}^2 & h_{t,im} + B_{t,i} B_{t,m} \sigma_{J,im} \\ h_{t,im} + B_{t,i} B_{t,m} \sigma_{J,im} & h_{t,mm} + B_{t,m}^2 \sigma_{J,m}^2 \end{pmatrix} \tag{2.26}$$

So an ex-post beta is

$$\beta_{t,i} = \begin{cases} \frac{h_{t,im} + \sigma_{J,im}}{h_{t,mm} + \sigma_{J,m}^2} & \text{both jump} \\ \frac{h_{t,im}}{h_{t,mm} + \sigma_{J,m}^2} & \text{only market jumps} \\ \frac{h_{t,im}}{h_{t,mm}} & \text{otherwise} \end{cases} \tag{2.27}$$

This result nicely agrees with how beta relates to systematic risk: when the market doesn't, there's no change in systematic risk, so the beta is not affected; if only the market jumps, then the stock's relative exposure to the market decreases and so does beta; if there's a co-jump, both market risk and stock risk increase, and the effect on beta depends on values in the jump size covariance matrix. Now systematic risk transfers through h_{im} and $\sigma_{J,im}$ when co-jumps occur. Since a single stock is usually

riskier than the market, ex-post beta is more likely to increase when co-jump occurs.

On the other hand, before knowing jump arrivals (integrate out $\boldsymbol{B}_t$), conditional covariances of $\tilde{\boldsymbol{r}}_t$ is a dynamic volatility component ($\boldsymbol{H}_t$) plus a constant correction of expected jump covariances. Given the rareness of jump events, this correction term is expected to be fairly small, and ex-ante beta should be relatively close to ex-post beta except for whenever market jumps occur. Additionally, because of the dominance of co-jumps among all jumps, the ex-post beta should be mostly greater than ex-ante beta when jump arrives. Figure 2.7 shows the plots of beta dynamics computed from bivariate models with excess returns of AXP and the market. These results confirm the above hypothesises.

Comparing to the MGARCH-jump model, the benchmark MGARCH-N model tries to fit the jump extremes into smoothly changing volatilities. Thus the whole volatility dynamic is contaminated by jumps, so is the beta dynamics. Empirically, the MGARCH-N model tends to overestimate beta in general by not separating jumps.

2.6.2 Impact on Value at Risk

Value at Risk (VaR) is an important measure that lends help in various investment decisions, especially in risk management. It's defined by a VaR probability α, which means the loss an investor may potentially face when the worst α circumstances happen. Consider an equally-weighted portfolio constructed by the five assets used in Section 2.5. The predictive VaR is computed as the following steps:

1. Simulate $\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}$ for $M = 10000$ times.

   (a) Propagate $\boldsymbol{H}_{t+1}$ and generate $\boldsymbol{e}_{1,t+1}$ from equation (2.3).

   (b) Generate $\boldsymbol{B}_{t+1}$ and $\boldsymbol{Y}_{t+1}$, and compute $\boldsymbol{e}_{2,t+1}$ from equation (2.5).

   (c) Compute $\boldsymbol{r}_{t+1} = \boldsymbol{\mu} + \boldsymbol{e}_{1,t+1} + \boldsymbol{e}_{2,t+1}$.

2. Collect all the $\boldsymbol{r}_{t+1}^{(i)}|I_t$ simulations, and compute the return of equally-weighted portfolio $r_{EW,t+1}^{(i)} = \boldsymbol{w}_{EW}'\boldsymbol{r}_{t+1}^{(i)}$, where $i = 1, \ldots, M$, and $\boldsymbol{w}_{EW} = (1/N, \ldots, 1/N)'$.

3. Find the $M\alpha$-th least value of $r_{EW,t+1}^{(i)}$, where α is the VaR probability.

Figure 2.8 plots the 10%, 5% and 1% predictive VaR respectively for both the MGARCH-jump model and the MGARCH-N model. In this 100 daily prediction periods, 10% predictive VaR's are almost the same for both models, with the solid and dashed grey lines moving along with each other. The MGARCH-N model starts to slightly underestimate the 5% predictive VaR relatively than the MGARCH-jump model, with the red dashed line lies just above the red solid line for most of the periods. As for the 1% predictive VaR, the MGARCH-jump model provides clearly more conservative predictions, with the blue solid line always stay below the blue dashed line. The MGARCH-jump model predicts around 0.2% more potential daily loss against the MGARCH-N model when the worst 1% scenarios occur. It shows that the MGARCH-jump model provides a similar density prediction to the MGARCH-N

model in general, and it's more conservative only in the far left tail in the return distribution.

2.7 Conclusion

This chapter proposes a multivariate GARCH-jump mixture model that is both parsimonious and flexible. The model consists of two major components: a smooth shock in return due to volatility dynamics governed by a VD-GARCH component, and a drastic shock due to jumps governed by a compounded multinomial component. This jump component is a Hadamard product of a multivariate normal variable, indicating jump sizes, and a multinomial realization, a vector of jump arrival indicators, from all different possible jump arrival combinations. This structure allows for both jump sizes and arrivals to be correlated respectively, and the first two conditional moments exhibit desirable properties especially when further conditional on jump arrivals. The element-by-element jump effect on both conditional mean and conditional covariance will be activated/deactivated by the corresponding jump arrival indicators.

Results estimated by MCMC method show the model is well identified, and strong cross-sectional correlation in both jump sizes and jump arrivals. When modelling individual stocks with corresponding industry and the market, we found most of the market jumps are co-jumps, and a considerable proportion of industry/stock jumps

are also co-jumps. The source of identified jumps is major unexpected events, including earning announcement surprises, bad public news, etc. As for jump sizes, individual stock, industry and the market always jump into the same direction; strong correlations are found both between stock and industry and between industry and the market, while only a moderate correlation between stock and the market. When modelling five stocks from different industries, most of the jumps are either mutual co-jumps or individual jumps. This shows low contagion effect across industries unless it's market-wise. Again, jump size correlations are all positive, and mostly high except for XOM from the oil industry.

The impact of jumps on return correlations is less trivial than that on return covariances, as it depends on the level of MGARCH correlations. When the MGARCH correlation is higher than the jump size correlation, jumps will decrease the overall correlation and increase the diversification benefits; and vice versa. This is especially of importance when an investor considers a portfolio diversification strategy.

To compare the proposed MGARCH-jump model with the benchmark MGARCH-N model, predictive likelihoods are computed for both models by rolling-forward out-of-sample forecasts. In all six cases, prediction very strongly supports the MGARCH-jump model with log-Bayes factor over the MGARCH-N model. This dominance does not consistently exist over time. Both models predict similarly in normal days, but the MGARCH-jump model can predict abnormal events better as opposed to the MGARCH-N model. A risk-averse investor cares more about such events than

regular days, and the MGARCH-jump model enables potentially better risk reduction strategies against these events in advance.

Beta dynamics can be extracted based on the MGARCH-jump model from conditional covariance matrices by definition. beta can be computed either ex-post (conditional on jump arrivals) or ex-ante (only conditional on time). These two methods provide relatively similar estimates, while beta extracted from the MGARCH-N model is more different because of the excess smoothing when ignoring jumps.

The MGARCH-jump model produces very similar predictive Value at Risk as the MGARCH-N model for a five-asset equally-weighted portfolio, and more conservative loss predictions only in the far left tail. Both models provide almost identical predictive VaR at a 10% level, and the MGARCH-jump model starts to show slightly more conservativeness at a 5% level. At a 1% level, the MGARCH-jump model predicts higher loss potential than the MGARCH-N model. This shows that the MGARCH-jump model does not severely overestimate the potential losses at relatively normal periods, while giving more conservative guidelines for truly disastrous events.

Table 2.1: Descriptive Statistics for Daily Returns

	GE	XOM	WMT	MSFT	AXP	MKT
Means	0.0372	0.0409	0.0421	0.0737	0.0379	0.0353
Std. Dev.	1.7608	1.4729	1.6679	2.0451	2.2313	1.1099
Skewness	0.0338	0.0657	0.1050	0.0217	0.0039	-0.3445
Ex. Kurtosis	8.4730	8.7859	3.9253	5.7624	7.8700	8.6008
Min	-13.6841	-15.0271	-10.5811	-16.9577	-19.3523	-9.4059
Max	17.9844	15.8631	10.5018	17.8692	18.7711	10.8753

Notes: From January 1, 1990 to December 31, 2016, 6805 observations.

TABLE 2.2: Estimates of Selected Stocks, Corresponding Industry and the Market

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_{1,t} + \boldsymbol{\epsilon}_{2,t},\ \boldsymbol{\epsilon}_{1,t} = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t,\ \boldsymbol{z}_t \sim NID\left(\boldsymbol{0}, \boldsymbol{I}\right),\ \boldsymbol{\epsilon}_{2,t} = \boldsymbol{J}_t - \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}$$

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}{}' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1},\ \boldsymbol{\epsilon}_{t-1} = \boldsymbol{r}_{t-1} - \boldsymbol{\mu}$$

$$\boldsymbol{J}_t = \boldsymbol{Y}_t \odot \boldsymbol{B}_t,\ \boldsymbol{Y}_t \sim N\left(\boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J\right),\ \boldsymbol{B}_t \sim multinomial\left(1, \boldsymbol{p}\right)$$

Parameter	GE		XOM		WMT	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
C_{11}	0.0604	(0.0307, 0.0873)	0.0432	(0.0120, 0.0683)	0.0261	(0.0068, 0.0440)
C_{21}	0.0312	(0.0008, 0.0547)	0.0180	(-0.0074, 0.0392)	0.0434	(0.0186, 0.0608)
C_{22}	0.0551	(0.0334, 0.0718)	0.0158	(0.0015, 0.0302)	0.0210	(0.0011, 0.0395)
C_{31}	0.0413	(0.0245, 0.0568)	0.0205	(-0.0023, 0.0402)	0.0295	(0.0061, 0.0461)
C_{32}	0.0203	(0.0028, 0.0330)	0.0217	(-0.0153, 0.0410)	0.0040	(-0.0194, 0.0256)
C_{33}	0.0230	(0.0032, 0.0335)	0.0194	(0.0012, 0.0362)	0.0159	(0.0010, 0.0285)
α_{STK}	0.2032	(0.1900, 0.2165)	0.1934	(0.1808, 0.2069)	0.1543	(0.1425, 0.1664)
α_{IND}	0.2022	(0.1894, 0.2152)	0.1901	(0.1802, 0.2010)	0.1773	(0.1676, 0.1872)
α_{MKT}	0.2002	(0.1877, 0.2130)	0.1984	(0.1856, 0.2122)	0.1898	(0.1772, 0.2035)
β_{STK}	0.9716	(0.9678, 0.9750)	0.9760	(0.9726, 0.9790)	0.9839	(0.9816, 0.9860)
β_{IND}	0.9732	(0.9697, 0.9764)	0.9775	(0.9749, 0.9797)	0.9791	(0.9769, 0.9813)
β_{MKT}	0.9727	(0.9691, 0.9759)	0.9752	(0.9717, 0.9783)	0.9772	(0.9738, 0.9801)
μ_{STK}	0.0298	(0.0008, 0.0587)	0.0239	(-0.0011, 0.0496)	0.0243	(-0.0047, 0.0533)
μ_{IND}	0.0314	(0.0067, 0.0560)	0.0205	(-0.0033, 0.0444)	0.0320	(0.0101, 0.0540)
μ_{MKT}	0.0307	(0.0127, 0.0490)	0.0321	(0.0140, 0.0502)	0.0336	(0.0156, 0.0513)
$p_{STK,IND,MKT}$	0.0954	(0.0716, 0.1234)	0.0943	(0.0695, 0.1217)	0.0958	(0.0735, 0.1213)
$p_{STK,IND,\overline{MKT}}$	0.0030	(0.0003, 0.0075)	0.0570	(0.0360, 0.0815)	0.0180	(0.0067, 0.0309)
$p_{STK,\overline{IND},MKT}$	0.0026	(0.0002, 0.0069)	0.0009	(0.0000, 0.0033)	0.0011	(0.0000, 0.0034)
$p_{STK,\overline{IND},\overline{MKT}}$	0.0529	(0.0370, 0.0715)	0.0071	(0.0016, 0.0151)	0.0047	(0.0003, 0.0120)
$p_{\overline{STK},IND,MKT}$	0.0032	(0.0001, 0.0097)	0.0026	(0.0001, 0.0079)	0.0056	(0.0003, 0.0153)
$p_{\overline{STK},IND,\overline{MKT}}$	0.0161	(0.0082, 0.0252)	0.0029	(0.0004, 0.0075)	0.0030	(0.0001, 0.0085)
$p_{\overline{STK},\overline{IND},MKT}$	0.0018	(0.0001, 0.0051)	0.0031	(0.0001, 0.0095)	0.0013	(0.0000, 0.0047)
$p_{\overline{STK},\overline{IND},\overline{MKT}}$	0.8250	(0.7843, 0.8605)	0.8320	(0.7936, 0.8678)	0.8705	(0.8423, 0.8962)
$\mu_{J,STK}$	0.0597	(-0.0843, 0.1985)	-0.1648	(-0.2880,-0.0476)	-0.2137	(-0.3785,-0.0529)
$\mu_{J,IND}$	-0.3825	(-0.5380,-0.2395)	-0.2500	(-0.3698,-0.1406)	-0.4222	(-0.5419,-0.3072)
$\mu_{J,MKT}$	-0.4654	(-0.5853,-0.3536)	-0.5963	(-0.7381,-0.4580)	-0.4652	(-0.5739,-0.3599)
$\sigma^2_{J,STK}$	3.7909	(3.0470, 4.6388)	2.1384	(1.7064, 2.6669)	4.3986	(3.5767, 5.3886)
$\sigma_{J,STK,IND}$	2.8112	(2.2472, 3.4705)	1.8279	(1.4466, 2.2952)	2.1381	(1.6897, 2.6688)
$\sigma^2_{J,IND}$	2.5340	(2.0338, 3.1581)	1.8704	(1.4839, 2.3497)	1.6121	(1.2676, 2.0193)
$\sigma_{J,STK,MKT}$	2.2686	(1.8405, 2.7742)	1.2118	(0.9410, 1.5322)	1.0306	(0.7271, 1.3770)
$\sigma_{J,IND,MKT}$	1.8730	(1.5009, 2.3271)	1.2394	(0.9736, 1.5563)	1.0919	(0.8466, 1.3901)
$\sigma^2_{J,MKT}$	1.5230	(1.2214, 1.8870)	1.0451	(0.8021, 1.3480)	1.0099	(0.7873, 1.2793)

Table 2.2: Estimates of Selected Stocks, Corresponding Industry and the Market (cont.)

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_{1,t} + \boldsymbol{\epsilon}_{2,t},\ \boldsymbol{\epsilon}_{1,t} = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t,\ \boldsymbol{z}_t \sim NID\left(\boldsymbol{0}, \boldsymbol{I}\right),\ \boldsymbol{\epsilon}_{2,t} = \boldsymbol{J}_t - \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}$$

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}{}' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1},\ \boldsymbol{\epsilon}_{t-1} = \boldsymbol{r}_{t-1} - \boldsymbol{\mu}$$

$$\boldsymbol{J}_t = \boldsymbol{Y}_t \odot \boldsymbol{B}_t,\ \boldsymbol{Y}_t \sim N\left(\boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J\right),\ \boldsymbol{B}_t \sim multinomial\left(1, \boldsymbol{p}\right)$$

Parameter	MSFT		AXP	
	Mean	0.95 DI	Mean	0.95 DI
C_{11}	0.0844	(0.0567, 0.1098)	0.0654	(0.0446, 0.0877)
C_{21}	0.0582	(0.0422, 0.0729)	0.0671	(0.0504, 0.0823)
C_{22}	0.0388	(0.0234, 0.0506)	0.0216	(0.0014, 0.0449)
C_{31}	0.0426	(0.0304, 0.0550)	0.0537	(0.0386, 0.0669)
C_{32}	0.0234	(0.0036, 0.0365)	-0.0022	(-0.0290, 0.0262)
C_{33}	0.0245	(0.0109, 0.0330)	0.0235	(0.0021, 0.0368)
α_{STK}	0.1809	(0.1670, 0.1956)	0.1909	(0.1754, 0.2069)
α_{IND}	0.1853	(0.1738, 0.1970)	0.1951	(0.1836, 0.2070)
α_{MKT}	0.1933	(0.1799, 0.2067)	0.1915	(0.1777, 0.2057)
β_{STK}	0.9762	(0.9722, 0.9799)	0.9746	(0.9704, 0.9783)
β_{IND}	0.9770	(0.9740, 0.9797)	0.9739	(0.9707, 0.9769)
β_{MKT}	0.9754	(0.9719, 0.9787)	0.9742	(0.9705, 0.9778)
μ_{STK}	0.0520	(0.0176, 0.0858)	0.0311	(-0.0043, 0.0656)
μ_{IND}	0.0393	(0.0131, 0.0652)	0.0376	(0.0134, 0.0618)
μ_{MKT}	0.0364	(0.0186, 0.0543)	0.0376	(0.0192, 0.0561)

Table 2.2: Estimates of Selected Stocks, Corresponding Industry and the Market (cont.)

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_{1,t} + \boldsymbol{\epsilon}_{2,t},\ \boldsymbol{\epsilon}_{1,t} = \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t,\ \boldsymbol{z}_t \sim NID\left(\boldsymbol{0}, \boldsymbol{I}\right),\ \boldsymbol{\epsilon}_{2,t} = \boldsymbol{J}_t - \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}$$

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}{}' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1},\ \boldsymbol{\epsilon}_{t-1} = \boldsymbol{r}_{t-1} - \boldsymbol{\mu}$$

$$\boldsymbol{J}_t = \boldsymbol{Y}_t \odot \boldsymbol{B}_t,\ \boldsymbol{Y}_t \sim N\left(\boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J\right),\ \boldsymbol{B}_t \sim multinomial\left(1, \boldsymbol{p}\right)$$

Parameter	MSFT		AXP	
	Mean	0.95 DI	Mean	0.95 DI
$p_{STK,IND,MKT}$	0.0595	(0.0419, 0.0781)	0.0787	(0.0602, 0.0997)
$p_{STK,IND,\overline{MKT}}$	0.0313	(0.0189, 0.0450)	0.0064	(0.0009, 0.0138)
$p_{STK,\overline{IND},MKT}$	0.0006	(0.0000, 0.0022)	0.0012	(0.0000, 0.0040)
$p_{STK,\overline{IND},\overline{MKT}}$	0.0014	(0.0000, 0.0049)	0.0696	(0.0521, 0.0887)
$p_{\overline{STK},IND,MKT}$	0.0054	(0.0003, 0.0146)	0.0023	(0.0001, 0.0074)
$p_{\overline{STK},IND,\overline{MKT}}$	0.0140	(0.0069, 0.0224)	0.0110	(0.0038, 0.0198)
$p_{\overline{STK},\overline{IND},MKT}$	0.0016	(0.0001, 0.0053)	0.0011	(0.0000, 0.0037)
$p_{\overline{STK},\overline{IND},\overline{MKT}}$	0.8862	(0.8648, 0.9057)	0.8297	(0.7974, 0.8575)
$\mu_{J,STK}$	-0.0007	(-0.2192, 0.2242)	-0.0939	(-0.2521, 0.0635)
$\mu_{J,IND}$	-0.3899	(-0.5268,-0.2587)	-0.2864	(-0.4345,-0.1428)
$\mu_{J,MKT}$	-0.4947	(-0.6175,-0.3782)	-0.4888	(-0.6176,-0.3642)
$\sigma^2_{J,STK}$	9.4538	(7.7795,11.4719)	5.9235	(4.9261, 7.0697)
$\sigma_{J,STK,IND}$	4.0416	(3.2766, 4.9676)	4.2522	(3.4928, 5.1996)
$\sigma^2_{J,IND}$	2.5496	(2.0470, 3.1482)	3.6641	(2.8846, 4.6942)
$\sigma_{J,STK,MKT}$	1.7621	(1.2730, 2.3417)	3.0789	(2.5331, 3.7240)
$\sigma_{J,IND,MKT}$	1.5444	(1.2142, 1.9376)	2.4323	(1.9411, 3.0436)
$\sigma^2_{J,MKT}$	1.2506	(0.9653, 1.5910)	1.9048	(1.5145, 2.3804)

Table 2.3: Jump probabilities for GE, XOM, WMT, MSFT and AXP with corresponding industry and the market

Panel A: marginal and joint probabilities

Probabilities		GE	XOM	WMT	MSFT	AXP
Marginal	Stk	0.1539	0.1593	0.1196	0.0928	0.1560
	Ind	0.1176	0.1569	0.1224	0.1102	0.0984
	Mkt	0.1030	0.1010	0.1038	0.0672	0.0833
Joint	Stk, Ind	0.0984	0.1513	0.1139	0.0908	0.0851
		(0.0181)	(0.0250)	(0.0146)	(0.0102)	(0.0153)
	Stk, Mkt	0.0980	0.0953	0.0969	0.0602	0.0800
		(0.0159)	(0.0161)	(0.0124)	(0.0062)	(0.0130)
	Ind, Mkt	0.0986	0.0969	0.1014	0.0650	0.0810
		(0.0121)	(0.0158)	(0.0127)	(0.0074)	(0.0082)
	Stk, Ind, Mkt	0.0954	0.0943	0.0958	0.0595	0.0787
		(0.0019)	(0.0025)	(0.0015)	(0.0007)	(0.0013)

Notes: Numbers in parentheses below the joint probabilities are the product of corresponding marginal probabilities.

Panel B: conditional probabilities

Stock	Probabilites	Stk,Ind,Mkt	Stk,Ind	Stk,Mkt	Ind,Mkt
GE	p (co-jump|mkt-jump)	0.9261	—	0.9516	0.9568
	p (co-jump|ind-jump)	0.8110	0.8363	—	0.8378
	p (co-jump|stk-jump)	0.6198	0.6392	0.6369	—
XOM	p (co-jump|mkt-jump)	0.9338	—	0.9432	0.9598
	p (co-jump|ind-jump)	0.6012	0.9647	—	0.6179
	p (co-jump|stk-jump)	0.5918	0.9496	0.5977	—
WMT	p (co-jump|mkt-jump)	0.9233	—	0.9337	0.9770
	p (co-jump|ind-jump)	0.7829	0.9300	—	0.8284
	p (co-jump|stk-jump)	0.8011	0.9517	0.8102	—
MSFT	p (co-jump|mkt-jump)	0.8855	—	0.8950	0.9664
	p (co-jump|ind-jump)	0.5402	0.8240	—	0.5896
	p (co-jump|stk-jump)	0.6414	0.9783	0.6483	—
AXP	p (co-jump|mkt-jump)	0.9452	—	0.9602	0.9723
	p (co-jump|ind-jump)	0.8004	0.8657	—	0.8233
	p (co-jump|stk-jump)	0.5046	0.5457	0.5126	—

Notes: Each column indicates a particular type of co-jumps. For example, column 3 shows conditional probabilities of stock-industry-market co-jumps for each stock.

Table 2.4: Jump size correlations for GE, XOM, WMT, MSFT and AXP with corresponding industry and the market

	GE	IND	MKT		XOM	IND	MKT
GE	1.0000	—	—	XOM	1.0000	—	—
IND	0.9070	1.0000	—	IND	0.9140	1.0000	—
MKT	0.9441	0.9534	1.0000	MKT	0.8106	0.8865	1.0000
	WMT	IND	MKT		MSFT	IND	MKT
WMT	1.0000	—	—	MSFT	1.0000	—	—
IND	0.8029	1.0000	—	IND	0.8232	1.0000	—
MKT	0.4890	0.8557	1.0000	MKT	0.5125	0.8649	1.0000
	AXP	IND	MKT				
AXP	1.0000	—	—				
IND	0.9127	1.0000	—				
MKT	0.9166	0.9207	1.0000				

TABLE 2.5: Estimates among GE, XOM, WMT, MSFT and AXP

Panel A: drift and GARCH parameters

Parameter		GE	XOM	WMT	MSFT	AXP
$\boldsymbol{\mu}$		0.0313	0.0328	0.0364	0.0672	0.0372
		(0.0015, 0.0613)	(0.0043, 0.0605)	(0.0064, 0.0666)	(0.0284, 0.1049)	(-0.0010, 0.0756)
$\boldsymbol{\alpha}$		0.1507	0.1516	0.1166	0.1363	0.1503
		(0.1381, 0.1644)	(0.1389, 0.1654)	(0.1054, 0.1502)	(0.1244, 0.1493)	(0.1391, 0.1627)
$\boldsymbol{\beta}$		0.9832	0.9825	0.9898	0.9855	0.9839
		(0.9797, 0.9860)	(0.9791, 0.9853)	(0.9829, 0.9916)	(0.9827, 0.9879)	(0.9812, 0.9862)
		1	2	3	4	5
$\boldsymbol{C}$	1	0.0868				
		(0.0668, 0.1049)				
	2	0.0532	0.0717			
		(0.0371, 0.0737)	(0.0515, 0.0898)			
	3	0.0317	0.0113	0.0196		
		(0.0207, 0.0573)	(-0.0223, 0.0225)	(0.0029, 0.0347)		
	4	0.0521	0.0123	0.0334	0.0206	
		(0.0366, 0.0713)	(-0.0057, 0.0275)	(0.0019, 0.0593)	(0.0010, 0.0481)	
	5	0.0600	0.0120	0.0333	0.0079	0.0249
		(0.0446, 0.0768)	(-0.0067, 0.0274)	(-0.0049, 0.0597)	(-0.0403, 0.0507)	(0.0014, 0.0535)

Notes: Numbers in parentheses are 0.95 DI of the parameters above them.

TABLE 2.5: Estimates among GE, XOM, WMT, MSFT and AXP (cont.)

Panel B: jump/co-jump probabilities

Parameter	Mean	0.95 DI	Parameter	Mean	0.95 DI
$p_{GE,XOM,WMT,MSFT,AXP}$	0.0170	(0.0076,0.0266)	$p_{\overline{GE},XOM,WMT,MSFT,AXP}$	0.0018	(0.0001,0.0051)
$p_{GE,XOM,WMT,MSFT,\overline{AXP}}$	0.0022	(0.0001,0.0071)	$p_{\overline{GE},XOM,WMT,MSFT,\overline{AXP}}$	0.0018	(0.0001,0.0058)
$p_{GE,XOM,WMT,\overline{MSFT},AXP}$	0.0034	(0.0002,0.0093)	$p_{\overline{GE},XOM,WMT,\overline{MSFT},AXP}$	0.0022	(0.0001,0.0074)
$p_{GE,XOM,WMT,\overline{MSFT},\overline{AXP}}$	0.0042	(0.0003,0.0108)	$p_{\overline{GE},XOM,WMT,\overline{MSFT},\overline{AXP}}$	0.0087	(0.0008,0.0231)
$p_{GE,XOM,\overline{WMT},MSFT,AXP}$	0.0024	(0.0001,0.0074)	$p_{\overline{GE},XOM,\overline{WMT},MSFT,AXP}$	0.0010	(0.0000,0.0036)
$p_{GE,XOM,\overline{WMT},MSFT,\overline{AXP}}$	0.0021	(0.0001,0.0064)	$p_{\overline{GE},XOM,\overline{WMT},MSFT,\overline{AXP}}$	0.0038	(0.0001,0.0126)
$p_{GE,XOM,\overline{WMT},\overline{MSFT},AXP}$	0.0030	(0.0001,0.0084)	$p_{\overline{GE},XOM,\overline{WMT},\overline{MSFT},AXP}$	0.0040	(0.0003,0.0105)
$p_{GE,XOM,\overline{WMT},\overline{MSFT},\overline{AXP}}$	0.0031	(0.0001,0.0106)	$p_{\overline{GE},XOM,\overline{WMT},\overline{MSFT},\overline{AXP}}$	0.0298	(0.0122,0.0521)
$p_{GE,\overline{XOM},WMT,MSFT,AXP}$	0.0020	(0.0001,0.0070)	$p_{\overline{GE},\overline{XOM},WMT,MSFT,AXP}$	0.0017	(0.0000,0.0056)
$p_{GE,\overline{XOM},WMT,MSFT,\overline{AXP}}$	0.0021	(0.0001,0.0068)	$p_{\overline{GE},\overline{XOM},WMT,MSFT,\overline{AXP}}$	0.0023	(0.0001,0.0076)
$p_{GE,\overline{XOM},WMT,\overline{MSFT},AXP}$	0.0020	(0.0001,0.0064)	$p_{\overline{GE},\overline{XOM},WMT,\overline{MSFT},AXP}$	0.0041	(0.0001,0.0143)
$p_{GE,\overline{XOM},WMT,\overline{MSFT},\overline{AXP}}$	0.0020	(0.0001,0.0065)	$p_{\overline{GE},\overline{XOM},WMT,\overline{MSFT},\overline{AXP}}$	0.0471	(0.0300,0.0660)
$p_{GE,\overline{XOM},\overline{WMT},MSFT,AXP}$	0.0030	(0.0001,0.0108)	$p_{\overline{GE},\overline{XOM},\overline{WMT},MSFT,AXP}$	0.0018	(0.0001,0.0059)
$p_{GE,\overline{XOM},\overline{WMT},MSFT,\overline{AXP}}$	0.0067	(0.0006,0.0208)	$p_{\overline{GE},\overline{XOM},\overline{WMT},MSFT,\overline{AXP}}$	0.0621	(0.0385,0.0849)
$p_{GE,\overline{XOM},\overline{WMT},\overline{MSFT},AXP}$	0.0037	(0.0001,0.0129)	$p_{\overline{GE},\overline{XOM},\overline{WMT},\overline{MSFT},AXP}$	0.0342	(0.0185,0.0496)
$p_{GE,\overline{XOM},\overline{WMT},\overline{MSFT},\overline{AXP}}$	0.0241	(0.0115,0.0357)	$p_{\overline{GE},\overline{XOM},\overline{WMT},\overline{MSFT},\overline{AXP}}$	0.7102	(0.6728,0.7454)

Table 2.5: Estimates among GE, XOM, WMT, MSFT and AXP (cont.)

Panel C: jump size parameters

Parameter		GE	XOM	WMT	MSFT	AXP
$\boldsymbol{\mu}_J$		0.0409	-0.2288	0.0713	0.4148	0.0279
		(-0.1949,0.2692)	(-0.4219,-0.0489)	(-0.1186,0.2588)	(0.1857, 0.6384)	(-0.2911,0.3439)
$\boldsymbol{\Sigma}_J$	GE	6.7958				
		(4.6207, 8.8016)				
	XOM	3.0414	4.6635			
		(0.1584, 5.4529)	(3.1805, 6.5901)			
	WMT	5.9970	2.1946	6.8466		
		(4.2706, 7.4735)	(0.4840, 4.0247)	(5.4093, 8.4263)		
	MSFT	6.6374	4.5465	6.3338	10.1653	
		(1.7305, 9.3556)	(1.7684, 7.0416)	(3.1860, 8.4018)	(8.3923,12.2379)	
	AXP	7.1474	4.1921	6.3004	8.9655	10.9217
		(2.6935,10.0838)	(1.1525, 6.6495)	(2.1256, 9.2238)	(2.5645,11.3860)	(7.8769,14.2148)

Notes: Numbers in parentheses are 0.95 DI of the parameters above them.

Table 2.6: Jump probabilities among GE, XOM, WMT, MSFT and AXP

Panel A: marginal and joint probabilities

Stock	GE	XOM	WMT	MSFT	AXP
Marginal probs	0.0832	0.0907	0.1047	0.1140	0.0875

Co-jump	Joint Pr	Product	Co-jump	Joint Pr	Product
All jump	0.0170	0.0000	GE,MSFT	0.0067	0.0001
GE,XOM,WMT,MSFT	0.0022	0.0000	GE,AXP	0.0037	0.0000
GE,XOM,WMT,AXP	0.0034	0.0000	XOM,WMT,MSFT,AXP	0.0018	0.0000
GE,XOM,WMT	0.0042	0.0000	XOM,WMT,MSFT	0.0018	0.0000
GE,XOM,MSFT,AXP	0.0024	0.0000	XOM,WMT,AXP	0.0022	0.0000
GE,XOM,MSFT	0.0021	0.0000	XOM,WMT	0.0087	0.0000
GE,XOM,AXP	0.0030	0.0000	XOM,MSFT,AXP	0.0010	0.0000
GE,XOM	0.0031	0.0000	XOM,MSFT	0.0038	0.0000
GE,WMT,MSFT,AXP	0.0020	0.0000	XOM,AXP	0.0040	0.0000
GE,WMT,MSFT	0.0021	0.0000	WMT,MSFT,AXP	0.0017	0.0000
GE,WMT,AXP	0.0020	0.0000	WMT,MSFT	0.0023	0.0000
GE,WMT	0.0020	0.0001	WMT,AXP	0.0041	0.0000
GE,MSFT,AXP	0.0030	0.0000	MSFT,AXP	0.0018	0.0000

Notes: Column "Product" is product of corresponding marginal probabilities.

Table 2.6: Jump probabilities among GE, XOM, WMT, MSFT and AXP (cont.)

Panel B: conditional probabilities

Type of co-jumps	Conditional on				
	GE-jump	XOM-jump	WMT-jump	MSFT-jump	AXP-jump
All jump	0.2041	0.1874	0.1622	0.1490	0.1941
GE,XOM,WMT,MSFT	0.0259	0.0238	0.0206	0.0189	—
GE,XOM,WMT,AXP	0.0409	0.0375	0.0325	—	0.0389
GE,XOM,WMT	0.0507	0.0465	0.0403	—	—
GE,XOM,MSFT,AXP	0.0294	0.0270	—	0.0215	0.0280
GE,XOM,MSFT	0.0255	0.0234	—	0.0186	—
GE,XOM,AXP	0.0365	0.0335	—	—	0.0347
GE,XOM	0.0378	0.0347	—	—	—
GE,WMT,MSFT,AXP	0.0243	—	0.0193	0.0178	0.0232
GE,WMT,MSFT	0.0257	—	0.0204	0.0188	—
GE,WMT,AXP	0.0240	—	0.0191	—	0.0228
GE,WMT	0.0243	—	0.0193	—	—
GE,MSFT,AXP	0.0366	—	—	0.0267	0.0348
GE,MSFT	0.0806	—	—	0.0588	—
GE,AXP	0.0445	—	—	—	0.0423
XOM,WMT,MSFT,AXP	—	0.0201	0.0174	0.0160	0.0208
XOM,WMT,MSFT	—	0.0195	0.0169	0.0155	—
XOM,WMT,AXP	—	0.0245	0.0212	—	0.0254
XOM,WMT	—	0.0961	0.0831	—	—
XOM,MSFT,AXP	—	0.0114	—	0.0090	0.0118
XOM,MSFT	—	0.0420	—	0.0334	—
XOM,AXP	—	0.0436	—	—	0.0452
WMT,MSFT,AXP	—	—	0.0161	0.0148	0.0193
WMT,MSFT	—	—	0.0223	0.0205	—
WMT,AXP	—	—	0.0396	—	0.0474
MSFT,AXP	—	—	—	0.0158	0.0206

Notes: Each number above is a conditional probability of a particular co-jump type defined by row conditional on a particular univariate jump defined by column.

Table 2.7: Jump size correlations among GE, XOM, WMT, MSFT and AXP

Stock	GE	XOM	WMT	MSFT	AXP
GE	1.0000	—	—	—	—
XOM	0.5403	1.0000	—	—	—
WMT	0.8792	0.3884	1.0000	—	—
MSFT	0.7986	0.6603	0.7592	1.0000	—
AXP	0.8296	0.5874	0.7286	0.8509	1.0000

Table 2.8: Log-predictive likelihoods comparison

	MGARCH-jump	MGARCH-N	Log-Bayes factor
GE,IND,MKT	-1190.2726	-1211.7256	21.4529
XOM,IND,MKT	-1234.9129	-1264.9846	30.0717
WMT,IND,MKT	-1193.7295	-1206.4306	12.7011
MSFT,IND,MKT	-1148.3214	-1181.2845	32.9631
AXP,IND,MKT	-1234.9474	-1296.7611	61.8137
GE,XOM,WMT,MSFT,AXP	-1586.1071	-1653.5347	67.4276

Notes: A log-Bayes factor greater than 6 shows that prediction favours MGARCH-jump model.

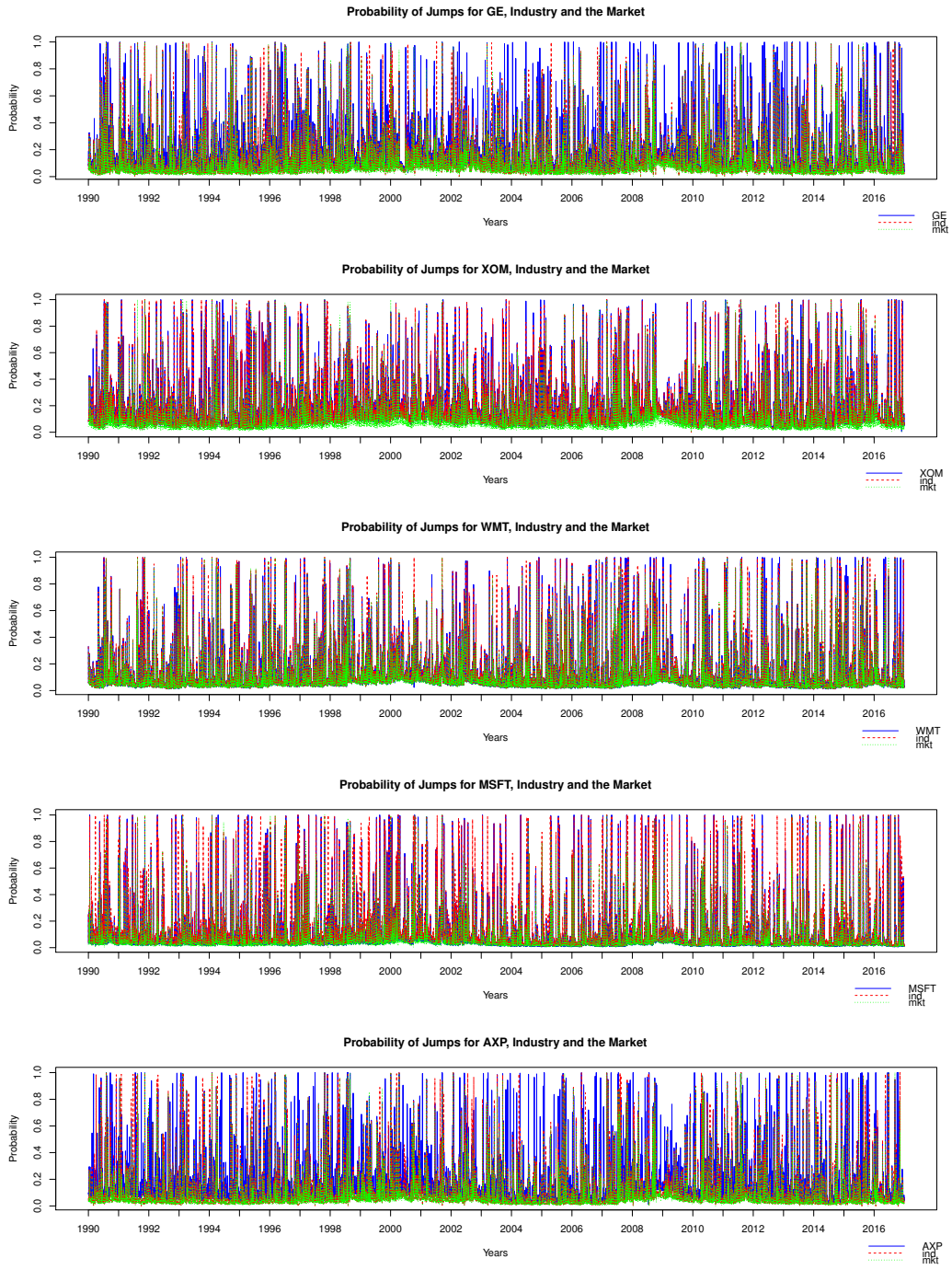


FIGURE 2.1: Jump arrivals for GE, XOM, WMT, MSFT and AXP with the corresponding industry and the market

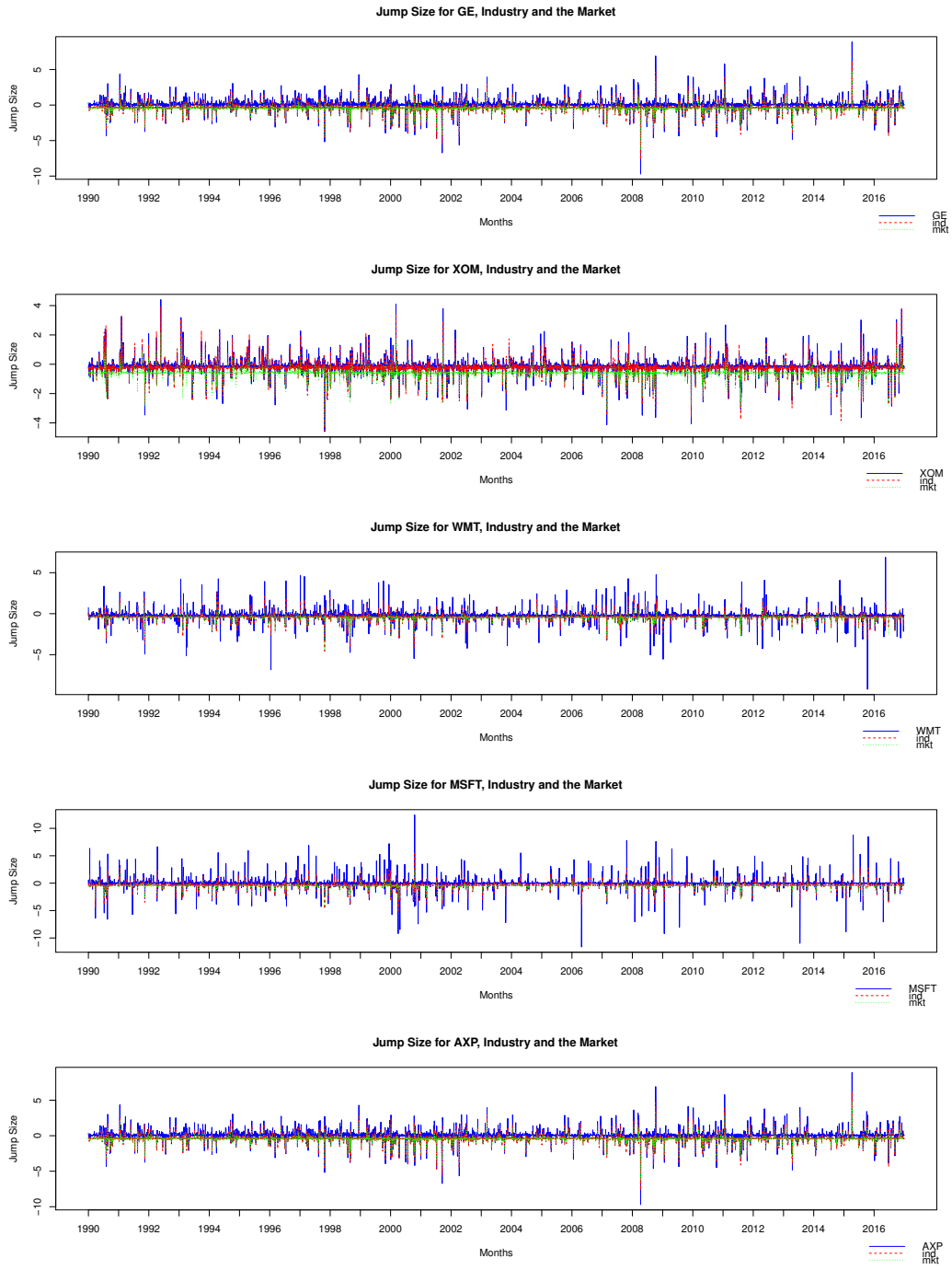


FIGURE 2.2: Jump sizes for GE, XOM, WMT, MSFT and AXP with the corresponding industry and the market

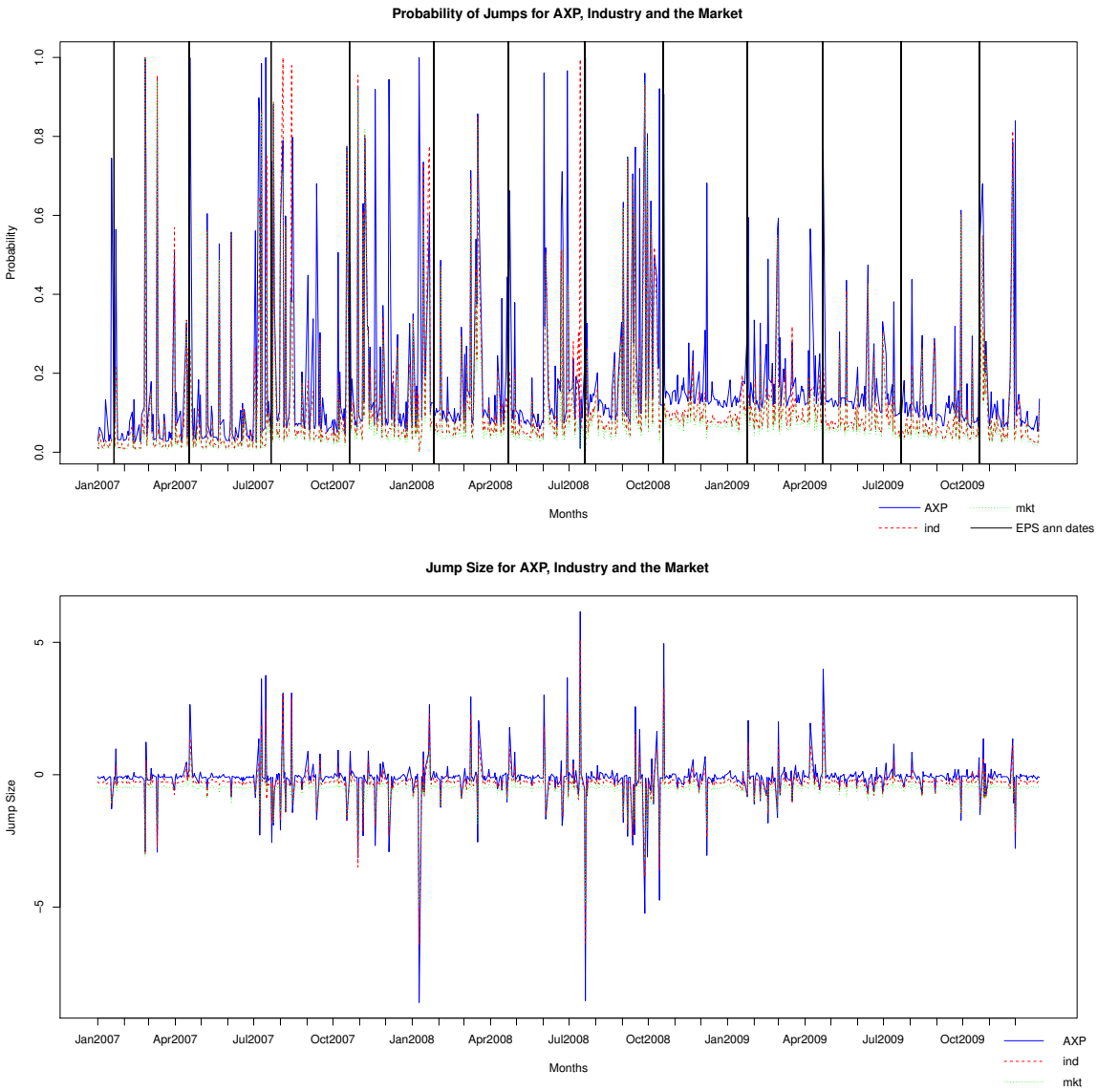


FIGURE 2.3: Jump probabilities and sizes for AXP over time from Jan 1, 2007 to Dec 31, 2009

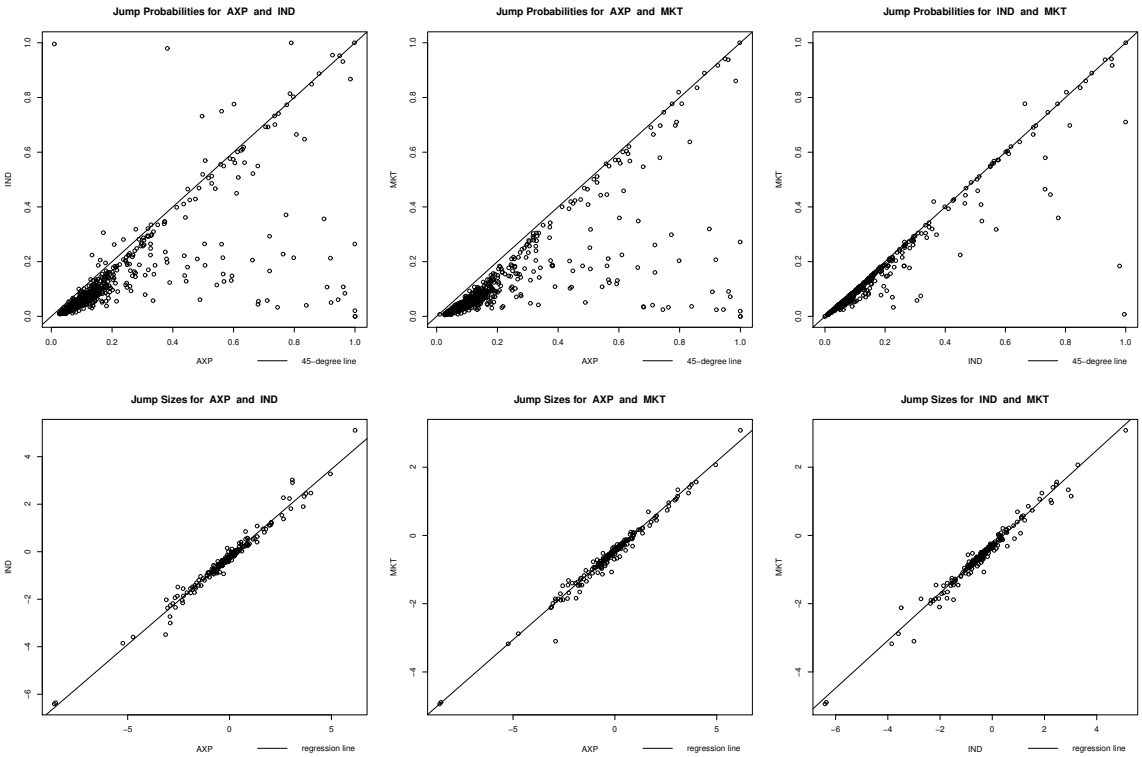


FIGURE 2.4: Jump probabilities and sizes for AXP crossover from Jan 1, 2007 to Dec 31, 2009

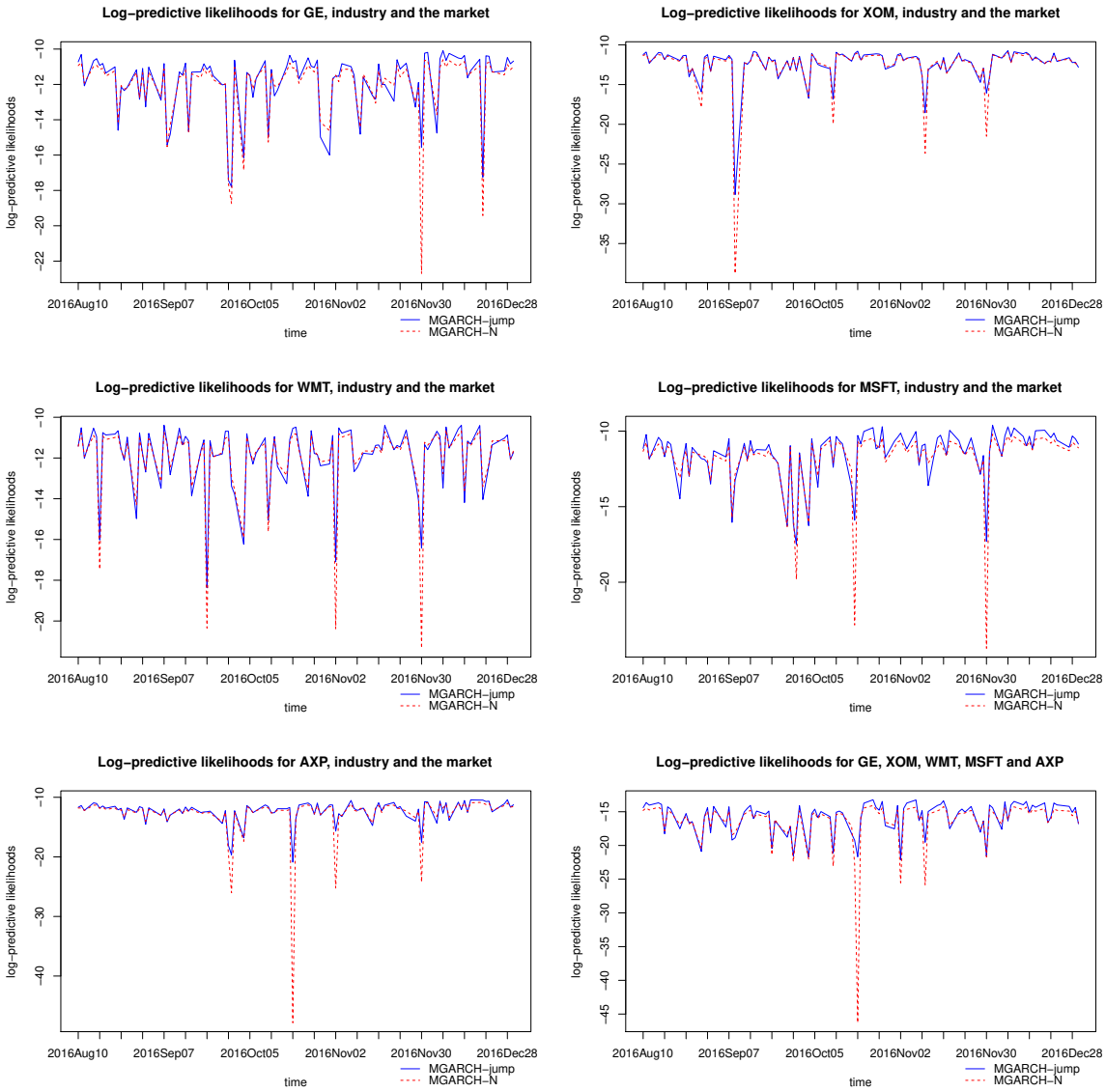


FIGURE 2.5: Log-predictive likelihoods of rolling-forward forecasts

Figure 2.6: Impact of Jumps on Correlation Dynamics over GARCH Component

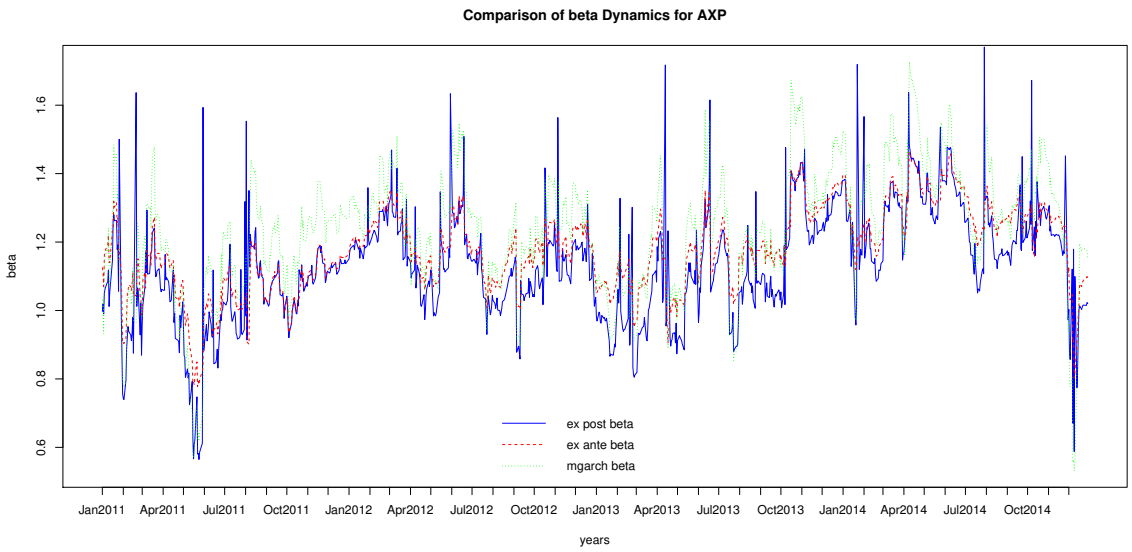


FIGURE 2.7: beta dynamics for AXP computed from MGARCH-jump model and MGARCH-N model

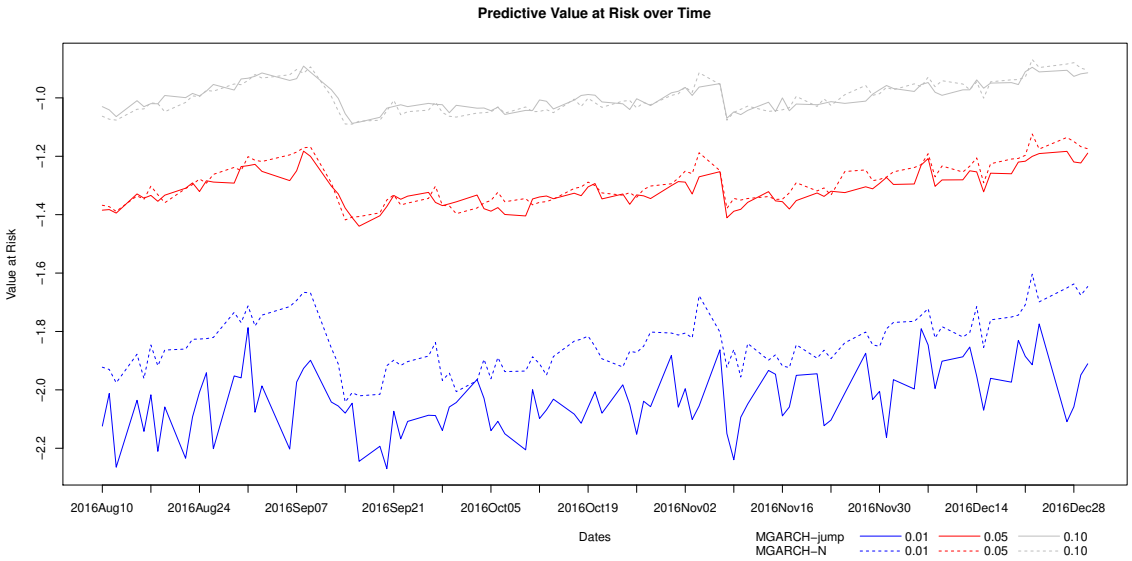


FIGURE 2.8: Predictive Value at Risk over time for a five-asset equally-weighted portfolio

Chapter 3

A Bayesian Semiparametric Stochastic Volatility Model with Markovian Mixtures

[1]

3.1 Introduction

Modelling financial and economic time series has always been a challenge due to their unknown and constantly changing underlying distributions. One way is to impose strong parametric structures into the model. Dynamic volatility models, such as generalized autoregressive conditional heteroskedasticity (GARCH, Bollerslev 1986)

[1]This chapter is a joint work with Dr. John Maheu and Dr. Qiao Yang.

and stochastic volatility (SV, Taylor 1982), are popular solutions. One distinguishable difference between SV and GARCH model is that the volatility dynamic in GARCH is deterministic conditionally, while it's latent and random in SV models.[2] Although they accommodate volatility persistence, a parametric assumption of normal or Student-t makes the innovation distribution static and symmetric.

Asymmetric and leptokurtic return distributions are well-acknowledged in many papers. Liesenfeld and Jung (2000) introduce a symmetrical scale mixture of normals on innovations. Chib et al. (2002) introduce an SV with Student-t innovation and a jump component. Durham (2006) decomposes the volatility component into two factors for more flexibility. Other alternatives include Barndorff-Nielsen (1997) using normal inverse Gaussian innovation, Mahieu and Schotman (1998) using various estimation approaches to study SV with mixture innovations.

Another approach to model asymmetric and leptokurtic returns are motivated by Dieobold (1986) and Lamoureux and Lastrapes (1990), who seek to approximate the conditional distribution from a structural break perspective. Their prime purpose is to incorporate high volatility persistence and structural breaks under a coherent framework. So et al. (1998) first introduce the Markov switching (MS) model into the SV model reparameterized by Kim et al. (1998), and Shibata and Watanabe (2005)

[2]For example, a standard SV model for asset returns is:

$$r_t = \mu + \exp\left(\frac{h_t}{2}\right) z_t, \quad z_t \overset{\text{iid}}{\sim} N(0,1)$$
$$h_t = \omega + \phi h_{t-1} + \sigma_v v_t, \quad v_t \overset{\text{iid}}{\sim} N(0,1)$$

employ MS on the standard SV parameterization. Kalimipalli and Susmel (2004) find regime switches in addition to SV dynamics in the short interest rates, and Vo (2009) finds a similar phenomenon on the crude oil market. However, their work does not provide evidence of out-of-sample performance for the MS-SV framework against fat-tailed innovation models. Ultimately, both approaches impose strong distributional assumptions that the data may not support.

In recent decades, the use of Bayesian nonparametric models has become increasingly popular. One of the most important such models, the Dirichlet process mixture (DPM) has been extensively used in many applications[3]. DPM model flexibly approximates the unknown conditional distribution with an infinite mixture of some known distribution (usually normal). Jensen and Maheu (2010) propose a Bayesian semiparametric model denoted as SV-DPM to improve the SV model. The SV-DPM allows for asymmetric and leptokurtic return distributions nonparametrically as SV-DPM contains both volatility persistence trough an autoregressive component and an infinite mixture in a unified framework. SV-DPM can capture conditional distributions with unknown shape and tails through its mixtures instead of a simple parametric assumption. Other applications with DPM include Delatola and Griffin (2013), Delatola, Griffin, et al. (2011), and Jensen and Maheu (2014). However, we show that the SV-DPM does not predict significantly better than simple parametric fat-tailed models, such as the SV-t model.[4]

[3]See examples in Chib and Hamilton (2002), Conley et al. (2008), Griffin and Steel (2004, 2011), Hirano (2002), Jensen (2004), Kacperczyk et al. (2013), and Tiwari et al. (1989), etc.

[4]See Section 3.5.

Virbickait and Lopes (2019) extend the SV-DPM model with an additional two-state MS mechanism. The conditional variance is decomposed into three components: a smooth change modelled by the SV, a nonparametric component modelled by the DPM, and a regime-switching modelled by the MS. This model is more flexible than the SV-DPM and allows the innovation distribution to be dynamic through a finite state MS framework.[5] A shortcoming of their approach is the fully parametric MS component has a pre-defined and fixed number of states. Essentially, mixing over ω_{s_t} and σ_{x_t} has a large degree of functional overlap if it is for modelling return distributions. It is perfectly fine when it comes to modelling squared log-return but not ideal when modelling log-returns directly. The DPM models unknown return distributions nonparametrically through an infinite mixture, but this mixture is still static over time.

Another popular Bayesian nonparametric model is the infinite hidden Markov model (IHMM) introduced by Beal et al. (2002). It is widely used in machine learning areas such as acoustic identification and video identification (Fox et al. 2011) and toward a board range of economic and financial time series in recent years.[6] IHMM also approximates the unknown conditional return distribution nonparametrically similarly

[5]For example:

$$\log r_t^2 = h_t + \sigma_{x_t} z_t,$$
$$h_t = \omega_{s_t} + \phi h_{t-1} + \sigma_v v_t, \quad v_t \overset{\text{iid}}{\sim} N(0,1) \quad s_t \sim \Pi,$$

where x_{s_t} is a DPM mixture and s_t follows a two-state Markov transition. h_t now is served with an intercept and an autoregressive time-varying feature, and the innovation is defined as a DPM mixture.

[6]Examples include Dufays (2016), Hou (2017), Jin and Maheu (2016), Jin et al. (2019), Jochmann (2015), Liu and Maheu (2018), Maheu and Yang (2016), Shi and Song (2015), and Song (2014).

as in DPM. Unlike the DPM model, the mixture weights in the IHMM are Markovian, which is constructed by a hierarchical Dirichlet process structure that consists of two layers of different Dirichlet processes. The IHMM can be seen as a regime-switching model with an infinite number of states. In each period, the return distribution is approximated by an infinite mixture as in the DPM, and the mixture in the next period depends on the choice of the state in the previous period through a Markovian transition, where the DPM sees the mixture constant for each period. Due to its unbounded state-space, the IHMM can accommodate both structural and recurrent changes in its unified framework.[7] However, a potential shortcoming is its lack of capturing smoothly changed volatilities, which is essentially important in modelling ARCH effect.

This chapter proposes a new Bayesian semiparametric model which integrate SV and IHMM denoted as SV-IHMM. The new model, SV-IHMM, nests the SV-DPM model, with more flexible time-varying mixtures. Although both SV-DPM and SV-IHMM allow for an unbounded state-space to model the return distribution, the mixture in the SV-DPM is static as the choice of states at one period is independent of the past. On the other hand, the SV-IHMM governs the mixture through a Markovian transition as the state for a particular period is affected by the last period state. This feature potentially allows for a much richer forecasting mechanism than the SV-DPM since the SV-IHMM can better approximate predictive distributions with more complex latent structures through a time-varying infinite mixture. We apply

[7] Song (2014).

the SV-IHMM on Apple Inc. stock returns, USD/CAD foreign exchange rates, oil price and monthly economic growths in the US. We find strong gains in log-predictive likelihoods against the SV-DPM and other competitive benchmark models. A larger and more dynamic state-space is documented with respect to the SV-DPM.

In the SV-t model and the SV-DPM model, all the changes in conditional return distribution are from the SV dynamics, while the SV-IHMM allows distributional features such as tails and shape to change independently of h_t over time. These distributional changes are achieved through the dynamic mixture weights in the IHMM component. Moreover, the SV-IHMM is so flexible that it can learn from data and capture potential skewness, fat-tails and structural breaks in conditional return distributions at the same time if there is any in the data. Empirical evidence shows that the SV-IHMM provides better out-of-sample density forecasts than all the benchmark models, especially a great improvement from the SV-t model compared to the SV-DPM model. And the conditional distribution of returns *is* changing over time for a wide range of applications.

This chapter is organized as follows. Section 3.2 illustrates the specification of the proposed SV-IHMM, along with the sampling algorithm and density forecast computation. Section 3.3 describes the details of the return series uses to test this model. Section 3.4 presents the posterior estimations of the SV-IHMM and two benchmark models. Section 3.5 further analyses its out-of-sample performance compared with multiple benchmark models. Section 3.6 tests the robustness of the estimates and

forecasts under different prior settings for hyper-parameters. And Section 3.7 concludes.

3.2 The Model

The proposed model, SV-IHMM, includes a parametric autoregressive component and a Bayesian nonparametric component. The autoregressive component, SV, captures the volatility smooth change. The Bayesian nonparametric component, IHMM, captures the potential regime switching in long-run expectation and volatility. The IHMM involves a hierarchical Dirichlet process (HDP) prior introduced by Teh et al. (2006). It can be written as a stick-breaking form.

$$\Gamma|\beta_0 \sim GEM\left(\beta_0\right), \quad \Pi_j|\Gamma, \alpha_0 \sim DP\left(\alpha_0, \Gamma\right), \quad j = 1, 2, \ldots,$$

where $\Gamma = (\gamma_1, \gamma_2, \ldots)'$, $\Pi_j = (\pi_{j1}, \pi_{j2}, \ldots)'$. GEM represents a Griffiths-Engen-McCloskey distribution, which is a stick-breaking process defined as following in Equation (3.1a). To be more specific, following Sethuraman (1994) and Teh et al. (2006),

$$\gamma_k = \hat{\gamma}_k \prod_{l=1}^{k-1} \left(1 - \hat{\gamma}_l\right), \qquad \hat{\gamma}_k \sim Beta\left(1, \beta_0\right), \tag{3.1a}$$

$$\pi_{jk} = \hat{\pi}_{jk} \prod_{l=1}^{k-1} \left(1 - \hat{\pi}_{jl}\right), \qquad \hat{\pi}_{jk} \sim Beta\left(\alpha_0 \gamma_k, \alpha_0 \left(1 - \sum_{l=1}^{k} \gamma_l\right)\right), \tag{3.1b}$$

where $\alpha_0 > 0$ and $\beta_0 > 0$ are concentration parameters. A lower value of these concentration parameters indicates a more concentrated HDP where only a few atoms (states) out of an infinite number of all are likely to occur, while the probability for the rest is close to zero.

The stick-breaking representation of SV-IHMM is: for $t = 1 : T$,

$$\Gamma|\beta_0 \sim GEM\left(\beta_0\right), \quad \Pi_j|\Gamma, \alpha_0 \sim DP\left(\alpha_0, \Gamma\right), \quad s_t|s_{t-1} \sim \Pi_{s_{t-1}}, \quad j = 1, 2, \ldots \tag{3.2a}$$

$$r_t = \mu_{s_t} + \sqrt{\omega_{s_t}} \exp(\frac{h_t}{2}) z_t \quad z_t \overset{\text{iid}}{\sim} N(0, 1) \tag{3.2b}$$

$$h_t = \phi h_{t-1} + \sigma_v v_t, \; v_t \overset{\text{iid}}{\sim} N(0, 1) \tag{3.2c}$$

$$\mu_j \sim N\left(b_0, B_0\right), \quad \omega_j \sim IG\left(v_0, s_0\right) \tag{3.2d}$$

where Π_j denotes the probability of switching from state j to another state, μ_{s_t} and ω_{s_t} are state-dependent parameters. $\Pi = (\Pi_1', \Pi_2', \ldots)'$ is an infinitely dimensional squared matrix of transition probabilities, and each row sums to one by construction.

The SV-IHMM allows approximating any unknown distribution by mixing an infinite number of normal distributions over μ_{s_t} and ω_{s_t}. The following is the mixture representation of equation (3.2b):

$$p(r_t|\mu_{s_t}, \omega_{s_t}, \Pi, s_{t-1}, h_t) = \sum_{k=1}^{\infty} \pi_{s_{t-1}k} N(r_t|\mu_k, \omega_k \exp(h_t)) \tag{3.3}$$

where k represents the state indicator. $\pi_{s_{t-1}k}$ is the weight assignment to different normal kernels, where weights for a given state change over time via the Markov chain

s_t. Note that in the SV-DPM model, $\pi_{s_{t-1}k} = \pi_k$ for all $t = 1, \ldots, T$, so the SV-DPM model is nested in the SV-IHMM.

Unlike a conventional SV model that attributes all volatility changes to an AR(1) random process, the SV-IHMM further allows for potential regime switching in volatility dynamics. Equation (3.2b) and (3.2c) can been written equivalently in the following way:

$$r_t = \mu_{s_t} + \exp(\frac{b_t}{2}) z_t \quad z_t \overset{\text{iid}}{\sim} N(0,1) \tag{3.4a}$$

$$b_t = \log \omega_{s_t} + \phi \left(b_{t-1} - \log \omega_{s_{t-1}} \right) + \sigma_v v_t \quad v_t \overset{\text{iid}}{\sim} N(0,1) \tag{3.4b}$$

where b_t represents total log-variances.

Compare to the conventional SV model, the basic form is conceptually the same. Conventional SV models treat the intercept of (3.4b) as a constant term, where the SV-IHMM assumes this term follow an infinite Markov switching structure. $\log \omega_{s_t}$ captures dynamic changes that the AR(1) component fails to. First, the infinite state dimension of SV-IHMM implicitly constructs variety of densities via combining any number of normal kernels. Second, the changes of mixtures between periods are linked via an infinite Markovian transition matrix rather than merely the volatility persistence. The key distinction between the SV-IHMM and the SV-DPM is the state dependence where the SV-DPM selects a state for one period independent of any other periods.

Similar to $\log \omega_{s_t}$, μ_{s_t} allows for changes in the long-run expected returns. Mixing over μ_{s_t} can potentially generate skewed conditional return distributions than the conventional SV model whenever supported by data. More importantly, the SV-IHMM enables richer dynamics by allowing the mixture to change over time compared to the SV-DPM model.

3.2.1 Benchmark Models

Several benchmark models are introduced for comparison. Two GARCH models are listed below and differ by innovations. Define the GARCH-N in the following way:

$$r_t = \mu + \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \epsilon_t \sim N(0,1). \tag{3.5}$$

GARCH-t replaces the $N(0, \sigma_t^2)$ with a Student-t distribution at a degree of freedom ν:

$$r_t = \mu + \sigma_t \epsilon_t, \quad \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \epsilon_t \sim t(\nu). \tag{3.6}$$

SV-N and SV-t correspond to normal and Student-t assumptions on the error term. SV-N is defined in the following way:

$$r_t = \mu + \exp(\frac{h_t}{2})\epsilon_t, \quad h_t = \omega + \phi h_{t-1} + \sigma_v v_t, \quad \epsilon_t \sim N(0,1). \tag{3.7}$$

Similarly, SV-t is defined with the t distribution in the following way:

$$r_t = \mu + \exp(\frac{h_t}{2})\epsilon_t, \quad h_t = \omega + \phi h_{t-1} + \sigma_v v_t, \quad \epsilon_t \sim t(\nu), \tag{3.8}$$

where z_t and ν_t of SV-N and SV-t are standard normal distributions. A vanilla infinite hidden Markov model is introduced here. Unlike SV-IHMM, the volatility of IHMM fully changes according to an infinite Markov transition. IHMM is defined in the following way:

$$\Gamma|\beta_0 \sim GEM\left(\beta_0\right), \quad \Pi_j|\Gamma, \alpha_0 \sim DP\left(\alpha_0, \Gamma\right), \quad j = 1, \ldots, \tag{3.9a}$$

$$s_t|s_{t-1} \sim \Pi_{s_{t-1}}, \quad r_t = \mu_{s_t} + \sigma_{s_t} z_t, \quad z_t \sim N(0,1), \tag{3.9b}$$

where Π is a Markov transition matrix with an infinite dimension. Both of μ_{s_t} and σ_{s_t} change according to Π. SV-DPM is introduced by Jensen and Maheu (2010), and it nests the nonparametric feature of the DPM and volatility persistence of the SV. It is defined in the following way:

$$\Gamma|\beta_0 \sim GEM\left(\beta_0\right), \quad s_t \sim multinomial(\Gamma), \tag{3.10a}$$

$$r_t = \mu_{s_t} + \sqrt{\omega_{s_t}} \exp(\frac{h_t}{2}) z_t, \quad z_t \sim N(0,1), \tag{3.10b}$$

$$h_t = \phi h_{t-1} + \sigma_v v_t, \quad v_t \sim N(0,1). \tag{3.10c}$$

Both of z_t and v_t are standard normal distributions. Changes of s_t at each period according to the same multinomial distribution and are determined independently across periods, while the SV-IHMM suggests the choice of s_t is dependent on the state of s_{t-1}. Therefore, the SV-IHMM nests the SV-DPM.

3.2.2 Hyper-Priors and Hierarchical Priors

This section defines the hyper-priors and hierarchical priors for the SV-IHMM and benchmark models. Hyper-priors for the infinite Markovian transition are formed by the HDP prior $\Gamma|\beta_0 \sim GEM(\beta_0)$ and $\Pi_s|\Gamma, \alpha_0 \sim DP(\alpha_0, \Gamma)$ discussed in the previous section. α_0 and β_0 govern the transition matrix Π. In order to keep the impact of priors minimum, rather than imposing constant value on α_0 and β_0, we use hyper-priors to allow learning and flexibility[8]:

$$\alpha_0 \sim Gamma(2, 8), \quad \beta_0 \sim Gamma(2, 8), \quad \mathrm{E}(\alpha_0) = \mathrm{E}(\beta_0) = 0.25. \tag{3.11}$$

Similarly, a hierarchical prior structure motivated by Song (2014) is applied for state-dependent parameters $\mu_{s_t} \sim N(b_0, B_0)$ and $\omega_{s_t} \sim IG(\nu_0, s_0)$. Rather than choosing fixed values of b_0, B_0, ν_0 and s_0, we let them learn from the state dynamics. This approach is appropriate for any state-dependent parameter models. Whenever a new state is introduced, the associated new μ_{s_t} and ω_{s_t} are drawn directly from the base measure, which contains informative b_0, B_0, ν_0 and s_0 updated by learning from data. The following are the hierarchical priors,[9]

$$b_0 \sim N(0, I), \quad B_0 \sim IW(I, 4), \quad v_0 \sim Exp(1), \quad s_0 \sim Gamma(5, 1), \tag{3.12}$$

where I is an identity matrix. ϕ and σ_v correspond to $N(0, 100)$ and $IG(5, 0.25)$.

[8] Fox et al. (2011).

[9] Appendix 3 shows the details of posteriors for the base measure parameters.

The hyper-priors and hierarchical priors of the IHMM in equation (3.9) are the same as the SV-IHMM except ω_{s_t} are replaced by σ_{s_t}, and ϕ and σ_v are no longer necessary.[10] We replicate the same hyper-priors and hierarchical priors in the SV-DPM as in the SV-IHMM. The key difference is that only one concentration parameter in the SV-DPM needs to be inferred.[11] $\beta_0 \sim Gamma(2, 8)$ is imposed. Let μ, ω, α follows $N(0, 1)$ in GARCH-N and GARCH-t. Similarly, let μ, ω, ϕ be $N(0, 1)$ and $\sigma_v \sim IG(5, 0.25)$ in both SV-N and SV-t. The prior for ν in the Student-t models is $U[2, 50]$.

3.2.3 Posterior Sampling

Sampling schemes for SV-IHMM consists of two parts. One is to sample the state-dependent parameters μ_{s_t}, ω_{s_t}, the state indicator s_t, transition matrix Π, concentration parameters of HDP (α_0 and β_0). The other part is to sample the stochastic volatility h_t and its parameters $\theta_h = \{\phi, \sigma_\nu\}$. These two parts are iterated via Markov chain Monte Carlo (MCMC).

This chapter adopts the beam sampler introduce by Van Gael et al. (2008) for the transition matrix, which is identical to the one in the IHMM. The beam sampler randomly generates an auxiliary variable (slice) that stochastically truncates

[10] IHMM tries to model the volatility persistence completely via Markov-switching.
[11] β_0 in equation (3.10a).

the infinitely dimensional transition matrix Π into a finite size, so that the forward-filtering backward-sampling (FFBS, Carter and Kohn 1994; Chib 1996; Frühwirth-Schnatter 1994) algorithm can be applied to sample the state indicators.[12] In the beam sampler, the auxiliary variable u_t (slice) is generated from a uniform distribution $u_t \sim U(0, \pi_{s_{t-1}s_t})$ for $t = 1, \ldots, T$. The number of truncated cluster K is updated to satisfy the condition $\max_{i\in\{1,\ldots,K\}}\{1 - \sum_{j=1}^{K} \pi_{i,j}\} < \min_{t\in\{1,\ldots,T\}}\{u_t\}$, so a finite number K can be truncated by slices. The value of K is determined by the transition matrix Π, which is implicitly regulated by α_0 and β_0. Greater values favour more frequent exploration of new states, but ultimately, the number of active states assigned with at least one observation is determined by the data assignments via FFBS.

After $s_{1:T}$ are sampled, we update K by excluding the states without any observation assignment. In terms of sampling h_t, a random length block-move Metropolis-Hastings (MH) sampler is used. It is introduced by Jensen and Maheu (2010) and built on Fleming and Kirby (2003). The block size is randomly drawn from a Poisson distribution with a preset hyperparameter $\lambda_h = 4$, and the expected block size is $\lambda_h + 1$. Once h_t are sampled, ϕ and σ_v can be easily sampled via conjugacy. $c_{1:K}$ is the colour ball count in the oracle urn which helps to sample the α_0 and β_0.[13] Appendix 3 describes the details of each sampling step. The entire posterior steps are summarized

[12]This is the exact method for sampling the IHMM. Another IHMM sampling approach is introduced by Fox et al. (2011), which uses a fixed large number of states that truncates the state dimension.

[13]Refer to Beal et al. (2002) for details.

in the following,

$p(u_{1:T}\|s_{1:T},\Pi)$	$p(s_{1:T}\|\Pi,u_{1:T},r_{1:T},h_{1:T},\Theta,\theta_h)$	$p(c_{1:K}\|s_{1:T},\Gamma,\alpha_0)$
$p(\Gamma\|s_{1:T},\beta_0,\alpha_0,c_{1:K})$	$p(\Pi\|s_{1:T},\Gamma,\alpha_0,c_{1:K})$	$p(\mu_{1:K},\omega_{1:K}\|r_{1:T},s_{1:T})$
$p(\alpha_0,\beta_0\|s_{1:T},c_{1:K})$	$p(h_{1:T}\|r_{1:T},s_{1:T},\Theta,\theta_h)$	$p(\phi\|h_{1:T},\sigma_v)$

Iterating the above steps and collect posterior draws for each parameter of interest. After dropping 20,000 burn-in draws, the posterior average or quantiles of each parameter of interest are computed from 20,000 draws. For example:

$$E(\alpha_0|r_{1:T}) = \frac{1}{N}\sum_{i=1}^{N}\alpha_0^{(i)}, \quad \text{for } t = 1,\ldots,T \tag{3.14}$$

where i indicates ith number of MCMC draws.

3.2.4 Density Forecast

Predictive likelihood measures the accuracy of density forecasts. Under the Bayesian approach, the predictive distribution integrates out all the parameter uncertainties. It can be generated by the following form,

$$p(r_{t+1}|r_{1:t}) = \int_\theta f(r_{t+1}|\theta,r_{1:t})p(\theta|r_{1:t})d\theta, \tag{3.15}$$

where θ is the parameter space of interest. $p(\theta|r_{1:t})$ is posterior which represents the parameter uncertainties and $p(r_{t+1}|r_{1:t})$ is predictive likelihood conditioned on θ and

past observations. Equation (3.15) marginalizes out the parameter uncertainties and implicitly describes the predictive distribution unconditional on parameters. This is essential in Bayesian approaches since the predictive likelihood not only measures the accuracy of predictive distribution but also assesses the significance of introducing extra parameters in out-of-sample forecasts. As opposed to point forecasts, density forecasts produce a whole predictive distribution, and it's further evaluated at the data realization for predictive likelihoods rather than a mere predictive expectation. In this case, root mean squared forecast error (RMSFE) does not fully reflect the full distributional performance of forecasts.

For computing the log-predictive likelihood (LPL), the first 20,000 MCMC draws are burn-in and the next 20,000 are used for the LPL computation. We apply a recursive method of predictive inference on an expanding window on a one period ahead basis. Let $\{\Theta^{(i)}, \Pi^{(i)}, s_t^{(i)}, \theta_h^{(i)}, \sigma_v^{(i)}\}$ be the ith posterior draw and $s_t^{(i)} \in \{1, \ldots, K^{(i)}\}$. $K^{(i)}$ is the number of active states. The predictive likelihood for r_{t+1} conditional on $r_{1:t}$ is computed as follows.

1. At each ith MCMC draw, simulate a state variable $s_{t+1}^{(i)}$ through $\Pi_{s_t}^{(i)}$ conditional on $s_t^{(i)}$.

2. If $s_{t+1}^{(i)} \leq K^{(i)}$, it implies that r_{t+1} belongs to an existing state, and the state-dependent parameter $\Theta_k^{(i)}$ has a $k \in \{1, \ldots, K^{(i)}\}$. Otherwise, r_{t+1} belongs to a new state $k = K^{(i)} + 1$, where $\Theta_k^{(i)}$ are drawn from the hierarchical prior: $N(b_0, B_0)$ and $IG(\nu_0, s_0)$.

$M = 20,000$ is the number of MCMC iterations for forecasting inference. The predictive likelihood at $t+1$ is computed over all MCMC draws.

$$p(r_{t+1}|r_{1:t}) \approx \frac{1}{M}\sum_{i=1}^{M} p(r_{t+1}|\Pi^{(i)}, s_t^{(i)}, \phi_h^{(i)}, \sigma_v^{(i)}, h_t^{(i)}) \tag{3.16}$$

The forecast performance of model A is computed by the logarithm of the joint predictive likelihood over the entire out-of-sample period.

$$LPL_A = \log p(r_{t_0:T}|r_{1:t}, \mathcal{M}_A) = \sum_{l=t_0}^{T} \log p(r_l|r_{1:l-1}, \mathcal{M}_A) \tag{3.17}$$

The out-of-sample period is $t_0 : T$. The outcome of LPL_A is the log-predictive likelihood (LPL) over the out-of-sample period for model A. The RMSFE is computed in a similar way[14].

$$\text{RMSFE} = \sqrt{\frac{\sum_{l=t_0}^{T}\left[r_l - E(r_l \mid r_{1:l-1})\right]^2}{T - t_0 + 1}} \tag{3.18}$$

The log-predictive Bayes factor (LBF) compares forecast performance between models. It is the difference between LPL values of two models. An LBF is greater than 5 suggests strong evidence that supports the first model. For illustration purposes, a log-predictive Bayes factor between models is defined as $LBF = LPL_A - LPL_B$. The Bayes factor is a real number which compares relative forecast performance at

[14] $E(r_{t+1}|r_{1:t}) = \frac{1}{M}\sum_{i=1}^{M} \mu_{s_{t+1}}^{(i)}$.

the corresponding out-of-sample period. Unfortunately, an LBF does not contain information of relative performance between model $\mathcal{M}_A$ and $\mathcal{M}_B$ at some sub-periods but the relative performance over the whole out-of-sample period. As a result, a cumulative log Bayes factor is constructed for addressing this issue. A cumulative LBF is a sequence of log-predictive BFs that cumulatively show the relative performance over time. For illustration purposes, the cumulative LBF between model $\mathcal{M}_A$ and $\mathcal{M}_B$ is the following,

$$CLBF_t = \sum_{l=t_0}^{t} [\log p(r_l|r_{1:l-1}, \mathcal{M}_A) - \log p(r_l|r_{1:l-1}, \mathcal{M}_B)] \quad \text{for } t = t_0, \ldots, T \tag{3.19}$$

3.3 Data

Four time-series are studied using the SV-IHMM and the benchmark models. The data covers equity, a commodity price, an exchange rate and a macroeconomic indicator. The common stock daily log returns of Apple Inc. (AAPL) is selected for equity, dated from December 15th, 1980 to December 29, 2017, with a sample size of 9,343. They are retrieved from the CRSP database.[15] For the foreign exchange market, we study the daily log changes of USD/CAD exchange rate dated from January 4th, 1971 to November 3rd, 2017. They are downloaded from the Federal Reserve Bank with a total of 11,308 observations. WTI crude oil log returns are selected for commodity, observed from January 2nd, 1986 to November 13, 2017. There are 8,037 daily observations and they are downloaded from the U.S. Energy Information Administration.

[15] AAPL has one of the largest capitalization and is most actively traded stock.

The monthly industrial production (IP) index is downloaded from the Federal Reserve Bank of St.Louis and log growth rates are computed. It measures the real output for all facilities. There are 1,205 observations and they are dated from March 1919 until July 2019. All of the log rates of changes are scaled by 100 for percentage values. Table 3.1 illustrates some descriptive statistics of the data. The equity and oil show much more volatile behaviour and skewness than the exchange rates and industrial growths.

3.4 Posterior Analysis

Table 3.2 summarizes posterior parameters of the SV-IHMM, SV-DPM and SV-t across four datasets.[16] Posterior average and 95% density intervals are computed. Posterior draws are sampled according to samplers and priors in Section 3.2.3 and Section 3.2.2. There are 20,000 burn-in draws and another 20,000 draws for constructing the posteriors. First, by introducing dynamic structures on the volatility component does not weaken the volatility persistence. This outcome stays robust across datasets. Second, we find that in general more states are estimated by the SV-IHMM than the SV-DPM.

Panel A of Table 3.2 summarizes the parameter posteriors of the SV-IHMM, SV-DPM and SV-t for AAPL returns. First, although regime switching is introduced

[16] All of the time-varying parameters are not illustrated, for example ω_{s_t} is too hard to report in the paper size since there is over 8000 of them.

through ω_{s_t} in SV-IHMM and SV-DPM, it does not weaken the high volatility persistence. A ϕ close to one indicates a high persistence in volatility. It is natural to generate more states in the SV-IHMM ($K = 7.9$) and SV-DPM ($K = 5.4$) since SV-IHMM suggests a more dynamic state-space than SV-DPM does. A small value of ν implies high excess kurtosis, which is consistent with Table 3.1.

Panel B of Table 3.2 illustrates the posteriors of parameters on USD and CAD exchange rates. Again, high volatility persistence is well captured through ϕ by the SV-IHMM, SV-DPM and SV-t. The introduction of the time-varying parameter, ω_{s_t}, does not show a significant impact on the magnitude of ϕ. A similar expansion on state space is documented when we move from SV-DPM (K=4) to SV-IHMM (K=8.5).

Posterior summaries on crude oil are illustrated in Panel C. Not surprisingly, the volatilities are highly persistent in all three models. The average of the state number between the SV-IHMM and SV-DPM only shows a marginal difference. Similar outcomes are documented in monthly IP growths at Panel D, volatility persistence remains about the same when we move from SV-t toward SV-IHMM. For IP growths, an additional lagged IP growth rate with state-dependent slope is regressed along with the intercept μ_{s_t}.

What really important is that the introduction of ω_{s_t} does not weaken the volatility persistence. In the next section, we study its out-of-sample forecast in a Bayesian way. It is critical to evaluate the importance of ω_{s_t} dynamics.

3.5 Out-of-Sample Forecasts

We perform one-period ahead out-of-sample forecasts recursively as indicated in Section 3.2.4. Two measurements are computed. One is the log-predictive likelihood for density forecasts, which evaluates the accuracy of an entire predictive distribution. The other is RMSFE for point forecasts, which essentially measures the predictive accuracy of the center of a predictive distribution. In other words, RMSFE of point forecast measurement assesses the first moment of a predictive distribution, while LPL investigates not only the first but also the second and higher moments. As a result, the differences in RMSFE among models are only marginal.

Tables 3.3 and 3.4 compare the LPL and RMSFE among all the models for the four datasets. We try to maximize the out-of-sample size so that the forecast performance is robust to time. But still, we do not calculate marginal likelihood since it requires all samples to be predicted and priors could have dominant impacts while the training sample is very small. In this chapter, the out-of-sample size contains at least 95% of the entire sample. For example, the out-of-sample is 9,000 out of 9434 for AAPL, 10,000 out of 11,308 for the exchange rates, 8,000 out of 8,037 for oil price and 1,165 out of 1,205 for industrial growths.

3.5.1 Log-Predictive Likelihoods

Tables 3.3 and 3.4 display the LPL, RMSFE and LBF, where the SV-IHMM is the base model for comparison.[17] Overall, the SV-IHMM provides the best forecast

[17] A positive value of Log-Bayes factor implies a favour of SV-IHMM.

result compared to all the benchmark models in terms of density forecast. The losses and gains in point forecasts (RMSFE) are very minor. Table 3.3 lists the out-of-sample forecast performance of AAPL returns and USD/CAD exchange rates. A significant gain against the SV-DPM (the 2nd best-performed model in terms of LPL) is documented. For the application of the exchange rates, 22 units of Bayes factor are found with respect to the SV-DPM (the 2nd best model by LPL). Table 3.4 illustrates the out-of-sample performance of crude oil and industrial growth. 16 units of log Bayes factor gain are generated relative to the SV-DPM (the 2nd best model). Similarly, the SV-IHMM shows the highest gain in terms of LPL for IP growth[18]. Gains in the RMSFE are minor in both datasets.

In short, the SV-IHMM shows superior density forecasts against all benchmarks at each dataset and it is robust to all datasets. The loss and gain of the SV-IHMM against the benchmarks are very minor for RMSFE. Our results suggest that the SV-DPM is a competitive model as it's consistently ranked as the 2nd best model in terms of the LPL. Compared to the SV-DPM, the SV-IHMM allows for the state mixture to be time-dependent. Secondly, appropriately modelling fat tails is important to achieve better density forecasts, although significant increases in LPL are found while a Gaussian innovation is replaced by Student-t (ie. GARCH-t and SV-t), using Student-t to capture fat tail becomes limited than the SV-IHMM, wherein more dynamic fat tails can be accommodated by time-varying mixtures.

[18]An additional AR(1) lagged IP growth with state-dependent slope is regressed.

3.5.2 Cumulative Log-Bayes Factor

A limitation is raised in Tables 3.3 and 3.4, where they only show the density forecasts according to the entire out-of-sample period. It does not tell the story whether any model shows superior performance at certain periods, but it delivers an overall inferior forecast performance. In other words, a subsample analysis is necessary to investigate whether or not the forecast performance is consistent and robust over time. Cumulative LBF is computed to address this issue. It is a sequence of LBFs according to equation (3.19). It enables us to investigate the predictive performance at subsample scenarios. We construct the cumulative LBF between the SV-IHMM (base model) and other top performed benchmark models across each dataset. If a monotonically increasing trend is found, the SV-IHMM continuously outperforms the benchmark models across the out-of-sample periods.

Figures 3.1 plots the cumulative LBF between the SV-IHMM and other top performed benchmark models for each dataset. Overall, each of the plots shows an increasing trend of cumulative LBF, which indicates a consistent and superior performance of SV-IHMM against other models. Figure 3.1a shows that the performance of AAPL returns for SV-IHMM against SV-DPM and SV-t are relatively stable compared to GARCH-t. A jump occurs at 2000 Internet bubble and 2008 Great Recession on GARCH-t indicates its poor performance on capturing these shocks with respect to SV-IHMM. Figure 3.1b plots for USD/CAD exchange rates. Again, robust and consistent gains of SV-IHMM against benchmark models are documented with respect to time. Figure 3.1c shows the results for daily oil returns. An increasing trend

suggests the consistent performance of SV-IHMM against the benchmarks.

Figure 3.1d is the cumulative LBF on IP growths. The nature of industrial production (IP) growths is very different from previous three datasets. IP growths are largely affected by economics states or endogenous shocks, such as economic policy, financial crisis or oil price[19], etc. Volatility persistence plays a less important role in IP growths than it does in previous applications. This is why IHMM shows an outstanding performanceăagainst GARCH-t and SV-t[20]. Two key findings are documented. First, in our application, we show the importance of modelling the time variation of state mixtures (i.e. the SV-IHMM versus the SV-DPM). The cumulative LBF of the SV-IHMM against the SV-DPM in Figure 3.1d shows that the Markov switching feature of the SV-IHMM contributes better density forecasts with a positive trend since 1987. Second, when comparing the SV-IHMM with the IHMM, we find that, before 2003, the volatility smooth change component (SV) plays an more important role than regime switching component (IHMM) with an increasing LBF of the SV-IHMM against the IHMM. However, the results are reversed after 2003 when data significantly supports the pure IHMM against the SV-IHMM. The decreasing LBF implies that the additional SV component is hurting the forecast power in this period. Nevertheless, the SV-IHMM outperforms all benchmark models in general, and underperforms the IHMM for IP only after 2003. A key message is that the SV-IHMM well accommodates and incorporates volatility smooth change (SV) and infinite Markov switching (IHMM) modelling in a unified framework.

[19]Maheu et al. (2019)
[20]Yang (2019)

3.5.3 Predictive Density

Predictive density curves can be constructed at each period and it represents the measure of the future uncertainties.[21] Different models could generate very different predictive densities on the same date. The LPLs show the accuracy of their predictive densities over the selected out-of-sample period. This section explains the source of gains in LPLs of SV-IHMM against benchmarks. SV-IHMM combines stochastic volatility, infinite mixtures and Markov switching in a unified framework. It is feasible for SV-IHMM to generate a board range of predictive curves, such as bell shape, fat tails, highly centralized and multiple modes. Figure 3.2 illustrates several examples of predictive densities on the selected dates across multiple models and datasets. The SV-IHMM produces substantial distributional variation for different times against SV-DPM where the mixture is static. For AAPL, USD/CAD and Oil, the log-predictive density curves instead of plain predictive densities are constructed due to its scale except for industrial production growths. For AAPL (Figure 3.2a), the top plot shows fatter tails with some degree of asymmetry for SV-IHMM relative to SV-DPM. The vertical line represents the ex-post return realization on that date. The bottom plot does not show much distinction for another date, showing that SV-IHMM can predict similarly as SV-DPM in some days and very differently on the other days. For USD/CAD in Figure 3.2b, a symmetrical curve is delivered with fatter tails. Such type of predictive curve suggests the future realization is likely to be at tails. The log-predictive density curves of oil are at the bottom left. Again, fat tails are produced by SV-IHMM compared to SV-DPM. Real industrial production growths are

[21] The predictive density curves we constructed are out-of-sample based.

illustrated in Figure 3.2d. The predictive density of IP is much more distinguishable than the other three datasets at the density level so that it is not necessary to take the logarithm. SV-IHMM produces an example of very dynamic non-bell shapes for predictive density curves. The top plot for August 1st, 2006 shows multiple modes by SV-IHMM with major weights at the center. In the bottom plot, SV-IHMM generates another multi-modal predictive curve, where one of the modes is relatively flat. It is impossible to see in a conventional model, such as GARCH and SV, to construct such an irregular predictive density curve. In summary, SV-IHMM shows that its flexibility not only could approach the data but also improves the forecast performance.

3.6 Sensitivity to Prior of Hyperparameters

The priors of precision parameters used for all computations are very informative and forecast lower probability to introduce new states.[22] The model performance is sometimes affected by the choice of hyper-parameter in the priors. This section applies loose hyper-priors on α_0 and β_0 and tests their impact on posterior estimation and forecast performance. Re-estimate and predict the SV-IHMM with a new hyper-parameter prior set: $\alpha_0 \sim Gamma(5,5)$ and $\beta_0 \sim Gamma(5,5)$. Table 3.5 compares the results of two different prior sets. The posterior estimates of the SV component are very similar between the two prior sets, and more states are used on average with the new prior set as expected. The out-of-sample forecast results of loose hyper-priors

[22] $\alpha_0 \sim Gamma(2,8)$ and $\beta_0 \sim Gamma(2,8)$ for SV-IHMM.

deteriorate from the tight hyper-priors only marginally except for industrial production growth rates, which shows significant improvements in density forecasts under loose hyper-priors.

The impact of setting loose hyper-priors on HDP can cause the model to introduce a larger state space to fit the data. Figure 3.3 illustrates the posteriors of the number of active states according to datasets and models (SV-IHMM, SV-DPM).[23] The top histogram shows the active state number posterior of AAPL for SV-DPM and SV-IHMM at tight and loose hyper-priors. As we let the hyper-priors to favour a large state space in both SV-IHMM and SV-DPM, a long tail extends to both ends and the center shifts rightwards, indicating that a board range of states are visited. For USD/CAD exchange rates, the corresponding posterior for SV-IHMM shifts at a much larger degree than SV-DPM. On the other side, the SV-IHMM shifts less distance than SV-DPM in oil prices. Actually, the posterior of the active state number for SV-IHMM in oil does not change as much as other datasets. Similar impacts are documented in industrial growth. As we can tell, imposing a loose hyper-prior can significantly change the state-space that model visits, but its influence on other parameters and out-of-sample forecast performances is only marginal.

[23]Posteriors are sampled with the full sample for each datasets. Each subfigure shows 2 posteriors histograms altogether.

3.7 Conclusion

This chapter proposes a new Bayesian semiparametric stochastic volatility model with Markovian mixtures. This model nests the SV-DPM model proposed by Jensen and Maheu (2010) and allows to investigate if the distribution of return innovations is changing over time. The empirical results suggest a positive answer to this question. The innovations are neither normal or Student-t, nor independently and identically distributed, no matter what type the asset is. This distributional time variation is of great importance and helps to explain why the SV-IHMM predicts better than the SV-t model and the SV-DPM model. The SV-IHMM combines the advantage of both the SV model and IHMM, and it's able to capture both the smooth changes of volatility and the changes in the unknown innovation distribution. In general, the SV-IHMM consistently outperforms all the benchmark models in general on out-of-sample prediction. Furthermore, the SV-IHMM is more adaptive than the iid-mixed SV-DPM model, especially in tail realizations. The results for the SV-IHMM are robust to different hyper-priors for the precision parameters.

Table 3.1: Descriptive Statistics

Returns	Mean	Median	StDev	Skew	Ex.Kurt	Min	Max
AAPL	0.0642	0.0000	2.9571	-1.7830	46.8446	-73.1248	28.6890
USD/CAD	-0.0009	0.0000	0.4074	0.1324	10.7373	-3.8070	5.0716
Crude oil	0.0099	0.0513	2.5245	-0.6538	13.6343	-40.6396	19.1506
IP index	0.2594	0.2818	1.8912	0.2989	10.6343	-10.9620	15.3200

Notes: AAPL, USD/CAD and Crude oil are daily log-returns, and the IP index is monthly log growths.

Table 3.2: Posterior Summary of Parameters

Panel A: AAPL(%)

	SV-IHMM		SV-DPM		SV-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
ϕ	0.9936	(0.9903, 0.9964)	0.9887	(0.9836, 0.9930)	0.9947	(0.9909, 0.9977)
σ_ν^2	0.0107	(0.0078, 0.0143)	0.0164	(0.0120, 0.0225)	0.0066	(0.0034, 0.0107)
ν					5.8864	(5.2184, 6.7063)
K	7.8989	(6.0000, 12.000)	5.3997	(3.0000, 9.0000)		

Panel B: USD/CAD(%)

	SV-IHMM		SV-DPM		SV-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
ϕ	0.9976	(0.9961, 0.9990)	0.9944	(0.9921, 0.9966)	0.9951	(0.9929, 0.9971)
σ_ν^2	0.0088	(0.0071, 0.0110)	0.0169	(0.0134, 0.0211)	0.0140	(0.0108, 0.0180)
ν					10.274	(8.4554, 12.694)
K	8.4898	(5.0000, 14.000)	4.0430	(2.0000, 7.0000)		

Panel C: Crude Oil(%)

	SV-IHMM		SV-DPM		SV-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
ϕ	0.9911	(0.9872, 0.9945)	0.9884	(0.9835, 0.9926)	0.9907	(0.9861, 0.9944)
σ_ν^2	0.0153	(0.0119, 0.0201)	0.0182	(0.0133, 0.0247)	0.0121	(0.0086, 0.0172)
ν					8.6553	(7.1545, 10.576)
K	6.0276	(5.0000, 9.000)	5.4026	(2.0000, 9.0000)		

Panel D: Industrial Production Growths(%)

	SV-IHMM		SV-DPM		SV-t	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
ϕ	0.9981	(0.9925, 0.9994)	0.9935	(0.9839, 0.9953)	0.9677	(0.9421, 0.9878)
σ_ν^2	0.0321	(0.0185, 0.0531)	0.0372	(0.0205, 0.0645)	0.1131	(0.0486, 0.1976)
ν					24.851	(6.9225, 48.620)
K	6.0276	(5.0000, 9.000)	5.4026	(2.0000, 9.0000)		

Table 3.3: Out-of-Sample Forecast Performance (Part 1)

	AAPL			USD/CAD		
	LPL	LBF	RMSFE	LPL	LBF	RMSFE
SV-IHMM	**-20736.70**	—	2.9403	**-3272.31**	—	0.4304
SV-DPM	-20756.95	20.25	2.9396	-3294.66	22.35	0.4305
SV-t	-20762.76	26.06	2.9397	-3301.91	29.60	0.4305
GARCH-t	-20806.00	69.29	2.9401	-3298.22	25.91	0.4305
IHMM	-20812.85	76.15	2.9427	-3369.79	97.48	0.4310
SV-N	-20894.22	157.5	2.9399	-3349.48	77.17	0.4305
GARCH-N	-21400.28	663.6	2.9399	-3566.49	294.18	0.4305
Notes:	9000 out-of-sample periods from Apr 26, 1982 to Dec 29, 2017.			10000 out-of-sample periods from Jun 10, 1976 to Nov 13, 2017.		

Table 3.4: Out-of-Sample Forecast Performance (Part 2)

	Crude Oil			Industrial Growth		
	LPL	LBF	RMSFE	LPL	LBF	RMSFE
SV-IHMM	**-17225.79**	—	2.5117	**-1616.40**	—	1.5282
SV-DPM	-17241.79	16.00	2.5117	-1633.73	17.33	1.5287
SV-t	-17254.04	28.25	2.5117	-1657.86	41.46	1.5290
GARCH-t	-17268.93	43.13	2.5127	-1663.39	46.99	1.5361
IHMM	-17282.15	56.36	2.5137	-1641.38	24.98	1.5345
SV-N	-17320.31	94.52	2.5117	-1658.84	42.44	1.5293
GARCH-N	-17523.57	297.8	2.5130	-1754.02	137.6	1.5359
Notes:	8000 out-of-sample periods from Feb 26, 1986 to Nov 13, 2017.			1165 out-of-sample periods from July 1922 to July 2019.		

Table 3.5: Robustness Test for Hyper-Prior

AAPL

	ϕ	σ_v^2	α_0	β_0	K	log PL	RMSE
$Gamma(5,5)$	0.9935	0.0116	3.7976	1.4712	9.7998	-20742.80	2.9407
$Gamma(2,8)$	0.9936	0.0107	1.5503	0.7339	7.8989	-20736.70	2.9403

FX

	ϕ	σ_v^2	α_0	β_0	K	log PL	RMSE
$Gamma(5,5)$	0.9977	0.0100	2.5569	2.0773	14.8310	-3271.67	0.4304
$Gamma(2,8)$	0.9976	0.0088	1.3126	0.7978	8.4898	-3272.31	0.4304

OIL

	ϕ	σ_v^2	α_0	β_0	K	log PL	RMSE
$Gamma(5,5)$	0.9912	0.0157	3.8383	1.2066	7.2819	-17229.74	2.5122
$Gamma(2,8)$	0.9911	0.0153	1.7355	0.5806	6.0276	-17225.79	2.5117

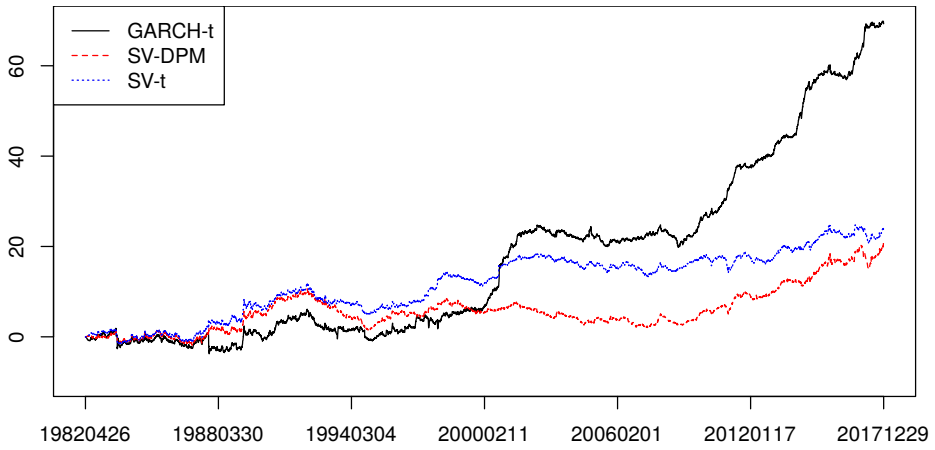
IP Index

	ϕ	σ_v^2	α_0	β_0	K	log PL	RMSE
$Gamma(5,5)$	0.9995	0.0140	0.4924	1.8955	10.09	-1522.3	1.5282
$Gamma(2,8)$	0.9994	0.0133	0.5391	0.7828	7.59	-1616.4	1.5282

Notes: Numbers under parameters are posterior averages for ϕ, σ_v^2, α_0, β_0 and K respectively.

FIGURE 3.1: Cumulative Log-Bayes Factor

(A) Equity: AAPL



(B) Exchange Rate: USD/CAD

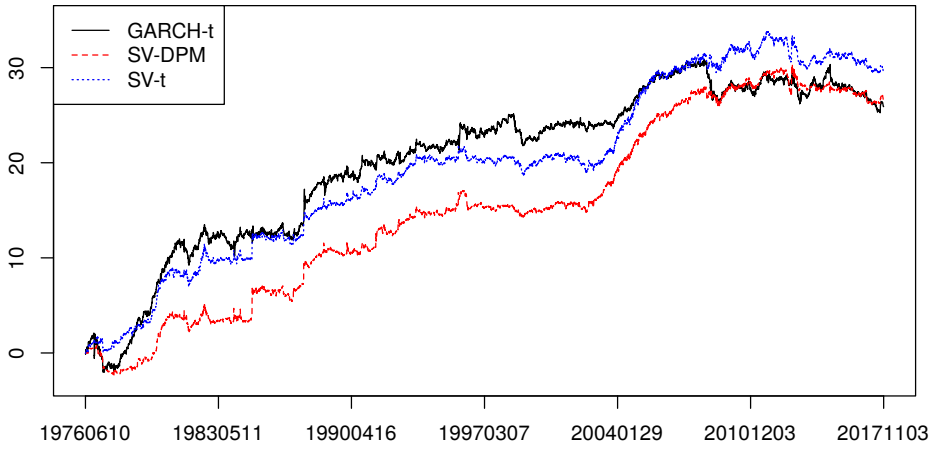
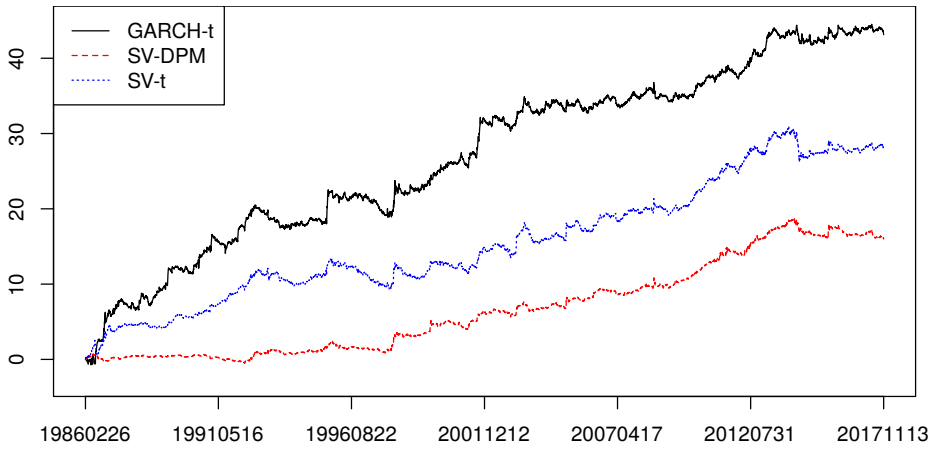


Figure 3.1: Cumulative Log-Bayes Factor (cont.)

(c) Commodity: OIL



(d) Macroeconomics: IP

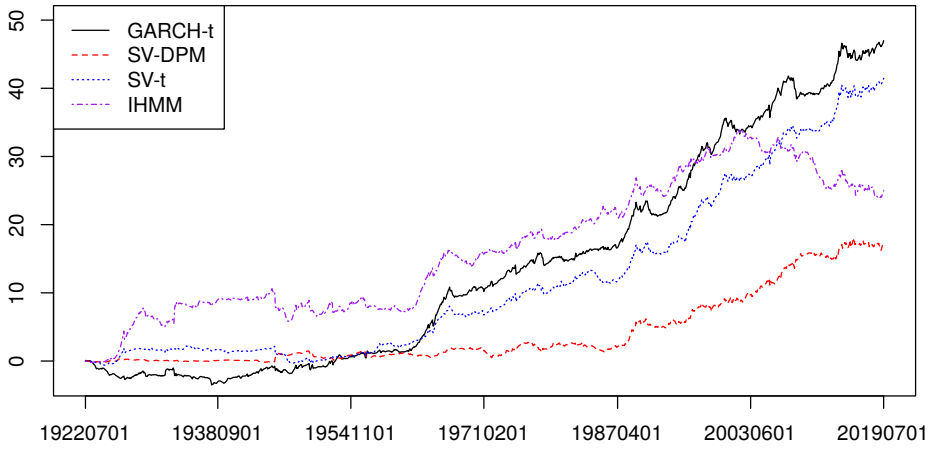
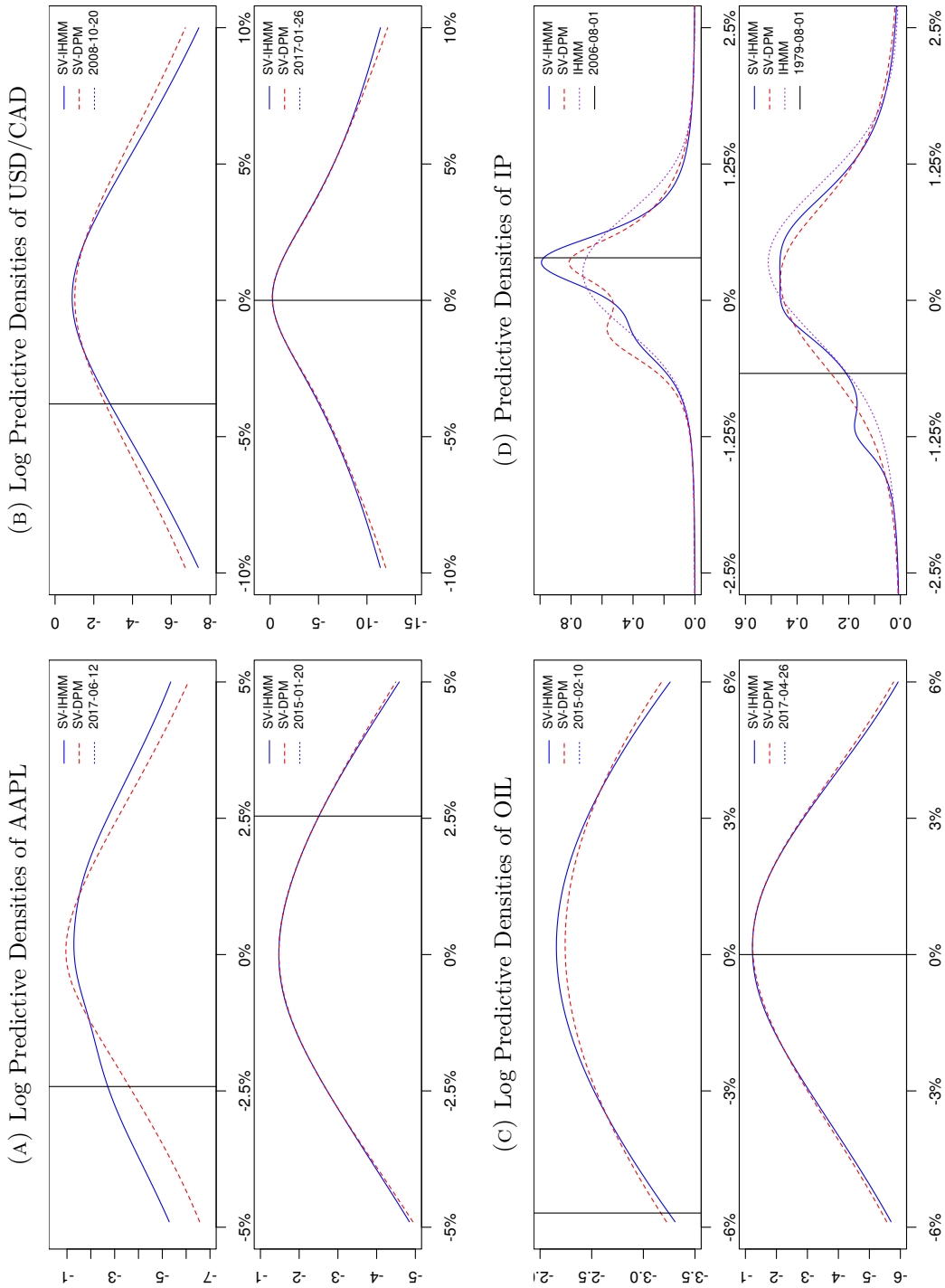


FIGURE 3.2: Predictive Densities Curve and Data Realizations



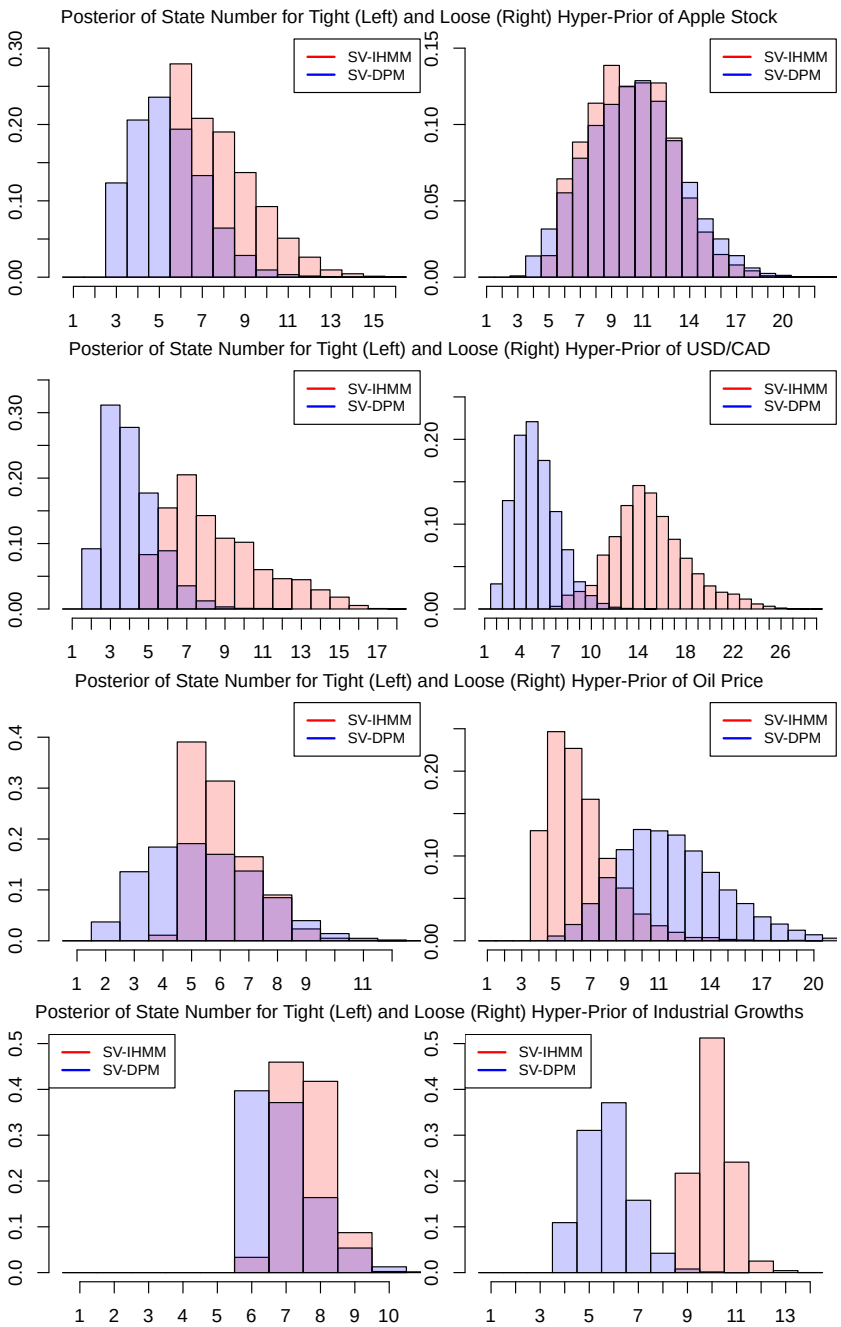
(A) Log Predictive Densities of AAPL

(B) Log Predictive Densities of USD/CAD

(C) Log Predictive Densities of OIL

(D) Predictive Densities of IP

FIGURE 3.3: Posterior of State Number



Notes: "Loose" indicates a prior of $Gamma(5,5)$ for both α_0 and β_0, and "Tight" indicates a prior of $Gamma(2,8)$ for both α_0 and β_0.

Chapter 4

Do Better Return Density Forecasts Lead to Economic Gains in Portfolio Allocation?

4.1 Introduction

Do econometric models with better density forecasts for returns translate into better portfolio allocations? To answer this question, this chapter proposes a new multivariate Bayesian semiparametric model that provides better density forecasts than benchmark models. This model approximates both the unknown conditional distribution and its short-term evolution nonparametrically through an infinite hidden Markov structure in addition to generalized autoregressive conditional heteroskedasticity (GARCH) dynamics. To test the model's economic performance in portfolio

allocation, this chapter maximizes an investor's expected utility by employing full information from predictive distributions, in contrast to the traditional mean-variance optimization. This portfolio problem will be sensitive to the shape of the predictive density of returns. The optimized portfolios (models) are then evaluated based on an ex-post performance fee. Empirical results show that a risk-averse investor is always willing to pay a significant positive management fee to switch from a benchmark model to the new model, regardless of investors utility type and risk-aversion level.

For better density forecasts, it is essential not only to properly approximate the stylized facts of conditional return distributions such as skewness and fat tails but also to approximate the time dependence of the whole conditional distribution. Most of the related work is done through parametric models such as dynamic volatility models (eg. GARCH) and finite-state Markov switching (MS) models. Examples include Fleming et al. (2001), Guidolin and Timmermann (2005, 2007), and Kalotychou et al. (2014) and Bollerslev et al. (2018). I show that both GARCH dynamics and an infinite Markov switching structure are empirically important.

The new model (MGARCH-IHMM) proposed in this chapter consists of a parametric multivariate GARCH (MGARCH) component and a Bayesian nonparametric infinite hidden Markov model (IHMM) component. The MGARCH component captures the strong dependence and smooth changes in conditional second moments. The IHMM component nests a finite-state MS model and captures the short-run changes in the conditional distribution. Further, the IHMM is a Bayesian nonparametric model that incorporates a hierarchical Dirichlet process, so each conditional return

distribution is a mixture of an infinite number of Gaussian kernels used to approximate distributional features like skewness, kurtosis and asymmetric tails. The mixing weights are allowed to change over time. To compare performance, this chapter selects three benchmark models: a multivariate GARCH (MGARCH) model with normal innovations, an MGARCH model with asymmetric volatility feedback and an IHMM without GARCH effects. The proposed MGARCH-IHMM exhibits a clear advantage in monthly density forecasts over all the benchmark models.

From a Bayesian perspective, maximizing expected utility is done with respect to the predictive density of returns. This integrates out all parameter and possible distributional uncertainties, so the portfolio allocation problem goes beyond the first two moments as implied in the classical Markowitz Portfolio Theory (MPT), which suggests a quadratic utility or normally distributed returns. Quadratic utility or quadratic approximation is usually considered as unrealistic, and evidence in Section 4.4.2 shows that the predictive return distributions from the Bayesian semi-parametric model are significantly skewed and leptokurtic. Consistent with these data features, the investor chooses her portfolio by optimizing her expected utility, where the problem, in general, depends on the whole predictive density and her utility function. Subsequently, the economic performance of the optimized portfolios is compared by evaluating the ex-post realized utilities. I show that a risk-averse investor is willing to pay at least 50 bps annually (and often more) to switch from a benchmark model to the MGARCH-IHMM. This result is robust to different utility functions and risk-aversion levels.

Currently, most of the work related to portfolio choice is done with parametric models and focuses on the point forecast of a few moments of returns and assumes either quadratic utilities or normal distributions. Fleming et al. (2001) are among the first who tested the relation between the moment forecasts, especially the volatility dynamics, and the portfolio allocation gain. Followed by Marquering and Verbeek (2004) and Kalotychou et al. (2014), they all optimize a mean-variance portfolio with a quadratic or approximately quadratic utility, while this chapter optimizes portfolios by maximizing the expected utility instead, and tests portfolio performances for robustness under three utility types: constant relative risk aversion (CRRA), constant absolute risk aversion (CRRA) and quadratic utility. Another shortfall of their work is that they only perform an in-sample test instead of out-of-sample, making their results questionable in practice. In this chapter, the portfolios are optimized with out-of-sample density forecasts and tested by ex-post realized returns.

None of the existing work considers the importance of modelling the whole return distribution flexibly for portfolio optimization problems. Bollerslev et al. (2018) test a mean-variance portfolio with realized covariances and compute performance fees under a CRRA utility, implying that the returns are normally distributed, which is against the findings in this chapter. Empirical results in Section 4.4.2 suggest that even the log return distributions are likely not elliptical, not to mention simple returns used in portfolio optimizations. Jondeau and Rockinger (2006) abandon mean-variance optimization by approximating a CARA utility with a fourth-order Taylor expansion, so the expected utility is also sensitive to skewness and kurtosis of return distributions.

This chapter computes expected utility using the exact utility function without any Taylor approximation and proposes a new econometric model by approximating the return distribution with a mixture of an infinite number of normal distributions.

Guidolin and Timmermann (2005, 2007) forecast the returns by a multivariate finite-state Markov switching (MS) model, and optimize the portfolio under CRRA utility. They provide a Monte Carlo method to compute the expected utility, but they only compare the realized utility itself of different portfolios, making the economic significance of their model unclear. This chapter builds on their work in several ways. First, their model only allows return distributions to change through a regime-switching scheme and does not incorporate GARCH dynamics as mine does. Moreover, my new model is a Bayesian semiparametric model that does not need to assume the number of states ex-ante as a finite-state MS model but estimates it jointly with other parameters. Second, this chapter computes expected utilities from the predictive densities of the model. Unlike Guidolin and Timmermann's method where the predictive samples are drawn from an asymptotic distribution estimated by MLE, the Bayesian method is finite-sample and essentially integrates out the distributional uncertainty of all parameters and latent variables. Last but not least, this chapter compares the economic performance between portfolios through an annualized break-even management fee motivated by Fleming et al., so the result is more intuitive than mere utility numbers.

Pettenuzzo and Timmermann (2011) and Pettenuzzo and Ravazzolo (2016) forecast the returns of a market portfolio using Bayesian methods and maximize the expected

utility of a CRRA form and compare portfolio performance by certainty equivalent returns (CER). However, CER is an ex-ante measure that assumes the return distribution predicted by an econometric model is correct, which is not guaranteed. The ex-post measure of break-even performance fee computed in this chapter is more appropriate and more realistic. Additionally, both of them only consider a combination of one risky asset and one risk-free asset, while this chapter focuses on a more general portfolio allocation problem with multiple risky assets and one risk-free asset.

This chapter is organized as follows. Section 4.2 illustrates the specification and the benefits of the proposed MGARCH-IHMM, along with the sampling algorithm and density forecast computation. Section 4.3 describes the details of the return series uses to test this model. Section 4.4 presents the posterior estimations of the MGARCH-IHMM and its out-of-sample forecasts. Section 4.5 compares the portfolio allocation performance of different econometric models under different utility function and risk-aversion levels. And Section 4.6 concludes.

4.2 MGARCH-IHMM Model

This chapter proposes a new Bayesian semiparametric model with an MGARCH component and an infinite hidden Markov model (IHMM) component. Jensen and Maheu (2013) propose an MGARCH model with Dirichlet process mixture (MGARCH-DPM model) that models the return innovations nonparametrically through a static infinite mixture. The new model proposed in this chapter extends this model by

replacing the DPM component into an infinite hidden Markov structure that allows the mixture to change over time. The proposed MGARCH-IHMM is more flexible and nests the MGARCH-DPM model as a special case. Let $\boldsymbol{r}_t$ be an $N \times 1$ vector of returns, and $\boldsymbol{r}_{1:T} = \{\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_T\}$. Define $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots\}$ as the set of state-dependent parameters, where $\boldsymbol{\Theta}_j = \left\{\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right\}$. The hierarchical representation of the model is

$$G_0|\beta_0, \Xi \sim DP\left(\beta_0, \Xi\right) \tag{4.1}$$

$$G_j|\alpha_0, G_0 \sim DP\left(\alpha_0, G_0\right), \quad j = 1, 2, \ldots \tag{4.2}$$

$$\boldsymbol{\Theta}_j|G_j \overset{\text{iid}}{\sim} G_j \tag{4.3}$$

$$\boldsymbol{r}_t|s_t, \boldsymbol{\Theta}, \mathcal{F}_{t-1} \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2'}\right) \tag{4.4}$$

$$\Xi = N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right) \times IW\left(\boldsymbol{\Sigma}_0, \nu + N\right), \ \nu > 0 \tag{4.5}$$

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot \left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1} \tag{4.6}$$

where $\mathcal{F}_{t-1}$ is the information set for $t-1$, $DP(\beta_0, \Xi)$ is a Dirichlet process with concentration parameter β_0 and base measure Ξ, where the concentration parameter represents how dispersed the draw is away from the base measure, and Ξ contains a normal prior $N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right)$ for $\boldsymbol{\mu}_{s_t}$ and an inverse Wishart prior $IW\left(\boldsymbol{\Sigma}_0, \nu + N\right)$ for $\boldsymbol{\Sigma}_{s_t}$. The Bayesian nonparametric component IHMM and the parametric component MGARCH of this model are linked through formula (4.4). s_t denotes the state/cluster for time t. s_t, $\boldsymbol{\mu}_{s_t}$ and $\boldsymbol{\Sigma}_{s_t}$ are determined by the IHMM component, and $\boldsymbol{H}_t$ is determined by the MGARCH component. $\boldsymbol{H}_t^{1/2}$ is the Cholesky decomposition of

$\boldsymbol{H}_t$. $\boldsymbol{\Sigma}_{s_t}$ is parametrized around identity matrix. The general level and long-run dynamics of conditional volatility is captured by the GARCH component, and $\boldsymbol{\Sigma}_{s_t}$ serves as an amplifier to boost up or shrink down the conditional covariance from $\boldsymbol{H}_t$. Clearly, when $\boldsymbol{\mu}_{s_t} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{s_t} = \boldsymbol{I}$, the MGARCH-IHMM reduces to a parametric MGARCH model.

The IHMM component of this model is a Bayesian nonparametric model employing a hierarchical Dirichlet process (HDP). (4.1) and (4.2) represents this HDP structure. There are an infinite number of states/clusters in the model, and the state-dependent parameter set $\boldsymbol{\Theta}_j$ is an iid draw from their corresponding probability measure G_j. Accordingly, there are also an infinite number of G_j's, and each one of them is drawn from a separate bottom layer Dirichlet process with precision parameter α_0 and base measure G_0. Further, G_0 is a draw from the top layer Dirichlet process with precision parameter β_0 and base measure Ξ. Ξ can be also seen as the prior of the state-dependent parameters $\boldsymbol{\Theta}$, a multivariate normal prior for $\boldsymbol{\mu}_{s_t}$ and an inverse Wishart prior for $\boldsymbol{\Sigma}_{s_t}$. The IHMM component is essentially an infinitely dimensional Markov switching (MS) model. It is designed to capture the sudden changes in the conditional distribution through a regime-switching scheme. This regime-switching behaviour is shown more clearly later in the stick-breaking representation. Because it's a Bayesian nonparametric model, one does not need to impose any distributional assumption to the conditional return distribution, so it can capture the features such as asymmetries and fat tails. Moreover, unlike another Bayesian nonparametric model DPM, where

the conditional distribution is static, IHMM also approximates the unknown short-term evolution of the conditional distributions nonparametrically.

The MGARCH component takes a variant of the diagonal BEKK-GARCH representation (Engle and Kroner 1995). $\boldsymbol{C}$ is an $N \times N$ lower triangular matrix, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are $N \times 1$ vectors, and $\odot$ is the Hadamard operator representing element-by-element multiplication. Parameter restriction of $\alpha_i^2 + \beta_i^2 < 1$ for all $i = 1, \ldots, N$ is imposed for stationarity in $\boldsymbol{H}_t$. Unlike traditional MGARCH models where the source of $\boldsymbol{H}_t$ variation is the last period return shock (the difference between the last period return and the conditional mean), the $\boldsymbol{H}_t$ dynamics in this model is determined by the change of returns around an additional parameter $\boldsymbol{\eta}$. This allows the model to capture the potential asymmetric volatility feedback effect. When $\boldsymbol{\eta} > \boldsymbol{\mu}_{s_{t-1}}$, for a $\boldsymbol{r}_{t-1}$ less than $\boldsymbol{\mu}_{s_{t-1}}$, $\|\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\| > \|\boldsymbol{r}_{t-1} - \boldsymbol{\mu}_{s_{t-1}}\|$, so this would increase $\boldsymbol{H}_t$ as in an asymmetric dynamic covariance (ADC) model (Kroner and Ng 1998). However, the MGARCH-IHMM does not enforce asymmetric volatility feedback or the sign of this asymmetry, but let it learn from data. Posterior estimates of Fama-French 5 industry portfolio returns show that η_i is generally greater than $\mu_{i,s_{t-1}}$ for each asset i.

This model nests many models as special cases. For example, when $\alpha_0 \to 0$, $G_j \to \delta_x$ where δ_x is a Dirac measure centred at x, and x is G_0 distributed. Then the MGARCH-IHMM reduces to the MGARCH-DPM model from Jensen and Maheu (2013) in this case. Additionally, if $\beta_0 \to 0$, the MGARCH-IHMM reduces to a parametric MGARCH model with normal innovation, and if $\beta_0 \to \infty$, the MGARCH-IHMM reduces to a parametric MGARCH model with Student's-t style innovation.

From a mixture perspective, this model can be rewritten into a stick-breaking representation (Sethuraman 1994; Teh et al. 2006):

$$\Gamma|\beta_0 \sim GEM\left(\beta_0\right) \tag{4.7}$$

$$\Pi_j|\alpha_0, \Gamma \sim DP\left(\alpha_0, \Gamma\right) \tag{4.8}$$

$$s_t|s_{t-1}, \Pi \sim \Pi_{s_{t-1}} \tag{4.9}$$

$$p\left(\boldsymbol{r}_t|\boldsymbol{\Theta}, \Pi, s_{t-1}\right) = \sum_{k=1}^{\infty} \pi_{s_{t-1}k} N\left(\boldsymbol{r}_t; \boldsymbol{\mu}_k, \boldsymbol{H}_t^{1/2} \boldsymbol{\Sigma}_k \boldsymbol{H}_t^{1/2'}\right) \tag{4.10}$$

$$\boldsymbol{\mu}_k \sim N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right), \quad \boldsymbol{\Sigma}_k \sim IW\left(\boldsymbol{\Sigma}_0, \nu + N\right) \tag{4.11}$$

where $\Gamma = (\gamma_1, \gamma_2, \dots)'$, $\Pi_j = (\pi_{j1}, \pi_{j2}, \dots)$, and $\boldsymbol{H}_t$ is defined as in (4.6). To be more specific,

$$\gamma_k = \hat{\gamma}_k \prod_{l=1}^{k-1} \left(1 - \hat{\gamma}_l\right), \qquad \hat{\gamma}_k \sim Beta\left(1, \beta_0\right) \tag{4.12}$$

$$\pi_{jk} = \hat{\pi}_{jk} \prod_{l=1}^{k-1} \left(1 - \hat{\pi}_{jl}\right), \qquad \hat{\pi}_{jk} \sim Beta\left(\alpha_0 \gamma_k, \alpha_0 \left(1 - \sum_{l=1}^{k} \gamma_l\right)\right). \tag{4.13}$$

$GEM(\beta_0)$[1] is a general stick-breaking process with a precision parameter β_0. As shown in equation (4.10), the conditional distribution of returns is a mixture of an infinite number of Gaussian kernels with a vector of weights Π_j. Π_j is the jth row of the infinitely dimensional squared transition matrix Π and a draw from a particular Dirichlet process. Because the IHMM is also a Bayesian nonparametric extension of the MS model, all the states can recur with certain probabilities. To ensure this state

[1]GEM stands for Griffiths, Engen, and McCloskey. See Pitman (2002) as an example.

recurrence, another Dirichlet process is required to make the atoms shared among all the bottom layer Dirichlet processes. π_{jk} indicates the probability of switching from state j to state k. Note that in the MGARCH-DPM model, $\pi_{s_{t-1}k} = \pi_k$ for all $t = 1, \ldots, T$, so the MGARCH-DPM model is nested in the MGARCH-IHMM. This scenario will occur when $\alpha_0 \to 0$. Please note that (4.13) is not applicable when α_0 approaches to 0, as it's derived from the formal definition of Dirichlet process where the precision parameter is strictly positive.

The state-dependent parameters $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ allow identifying potential regime switches. Furthermore, mixing over $\boldsymbol{\mu}_k$ also generates possible skewness in conditional return distribution, and mixing over $\boldsymbol{\Sigma}_k$ generates kurtosis. Since the mixture weights are Markovian, the unknown conditional distribution is allowed to change over time in an unknown pattern. In summary, the proposed MGARCH-IHMM retains all the advantages from both the MGARCH model and IHMM. It is designed to capture the highly persistent long-run volatility dynamics and potential drastic regime switches. Additionally, it approximates both the shape and the short-term evolution of the unknown conditional distributions nonparametrically.

4.2.1 Hierarchical Priors

Prior settings are very important in Bayesian inferences, especially when estimating Bayesian nonparametric models where the number of useful states is estimated jointly with other parameters. Each state-dependent parameter is estimated based on the subsamples assigned to the corresponding state. For states with fewer observations,

parameter posteriors can be heavily influenced by priors compared to states with more observations where posteriors are dominated by the likelihood. On the other hand, whenever a new state is introduced, the parameters are drawn directly from the priors. Priors far away from the support of data can easily make the new state wasted.

This problem can be successfully solved by introducing hierarchical priors, a prior of the base measure parameters. The base measure parameters are now estimated from the state-dependent parameters instead of preset as constants. This allows the base measure to learn from data, so the state-dependent parameters, especially those of a new state, can always land in a reasonable region, and then greatly improves both the in-sample fits and out-of-sample forecasts.

Consider a set of hierarchical priors motivated by Song (2014) and Maheu and Yang (2016):

$$\boldsymbol{b}_0 \sim N\left(\boldsymbol{h}_0, \boldsymbol{H}_0\right), \quad \boldsymbol{B}_0 \sim IW\left(\boldsymbol{A}_0, a_0\right), \quad \boldsymbol{\Sigma}_0 \sim W\left(\boldsymbol{C}_0, d_0\right), \quad \nu \sim Exp\left(g_0\right). \tag{4.14}$$

Then $\boldsymbol{b}_0$, $\boldsymbol{B}_0$, $\boldsymbol{\Sigma}_0$ and ν are drawn conditional on both hierarchical priors and the corresponding parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$).[2]

4.2.2 Covariance Targeting

In the MGARCH component, $\boldsymbol{C}$ has $N(N+1)/2$ parameters to estimate, while $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ all have N parameters respectively. Obviously, the number of parameters

[2]See Appendix 4 for more details.

grows quadratically in $\boldsymbol{C}$ and linearly in $\boldsymbol{\theta}_H = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}\}$. Hence, by targeting the symmetric $\boldsymbol{CC}'$ matrix instead of estimating it, one can greatly reduce the total number of parameters in estimation. Let $\bar{\boldsymbol{\mu}}$ be the sample unconditional expectation and $\bar{\boldsymbol{H}}$ be the sample unconditional covariance matrix. In a reduced form of the MGARCH-IHMM where $\boldsymbol{\Sigma}_t = \boldsymbol{I}$ for all t, referring to equation (4.6), the unconditional expectation of $\boldsymbol{H}_t$ is

$$\begin{aligned}
\mathrm{E}\left(\boldsymbol{H}_t\right) &= \boldsymbol{CC}' + \boldsymbol{\alpha\alpha}' \odot \mathrm{E}\left[\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \boldsymbol{\eta}\right)'\right] + \boldsymbol{\beta\beta}' \odot \mathrm{E}\left(\boldsymbol{H}_{t-1}\right) \\
&= \boldsymbol{CC}' + \boldsymbol{\alpha\alpha}' \odot \mathrm{E}\left[\left(\boldsymbol{r}_{t-1} - \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)\left(\boldsymbol{r}_{t-1} - \bar{\boldsymbol{\mu}} + \bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)'\right] + \boldsymbol{\beta\beta}' \odot \mathrm{E}\left(\boldsymbol{H}_{t-1}\right) \\
&= \boldsymbol{CC}' + \boldsymbol{\alpha\alpha}' \odot \bar{\boldsymbol{H}} + \boldsymbol{\alpha\alpha}' \odot \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)\left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)' + \boldsymbol{\beta\beta}' \odot \mathrm{E}\left(\boldsymbol{H}_{t-1}\right).
\end{aligned}$$

Further assuming $\mathrm{E}\left(\boldsymbol{H}_t\right) = \bar{\boldsymbol{H}}$ for all $t = 1, \ldots, T$, we have

$$\boldsymbol{CC}' = \bar{\boldsymbol{H}} \odot \left[\mathbf{1} - \boldsymbol{\alpha\alpha}' - \boldsymbol{\beta\beta}'\right] - \boldsymbol{\alpha\alpha}' \odot \left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)\left(\bar{\boldsymbol{\mu}} - \boldsymbol{\eta}\right)' \tag{4.15}$$

where $\mathbf{1}$ is a $N \times N$ matrix with all the elements are all 1. Note that any draw of $\boldsymbol{\theta}_H$ from the posterior that results in non-positive definite $\boldsymbol{CC}'$ is rejected.

4.2.3 Sampling Algorithm

The MGARCH-IHMM is estimated through an MCMC algorithm. For the Bayesian nonparametric component (IHMM), I employ the beam sampler introduced by Van Gael et al. (2008) (see also Fox et al. 2011; Maheu and Yang 2016). Similar to the slice sampler for DPM model, the beam sampler partitions the infinite number of states in

IHMM into a finite set of "major" states with observations assigned and an additional "remaining" state without any observation assigned. Whenever introducing a new state, the corresponding data density parameters are drawn from the hierarchical priors directly since there's no data assigned to that state yet. Let state R be the "remaining" state, then $\Gamma = (\gamma_1, \ldots, \gamma_K, \gamma_R)'$ and $\Pi_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jR})$, where $\gamma_R = \sum_{k=K+1}^{\infty} \gamma_k = 1 - \sum_{k=1}^{K} \gamma_k$ and $\pi_{jR} = \sum_{k=K+1}^{\infty} \pi_{jk} = 1 - \sum_{k=1}^{K} \pi_{jk}$. The sampling steps are:

1. sample the auxiliary slice variable $u_{1:T}|\Gamma, \Pi$.

2. update K. If a new "major" state is introduced, draw the corresponding parameters and transition probability from the prior. The transition matrix now has an additional column and row.

3. forward filter backward sampler for the state variable $s_{1:T}|\boldsymbol{r}_{1:T}, u_{1:T}, \Gamma, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:T}$.

4. simulate the coloured ball counts in the "oracle" urn $c_{1:K}|s_{1:T}, \Gamma, \alpha_0$.[3]

5. sample β_0 and α_0 following Fox et al. (2011).

6. sample $\Gamma|c_{1:K}, \beta_0$.

7. sample $\Pi|n_{1:K,1:K}, \Gamma, \alpha_0$.

8. sample the state-dependent parameters $\boldsymbol{\Theta}|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{H}_{1:T}$.

[3]The "oracle" urn is a special urn in the hierarchical Pòlya urn scheme introduced by Beal et al. (2002). Please refer to it for details.

9. sample hierarchical priors.

   (a) sample $\boldsymbol{b}_0|\boldsymbol{\mu}_{1:K}, \boldsymbol{B}_0, \boldsymbol{h}_0, \boldsymbol{H}_0$.

   (b) sample $\boldsymbol{B}_0|\boldsymbol{\mu}_{1:K}, \boldsymbol{b}_0, a_0, \boldsymbol{A}_0$.

   (c) sample $\nu|\sigma^2_{1:K}, s_0, g_0$.

   (d) sample $\boldsymbol{\Sigma}_0|\boldsymbol{\Sigma}_{1:K}, v_0, \boldsymbol{C}_0, d_0$.

10. sample GARCH parameters $\boldsymbol{\theta}_H = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}\}\,|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{\Theta}$. Apply a block-move random-walk Metropolis-Hastings algorithm to sample $\boldsymbol{\theta}_H$.

Repeat the above steps, discard the first M_0 samples as burn-in, and collect the following M samples. Details of the sampling algorithm can be found in Appendix 4. After simulating all the MCMC samples, the posterior mean of each parameter and latent variable can be computed by

$$\mathrm{E}(\theta|\boldsymbol{r}_{1:T}) \approx \frac{1}{M}\sum_{i=1}^{M}\theta^{(i)},$$

where $\theta^{(i)}$ is the ith MCMC draw of the given parameter θ.

4.2.4 Predictive Likelihood

Define the predictive likelihood for a particular out-of-sample period as $p\left(\boldsymbol{r}_{t+1}|\mathcal{F}_t\right)$, where $\mathcal{F}_t$ is the information set for time t. This predictive likelihood is computed as follows:

1. Estimate the model for $\boldsymbol{r}_{1:t}$ and collect M posterior samples for all the parameters $\left\{\boldsymbol{\Theta}^{(i)}, \boldsymbol{\theta}_H^{(i)}, \Pi^{(i)}, s_{1:t}^{(i)}, K^{(i)}\right\}_{i=1}^{M}$ as described in Section 4.2.3.

2. Evaluate the predictive likelihood for the realized return $\boldsymbol{r}_{t+1}$ conditional on every MCMC sample

$$\begin{aligned} p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}^{(i)}, \boldsymbol{\theta}_H^{(i)}, \Pi^{(i)}, s_t^{(i)}) = \sum_{k=1}^{K^{(i)}} \pi^{(i)}_{s_t^{(i)}k} p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}_k^{(i)}, \boldsymbol{\theta}_H^{(i)}) \\ + \pi^{(i)}_{s_t^{(i)}R} p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}_R^{(i)}, \boldsymbol{\theta}_H^{(i)}) \\ \boldsymbol{\Theta}_R^{(i)} \sim N\left(\boldsymbol{b}_0^{(i)}, \boldsymbol{B}_0^{(i)}\right) \times IW\left(\boldsymbol{\Sigma}_0^{(i)}, \nu^{(i)} + N\right). \end{aligned}$$

   Note that $K^{(i)}$ is the number of alive clusters (the "major" states) for the ith MCMC sample. The corresponding state-dependent parameters for the extra cluster (the "remaining" state) are drawn directly from the base measure.

3. Average out the conditional predictive likelihoods with respect to the MCMC draws

$$p(\boldsymbol{r}_{t+1}|\mathcal{F}_t) \approx \frac{1}{M} \sum_{i=1}^{M} p(\boldsymbol{r}_{t+1}|\boldsymbol{r}_{1:t}, \boldsymbol{\Theta}^{(i)}, \boldsymbol{\theta}_H^{(i)}, \Pi^{(i)}, s_t^{(i)})$$

Then the log-predictive likelihood over an out-of-sample period $t+1:T$ is

$$\log\left(p(\boldsymbol{r}_{t+1:T}|\mathcal{F}_t)\right) = \sum_{l=t}^{T} \log\left(p(\boldsymbol{r}_{l+1}|\mathcal{F}_l)\right)$$

The log-predictive likelihood is essentially evaluating the predictive density at the

actual return realization point. It can serve as a metric when comparing the out-of-sample performance among different models. It measures how likely the actually out-of-sample returns can be realized given the predictive densities forecasted by the provided model. To compare, one can compute the difference of log-predictive likelihoods between two models, also called the log-Bayes factor. A log-Bayes factor greater than 5 is usually considered as strong evidence supporting one model over another.

4.2.5 Benchmark Models

IHMM A fully nonparametric multivariate model with regime switching is specified as

$$\Gamma|\beta_0 \sim GEM\left(\beta_0\right), \quad \Pi_j|\alpha_0, \Gamma \sim DP\left(\alpha_0, \Gamma\right)$$

$$s_t|s_{t-1}, \Pi \sim \Pi_{s_{t-1}}$$

$$\boldsymbol{r}_t|s_t, \boldsymbol{\Theta}, \mathcal{F}_{t-1} \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{\Sigma}_{s_t}\right)$$

$$\boldsymbol{\mu}_s \sim N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right), \quad \boldsymbol{\Sigma}_s \sim IW\left(\boldsymbol{V}, \nu + N\right)$$

$$\boldsymbol{b}_0 \sim N\left(\boldsymbol{h}_0, \boldsymbol{H}_0\right), \quad \boldsymbol{B}_0 \sim IW\left(\boldsymbol{A}_0, a_0\right), \quad \boldsymbol{\Sigma}_0 \sim W\left(\boldsymbol{C}_0, d_0\right), \quad \nu \sim Exp\left(g_0\right)$$

This model is similar to the MGARCH-IHMM but without the MGARCH component. So this model is nested in the MGARCH-IHMM when $\boldsymbol{H}_t = \boldsymbol{I}$ for all t. The same set of hierarchical priors are also employed in this model.

MGARCH-N A fully parametric multivariate GARCH model with normal innovation is specified as

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t, \quad \boldsymbol{z}_t \overset{\text{iid}}{\sim} N(\boldsymbol{0}, \boldsymbol{I})$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\boldsymbol{r}_{t-1} - \boldsymbol{\mu})(\boldsymbol{r}_{t-1} - \boldsymbol{\mu})' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

where $\boldsymbol{C}\boldsymbol{C}'$ is defined through covariance targeting as

$$\boldsymbol{C}\boldsymbol{C}' = \bar{\boldsymbol{H}} \odot (\boldsymbol{1} - \boldsymbol{\alpha}\boldsymbol{\alpha}' - \boldsymbol{\beta}\boldsymbol{\beta}')$$

MGARCH-A A fully parametric asymmetric multivariate GARCH model with normal innovation is specified as

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t, \quad \boldsymbol{z}_t \overset{\text{iid}}{\sim} N(\boldsymbol{0}, \boldsymbol{I})$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\boldsymbol{r}_{t-1} - \boldsymbol{\eta})(\boldsymbol{r}_{t-1} - \boldsymbol{\eta})' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

where $\boldsymbol{C}\boldsymbol{C}'$ is targeted as in equation (4.15). Unlike the MGARCH-N model, this model can also capture the potential volatility feedback asymmetry.

4.3 Data

The data used in this chapter include monthly returns of the Fama-French 5 industry portfolios, consists of Consumer, Manufacture, High Tech, Health and Other, and

the 1-month Treasury bill rate from Kenneth French website. All the returns range from July 1926 to August 2018 at a monthly frequency, 1106 observations in total. Furthermore, all the returns are converted into log-returns for continuous compounding and scaled by 100 for percentage values.

Table 4.1 is approximately here.

Panel A of Table 4.1 illustrates some univariate descriptive statistics of the five industry portfolio log-returns. The expected returns are around 0.7% – 0.9%, and standard deviation around 1.09% – 1.26%. All the industries are negatively skewed and leptokurtic. Panel B shows that all the five industries are highly but not perfectly correlated, so the diversification benefit still exists. Moreover, as discussed later in Section 4.5, if an investor's utility is not quadratic, diversification benefits are not solely determined by correlations.[4]

4.4 Model Performance

4.4.1 Estimation

In all three models, the first 20000 iterations are discarded as burn-in. Further, the MCMC chain is thinned by 10 to reduce dependence, and there are 20000 MCMC samples collected for posterior inference. This section summarizes the results of full sample estimates consisting of 1106 observations.

[4]Also see the simulation experiment in Jondeau and Rockinger (2006) where the sensitivities to skewness and kurtosis are tested for portfolio allocations.

Table 4.2 is approximately here.

Panel A of Table 4.2 lists the posterior estimates of non-state-dependent parameters for three models, and the assets in $\boldsymbol{r}_t$ are indexed in the same order as in Table 4.1. The MGARCH-IHMM, the MGARCH-N model and the MGARCH-A model all have high $\boldsymbol{\beta}$ that are above 0.93, and low $\boldsymbol{\alpha}$ that are below 0.3 in general. As mentioned in Section 4.2, $\boldsymbol{\eta}$ helps to capture the asymmetric volatility feedback. Comparing the $\boldsymbol{\eta}$ estimates and the posterior means of $\boldsymbol{\mu}$ over time in Panel B, η_i's are in general greater than the corresponding μ_i. It shows that the Fama-French 5 industry portfolio returns usually exhibit negative feedback to their volatility (a return lower than expectation leading to higher volatility next period). This is also the case for the MGARCH-A model, where η_i's are greater than the corresponding μ_i except for the 3rd portfolio (High Tech). The MGARCH persistences $(\alpha_i^2 + \beta_i^2)$ in the MGARCH-IHMM are almost the same as those in the MGARCH-N model and the MGARCH-A model, although each α_i's and β_i's are a little different. However, the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ estimates are almost identical between the MGARCH-N and the MGARCH-A, so the difference in the MGARCH-IHMM is unlikely due to the asymmetric volatility feedback but the Bayesian nonparametric IHMM component. On the other hand, the 0.95 density intervals for the GARCH parameters are slightly wider in the MGARCH-IHMM than those in the MGARCH-N model and the MGARCH-A model. For example, the 0.95 density interval for α_1 is $(0.1932, 0.2493)$ in the MGARCH-IHMM, $(0.1932, 0.2493)$ in the MGARCH-N model and $(0.2392, 0.2795)$ in the MGARCH-A model. The posterior for GARCH parameters in the MGARCH-IHMM are in general more dispersed probably due to the complexity of the model.

Comparing the semiparametric model MGARCH-IHMM and the nonparametric model IHMM, the IHMM has higher concentration parameters ($\alpha_0 = 0.8150$ and $\beta_0 = 0.8207$ on average) than the MGARCH-IHMM ($\alpha_0 = 1.8456$ and $\beta_0 = 0.9249$ on average) but a lower number of active states ($K = 8.3095$ compared to $K = 9.9907$ on average).

Panel C lists the posterior estimates for the hierarchical prior parameters. The IHMM suggests a more concentrated base measure for $\boldsymbol{\mu}_s \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0)$ with higher mean (0.58 – 0.88 in $\boldsymbol{b}_0$) and lower variance (0.34 – 0.81 for diagonal elements in $\boldsymbol{B}_0$), while in the MGARCH-IHMM, the base measure for the mean parameter is closer at zero (0.06 – 0.54 in $\boldsymbol{b}_0$) and more disperse (0.58 – 1.33 for diagonal elements in $\boldsymbol{B}_0$).

For the state-dependent covariance $\boldsymbol{\Sigma}_s$, we find the degree of freedom ν is higher in the MGARCH-IHMM (1.35 compared to 0.05 in the IHMM), indicating that the hierarchical prior for the IHMM is much more dispersed. Moreover, in the scale parameter $\boldsymbol{\Sigma}_0$, the MGARCH-IHMM gives lower diagonal elements (1.78 – 2.56) compared to those in the IHMM (4.29 – 5.07) and lower covariances (-0.03 – 0.31 compared to 0.17 – 0.66 in the IHMM). This is because, in the MGARCH-IHMM, the state-dependent covariance parameter captures the regime-switching effect *around* the GARCH dynamics, while in the IHMM, all the covariance changes must be captured by $\boldsymbol{\Sigma}_s$. This behaviour is also shown in Section 4.4.2.

Figure 4.1 is approximately here.

In addition to the GARCH dynamics, the Fama-French 5 industry portfolios still exhibit clear regime-switching behaviour, implying that both long-run smooth changes and short-run regime switches are important. Figure 4.1 plots the heat map generated by the MGARCH-IHMM. This heat map shows the empirical probability that two periods share the same state. The redder the colour, the higher the probability of sharing states. There are three dominating states: an "ordinary" state, a "bear" state with low expected return and high volatility and a "mild" state with low expected return and low volatility. Most periods fall in the "ordinary" state, but there are several exceptions. For example, the "bear" state is shared by multiple eras, including but not limited to the Great Depression identified from May 1928 to November 1932, the Oil Crisis identified from September 1973 to February 1975, the Tech Bubble burst identified from May 1998 to March 2001, and the recent financial crisis identified from November 2007 to April 2009. The "mild" state is also shared by several eras, mostly the slow down within the economic expansion after World War II, identified as of November 1947 – December 1948, March 1951 – November 1953, and March 1956 – October 1957. Note that these are states identified *in addition to* the GARCH effect, indicating that the regular MGARCH-N model is insufficient especially during crises.

Figure 4.2 is approximately here.

Figure 4.2 plots the posterior means of some time-varying variables for the return of the Consumer industry portfolio, with the previously mentioned major "bear" and "mild" periods shaded in red and green respectively. The "bear" state (red) is accompanied by low expected return and high volatility, which represents slow down

in investors' wealth growth and high investment uncertainty. In contrast, the "mild" state (green) not only has a low expected return but also low volatility. This suggests that the market is rather calm and confident even the wealth growth is low. Because the state switching captures dynamics in addition to the GARCH recursion if the MGARCH component is sufficient for conditional covariance dynamics, the state-dependent covariance $\boldsymbol{\Sigma}_t$ will be an identity matrix, which clearly is not the case here. Therefore, the "bear" state, where the state-dependent variance is considerably greater than one (the horizontal dashed line in the second graph), compensates for the extra volatility that the MGARCH component fails to capture, and the "mild" state, where the state-dependent variance is lower than one, shrinks down the overall conditional variance of return. As for the $\boldsymbol{H}_t$ dynamics, the "bear" state is usually when $\boldsymbol{H}_t$ is increasing, and the "mild" state is usually when $\boldsymbol{H}_t$ is decreasing. It implies that these states are identified because $\boldsymbol{H}_t$ is not climbing or declining fast enough, and an additional multiplier $\boldsymbol{\Sigma}_t$ is required to boost the conditional covariances.

The last graph of Figure 4.2 plots the posterior means of overall conditional covariances estimated from each model, namely $\mathrm{E}\left(\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_k\boldsymbol{H}_t^{1/2'}|\text{MGARCH-IHMM}\right)$, $\mathrm{E}\left(\boldsymbol{\Sigma}_{s_t}|\text{IHMM}\right)$, $\mathrm{E}\left(\boldsymbol{H}_t|\text{MGARCH-N}\right)$ and $\mathrm{E}\left(\boldsymbol{H}_t|\text{MGARCH-A}\right)$. The conditional variance estimated from the MGARCH-IHMM is the most flexible one among all the selected models with greater range and smoother change.

Figure 4.3 is approximately here.

Figure 4.3 plots the dynamic of second moment matrices as a whole through log determinants. The upper graph plots the log determinants of the state-dependent

parameter $\mathbf{\Sigma}_{s_t}$. As discussed earlier, $\mathbf{\Sigma}_{s_t}$ is parametrized around identity matrix whose log determinant is 0. If $\mathbf{\Sigma}_{s_t}$ equals to identity, the conditional covariance dynamics of the MGARCH-IHMM reduces to the same as the MGARCH-A model. From the graph, the log determinants are greater than 0 in the "bear" state and less than 0 in the "mild" state, indicating that the regime-switching behaviour not only occurs in the diagonal elements but also the whole matrix. The lower graph plots the log determinant of the overall conditional covariance matrix defined above. For the IHMM, the log determinant of covariances change drastically and switch among different levels, while for the MGARCH-N and MGARCH-A model, the log determinant can change gradually, but the range that it can change is confined. In the MGARCH-IHMM, the log determinant of the second moments can also change gradually, and it can change more freely to a wider range than that in the MGARCH-N or the MGARCH-A model.

4.4.2 Forecast

Table 4.3 is approximately here.

In Bayesian econometrics, one usually compares forecasts from different models by log predictive likelihoods or log Bayes factors as discussed in Section 4.2.4. Empirical results show that the MGARCH-IHMM outperforms the benchmark IHMM, MGARCH-N and MGARCH-A model in terms of density forecasts. A recursive prediction is performed for 200 out-of-sample periods from January 2002 to August 2018 by re-estimating each model for each period. Table 4.3 compares the performance of these recursive forecasts among the four models. The difference in log-predictive

likelihoods between two models represents a log Bayes factor, and a number greater than 6 is usually considered as strong evidence that one model predicts better than the other. The MGARCH-IHMM produces a log-predictive likelihood of -2409.809, 27.4454 higher than the log-predictive likelihood of the IHMM at -2437.254, 27.3832 higher than that of the MGARCH-N at -2437.192, and 22.3156 higher than that of the MGARCH-A at -2432.124. However, the MGARCH-IHMM doesn't show an advantage in point forecast. All four models yield similar root mean squared forecast error (RMSFE)[5], with 15.4987 for the MGARCH-IHMM, 15.4810 for the IHMM, 15.6134 for the MGARCH-N model and 15.6100 for the MGARCH-A model.

Figure 4.4 is approximately here.

Figure 4.5 is approximately here.

Figure 4.4 plots the cumulative log-Bayes factors over the out-of-sample periods. The MGARCH-IHMM generally outperforms the IHMM, the MGARCH-N model and the MGARCH-A model with all three lines trending upwards, especially during the recent financial crisis in though out 2018. There is also a short exception around 2015 and 2016 when all three benchmark models outperform the MGARCH-IHMM probably because the returns are relatively stable, and a simple model is sufficient in this period. The MGARCH-IHMM can predict a very skewed density out-of-sample. Figure 4.5 plots the marginal predictive density of the Health portfolio for October 2018 by different models, with the actual realized return being the vertical line. The density predicted by the MGARCH-IHMM is clearly more negatively skewed than

[5]Forecast error is computed from the average of the difference between predicted mean of $\boldsymbol{r}_t$ and the actual realized returns across five industry portfolios.

those by other models during the 2018 financial crisis and allows this large negative return to occur with a higher probability.

Figure 4.6 is approximately here.

Figure 4.7 is approximately here.

Mardia (1970) proposes a multivariate measure of skewness and kurtosis. A multivariate normal distribution gives a Mardia skewness of 0 and a Mardia kurtosis of $N(N+2)\frac{M-1}{M+1}$ where N is the dimension of some distribution and M is the sample size of this distribution. Since the Mardia skewness is always non-negative, it does not distinguish the direction of skewness, and the higher the measure, the severe a distribution deviates from normality. A Mardia kurtosis higher than $N(N+2)\frac{M-1}{M+1}$, which is approximately 35 for my 5-variate example with 20000 predictive draws for each period, indicates the distribution is leptokurtic, and a number lower than 35 indicates the distribution is platykurtic. Figure 4.6 and 4.7 plot the Mardia skewness and kurtosis of the log returns and simple returns predicted from the MGARCH-IHMM respectively over each out-of-sample period. Clearly, none of the 0.95 density intervals covers the normality condition. The MGARCH-IHMM predicts skewed and leptokurtic log-return distribution in every out-of-sample period, and the simple return distributions used in portfolio allocation is even more skewed and leptokurtic. These results imply that the mean-variance optimization should not apply here unless an investor's utility function is quadratic.

4.5 General Portfolio Optimization

Clearly, the proposed MGARCH-IHMM is more flexible and can provide better density forecasts than all the benchmark models, but does it translate into actual economic gains in portfolio allocation? As discussed at the beginning of this chapter, in most research, a portfolio is optimized on a mean-variance space, which means that a risk-averse investor wants to maximize the expected return while minimizing the variance of her portfolio. Here, the "risk" is defined as variance or standard deviation of asset returns, and ignore any higher moments or the whole return distribution. This implies that either the investor's utility function is quadratic or approximately quadratic with respect to her wealth, or the asset returns are multivariate elliptically distributed.

On one hand, the stock returns are better approximated by a mixture of normals as discussed in Section 4.4.1, so ellipticity is not guaranteed. Moreover, After converted into simple returns, all the predictive distributions are heavily skewed as shown in Section 4.4.2. On the other hand, a quadratic utility is usually considered as unrealistic due to its increasing absolute risk-aversion. Constant relative risk-aversion (CRRA) utility, whose absolute risk-aversion is decreasing when wealth increases, is relatively more realistic. To incorporate CRRA utility with general return distributions, we need to consider a more general portfolio optimization problem.

4.5.1 Dynamic Optimal Portfolio Weights

Consider a rational and risk-averse investor whose utility function is $U(W)$, where W denotes her wealth. She distributes her wealth into N risky asset[6] and one risk-free asset. Without loss of generality, assume her wealth set as 1 at the beginning of each period, and her wealth at the end of the period t is

$$W_t = 1 + \boldsymbol{w}_t' \boldsymbol{R}_t + (1 - \boldsymbol{w}_t' \boldsymbol{\iota}) R_{f,t}, \tag{4.16}$$

where $\boldsymbol{w}_t$ represents the vector of portfolio weights of risky assets, $\boldsymbol{R}_t$[7] is a vector of simple returns of risky assets, $\boldsymbol{\iota}$ is a vector of 1, and $R_{f,t}$ is the simple return of the risk-free asset.

Suppose the investor rebalances her portfolio at the end of each period without any transaction cost, and therefore there is a separate optimization problem for each period whose decision is independent to all other periods. For a particular period t, in order to obtain the optimal portfolio weight, the investor maximizes her conditional

[6] $N = 5$. Consumer, Manufacture, High Tech, Health and Other portfolio from Fama-French 5 industry portfolios.

[7] For each element in $\boldsymbol{R}_t$, $R_{i,t} = \exp(r_{i,t}/100) - 1$.

expected utility

$$\begin{aligned} \max_{\boldsymbol{w}_t} \quad & \mathrm{E}\left[U(W_t)|\mathcal{F}_{t-1}\right] \\ & \approx \frac{1}{M}\sum_{i=1}^{M} U(W_t^{(i)}) \\ \text{where} \quad & W_t^{(i)} = 1 + \boldsymbol{w}_t'\boldsymbol{R}_t^{(i)} + (1 - \boldsymbol{w}_t'\boldsymbol{\iota})R_{f,t}. \end{aligned} \tag{4.17}$$

The superscript (i) corresponds to the ith draw of the simulated predictive returns for time t. No constraint on $\boldsymbol{w}_t$ is required since the rest of the wealth is distributed into the risk-free asset.

Analytical solution for this problem is in general unavailable, so it needs to be found numerically. Note that if the utility function $U(W_t^{(i)})$ is strictly concave with respect to $W_t^{(i)}$, then it's also strictly concave with respect to $\boldsymbol{w}_t$, and so is its empirical expectation. This ensures the existence of a unique solution in the above optimization problem.

4.5.2 Break-even Management Fees

Because the notion "risk" is not measured by the second moment of return distribution but the uncertainty of returns, Sharpe ratio, which measures risk by standard deviation, may not be desirable criteria to compare portfolio allocation performance. Instead, this chapter implements a more direct method, by computing the break-even management fee that an investor is willing to pay for switching from one model to

another (e.g. Bollerslev et al. 2018; Fleming et al. 2001) to compare the economic performance of different portfolios.

At the end of period $t-1$, the investor rebalances her portfolio as the optimal weights $\boldsymbol{w}_t$ found above, and carries this portfolio into the next period when the actual return $\boldsymbol{R}_t$ is realized. One can easily compute her realized wealth W_t and the realized utility $U(W_t)$ for each out-of-sample period. The break-even management fee the investor willing to pay (C) for switching from model $\mathcal{M}_2$ to model $\mathcal{M}_1$ is given by equating ex-post utilities as

$$\sum_{l=t+1}^{T} U(W_l - C|\mathcal{M}_1) = \sum_{l=t+1}^{T} U(W_l|\mathcal{M}_2). \tag{4.18}$$

Note that her initial wealth is always 1 for each period, so C can either be seen as a dollar fee or a return fee.

4.5.3 Empirical Results

I optimize the portfolio under CRRA utility, CARA utility and quadratic utility respectively. The utility functions are specified as

CRRA utility

$$U(W) = \frac{W^{1-a}}{1-a},$$

CARA utility

$$U(W) = 1 - e^{-aW},$$

quadratic utility [8]

$$U(W) = W - \frac{a}{20(1+a)}W^2,$$

where a is the risk-aversion parameter. However, it doesn't necessarily indicates relative risk aversion (RRA) or absolute risk aversion (ARA) for each utility function. For example, $RRA_{CRRA} = a$, $ARA_{CRRA} = \frac{a}{W}$, $RRA_{CARA} = aW$, $ARA_{CARA} = a$, $RRA_{quad} = \frac{a}{10+9a}W$ and $ARA_{quad} = \frac{a}{10+9a}$.

For each utility function, the portfolio is optimized when $a = \{2, 4, 6\}$ for all three econometric models. A quasi-Newton method is used for each optimization, except for quadratic utility where an analytical solution exists.[9] A Brent-Dekker method is used to find the scalar root in equation (4.18).

Table 4.4 is approximately here.

Table 4.5 is approximately here.

Table 4.6 is approximately here.

Table 4.4 – 4.6 show the annualized fee that an investor is willing to pay for switching from one econometric model to another over the out-of-sample period from January 2002 to August 2018. For both CRRA and CARA utility, the investor is always willing to pay a positive fee to switch from MGARCH-N to the proposed Bayesian semiparametric model. Depending on the level of risk-aversion, a CRRA investor is

[8] A more commonly seen representation is $U(W) = W - \frac{a}{2}W^2$, but it's maximized at $W = \frac{1}{a} < 1$ when $a > 1$. This will make the investor better-off by losing money when the initial wealth is 1. With my setting, the utility is maximized at $W = \frac{10(1+a)}{a} > 0$ when $a > 0$.

[9] It's easy to prove that the analytical solution for quadratic utility is the same as the solution derived from the Two-Fund Separation Theorem.

willing to pay an annual fee of 1.7788% – 4.2803% for switching from the MGARCH-N model to the MGARCH-IHMM and a fee of 2.4784% – 6.4765% for switching from the MGARCH-A model to the MGARCH-IHMM. Similarly, a CARA investor is willing to pay an annual fee of 1.8134% – 5.5067% for switching from the MGARCH-N model to the MGARCH-IHMM and a fee of 2.5070% – 7.6793% for switching from the MGARCH-A model to the MGARCH-IHMM. As for quadratic utility, an investor is willing to pay an annual fee of 1.7845% – 5.4120% for switching from the MGARCH-N model to the MGARCH-IHMM, and a fee of 2.4053% – 7.3754% for switching from the MGARCH-A model to the MGARCH-IHMM. On the other hand, a risk-averse investor is also willing to pay for a positive fee for switching from IHMM to MGARCH-IHMM, an annual fee of 0.5575% – 0.8600% with CRRA utility, 0.5902% – 1.7668% with CARA utility and 0.5011% – 1.4874% with quadratic utility. Clearly, no matter whether the utility is CRRA, CARA or quadratic, a risk-averse investor always prefers a more sophisticated MGARCH-IHMM than a less sophisticated benchmark model.

The fee decreases when the investor is more risk-averse. Specifically, when a increases from 2 to 6 under CRRA utility, the fee decreases from 0.8600% to 0.5575% annually in general for switching from the IHMM to the MGARCH-IHMM, from 4.2803% to 1.7788% annually for switching from the MGARCH-N model to the MGARCH-IHMM, and from 6.4765% to 2.4784% for switching from the MGARCH-A model to the MGARCH-IHMM. Similarly, under CARA utility, the fee decreases from 1.7668% to 0.5902% for switching from the IHMM to the MGARCH-IHMM, a decrease from

5.5067% to 1.8134% for switching from the MGARCH-N model to the MGARCH-IHMM, and a decrease from 7.6793% to 2.5070% for switching from the MGARCH-A model to the MGARCH-IHMM. And if the investor is quadratic, the fee decreases from 1.4874% to 0.5011% for switching from the IHMM to the MGARCH-IHMM, decreases from 5.4120% to 1.7845% for switching from the MGARCH-N model to the MGARCH-IHMM and decreases from 7.3754% to 2.4053% for switching from the MGARCH-A model to the MGARCH-IHMM.

Figure 4.8 is approximately here.

The decrease in fees is related to the positions on risky assets that the investor invests in. Both positive weights (long position) and negative weights (short position) for risky assets represent the investor's position in risky asset trading, so I report the sum of the absolute value of weights across risky assets as a proxy for the risky position. As stated in Table 4.4 – 4.6, the average risky position over the out-of-sample period monotonically decreases when the investor is more risk-averse. Figure 4.8 also confirms this observation that dynamics of risky position involvement shifts down when the investor becomes more risk-averse. This means an investor would put less of her wealth in risky assets when she is more risk-averse, and the diversification benefit is not large enough to overcome the greater penalty in risk. If one further separates the risky positions into long positions and short positions, both the long side and short side move towards zero as listed in Table 4.4 – 4.6.

Table 4.4 – 4.6 also compares the performance of MGARCH-IHMM portfolio against mixing an equally-weighted risky portfolio with a risk-free asset (EW+RF) and a portfolio of an equally weighted portfolio that consists of Fama-French 5 industry portfolios and a risk-free asset (EW). For the EW+RF, I assume that the risky portfolio returns are independently and identically distributed (i.i.d.), and the weights between risky and risk-free components are optimized accordingly by maximizing expected utility in the beginning of the investment horizon and hold it afterwards. For the EW, it follows an 1/N buy-and-hold strategy, and each asset, including the risk-free asset, has a constant weight of 1/6. In both cases, an investor makes positive economic gain by switching from a holding EW+RF portfolio or a EW portfolio regardless the utility type and risk aversion level, and these gains diminish when the risk aversion level increases. The pattern is consistent with those discussed previously.

4.6 Conclusion

This chapter investigates the relationship between accurate return density forecasts and whether they lead to economic gains in portfolio choice. The results show that under a relatively realistic CRRA utility, a positive economic gain does not have to be generated from an increase in log-predictive likelihood unless there's an improvement in tail forecasts.

A new multivariate Bayesian semiparametric model (MGARCH-IHMM) is proposed in this chapter. The additional IHMM component successfully captures the regime-switching phenomenon around GARCH dynamics. Several occurrences "bear" state and "mild" state are identified apart from the "ordinary" state. The state-dependent covariance parameter serves as a multiplier to the GARCH covariance, helping to boost up covariance when it's not increasing fast enough and shrink it down when it's not decreasing fast enough.

The MGARCH-IHMM shows clear advantage in density forecasts against the benchmark IHMM, MGARCH-N model and MGARCH-A model, especially during the financial crisis in 2009. By adding the IHMM component, the MGARCH-IHMM is capable to predict skewed and leptokurtic return distributions compared to the MGARCH-N and MGARCH-A model. Compared to the IHMM, the predicted return distribution is generally more skewed but less leptokurtic.

Since the commonly used Markowitz Portfolio Theory implies that the investor's utility is quadratic, which has increasing absolute risk-aversion with wealth, a more general portfolio optimization problem, which maximizes the expected utility, is implemented instead. It's easy to prove that if the investor's utility is quadratic, the general optimization problem is equivalent to the Markowitz framework.

Empirical results show that a risk-averse investor is always willing to pay a positive fee to switch to the more sophisticated Bayesian semiparametric model MGARCH-IHMM with better density forecasts. Moreover, the more risk-averse the investor is,

the less the financial benefit for the superior model, because the investor is moving away from the risky assets when more risk-averse. These results are robust to switching from holding equally-weighted portfolios.

Table 4.1: Descriptive Statistics of the Industry Portfolio Returns

Panel A: Univariate statistics

Industry	Mean	Median	StDev	Skew	Ex.Kurt	Min	Max
Consumer	0.8673	1.2324	5.2503	-0.6087	7.0192	-33.6592	36.2349
Manufacture	0.8133	1.2571	5.4590	-0.4372	7.7299	-36.9037	36.1374
High Tech	0.8111	1.1978	5.5790	-0.6676	4.0792	-31.1702	29.1475
Health	0.9323	1.0989	5.5558	-0.6482	7.6243	-41.6728	31.5759
Other	0.7160	1.2571	6.3188	-0.2661	8.3730	-35.7104	46.2160

Notes: From July 1926 to August 2018, 1106 observations.

Panel B: Correlations

	Consumer	Manufacture	High Tech	Health	Other
Consumer	1.0000	0.8746	0.8126	0.7817	0.8825
Manufacture	0.8746	1.0000	0.8082	0.7438	0.8916
High Tech	0.8126	0.8082	1.0000	0.7077	0.7999
Health	0.7817	0.7438	0.7077	1.0000	0.7401
Other	0.8825	0.8916	0.7999	0.7401	1.0000

Notes: From July 1926 to August 2018, 1106 observations.

Table 4.2: Posterior Estimates

MGARCH-IHMM:

$$\Gamma|\beta_0 \sim GEM(\beta_0), \quad \Pi_j|\alpha_0, \Gamma \sim DP(\alpha_0, \Gamma)$$
$$s_t|s_{t-1}, \Pi \sim \Pi_{s_{t-1}}$$
$$\boldsymbol{r}_t|\boldsymbol{\Theta}, \boldsymbol{H}_t, \Pi, s_t \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_{s_t}\boldsymbol{H}_t^{1/2'}\right)$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\boldsymbol{r}_{t-1} - \boldsymbol{\eta})(\boldsymbol{r}_{t-1} - \boldsymbol{\eta})' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

IHMM:

$$\Gamma|\beta_0 \sim GEM(\beta_0), \quad \Pi_j|\alpha_0, \Gamma \sim DP(\alpha_0, \Gamma)$$
$$s_t|s_{t-1}, \Pi \sim \Pi_{s_{t-1}}$$
$$\boldsymbol{r}_t|s_t, \boldsymbol{\Theta}, \mathcal{F}_{t-1} \sim N\left(\boldsymbol{\mu}_{s_t}, \boldsymbol{\Sigma}_{s_t}\right)$$

MGARCH-N:

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\boldsymbol{r}_{t-1} - \boldsymbol{\mu})(\boldsymbol{r}_{t-1} - \boldsymbol{\mu})' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

MGARCH-A:

$$\boldsymbol{r}_t = \boldsymbol{\mu} + \boldsymbol{H}_t^{1/2}\boldsymbol{z}_t$$
$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{\alpha}\boldsymbol{\alpha}' \odot (\boldsymbol{r}_{t-1} - \boldsymbol{\eta})(\boldsymbol{r}_{t-1} - \boldsymbol{\eta})' + \boldsymbol{\beta}\boldsymbol{\beta}' \odot \boldsymbol{H}_{t-1}$$

Panel A: MGARCH and IHMM parameters

	MGARCH-IHMM		IHMM		MGARCH-N		MGARCH-A	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
μ_1					0.7505	(0.5049, 0.9951)	0.7905	(0.5403, 1.0397)
μ_2					0.7338	(0.4897, 0.9757)	0.7570	(0.5091, 1.0031)
μ_3					0.7759	(0.5276, 1.0236)	0.7776	(0.5279, 1.0273)
μ_4					0.8357	(0.5644, 1.1089)	0.8467	(0.5789, 1.1154)
μ_5					0.6025	(0.3130, 0.8872)	0.6727	(0.3850, 0.9561)
α_1	0.2209	(0.1932, 0.2493)			0.2657	(0.2461, 0.2859)	0.2594	(0.2392, 0.2795)
α_2	0.2279	(0.2002, 0.2558)			0.2727	(0.2532, 0.2924)	0.2684	(0.2485, 0.2882)
α_3	0.2381	(0.2067, 0.2721)			0.3015	(0.2746, 0.3279)	0.3009	(0.2750, 0.3272)
α_4	0.2474	(0.2128, 0.2840)			0.2810	(0.2550, 0.3082)	0.2768	(0.2507, 0.3035)
α_5	0.2264	(0.1954, 0.2580)			0.2850	(0.2636, 0.3067)	0.2795	(0.2575, 0.3017)
β_1	0.9541	(0.9421, 0.9646)			0.9453	(0.9359, 0.9536)	0.9470	(0.9379, 0.9554)
β_2	0.9581	(0.9467, 0.9676)			0.9465	(0.9376, 0.9544)	0.9477	(0.9390, 0.9556)
β_3	0.9498	(0.9362, 0.9615)			0.9338	(0.9210, 0.9459)	0.9336	(0.9206, 0.9453)
β_4	0.9404	(0.9202, 0.9567)			0.9335	(0.9183, 0.9465)	0.9352	(0.9204, 0.9481)
β_5	0.9567	(0.9439, 0.9672)			0.9383	(0.9271, 0.9483)	0.9399	(0.9288, 0.9499)
η_1	1.9503	(0.9871, 3.0028)					1.3661	(0.8317, 1.9084)
η_2	1.4279	(0.5337, 2.3317)					1.0287	(0.5172, 1.5428)
η_3	1.0370	(0.0684, 2.0076)					0.6164	(0.1141, 1.1101)
η_4	1.3769	(0.3577, 2.4393)					0.9352	(0.3673, 1.4940)
η_5	2.1781	(1.1267, 3.3298)					1.6148	(1.0230, 2.2265)
α_0	1.8456	(1.0777, 2.8408)	0.8150	(0.4842, 1.2485)				
β_0	0.9249	(0.3426, 1.7351)	0.8207	(0.3271, 1.5020)				
K	9.9907	(6.0000,15.0000)	8.3095	(6.0000,11.0000)				

Table 4.2: Posterior Estimates (cont.)

Panel B: Summary statistics of Posterior means of $\boldsymbol{\mu}_{s_t}$ over time for the MGARCH-IHMM

	Min	25%	Median	Mean	75%	Max
μ_1	-1.4943	0.3654	1.0701	0.8951	1.4916	1.7938
μ_2	-0.1986	0.6465	0.9914	0.8916	1.2148	1.3701
μ_3	-0.2149	0.6480	1.0465	0.9201	1.2652	1.4121
μ_4	-0.0007	0.7195	1.0922	0.9951	1.3564	1.5852
μ_5	-0.9861	0.2398	0.9761	0.7787	1.4105	1.7246

Table 4.2: Posterior Estimates (cont.)

$$\boldsymbol{\mu}_s \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0), \quad \boldsymbol{\Sigma}_s \sim IW(\boldsymbol{\Sigma}_0, \nu)$$

Panel C: Base measure parameters

	MGARCH-IHMM			IHMM		
	Mean	StDev	0.95 DI	Mean	StDev	0.95 DI
$b_{0,1}$	0.1180	0.4319	(-0.7562,0.9336)	0.7577	0.3496	(-0.0351,1.3517)
$b_{0,2}$	0.5153	0.3256	(-0.1657,1.1298)	0.8490	0.2861	(0.2403,1.3768)
$b_{0,3}$	0.4966	0.3298	(-0.1826,1.1121)	0.8814	0.3103	(0.2168,1.4522)
$b_{0,4}$	0.5395	0.3656	(-0.2404,1.1974)	0.7116	0.3399	(-0.0285,1.3080)
$b_{0,5}$	0.0608	0.4255	(-0.8104,0.8617)	0.5850	0.3917	(-0.2550,1.2992)
$B_{0,11}$	1.3330	1.0397	(0.2960,3.9064)	0.6726	0.6597	(0.1113,2.4355)
$B_{0,21}$	0.5585	0.6283	(-0.0966,2.1085)	0.2195	0.3648	(-0.1644,1.1707)
$B_{0,22}$	0.5795	0.5672	(0.1258,1.8940)	0.3455	0.3080	(0.0874,1.1321)
$B_{0,31}$	0.6021	0.6319	(-0.0617,2.1796)	0.2242	0.3935	(-0.2475,1.2501)
$B_{0,32}$	0.3558	0.4506	(-0.0797,1.4463)	0.1709	0.2935	(-0.1319,0.9225)
$B_{0,33}$	0.6069	0.5247	(0.1330,1.9401)	0.4385	0.4254	(0.0977,1.5418)
$B_{0,41}$	0.6015	0.7185	(-0.1589,2.3735)	0.3762	0.5332	(-0.1306,1.7932)
$B_{0,42}$	0.3508	0.5185	(-0.1251,1.5306)	0.1869	0.3175	(-0.1662,0.9954)
$B_{0,43}$	0.3799	0.5072	(-0.1093,1.6382)	0.1627	0.3450	(-0.2892,1.0200)
$B_{0,44}$	0.6864	0.7037	(0.1286,2.3744)	0.5775	0.5269	(0.1159,1.9580)
$B_{0,51}$	0.9640	0.8929	(0.0480,3.2185)	0.5051	0.5916	(-0.0600,2.0713)
$B_{0,52}$	0.5597	0.6645	(-0.0514,2.1663)	0.2636	0.4070	(-0.1714,1.2878)
$B_{0,53}$	0.5789	0.6328	(-0.0444,2.1562)	0.2479	0.4361	(-0.3057,1.3409)
$B_{0,54}$	0.5823	0.7071	(-0.1013,2.3069)	0.4155	0.5233	(-0.1694,1.7658)
$B_{0,55}$	1.2008	1.0120	(0.2572,3.7758)	0.8085	0.7088	(0.1571,2.6475)

Table 4.2: Posterior Estimates (cont.)

$$\boldsymbol{\mu}_s \sim N(\boldsymbol{b}_0, \boldsymbol{B}_0), \quad \boldsymbol{\Sigma}_s \sim IW(\boldsymbol{\Sigma}_0, \nu)$$

Panel C: Base measure parameters (cont.)

	MGARCH-IHMM			IHMM		
	Mean	StDev	0.95 DI	Mean	StDev	0.95 DI
$\Sigma_{0,11}$	2.3833	0.5620	(1.4023,3.5985)	4.2914	0.9690	(2.6173,6.3997)
$\Sigma_{0,21}$	0.0412	0.3748	(-0.6971,0.7752)	0.3854	0.6763	(-0.9360,1.7161)
$\Sigma_{0,22}$	2.5218	0.6399	(1.4169,3.9066)	4.3922	0.9835	(2.6620,6.4749)
$\Sigma_{0,31}$	-0.0320	0.3227	(-0.6486,0.6185)	0.3336	0.6889	(-0.9869,1.7246)
$\Sigma_{0,32}$	0.1146	0.3354	(-0.5393,0.7804)	0.4914	0.7091	(-0.8762,1.9186)
$\Sigma_{0,33}$	1.7873	0.4730	(0.9825,2.8158)	4.7363	1.1201	(2.7314,7.1236)
$\Sigma_{0,41}$	-0.0151	0.3791	(-0.7597,0.7433)	0.2988	0.7264	(-1.1183,1.7483)
$\Sigma_{0,42}$	0.1092	0.3971	(-0.6768,0.8855)	0.2506	0.7223	(-1.1710,1.6699)
$\Sigma_{0,43}$	0.3070	0.3506	(-0.3739,1.0130)	0.2798	0.7458	(-1.1821,1.7570)
$\Sigma_{0,44}$	2.3682	0.6030	(1.3251,3.6773)	5.0745	1.1458	(3.0563,7.5358)
$\Sigma_{0,51}$	-0.1025	0.3790	(-0.8555,0.6384)	0.3202	0.7035	(-1.0551,1.7159)
$\Sigma_{0,52}$	-0.0893	0.3813	(-0.8454,0.6683)	0.6659	0.7267	(-0.7114,2.1641)
$\Sigma_{0,53}$	-0.0053	0.3397	(-0.6840,0.6629)	0.4101	0.7342	(-1.0161,1.8927)
$\Sigma_{0,54}$	-0.0550	0.4094	(-0.8510,0.7676)	0.1746	0.7518	(-1.2848,1.6517)
$\Sigma_{0,55}$	2.5568	0.6044	(1.5036,3.8709)	4.8151	1.1132	(2.8719,7.2246)
ν	1.3544	0.5679	(0.3271,2.5261)	0.0528	0.0515	(0.0012,0.1914)

Table 4.3: Log-Predictive Likelihoods and Log-Bayes Factors

Model	log PL	log BF	RMSFE
MGARCH-IHMM	-2409.809	—	15.4987
IHMM	-2437.254	27.4454	15.4810
MGARCH-N	-2437.192	27.3832	15.6134
MGARCH-A	-2432.124	22.3156	15.6100

Note: Training samples are from July 1926 to December 2001, and out-of-sample periods are from January 2002 to August 2018.

Table 4.4: Annualized Fees a CRRA Investor Is Willing to Pay

$$W = 1 + \boldsymbol{w}^{*\prime}\boldsymbol{R} + (1 - \boldsymbol{w}^{*\prime}\boldsymbol{\iota})R_f$$

CRRA utility: $U(W) = \frac{1}{1-a}W^{1-a}$

Break-even Fee	$a = 2$	$a = 4$	$a = 6$
IHMM	0.8600%	0.7559%	0.5575%
MGARCH-N	4.2803%	2.5936%	1.7788%
MGARCH-A	6.4765%	3.6583%	2.4784%
EW+RF	11.2294%	5.3389%	3.1819%
EW	13.8218%	4.9366%	3.0087%
Average risky position: $\mathrm{E}\left[\sum \lvert w_i \rvert\right]$			
MGARCH-IHMM	6.1555	3.3996	2.3251
IHMM	7.1494	3.9141	2.6661
MGARCH-N	7.3975	3.7256	2.4859
MGARCH-A	6.6294	3.3362	2.2258
Average long position: $\mathrm{E}\left[\sum w_i \mathbb{1}(w_i > 0)\right]$			
MGARCH-IHMM	4.4783	2.4878	1.7070
IHMM	4.9147	2.7100	1.8484
MGARCH-N	5.0841	2.5603	1.7083
MGARCH-A	4.6908	2.3606	1.5749
Average short position: $\mathrm{E}\left[\sum w_i \mathbb{1}(w_i < 0)\right]$			
MGARCH-IHMM	-1.6772	-0.9118	-0.6181
IHMM	-2.2346	-1.2040	-0.8176
MGARCH-N	-2.3135	-1.1653	-0.7776
MGARCH-A	-1.9386	-0.9756	-0.6509

Notes: IHMM means an investor switches from IHMM to MGARCH-IHMM. EW+RF indicates mixing an equally-weighted portfolio with the risk-free asset assuming i.i.d., and EW indicates an equally-weighted portfolio of Fama-French 5 industry portfolios and a risk-free asset (each has a weight of 1/6).

Table 4.5: Annualized Fees a CARA Investor Is Willing to Pay

$$W = 1 + \boldsymbol{w}^{*\prime}\boldsymbol{R} + (1 - \boldsymbol{w}^{*\prime}\boldsymbol{\iota})R_f$$

CARA utility: $U(W) = 1 - e^{-aW}$

Break-even Fee	$a = 2$	$a = 4$	$a = 6$
IHMM	1.7668%	0.8834%	0.5902%
MGARCH-N	5.5067%	2.7256%	1.8134%
MGARCH-A	7.6793%	3.7783%	2.5070%
EW+RF	12.5250%	5.4696%	3.2132%
EW	15.0602%	5.2209%	3.0783%
Average risky position: $\mathrm{E}\left[\sum \lvert w_i \rvert\right]$			
MGARCH-IHMM	7.2096	3.6048	2.4032
IHMM	8.2119	4.1060	2.7372
MGARCH-N	7.4481	3.7241	2.4827
MGARCH-A	6.6677	3.3339	2.2225
Average long position: $\mathrm{E}\left[\sum w_i \mathbb{1}(w_i > 0)\right]$			
MGARCH-IHMM	5.3132	2.6566	1.7711
IHMM	5.7005	2.8502	1.9001
MGARCH-N	5.1182	2.5591	1.7061
MGARCH-A	4.7177	2.3589	1.5726
Average long position: $\mathrm{E}\left[\sum w_i \mathbb{1}(w_i < 0)\right]$			
MGARCH-IHMM	-1.8964	-0.9482	-0.6321
IHMM	-2.5114	-1.2557	-0.8371
MGARCH-N	-2.3299	-1.1650	-0.7766
MGARCH-A	-1.9500	-0.9750	-0.6500

Notes: IHMM means an investor switches from IHMM to MGARCH-IHMM. EW+RF indicates mixing an equally-weighted portfolio with the risk-free asset assuming i.i.d., and EW indicates an equally-weighted portfolio of Fama-French 5 industry portfolios and a risk-free asset (each has a weight of 1/6).

Table 4.6: Annualized Fees a Quadratic Investor Is Willing to Pay

$$W = 1 + \boldsymbol{w}^{*\prime}\boldsymbol{R} + (1 - \boldsymbol{w}^{*\prime}\boldsymbol{\iota})R_f$$

Quadratic utility: $U(W) = W - \frac{a}{20(1+a)}W^2$

Break-even Fee	$a = 2$	$a = 4$	$a = 6$
IHMM	2.1846%	1.0942%	0.7296%
MGARCH-N	10.0166%	4.9282%	3.2673%
MGARCH-A	12.7547%	6.2312%	4.1197%
EW+RF	24.9890%	11.2755%	6.9881%
EW	31.8336%	10.6412%	4.2617%
Average risky position: $\mathrm{E}\left[\sum \lvert w_i \rvert\right]$			
MGARCH-IHMM	6.7255	3.3569	2.2341
IHMM	7.5993	3.7928	2.5240
MGARCH-N	6.9822	3.4853	2.3197
MGARCH-A	6.2667	3.1283	2.0822
Average long position: $\mathrm{E}\left[\sum w_i \mathbb{1}(w_i > 0)\right]$			
MGARCH-IHMM	4.9754	2.4834	1.6527
IHMM	5.2651	2.6279	1.7488
MGARCH-N	4.7981	2.3951	1.5940
MGARCH-A	4.4342	2.2134	1.4732
Average short position: $\mathrm{E}\left[\sum w_i \mathbb{1}(w_i < 0)\right]$			
MGARCH-IHMM	-1.7501	-0.8735	-0.5813
IHMM	-2.3342	-1.1649	-0.7752
MGARCH-N	-2.1840	-1.0903	-0.7257
MGARCH-A	-1.8326	-0.9149	-0.6090

Notes: IHMM means an investor switches from IHMM to MGARCH-IHMM. EW+RF indicates mixing an equally-weighted portfolio with the risk-free asset assuming i.i.d., and EW indicates an equally-weighted portfolio of Fama-French 5 industry portfolios and a risk-free asset (each has a weight of 1/6).

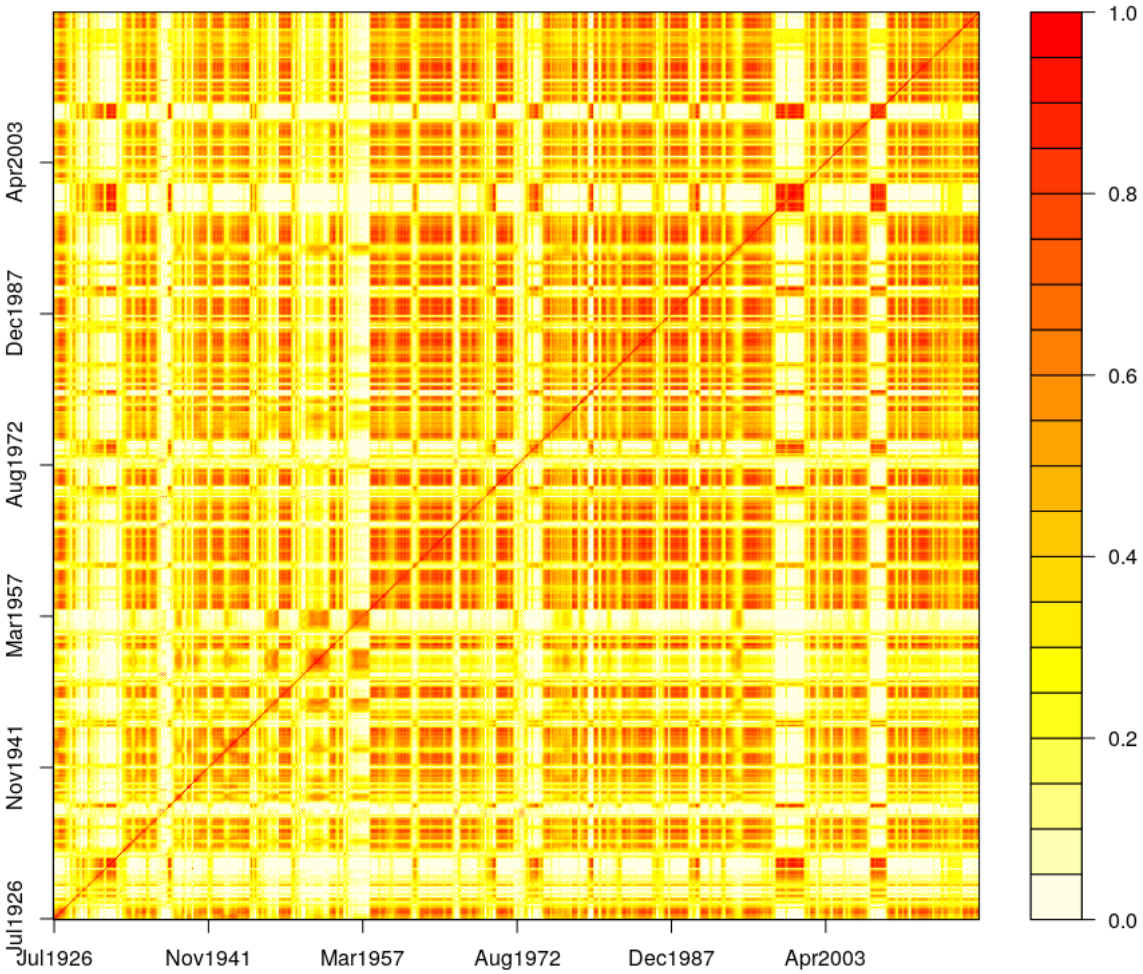
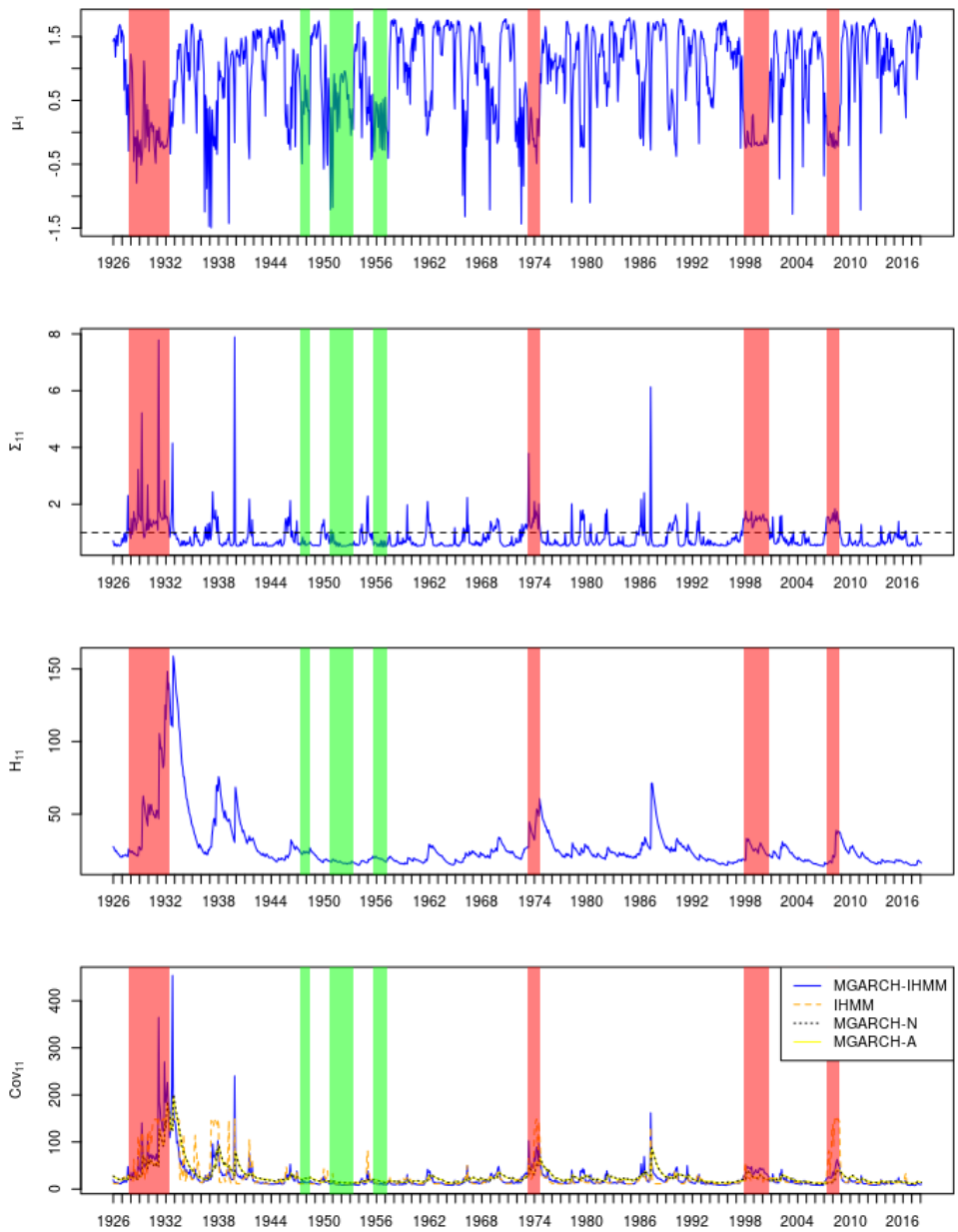
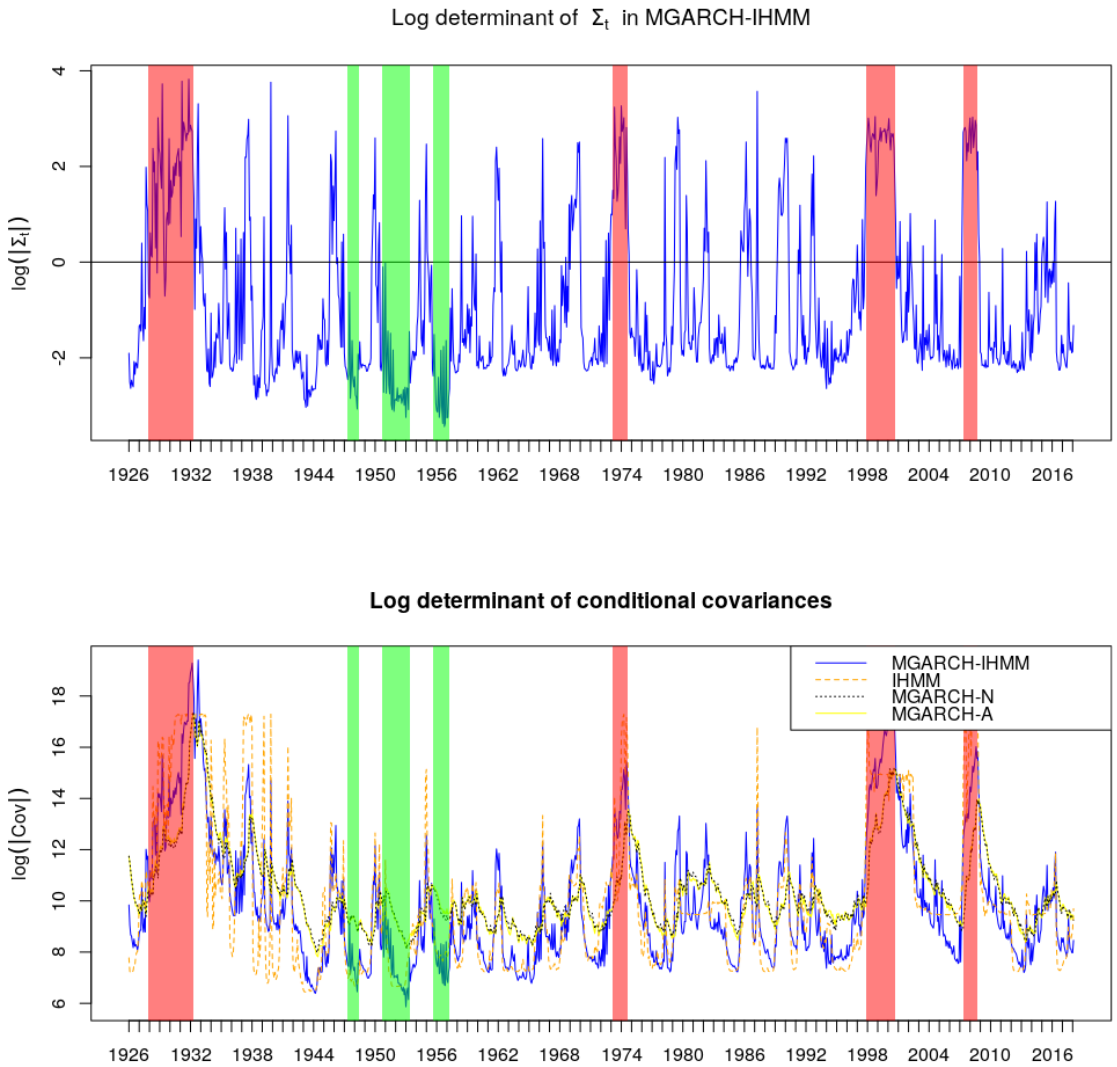


FIGURE 4.1: Heat Map of States Estimated by the MGARCH-IHMM



$$\mathrm{Cov}_{\text{MGARCH-IHMM}} = \mathrm{E}(\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_k\boldsymbol{H}_t^{1/2'}),\ \mathrm{Cov}_{\text{IHMM}} = \mathrm{E}(\boldsymbol{\Sigma}_{s_t}),$$
$$\mathrm{Cov}_{\text{MGARCH-N}} = \mathrm{E}(\boldsymbol{H}_t),\ \mathrm{Cov}_{\text{MGARCH-A}} = \mathrm{E}(\boldsymbol{H}_t)$$

FIGURE 4.2: Posterior Means of the Time-Varying Parameters over Time



$$\mathrm{Cov}_{\text{MGARCH-IHMM}} = \mathrm{E}(\boldsymbol{H}_t^{1/2}\boldsymbol{\Sigma}_k\boldsymbol{H}_t^{1/2'}), \quad \mathrm{Cov}_{\text{IHMM}} = \mathrm{E}(\boldsymbol{\Sigma}_{s_t}),$$
$$\mathrm{Cov}_{\text{MGARCH-N}} = \mathrm{E}(\boldsymbol{H}_t), \quad \mathrm{Cov}_{\text{MGARCH-A}} = \mathrm{E}(\boldsymbol{H}_t)$$

Figure 4.3: Log Determinants of the Posterior Means of the Second Moments over Time

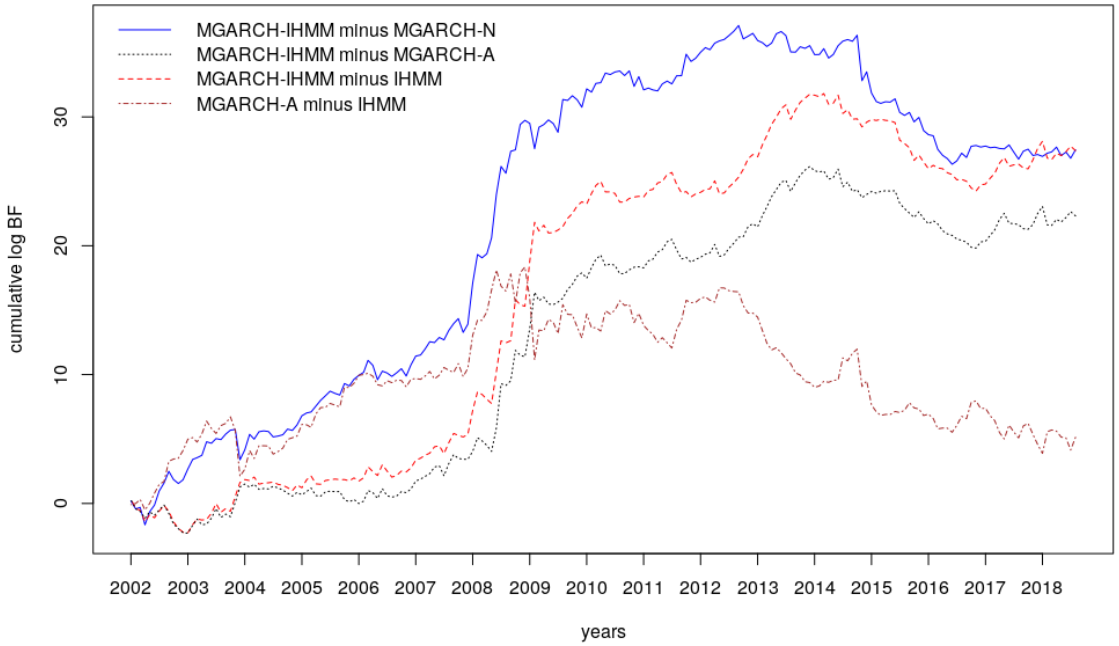


Figure 4.4: Cumulative Log-Bayes Factors over Time

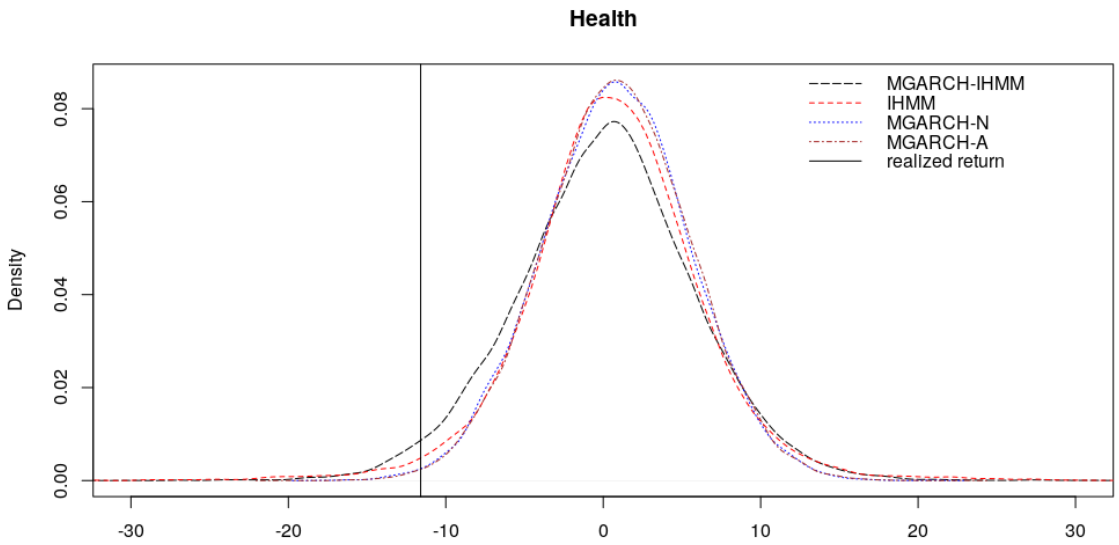


FIGURE 4.5: The Predictive Density of the Health Portfolio Return for October 2018

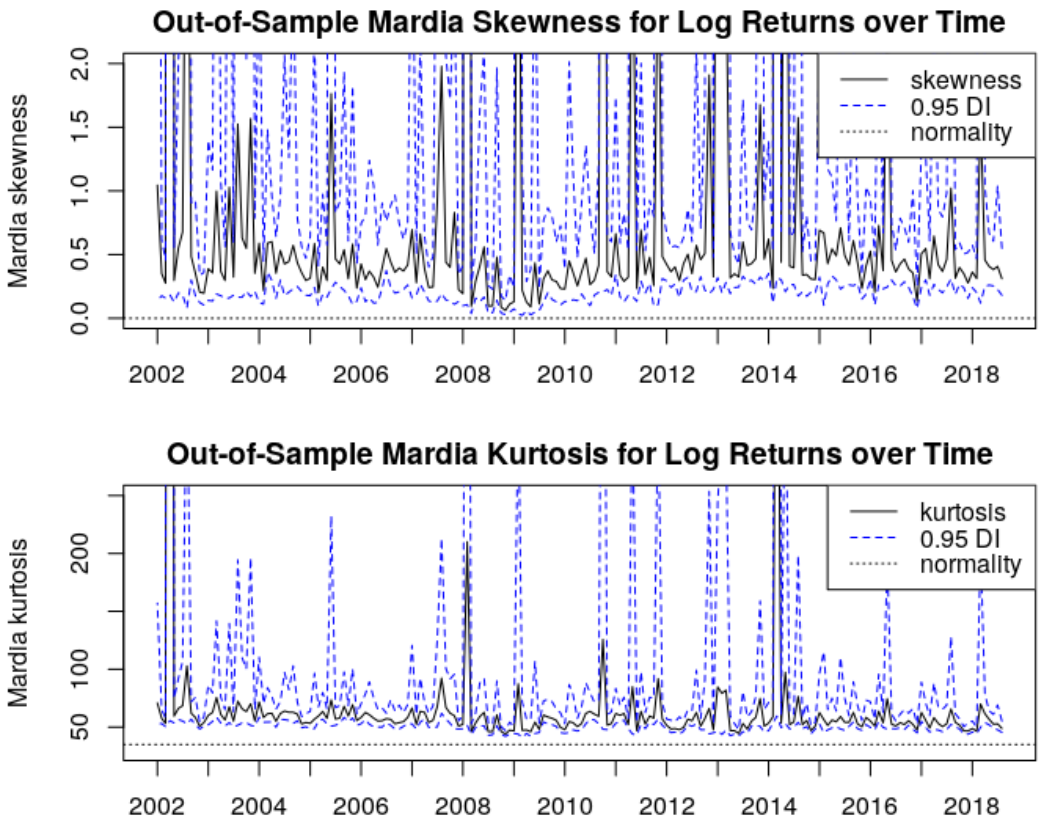


Figure 4.6: Out-of-Sample Mardia Skewness and Kurtosis over Time for Log Return Predictive Distributions from MGARCH-IHMM

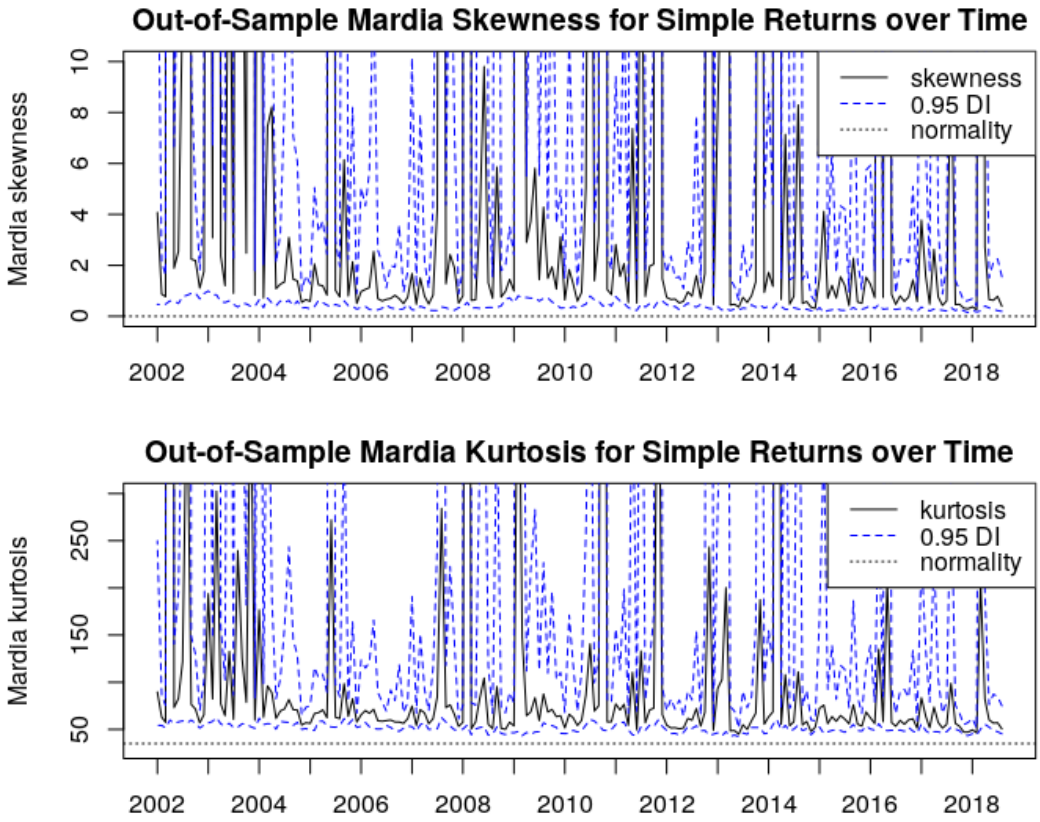


Figure 4.7: Out-of-Sample Mardia Skewness and Kurtosis over Time for Simple Return Predictive Distributions from MGARCH-IHMM

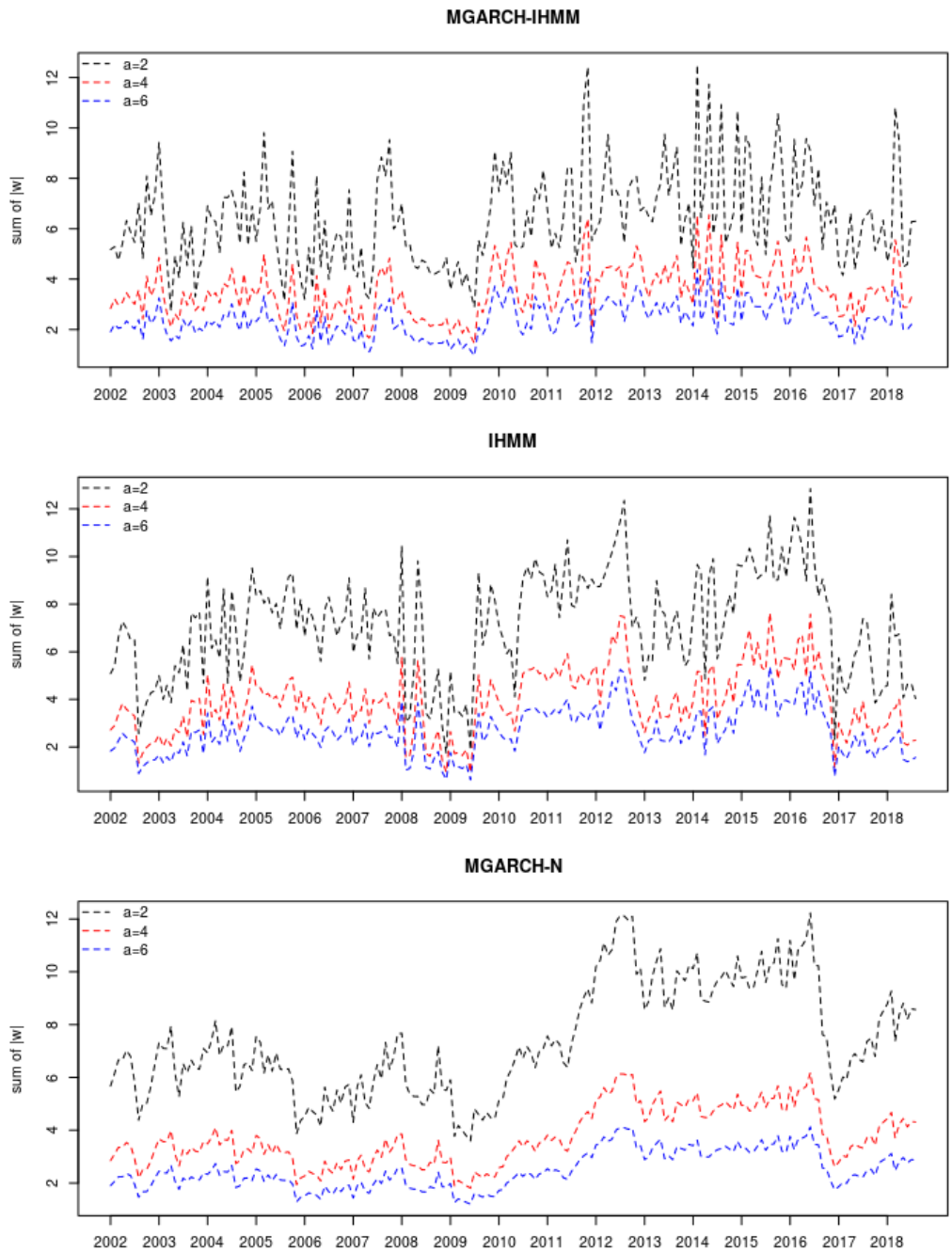


FIGURE 4.8: Risky Positions in the Portfolio Optimized with CRRA Utility over Time

Chapter 5

Conclusion

This thesis contributes to the literature on three important topics regarding forecasting return densities with a particular focus on using mixture models. Unlike traditional forecast methods dedicating on few predictive moments, density forecasts try to improve the predictive distribution as a whole. Predictive distribution is relevant to many risk-related investment decisions such as Value at Risk (VaR), portfolio management, etc.

Mixture models consist of two or more kernels and mix them with a set of selected weights. These weights can be either static or dynamic over time. This thesis consists of three essays. Each proposes a new mixture model. Chapter 2 proposes a finite, static mixture model and Chapter 3 and 4 propose infinite, dynamic mixture models. Each model is supported by empirical results from real-world data respectively.

Chapter 2 introduces a new parsimonious and flexible multivariate GARCH-jump mixture model (MGARCH-jump model). It has two components. One is a multivariate GARCH (MGARCH) component that incorporates the vector-diagonal representation of the BEKK-GARCH model. The other is a compounded jump component that can be further factored into a latent jump size variable and a latent jump arrival variable. Jump sizes are multivariate normally distributed and jump arrivals are multinomially distributed. With this setup, both jump sizes and jump arrivals are allowed to be correlated among assets. Jump effects to the first two conditional moments are activated/deactivated when corresponding jumps/co-jumps occur.

Results for five trivariate regressions and one 5-dimensional regression are provided. Each of the trivariate regressions consists of returns of an individual stock, the corresponding industry and the market. The 5-dimensional regression estimates the model with the five individual stocks used in the trivariate regressions. I find that co-jump is a phenomenon that should not be ignored. Most of the jumps are essentially co-jumps and the jump sizes are strongly positively correlated. In all of the above datasets, the MGARCH-jump model forecasts significantly better than the benchmark MGARCH model with normal innovations (MGARCH-N model).

Two investment-related applications are also provided. The first one extracts beta dynamics from the estimated models. Results suggest that the conventional MGARCH-N tend to smooth out jump events into volatility dynamics and provide very different beta dynamics than the MGARCH-jump model. The second application is the VaR estimated by these two models over time. The estimated VaR dynamics are

almost the same for both models at a high risk tolerance and much more conservative for the MGARCH-jump model at a low risk tolerance.

Chapter 3 is joint work with John Maheu and Qiao Yang. In Chapter 3 we extend the Bayesian semiparametric stochastic volatility model (SV-DPM) proposed by Jensen and Maheu (2010) by replacing the Dirichlet process mixture component with an infinite hidden Markov model (IHMM) component. This extension not only retains the nonparametric property of the SV-DPM in modelling innovation distributions but also allows the innovation distributions to be changed in a Markovian way over time.

The proposed SV-IHMM is valid throughout equity returns, foreign exchange returns, commodity returns and macroeconomic growths as shown by empirical results. In a full sample posterior analysis, the SV parameters are estimated with reasonable values accompanied by many states used to approximate the return innovations. In out-of-sample forecasts, the SV-IHMM performs significantly better than all the benchmark models, including the already powerful SV-DPM model and SV-t model. Predictive density results show that the innovation distributions are very likely to be dynamic rather than a static one.

Chapter 4 answers the question of whether better return density forecasts translate into higher economic gains in portfolio allocations. Compared to the past work that mainly concentrated on improving volatility forecasts, this chapter is among the few first who shifts the scope to a distributional framework. It proposes a new multivariate Bayesian semiparametric model that has two components: an MGARCH component

that allows for asymmetric volatility feedbacks and an IHMM component. Similarly, this model captures smooth volatility change through the MGARCH component and approximates the unknown and dynamic innovation distribution nonparametrically.

Empirical results indicate a very strong regime-switching behaviour in addition to the MGARCH dynamics, suggesting that a richer structure is required to replenish the MGARCH. Moreover, the conditional covariance is amplified by the state-dependent component when the MGARCH covariance is increasing and vice versa. The MGARCH-IHMM forecasts significantly better than the benchmark models in terms of log predictive likelihoods.

This improvement in density forecasts does translate into economic gains in portfolio allocations. The performance is evaluated in the following way. First, solve a utility-based portfolio allocation problem using all the information from a predictive distribution. Second, carry this optimized portfolio over the next period and obtain realized wealth and utilities. Finally, compute the break-even management fee that a risk-averse investor is willing to pay for switching from one model to another. Results show that an investor is always willing to pay a significant positive fee to switch from any benchmark model to the MGARCH-IHMM. This gain is robust to three commonly used utility functions and different risk-aversion parameters.

In summary, this research emphasizes the importance of return density forecasts with a particular focus on employing mixture models. This thesis explores two big categories of mixture modelling: jump models and Bayesian nonparametric models.

Both categories of mixture models can effectively improve forecast accuracy. This improvement is meaningful both statistically and economically to financial investment practices.

Appendices

Appendix 1 Chapter 2: Proof of Conditional Moments of $\boldsymbol{J}_t$

Proof. First, prove $\mathrm{E}\left(\boldsymbol{J}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right)$:

$$\mathrm{E}\left(\boldsymbol{J}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\mu}_J \odot \mathrm{E}\left(\boldsymbol{B}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right)$$
$$\mathrm{E}\left(\boldsymbol{B}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) = \boldsymbol{\Omega}_B{}'\boldsymbol{p} = \sum_{j=1}^{2^N} \boldsymbol{B}_t^{(j)} p_j$$

Then, prove $\mathrm{Cov}\left(\boldsymbol{J}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right)$:

$$\begin{aligned}
\mathrm{Cov}\left(\boldsymbol{J}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) &= \mathrm{E}\left[\left(\boldsymbol{Y}_t \odot \boldsymbol{B}_t\right)\left(\boldsymbol{Y}_t \odot \boldsymbol{B}_t\right)'|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right] \\
&\qquad - \mathrm{E}\left(\boldsymbol{Y}_t \odot \boldsymbol{B}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right)\mathrm{E}\left(\boldsymbol{Y}_t \odot \boldsymbol{B}_t|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right)' \\
&= \mathrm{E}\left(\boldsymbol{Y}_t\boldsymbol{Y}_t{}' \odot \boldsymbol{B}_t\boldsymbol{B}_t{}'|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) - \left(\boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}\right)\left(\boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}\right)' \\
&= \mathrm{E}\left(\boldsymbol{Y}_t\boldsymbol{Y}_t{}' \odot \boldsymbol{B}_t\boldsymbol{B}_t{}'|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) - \boldsymbol{\mu}_J\boldsymbol{\mu}_J{}' \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}\boldsymbol{p}'\boldsymbol{\Omega}_B
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}\left(\boldsymbol{Y}_t\boldsymbol{Y}_t{}' \odot \boldsymbol{B}_t\boldsymbol{B}_t{}'|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) &= \mathrm{E}\left[\mathrm{E}\left(\boldsymbol{Y}_t\boldsymbol{Y}_t{}' \odot \boldsymbol{B}_t\boldsymbol{B}_t{}'|\boldsymbol{B}_t, \boldsymbol{\Theta}\right)|\boldsymbol{r}_{1:t-1}\right] \\
&= \mathrm{E}\left[\mathrm{Cov}\left(\boldsymbol{Y}_t \odot \boldsymbol{B}_t|\boldsymbol{B}_t, \boldsymbol{\Theta}\right)\right. \\
&\qquad \left. +\mathrm{E}\left(\boldsymbol{Y}_t \odot \boldsymbol{B}_t|\boldsymbol{B}_t, \boldsymbol{\Theta}\right)\mathrm{E}\left(\boldsymbol{Y}_t \odot \boldsymbol{B}_t|\boldsymbol{B}_t, \boldsymbol{\Theta}\right)'|\boldsymbol{r}_{1:t-1}\right] \\
&= \mathrm{E}\left[\boldsymbol{\Sigma}_J \odot \boldsymbol{B}_t\boldsymbol{B}_t{}' + \boldsymbol{\mu}_J\boldsymbol{\mu}_J{}' \odot \boldsymbol{B}_t\boldsymbol{B}_t{}'|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right] \\
&= \left(\boldsymbol{\Sigma}_J + \boldsymbol{\mu}_J\boldsymbol{\mu}_J{}'\right) \odot \mathrm{E}\left(\boldsymbol{B}_t\boldsymbol{B}_t{}'|\boldsymbol{\Theta}, \boldsymbol{r}_{1:t-1}\right) \\
&= \left(\boldsymbol{\Sigma}_J + \boldsymbol{\mu}_J\boldsymbol{\mu}_J{}'\right) \odot \left(\sum_{j=1}^{2^N} p_j\boldsymbol{\Omega}_j\boldsymbol{\Omega}_j{}'\right) \qquad \square
\end{aligned}
$$

Appendix 2 Chapter 2: Sampling Algorithms

In each MCMC iteration,

1. $\boldsymbol{\mu}|\boldsymbol{r}_{1:T}, \boldsymbol{H}_{1:T}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}, \boldsymbol{p}$. With everything else given, it's nothing more than a linear model with normal innovation, so the standard conjugate Gibbs result can be applied. Assuming $\boldsymbol{\mu}$ has a normal prior $N\left(\boldsymbol{b}_\mu, \boldsymbol{B}_\mu\right)$, let $\boldsymbol{T}_\mu = \boldsymbol{B}_\mu^{-1}$, then

$$
\begin{aligned}
\boldsymbol{\mu}|\boldsymbol{r}_{1:T}, \boldsymbol{H}_{1:T}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}, \boldsymbol{p} &\sim N\left(\boldsymbol{M}_\mu, \boldsymbol{V}_\mu\right) \\
\boldsymbol{M}_\mu &= \boldsymbol{V}_\mu\left[\sum_{t=1}^{T}\left(\boldsymbol{H}_t + \boldsymbol{B}_t\boldsymbol{B}_t{}' \odot \boldsymbol{\Sigma}_J\right)^{-1}\boldsymbol{r}_t^* + \boldsymbol{T}_\mu\boldsymbol{b}_\mu\right] \\
\boldsymbol{V}_\mu &= \left[\sum_{t=1}^{T}\left(\boldsymbol{H}_t + \boldsymbol{B}_t\boldsymbol{B}_t{}' \odot \boldsymbol{\Sigma}_J\right)^{-1} + \boldsymbol{T}_\mu\right]^{-1}
\end{aligned}
$$

where $\boldsymbol{r}_t^* = \boldsymbol{r}_t - \boldsymbol{\mu}_J \odot \left(\boldsymbol{B}_t - \boldsymbol{\Omega}_B{}'\boldsymbol{p}\right)$.

2. $\boldsymbol{\theta}_H|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}, \boldsymbol{p}$, where $\boldsymbol{\theta}_H = (\boldsymbol{C}, \boldsymbol{\alpha}, \boldsymbol{\beta})'$. The posterior is

$$p\left(\boldsymbol{\theta}_H|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}, \boldsymbol{p}\right) \propto \prod_{t=1}^{T} p\left(\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}\right) p\left(\boldsymbol{\theta}_H\right)$$

$$\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p} \sim N\left(\boldsymbol{\mu} + \boldsymbol{\mu}_J \odot \left(\boldsymbol{B}_t - \boldsymbol{\Omega}_B{}'\boldsymbol{p}\right), \boldsymbol{H}_t + \boldsymbol{B}_t\boldsymbol{B}_t{}' \odot \boldsymbol{\Sigma}_J\right)$$

where $\boldsymbol{H}_t$ follows equation (2.6). Apply a standard random-walk Metropolis-Hastings (MH) algorithm.

3. $\boldsymbol{B}_t|\boldsymbol{r}_t, \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{p}$. There are 2^N different possible realizations of $\boldsymbol{B}_t$, and the posterior is

$$p\left(\boldsymbol{B}_t|\boldsymbol{r}_t, \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{p}\right) = \frac{p\left(\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}\right) p\left(\boldsymbol{B}_t|\boldsymbol{p}\right)}{\int p\left(\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}\right) p\left(\boldsymbol{B}_t|\boldsymbol{p}\right) \, \mathrm{d}\boldsymbol{B}_t}$$

$$p\left(\boldsymbol{B}_t|\boldsymbol{p}\right) = \prod_{i=1}^{2^N} p_{t,i}^{x_i}$$

where $x_i = \delta\left(\boldsymbol{B}_t, \boldsymbol{\Omega}_i\right)$. Here, x_i indicates whether the ith row of $\boldsymbol{\Omega}_B$, $\boldsymbol{\Omega}_i$, is realized.

4. $\boldsymbol{p}|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{H}_{1:T}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}$. Assuming $\boldsymbol{p}$ has a Dirichlet prior $Dir(a_1, \ldots, a_{2^N})$, the posterior is

$$p\left(\boldsymbol{p}|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{H}_{1:T}, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_{1:T}\right) \propto \prod_{t=1}^{T} p\left(\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}\right) p\left(\boldsymbol{B}_{1:T}|\boldsymbol{p}\right) p\left(\boldsymbol{p}\right)$$

An asymmetric MH sampler instead of Gibbs need be applied. Since $B_{t,i}$'s are iid conditional on p_i, one asymmetric proposal density is the conjugate posterior

of multinomial distribution:

$$\boldsymbol{p}' \sim Dir\left(a_i + \sum_{t=1}^{T} x_{t,i}\right), \quad i \in \left\{1, \ldots, 2^N\right\}$$

and accept $\boldsymbol{p}'$ with probability

$$\alpha\left(\boldsymbol{p}^{(i)}, \boldsymbol{p}'\right) = \min\left\{1, \frac{\prod_{t=1}^{T} p\left(\boldsymbol{r}_t | \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}'\right)}{\prod_{t=1}^{T} p\left(\boldsymbol{r}_t | \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}^{(i)}\right)}\right\}$$

5. $\boldsymbol{Y}_t | \boldsymbol{r}_t, \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}$. After simple transformation, conjugate Gibbs result can be applied:

$$\boldsymbol{Y}_t | \boldsymbol{r}_t, \boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p} \sim N\left(\boldsymbol{M}_{Y,t}, \boldsymbol{V}_{Y,t}\right)$$

where

$$\begin{aligned}
\boldsymbol{M}_{Y,t} &= \boldsymbol{V}_{Y,t}\left[\boldsymbol{B}_t \odot \boldsymbol{H}_t^{-1}\left(\boldsymbol{r}_t - \boldsymbol{\mu} + \boldsymbol{\mu}_J \odot \boldsymbol{\Omega}_B{}'\boldsymbol{p}\right) + \boldsymbol{\Sigma}_J^{-1}\boldsymbol{\mu}_J\right] \\
\boldsymbol{V}_{Y,t} &= \left(\boldsymbol{B}_t \boldsymbol{B}_t{}' \odot \boldsymbol{H}_t^{-1} + \boldsymbol{\Sigma}_J^{-1}\right)^{-1}
\end{aligned}$$

6. $\boldsymbol{\mu}_J|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{H}_{1:T}, \boldsymbol{\Sigma}_J, \boldsymbol{Y}_{1:T}, \boldsymbol{B}_{1:T}, \boldsymbol{p}$. Assume a prior of $\boldsymbol{\mu}_J \sim N(\boldsymbol{b}_{\mu_J}, \boldsymbol{B}_{\mu_J})$, then the posterior is

$$
\begin{aligned}
&p\left(\boldsymbol{\mu}_J|\boldsymbol{r}_{1:T}, \boldsymbol{\mu}, \boldsymbol{H}_{1:T}, \boldsymbol{\Sigma}_J, \boldsymbol{Y}_{1:T}, \boldsymbol{B}_{1:T}, \boldsymbol{p}\right) \\
&\qquad \propto \prod_{t=1}^{T} p\left(\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}\right) p\left(\boldsymbol{Y}_{1:T}|\boldsymbol{\mu}_J, \boldsymbol{\Sigma}_J\right) p\left(\boldsymbol{\mu}_J\right)
\end{aligned}
$$

Similarly, a conjugate proposal density can be applied:

$$
\begin{aligned}
\boldsymbol{\mu}_J{}' &\sim N\left(\boldsymbol{M}_{\mu_J}, \boldsymbol{V}_{\mu_J}\right) \\
\boldsymbol{M}_{\mu_J} &= \boldsymbol{V}_{\mu_J}\left(\boldsymbol{\Sigma}_J^{-1}\sum_{t=1}^{T}\boldsymbol{Y}_t + \boldsymbol{B}_{\mu_J}^{-1}\boldsymbol{b}_{\mu_J}\right) \\
\boldsymbol{V}_{\mu_J} &= \left(T\boldsymbol{\Sigma}_J^{-1} + \boldsymbol{B}_{\mu_J}^{-1}\right)^{-1}
\end{aligned}
$$

accept $\boldsymbol{\mu}_J{}'$ with probability

$$
\alpha\left(\boldsymbol{\mu}_J{}^{(i)}, \boldsymbol{\mu}_J{}'\right) = \min\left\{1, \frac{\prod_{t=1}^{T} p\left(\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J{}', \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}\right)}{\prod_{t=1}^{T} p\left(\boldsymbol{r}_t|\boldsymbol{\mu}, \boldsymbol{H}_t, \boldsymbol{\mu}_J{}^{(i)}, \boldsymbol{\Sigma}_J, \boldsymbol{B}_t, \boldsymbol{p}\right)}\right\}
$$

7. $\boldsymbol{\Sigma}_J|\boldsymbol{\mu}_J, \boldsymbol{Y}_{1:T}$. Assume a prior of $\boldsymbol{\Sigma}_J \sim IW\left(\nu_p, \boldsymbol{V}_p\right)$, then apply the standard conjugate Gibbs result

$$
\begin{aligned}
\boldsymbol{\Sigma}_J|\boldsymbol{\mu}_J, \boldsymbol{Y}_{1:T} &\sim IW\left(\nu_J, \boldsymbol{V}_J\right) \\
\nu_J &= T + \nu_p \\
\boldsymbol{V}_J &= \sum_{t=1}^{T}\left(\boldsymbol{Y}_t - \boldsymbol{\mu}_J\right)\left(\boldsymbol{Y}_t - \boldsymbol{\mu}_J\right)' + \boldsymbol{V}_p
\end{aligned}
$$

Appendix 3 Chapter 3: Sampling Algorithms

1. sample $u_{1:T}|\Gamma, \Pi$. The auxiliary slice variable $U = \{u_t\}_{t=1}^{T}$ is drawn by $u_1 \sim U(0, \gamma_{s_1})$ and $u_t \sim U\left(0, \pi_{s_{t-1}s_t}\right)$.

2. update K. Similar to DPM model, if K doesn't meet the condition

$$\min \{u_t\}_{t=1}^{T} > \max \{\pi_{jR}\}_{j=1}^{K} \tag{1}$$

   then K needs to be increased by 1 ($K' = K + 1$), and all parameters need to be drawn from the base measure. In addition, since a new "major" state is introduced, Γ and Π also need to be updated accordingly:

   (a) $\Theta_{K'} \sim H$;

   (b) draw $v \sim Beta(1, \beta_0)$, then update $\Gamma = (\gamma_1, \ldots, \gamma_K, \gamma_{K'}, \gamma_R)'$, where $\gamma_{K'} = v\gamma_R$ and $\gamma_R = (1 - v)\gamma_R$;

   (c) draw $v_j \sim Beta(\alpha_0\gamma_{K'}, \alpha_0\gamma_R)$, update $\Pi_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jK'}, \pi_{jR})$ for $j = 1, \ldots, K$, where $\pi_{jK'} = v\pi_{jR}$ and $\pi_{jR} = (1 - v)\pi_{jR}$;

   (d) draw the K'th row of Π, $\Pi_{K'}$, by $\Pi_{K'} \sim Dir(\alpha_0\gamma_1, \ldots, \alpha_0\gamma_K, \alpha_0\gamma_{K'}, \alpha_0\gamma_R)$.

   Repeat the above steps until inequality (1) holds.

3. forward filter for $s_{1:T}|r_{1:T}, u_{1:T}, \Gamma, \Pi, \Theta, h_{1:T}$. Iterating the following steps forward from 1 to T:

(a) prediction step for initial state s_1:

$$p(s_1 = k|u_1, \Gamma) \propto \mathbb{1}\left(u_1 < \gamma_k\right), \qquad k = 1, \ldots, K \tag{2}$$

for the following states $s_{2:T}$:

$$\begin{aligned} &p(s_t = k|r_{1:t-1}, u_{1:t}, \Pi, \Theta, h_{1:t-1}) \\ &\qquad \propto \sum_{j=1}^{K} \mathbb{1}\left(u_t < \pi_{jk}\right) p\left(s_{t-1} = j|r_{1:t-1}, u_{1:t-1}, \Pi, \Theta, h_{1:t-1}\right) \end{aligned} \tag{3}$$

(b) update step for $s_{1:T}$:

$$\begin{aligned} &p\left(s_t = k|r_{1:t}, u_{1:t}, \Pi, \Theta, h_{1:t}\right) \\ &\qquad \propto p\left(r_t|r_{t-1}, \theta_k, h_t\right) p\left(s_t = k|r_{1:t-1}, u_{1:t}, \Pi, \Theta, h_{1:t-1}\right) \end{aligned} \tag{4}$$

4. backward sampler for $s_{1:T}|r_{1:T}, u_{1:T}, \Pi, \Theta, h_{1:T}$. Sample the states $s_{1:T}$ using the previously filtered values backward from T to 1:

   (a) for the terminal state s_T directly from $p\left(s_T|r_{1:T}, u_{1:T}, \Pi, \Theta, h_{1:T}\right)$

   (b) for the rest states,

$$\begin{aligned} &p\left(s_t = k|s_{t+1} = j, r_{1:t}, u_{1:t+1}, \Pi, \Theta, h_{1:T}\right) \\ &\qquad \propto \mathbb{1}\left(u_{t+1} < \pi_{kj}\right) p\left(s_t = k|r_{1:t}, u_{1:t}, \Pi, \Theta, h_{1:T}\right) \end{aligned} \tag{5}$$

5. sample $c_{1:K}|s_{1:T}, \Gamma, \alpha_0$. $c_{1:K}$ is essential for sampling Γ, and it counts balls in different colour in the "oracle" urn.[1] The "oracle" urn is only involved when a new state is drawn from the regular urn, and sampling c_k directly would be difficult. Fox et al. (2011) propose to simulate c_k from the original Pòlya urn scheme instead of to sample it.

   (a) count the number of each transition type, n_{jk}, for times from state j switching to state k.

   (b) simulate an auxiliary trail variable $x_i \sim Bernoulli\left(\frac{\alpha_0\gamma_k}{i-1+\alpha_0\gamma_k}\right)$, for $i = 1, \ldots, n_{jk}$. If the trial is successful, an "oracle" urn step in involved at the ith step toward n_{jk} and we increase the corresponding "oracle" counts, o_{jk}, by one.

   (c) $c_k = \sum_{j=1}^{K} o_{jk}$.

6. sample β_0. Following Fox et al. (2011) and Maheu and Yang (2016), assume a Gamma hyper-prior $\beta_0 \sim Gamma\,(a_1, b_1)$, and let $c = \sum_{j=1}^{K} c_j$,

   (a) $\nu \sim Bernoulli\left(\frac{c}{c+\beta_0}\right)$

   (b) $\lambda \sim Beta\,(\beta_0 + 1, c)$

   (c) $\beta_0 \sim Gamma\,(a_1 + K - \nu, b_1 - \log\lambda)$

[1] Refer to Beal et al. (2002) for details.

7. sample α_0. Following Fox et al. (2011), assume a Gamma hyper-prior $\alpha_0 \sim Gamma\left(a_2, b_2\right)$, and let $n_j = \sum_{k=1}^{K} n_{jk}$,

   (a) $\nu_j \sim Bernoulli\left(\frac{n_j}{n_j+\alpha_0}\right)$

   (b) $\lambda_j \sim Beta\left(\alpha_0 + 1, n_j\right)$

   (c) $\alpha_0 \sim Gamma\left(a_2 + c - \sum_{j=1}^{K} \nu_j, b_2 - \sum_{j=1}^{K} \log\left(\lambda_j\right)\right)$

8. sample $\Gamma | c_{1:K}, \beta_0$. Given the "oracle" urn counts $c_{1:K}$ and the property of Dirichlet process, the conjugate posterior is

$$\Gamma | c_{1:K}, \beta_0 \sim Dir\left(c_1, \ldots, c_K, \beta_0\right) \tag{6}$$

9. sample $\Pi | n_{1:K,1:K}, \Gamma, \alpha_0$. Similarly, the conjugate posterior of Π_j is

$$\Pi_j | n_{j,1:K}, \Gamma, \alpha \sim Dir\left(\alpha_0\gamma_1 + n_{j1}, \ldots, \alpha_0\gamma_K + n_{jK}, \alpha_0\gamma_R\right) \tag{7}$$

10. sample $\Theta | r_{1:T}, s_{1:T}, h_{1:T}$. Define $Y_k \equiv \left(e^{-\frac{h_t}{2}} r_t | s_t = k\right)_{t=2}^{T}$, $X_k \equiv \left(e^{-\frac{h_t}{2}} | s_t = k\right)_{t=2}^{T}$. The linear model is now

$$Y_k = X_k \mu_k + \omega_k \epsilon_k, \quad \epsilon_k \sim N\left(0, I\right) \tag{8}$$

The posteriors are

$$p\left(\mu_k|Y_k,\omega_k\right) \sim \prod_{t:s_t=k} p\left(r_t|\mu_k,\omega_k\right) p\left(\mu_k\right) \tag{9}$$

$$\sim N\left(M_\mu, V_\mu\right) \tag{10}$$

where

$$M_\mu = V_\mu \left(\omega_k^{-1} X_k' Y_k + B_0^{-1} b_0\right) \tag{11}$$

$$V_\mu = \left(\omega_k^{-1} X_k' X_k + B_0^{-1}\right)^{-1} \tag{12}$$

and

$$p\left(\omega_k|Y_{,}\mu_k\right) \propto \prod_{t:s_t=k} p\left(r_t|\mu_k,\omega_k\right) p\left(\omega_k\right) \tag{13}$$

$$\sim IG\left(\bar{v},\bar{s}\right) \tag{14}$$

where

$$\bar{v} = \frac{T_k}{2} + v_0 = \frac{1}{2}\sum_{t=1}^{T} \mathbb{1}\left(s_t = k\right) + v_0 \tag{15}$$

$$\bar{s} = \frac{1}{2}\left(Y_k - X_k\mu_k\right)'\left(Y_k - X_k\mu_k\right) + s_0 \tag{16}$$

11. sample hierarchical priors.

(a) sample $b_0|\mu_{1:K}, B_0, h_0, H_0 \sim N(\mu_b, \Sigma_b)$, where

$$\mu_b = \Sigma_b \left(B_0^{-1} \sum_{k=1}^{K} \mu_k + H_0^{-1} h_0 \right) \tag{17}$$

$$\Sigma_b = \left(K B_0^{-1} + H_0^{-1} \right)^{-1} \tag{18}$$

(b) sample $B_0|\mu_{1:K}, b_0, a_0, A_0 \sim IW(\Omega_B, \omega_b)$, where

$$\omega_b = K + a_0 \tag{19}$$

$$\Omega_B = \sum_{k=1}^{K} (\mu_k - b_0)(\mu_k - b_0)' + A_0 \tag{20}$$

(c) sample $s_0|\sigma^2_{1:K}, v_0, c_0, d_0 \sim Gamma(c_s, d_s)$, where

$$c_s = K v_0 + c_0 \tag{21}$$

$$d_s = \sum_{k=1}^{K} \sigma_k^{-2} + d_0 \tag{22}$$

(d) sample $v_0|\sigma^2_{1:K}, s_0, g_0$. There's no easily applicable conjugate prior for v_0 so a Metropolis-Hastings step needs to be applied. Implement a Gamma proposal following Maheu and Yang (2016):

$$v_0'|v_0 \sim Gamma\left(\tau, \frac{\tau}{v_0}\right) \tag{23}$$

and the acceptance rate is

$$\min\left\{1, \frac{p\left(v_0'|\sigma_{1:K}^2, s_0, g_0\right)/q\left(v_0'|v_0\right)}{p\left(v_0|\sigma_{1:K}^2, s_0, g_0\right)/q\left(v_0|v_0'\right)}\right\} \tag{24}$$

12. $\theta_h|h_{1:T}$. Equation (3.2c) is simply a linear regression model. Assuming conjugate prior $\beta \sim N\left(b_h, B_h\right)$, posterior is

$$\delta|\sigma_v, h_{1:T} \sim N\left(M, V\right) \tag{25}$$

$$M = V\left(\sigma_v^{-2}\sum_{t=1}^{T-1} h_t h_{t+1} + b_h B_h^{-1}\right) \tag{26}$$

$$V = \left(\sigma_v^{-2}\sum_{t=1}^{T-1} h_t^2 + B_h^{-1}\right)^{-1} \tag{27}$$

Based on the above linear regression model with conjugate prior $\sigma_v^2 \sim IG\left(v_h, s_h\right)$, the posterior is

$$\sigma_v^2|\delta, h_{1:T} \sim IG\left(\frac{T}{2} + v_h, \frac{\sum_{t=1}^{T-1}\left(h_{t+1} - \delta h_t\right)^2}{2} + s_h\right) \tag{28}$$

13. sample $h_t|h_{-t}, r_{1:T}, \Theta, s_{1:T}$. Use the block Metropolis-Hastings (MH) sampler as in Jensen and Maheu (2010) with random block size $k = Poisson\left(\lambda_h\right) + 1$. Proposal density is derived by approximating the autoregressive coefficient to 1. This approximation provides an analytic inversion of the covariance matrix.

Draw $h'_{(t,\tau)}$ from proposal density

$$g\left(h_{(t,\tau)}|\cdots\right) = N\left(h_{(t,\tau)}; M_h - 0.5V_h\left(\iota - \tilde{y}\right), V_h\right) \tag{29}$$

where

$$\tilde{y}_i = \frac{\left(r_i - \mu_{s_i}\right)^2}{\omega_{s_i}} \exp\left(-M_{h,i}\right) \tag{30}$$

$$M_{h,i} = \frac{\left(k+1-i\right)h_{t-1} + ih_{\tau+1}}{k+1},\ i = 1, 2, \ldots, k \tag{31}$$

$$V_{h,ij} = \sigma_v^2 \frac{\min\left(i,j\right)\left(1+k\right) - ij}{k+1} \tag{32}$$

$$V_{h,ij}^{-1} = \begin{cases} 2\sigma_v^2 & i = j \\ -\sigma_v^2 & j = i \pm 1 \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

Accept $h'_{(t,\tau)}$ with probability

$$\min\left\{1, \frac{p\left(h'_{(t,\tau)}|r_{1:T}, h_{-(t,\tau)}, \Theta, s_{1:T}\right)/g\left(h'_{(t,\tau)}|h_{-(t,\tau)}\right)}{p\left(h_{(t,\tau)}|r_{1:T}, h_{-(t,\tau)}, \Theta, s_{1:T}\right)/g\left(h_{(t,\tau)}|h_{-(t,\tau)}\right)}\right\} \tag{34}$$

Appendix 4 Chapter 4: Sampling Algorithms

Recall that $\Gamma = (\gamma_1, \ldots, \gamma_K, \gamma_R)'$ and $\Pi_j = (\pi_{j1}, \ldots, \pi_{jK}, \pi_{jR})$, where

$$\gamma_R = \sum_{k=K+1}^{\infty} \gamma_k = 1 - \sum_{k=1}^{K} \gamma_k,$$
$$\pi_{jR} = \sum_{k=K+1}^{\infty} \pi_{jk} = 1 - \sum_{k=1}^{K} \pi_{jk}.$$

The sampling steps are:

1. sample $u_{1:T}|\Gamma, \Pi$. The auxiliary slice variable $U = \{u_t\}_{t=1}^{T}$ is drawn by $u_1 \sim U(0, \gamma_{s_1})$ and $u_t \sim U\left(0, \pi_{s_{t-1}s_t}\right)$.

2. update K. Similar to DPM model, if K doesn't meet the condition

$$\min \{u_t\}_{t=1}^{T} > \max \{\pi_{jR}\}_{j=1}^{K} \tag{35}$$

   then K needs to be increased by 1 ($K' = K + 1$), and all parameters need to be drawn from the base measure. In addition, since a new "major" state is introduced, Γ and Π also need to be updated accordingly:

   (a) $\mathbf{\Theta}_{K'} \sim H$;

   (b) draw $v \sim Beta(1, \beta_0)$, then update $\Gamma = (\gamma_1, \ldots, \gamma_K, \gamma_{K'}, \gamma_R)'$, where $\gamma_{K'} = v\gamma_R$ and $\gamma_R = (1 - v)\gamma_R$;

(c) draw $v_j \sim Beta\left(\alpha_0\gamma_{K'}, \alpha_0\gamma_R\right)$, update $\Pi_j = \left(\pi_{j1}, \ldots, \pi_{jK}, \pi_{jK'}, \pi_{jR}\right)$ for $j = 1, \ldots, K$, where $\pi_{jK'} = v\pi_{jR}$ and $\pi_{jR} = (1 - v)\,\pi_{jR}$;

(d) draw the K'th row of Π, $\Pi_{K'}$, by $\Pi_{K'} \sim Dir\left(\alpha_0\gamma_1, \ldots, \alpha_0\gamma_K, \alpha_0\gamma_{K'}, \alpha_0\gamma_R\right)$.

Repeat the above steps until inequality (35) holds.

3. forward filter for $s_{1:T}|\boldsymbol{r}_{1:T}, u_{1:T}, \Gamma, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:T}$. Iterating the following steps forward from 1 to T:

   (a) prediction step for initial state s_1:

$$p(s_1 = k|u_1, \Gamma) \propto \mathbb{1}\left(u_1 < \gamma_k\right), \qquad k = 1, \ldots, K \tag{36}$$

   for the following states $s_{2:T}$:

$$\begin{aligned} &p(s_t = k|\boldsymbol{r}_{1:t-1}, u_{1:t}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t-1}) \\ &\qquad \propto \sum_{j=1}^{K} \mathbb{1}\left(u_t < \pi_{jk}\right) p\left(s_{t-1} = j|\boldsymbol{r}_{1:t-1}, u_{1:t-1}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t-1}\right) \end{aligned} \tag{37}$$

(b) update step for $s_{1:T}$:

$$p\left(s_t = k|\boldsymbol{r}_{1:t}, u_{1:t}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \propto p\left(\boldsymbol{r}_t|\boldsymbol{r}_{t-1}, \boldsymbol{\Theta}_k, \boldsymbol{H}_t\right) p\left(s_t = k|\boldsymbol{r}_{1:t-1}, u_{1:t}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t-1}\right) \tag{38}$$

4. backward sampler for $s_{1:T}|\boldsymbol{r}_{1:T}, u_{1:T}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:T}$. Sample the states $s_{1:T}$ using the previously filtered values backward from T to 1:

   (a) for the terminal state s_T directly from $p\left(s_T|\boldsymbol{r}_{1:T}, u_{1:T}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:T}\right)$

   (b) for the rest states,

$$p\left(s_t = k|s_{t+1} = j, \boldsymbol{r}_{1:t}, u_{1:t+1}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \propto \mathbb{1}\left(u_{t+1} < \pi_{kj}\right) p\left(s_t = k|\boldsymbol{r}_{1:t}, u_{1:t}, \Pi, \boldsymbol{\Theta}, \boldsymbol{H}_{1:t}\right) \tag{39}$$

5. sample $c_{1:K}|s_{1:T}, \Gamma, \alpha_0$. $c_{1:K}$ is essential for sampling Γ, and it counts balls in different colour in the "oracle" urn.[2] The "oracle" urn is only involved when a new state is drawn from the regular urn, and sampling c_k directly would be difficult. Fox et al. (2011) propose to simulate c_k from the original Pòlya urn scheme instead of to sample it.

[2]Refer to Beal et al. (2002) for details.

(a) count the number of each transition type, n_{jk}, for times from state j switching to state k.

(b) simulate an auxiliary trail variable $x_i \sim Bernoulli\left(\frac{\alpha_0\gamma_k}{i-1+\alpha_0\gamma_k}\right)$, for $i = 1, \ldots, n_{jk}$. If the trial is successful, an "oracle" urn step in involved at the ith step toward n_{jk} and we increase the corresponding "oracle" counts, o_{jk}, by one.

(c) $c_k = \sum_{j=1}^{K} o_{jk}$.

6. sample β_0. Following Fox et al. (2011) and Maheu and Yang (2016), assume a Gamma hyper-prior $\beta_0 \sim Gamma\,(a_1, b_1)$, and let $c = \sum_{j=1}^{K} c_j$,

   (a) $\nu \sim Bernoulli\left(\frac{c}{c+\beta_0}\right)$

   (b) $\lambda \sim Beta\,(\beta_0 + 1, c)$

   (c) $\beta_0 \sim Gamma\,(a_1 + K - \nu, b_1 - \log\lambda)$

7. sample α_0. Following Fox et al. (2011), assume a Gamma hyper-prior $\alpha_0 \sim Gamma\,(a_2, b_2)$, and let $n_j = \sum_{k=1}^{K} n_{jk}$,

   (a) $\nu_j \sim Bernoulli\left(\frac{n_j}{n_j+\alpha_0}\right)$

   (b) $\lambda_j \sim Beta\,(\alpha_0 + 1, n_j)$

   (c) $\alpha_0 \sim Gamma\left(a_2 + c - \sum_{j=1}^{K} \nu_j, b_2 - \sum_{j=1}^{K} \log(\lambda_j)\right)$

8. sample $\Gamma|c_{1:K}, \beta_0$. Given the "oracle" urn counts $c_{1:K}$ and the property of Dirichlet process, the conjugate posterior is

$$\Gamma|c_{1:K}, \beta_0 \sim Dir\left(c_1, \ldots, c_K, \beta_0\right) \tag{40}$$

9. sample $\Pi|n_{1:K,1:K}, \Gamma, \alpha_0$. Similarly, the conjugate posterior of Π_j is

$$\Pi_j|n_{j,1:K}, \Gamma, \alpha \sim Dir\left(\alpha_0\gamma_1 + n_{j1}, \ldots, \alpha_0\gamma_K + n_{jK}, \alpha_0\gamma_R\right) \tag{41}$$

10. sample $\boldsymbol{\Theta}|\boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{H}_{1:T}$. Assume conjugate priors $\boldsymbol{\mu} \sim N\left(\boldsymbol{b}_0, \boldsymbol{B}_0\right)$, $\boldsymbol{\Sigma} \sim IW\left(\boldsymbol{\Sigma}_0, \nu + N\right)$. Define $\boldsymbol{Y}_k \equiv \left\{\boldsymbol{H}_t^{-1/2}\boldsymbol{r}_t|s_t = k\right\}_{t=2}^T$, $\boldsymbol{X}_k \equiv \left\{\boldsymbol{H}_t^{-1/2}|s_t = k\right\}_{t=2}^T$. The linear model is now

$$\boldsymbol{Y}_k = \boldsymbol{X}_k\boldsymbol{\mu}_k + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_k\right) \tag{42}$$

The posteriors are

$$p\left(\boldsymbol{\mu}_k|\boldsymbol{Y}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_{1:T}\right) \sim \prod_{t:s_t=k} p\left(\boldsymbol{Y}_t|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_t\right) p\left(\boldsymbol{\mu}_k\right) \tag{43}$$

$$\sim N\left(\boldsymbol{M}_\mu, \boldsymbol{V}_\mu\right) \tag{44}$$

where

$$\boldsymbol{M}_\mu = \boldsymbol{V}_\mu \left(\sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2\prime} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{H}_t^{-1/2} \boldsymbol{r}_t + \boldsymbol{B}_0^{-1} \boldsymbol{b}_0 \right) \tag{45}$$

$$\boldsymbol{V}_\mu = \left(\sum_{t=1}^{T} \boldsymbol{H}_t^{-1/2\prime} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{H}_t^{-1/2} + \boldsymbol{B}_0^{-1} \right)^{-1} \tag{46}$$

and

$$p\left(\boldsymbol{\Sigma}_k | \boldsymbol{Y}_k, \boldsymbol{\mu}_k, \boldsymbol{H}_{1:T}\right) \propto \prod_{t:s_t=k} p\left(\boldsymbol{r}_t | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \boldsymbol{H}_t\right) p\left(\boldsymbol{\Sigma}_k\right) \tag{47}$$

$$\sim IW\left(\bar{\boldsymbol{\Sigma}}, \bar{\nu} + N\right) \tag{48}$$

where

$$\bar{\nu} = T_k + \nu = \sum_{t=1}^{T} \mathbb{1}\left(s_t = k\right) + \nu \tag{49}$$

$$\bar{\boldsymbol{\Sigma}} = \sum_{t:s_t=k} \boldsymbol{H}_t^{-1/2} \left(\boldsymbol{r}_t - \boldsymbol{\mu}_k\right) \left(\boldsymbol{r}_t - \boldsymbol{\mu}_k\right)' \boldsymbol{H}_t^{-1/2\prime} + \boldsymbol{\Sigma}_0 \tag{50}$$

11. sample hierarchical priors.

    (a) sample $\boldsymbol{b}_0 | \boldsymbol{\mu}_{1:K}, \boldsymbol{B}_0, \boldsymbol{h}_0, \boldsymbol{H}_0 \sim N\left(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b\right)$, where

$$\boldsymbol{\mu}_b = \boldsymbol{\Sigma}_b \left(\boldsymbol{B}_0^{-1} \sum_{k=1}^{K} \boldsymbol{\mu}_k + \boldsymbol{H}_0^{-1} \boldsymbol{h}_0 \right) \tag{51}$$

$$\boldsymbol{\Sigma}_b = \left(K \boldsymbol{B}_0^{-1} + \boldsymbol{H}_0^{-1} \right)^{-1} \tag{52}$$

(b) sample $\boldsymbol{B}_0|\boldsymbol{\mu}_{1:K}, \boldsymbol{b}_0, a_0, \boldsymbol{A}_0 \sim IW\left(\boldsymbol{\Omega}_B, \omega_b\right)$, where

$$\omega_b = K + a_0 \tag{53}$$

$$\boldsymbol{\Omega}_B = \sum_{k=1}^{K} \left(\boldsymbol{\mu}_k - \boldsymbol{b}_0\right)\left(\boldsymbol{\mu}_k - \boldsymbol{b}_0\right)' + \boldsymbol{A}_0 \tag{54}$$

(c) sample $\nu|\sigma^2_{1:K}, s_0, g_0$. There's no easily applicable conjugate prior for ν so a Metropolis-Hastings step needs to be applied. Implement a Gamma proposal following Maheu and Yang (2016):

$$\nu'|\nu \sim Gamma\left(\tau, \frac{\tau}{\nu}\right) \tag{55}$$

and the acceptance rate is

$$\min\left\{1, \frac{p\left(\nu'|\boldsymbol{\Sigma}_{1:K}, s_0, g_0\right)/q\left(\nu'|\nu\right)}{p\left(\nu|\boldsymbol{\Sigma}_{1:K}, s_0, g_0\right)/q\left(\nu|\nu'\right)}\right\} \tag{56}$$

(d) sample $\boldsymbol{\Sigma}_0|\boldsymbol{\Sigma}_{1:K}, v_0, \boldsymbol{C}_0, d_0 \sim W\left(\boldsymbol{C}_s, d_s\right)$, where

$$\boldsymbol{C}_s = \left(\sum_{k=1}^{K} \boldsymbol{\Sigma}_k^{-1} + \boldsymbol{C}_0^{-1}\right)^{-1} \tag{57}$$

$$d_s = K\left(\nu + N\right) + d_0 \tag{58}$$

12. sample GARCH parameters $\boldsymbol{\theta}_H = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}\} | \boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{\Theta}$. With normal prior $\boldsymbol{\theta}_H \sim N(\mathbf{0}, \boldsymbol{I})$, the posterior is

$$p(\boldsymbol{\theta}_H | \boldsymbol{r}_{1:T}, s_{1:T}, \boldsymbol{\Theta}) \sim \prod_{t=1}^{T} p(\boldsymbol{r}_t | \boldsymbol{\Theta}, \boldsymbol{H}_t) \, p(\boldsymbol{\theta}_H) \tag{59}$$

Apply a random-walk Metropolis-Hastings algorithm to sample $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and $\boldsymbol{C}$ is jointly targeted.

Bibliography

At-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics* 117(3), 585–606.

At-Sahalia, Y. and Hurd, T. R. (2015). Portfolio choice in markets with contagion. *Journal of Financial Econometrics* 14(1), 1–28.

At-Sahalia, Y. and Xiu, D. (2016). Increased correlation among asset classes: Are volatility or jumps to blame, or both? *Journal of Econometrics* 194(2), 205–219.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance* 13(1), 65–77.

Bali, T. G., Engle, R. F., and Tang, Y. (2017). Dynamic conditional beta is alive and well in the cross section of daily stock returns. *Management Science* 63(11), 3760–3779.

Ball, C. A. and Torous, W. N. (1983). A simplified jump process for common stock returns. *Journal of Financial and Quantitative Analysis* 18(1), 53–65.

Bandi, F. M. and Renò, R. (2016). Price and volatility co-jumps. *Journal of Financial Economics* 119(1), 107–146.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24(1), 1–13.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *The Review of Financial Studies* 9(1), 69–107.

Bates, D. S. (2000). Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* 94(1), 181–238.

Beal, M. J., Ghahramani, Z., and Rasmussen, C. E. (2002). The infinite hidden Markov model. In: *Advances in Neural Information Processing Systems*, 577–584.

Bibinger, M. and Winkelmann, L. (2015). Econometrics of co-jumps in high-frequency data with noise. *Journal of Econometrics* 184(2), 361–378.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.

Bollerslev, T., Law, T. H., and Tauchen, G. (2008). Risk, jumps, and diversification. *Journal of Econometrics* 144(1), 234–256.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2018). Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *Journal of Econometrics* 207(1), 71–91.

Bormetti, G., Calcagnile, L. M., Treccani, M., Corsi, F., Marmi, S., and Lillo, F. (2015). Modelling systemic price cojumps with Hawkes factor models. *Quantitative Finance* 15(7), 1137–1156.

Caporin, M., Kolokolov, A., and Renò, R. (2017). Systemic co-jumps. *Journal of Financial Economics* 126(3), 563–591.

Caporin, M., Rossi, E., and Magistris, P. S. d. (2014). Volatility jumps and their economic determinants. *Journal of Financial Econometrics* 14(1), 29–80.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81(3), 541–553.

Chan, W. H. and Maheu, J. M. (2002). Conditional jump dynamics in stock market returns. *Journal of Business & Economic Statistics* 20(3), 377–389.

Chan, W. H. and Feng, L. (2012). Time-varying jump risk premia in stock index futures returns. *Journal of Futures Markets* 32(7), 639–659.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* 75(1), 79–97.

Chib, S. and Hamilton, B. H. (2002). Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics* 110(1), 67–89.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108(2), 281–316.

Chorro, C., Ielpo, F., and Sévi, B. (2017). The contribution of jumps to forecasting the density of returns.

Chua, C. L. and Tsiaplias, S. (2019). Information flows and stock market volatility. *Journal of Applied Econometrics* 34(1), 129–148.

Conley, T. G., Hansen, C. B., McCulloch, R. E., and Rossi, P. E. (2008). A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics* 144(1), 276–305.

Corrado, C. J. and Su, T. (1996). Skewness and kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial Research* 19(2), 175–192.

Delatola, E.-I. and Griffin, J. E. (2013). A Bayesian semiparametric model for volatility with a leverage effect. *Computational Statistics & Data Analysis* 60, 97–110.

Delatola, E.-I., Griffin, J. E., et al. (2011). Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis* 6(4), 901–926.

Dieobold, F. X. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Reviews* 5(1), 51–56.

Ding, Z. and Engle, R. F. (2001). Large scale conditional covariance matrix modeling, estimation and testing.

Dufays, A. (2016). Infinite-State Markov-Switching for Dynamic Volatility. *Journal of Financial Econometrics* 14(2), 418–460.

Durham, G. B. (2006). Monte Carlo methods for estimating, smoothing, and filtering one-and two-factor stochastic volatility models. *Journal of Econometrics* 133(1), 273–305.

Engle, R. F. (2016). Dynamic conditional beta. *Journal of Financial Econometrics* 14(4), 643–667.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11(1), 122–150.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *The Journal of Finance* 59(3), 1367–1403.

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns. *The Journal of Finance* 58(3), 1269–1300.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business* 38(1), 34–105.

Fleming, J. and Kirby, C. (2003). A closer look at the relation between GARCH and stochastic autoregressive volatility. *Journal of Financial Econometrics* 1(3), 365–419.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The economic value of volatility timing. *The Journal of Finance* 56(1), 329–352.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *The Annals of Applied Statistics*, 1020–1056.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15(2), 183–202.

Gilder, D., Shackleton, M. B., and Taylor, S. J. (2014). Cojumps in stock prices: Empirical evidence. *Journal of Banking & Finance* 40, 443–459.

Gobbi, F. and Mancini, C. (2007). Diffusion covariation and co-jumps in bidimensional asset price processes with stochastic volatility and infinite activity Lévy jumps. *arXiv preprint arXiv:0705.1268*.

Griffin, J. E. and Steel, M. F. (2004). Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics* 123(1), 121–152.

Griffin, J. E. and Steel, M. F. (2011). Stick-breaking autoregressive processes. *Journal of Econometrics* 162(2), 383–396.

Guidolin, M. and Timmermann, A. (2005). *Strategic Asset Allocation and Consumption Decisions under Multivariate Regime Switching*. Tech. rep. Federal Reserve Bank of St, Louis Working Paper 2005-002A.

Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control* 31(11), 3503–3544.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1), 83–90.

Hirano, K. (2002). Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica* 70(2), 781–799.

Hou, C. (2017). Infinite hidden Markov switching VARs with application to macroeconomic forecast. *International Journal of Forecasting* 33(4), 1025–1043.

Jacod, J., Klüppelberg, C., and Müller, G. (2017). Testing for non-correlation between price and volatility jumps. *Journal of Econometrics* 197(2), 284–297.

Jensen, M. J. (2004). Semiparametric Bayesian inference of long-memory stochastic volatility models. *Journal of Time Series Analysis* 25(6), 895–922.

Jensen, M. J. and Maheu, J. M. (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics* 157(2), 306–316.

Jensen, M. J. and Maheu, J. M. (2013). Bayesian semiparametric multivariate GARCH modeling. *Journal of Econometrics* 176(1), 3–17.

Jensen, M. J. and Maheu, J. M. (2014). Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics* 178, 523–538.

Jin, X. and Maheu, J. M. (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics* 192(1), 19–39.

Jin, X., Maheu, J. M., and Yang, Q. (2019). Bayesian parametric and semiparametric factor models for large realized covariance matrices. *Journal of Applied Econometrics* 34(5), 641–660.

Jochmann, M. (2015). Modeling U.S. Inflation Dynamics: A Bayesian Nonparametric Approach. *Econometric Reviews* 34(5), 537–558.

Jondeau, E. and Rockinger, M. (2006). Optimal portfolio allocation under higher moments. *European Financial Management* 12(1), 29–55.

Jorion, P. (1988). On jump processes in the foreign exchange and stock markets. *The Review of Financial Studies* 1(4), 427–445.

Kacperczyk, M., Damien, P., and Walker, S. G. (2013). A new class of Bayesian semi-parametric models with applications to option pricing. *Quantitative Finance* 13(6), 967–980.

Kalimipalli, M. and Susmel, R. (2004). Regime-switching stochastic volatility and short-term interest rates. *Journal of Empirical Finance* 11(3), 309–329.

Kalotychou, E., Staikouras, S. K., and Zhao, G. (2014). The role of correlation dynamics in sector allocation. *Journal of Banking & Finance* 48, 1–12.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies* 65(3), 361–393.

Kon, S. J. (1984). Models of stock returns – a comparison. *The Journal of Finance* 39(1), 147–165.

Kroner, K. F. and Ng, V. K. (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies* 11(4), 817–844.

Lamoureux, C. G. and Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics* 8(2), 225–234.

Laurini, M. P. and Mauad, R. B. (2015). A common jump factor stochastic volatility model. *Finance Research Letters* 12, 2–10.

Liesenfeld, R. and Jung, R. C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* 15(2), 137–160.

Liu, J. and Maheu, J. M. (2018). Improving Markov switching models using realized variance. *Journal of Applied Econometrics* 33(3), 297–318.

Maheu, J. M. and McCurdy, T. H. (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *The Journal of Finance* 59(2), 755–793.

Maheu, J. M., McCurdy, T. H., and Zhao, X. (2013). Do jumps contribute to the dynamics of the equity premium? *Journal of Financial Economics* 110(2), 457–477.

Maheu, J. M., Song, Y., and Yang, Q. (2019). Oil price shocks and economic growth: The volatility link. *International Journal of Forecasting*.

Maheu, J. M. and Yang, Q. (2016). An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance* 38, 202–220.

Mahieu, R. J. and Schotman, P. C. (1998). An empirical application of stochastic volatility models. *Journal of Applied Econometrics* 13(4), 333–360.

Mancini, C. and Gobbi, F. (2012). Identifying the brownian covariation from the co-jumps given discrete observations. *Econometric Theory* 28(2), 249–273.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business* 36(4), 394–419.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57(3), 519–530.

Marquering, W. and Verbeek, M. (2004). The economic value of predicting stock index returns and volatility. *Journal of Financial and Quantitative Analysis* 39(2), 407–429.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3(1-2), 125–144.

Mills, T. C. (1995). Modelling skewness and kurtosis in the London Stock Exchange FT-SE index return distributions. *The Statistician*, 323–332.

Nieuwland, F. G., Verschoor, W. F., and Wolff, C. C. (1994). Stochastic trends and jumps in EMS exchange rates. *Journal of International Money and Finance* 13(6), 699–727.

Novotný, J. and Urga, G. (2018). Testing for co-jumps in financial markets. *Journal of Financial Econometrics* 16(1), 118–128.

Oldfield, G. S., Rogalski, R. J., and Jarrow, R. A. (1977). An autoregressive jump process for common stock returns. *Journal of Financial Economics* 5(3), 389–418.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics* 63(1), 3–50.

Peiró, A. (1994). The distribution of stock returns: international evidence. *Applied Financial Economics* 4(6), 431–439.

Peiró, A. (1999). Skewness in financial returns. *Journal of Banking & Finance* 23(6), 847–862.

Pettenuzzo, D. and Ravazzolo, F. (2016). Optimal Portfolio Choice Under Decision-Based Model Combinations. *Journal of Applied Econometrics* 31(7), 1312–1332.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics* 164(1), 60–78.

Pitman, J. (2002). Poisson-Dirichlet and GEM invariant distributions for split-and-merge transformations of an interval partition. *Combinatorics, Probability and Computing* 11(5), 501–514.

Praetz, P. D. (1972). The distribution of share price changes. *The Journal of Business*, 49–55.

Press, S. J. (1967). A compound events model for security prices. *Journal of Business*, 317–335.

Rambaldi, M., Pennesi, P., and Lillo, F. (2014). Modeling FX market activity around macroeconomic news: a Hawkes process approach. arXiv preprint. *arXiv preprint arXiv:1405.6047* 67(75), 210.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 639–650.

Shi, S. and Song, Y. (2015). Identifying Speculative Bubbles Using an Infinite Hidden Markov Model. *Journal of Financial Econometrics* 14(1), 159–184.

Shibata, M. and Watanabe, T. (2005). Bayesian analysis of a Markov switching stochastic volatility model. *Journal of the Japan Statistical Society* 35(2), 205–219.

So, M. E. P., Lam, K., and Li, W. K. (1998). A stochastic volatility model with Markov switching. *Journal of Business & Economic Statistics* 16(2), 244–253.

Song, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics* 29(5), 825–842.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice* 1, 203–226.

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.

Tiwari, R. C., Jammalamadaka, S., and Chib, S. (1989). Bayes prediction density and regression estimation – A semiparametric approach. In: *Semiparametric and Nonparametric Econometrics*. Springer, 87–100.

Todorov, V. and Bollerslev, T. (2010). Jumps and betas: A new framework for disentangling and estimating systematic risks. *Journal of Econometrics* 157(2), 220–235.

Van Gael, J., Saatci, Y., Teh, Y. W., and Ghahramani, Z. (2008). Beam sampling for the infinite hidden Markov model. In: *Proceedings of the 25th International Conference on Machine Learning*. ACM, 1088–1095.

Virbickait, A. and Lopes, H. F. (2019). Bayesian semiparametric Markov switching stochastic volatility model. *Applied Stochastic Models in Business and Industry* 35(4), 978–997.

Vlaar, P. J. and Palm, F. C. (1993). The message in weekly exchange rates in the European monetary system: mean reversion, conditional heteroscedasticity, and jumps. *Journal of Business & Economic Statistics* 11(3), 351–360.

Vo, M. T. (2009). Regime-switching stochastic volatility: Evidence from the crude oil market. *Energy Economics* 31(5), 779–788.

Winkelmann, L., Bibinger, M., and Linzert, T. (2016). ECB monetary policy surprises: identification through cojumps in interest rates. *Journal of Applied Econometrics* 31(4), 613–629.

Yang, Q. (2019). Stock Returns and Real Growth: A Bayesian Nonparametric Approach. *Journal of Empirical Finance* 38, 202–220.

Zhang, S., Zhou, Q. M., Zhu, D., and Song, P. X.-K. (2019). Goodness-of-Fit Test In Multivariate Jump Diffusion Models. *Journal of Business & Economic Statistics* 37(2), 275–287.