Dissertations and Theses Collection (Open Access)   Dissertations and Theses

7-2020

Essays on multivariate stochastic volatility models

Han CHEN
Singapore Management University

Citation

SINGAPORE MANAGEMENT UNIVERSITY

PHD DISSERTATION

Essays on Multivariate Stochastic Volatility Models

Han (Hank) Chen

supervised by

Professor JUN YU

2020

Essays On Multivariate Stochastic Volatility Models

Han (Hank) Chen

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial

Fulfilment

of the Requirements for the Degree of PhD in Economics

2020

_____

Supervisor of Dissertation

_____

PhD in Economics, Programme Director

Dissertation Committee:

Jun Yu (Supervisor/Chair)
Lee Kong Chian Professor of Economics and Finance
School of Economics and Lee Kong Chian School of Business
Singapore Management University

Yichong Zhang
Assistant Professor of Economics
Singapore Management University

Daniel Preve
Associate Professor of Economics
Singapore Management University

Tao Zeng (External member)
Associate Professor of Economics
Zhejiang University

Abstract

In this dissertation, I have made several contributions to the literature on the multivariate stochastic volatility model.

First, I have considered a new multivariate stochastic volatility (MSV) model based on a recently proposed novel parameterization of the correlation matrix. This modeling design is a generalization of Fisher's z-transformation to the high-dimensional case. It is fully flexible as the validity of the resulting correlation matrix is guaranteed automatically. It allows me to completely separate the driving factors of volatilities and correlations. To conduct an econometric analysis of the proposed model, I develop a new Bayesian method that relies on the Markov Chain Monte Carlo (MCMC) tool. For the latent variables, the traditional single-move or multi-move sampler is replaced by a novel technique called Particle Gibbs Ancestor Sampling (PGAS), which is built upon the Sequential Monte Carlo (SMC) method. Simulation results indicate that our algorithm performs well when a small number of particles are used. Empirical studies based on the exchange rate returns and equity returns are considered and reveal some interesting empirical results.

Second, I further develop a multivariate stochastic volatility model with intraday realized measures. A simple and consistent estimation technique is developed. The problem of under-identification is discussed. A two-stage approach is introduced to address the problem. A simulation study shows that the proposed method works well in finite samples. The new model is then implemented using two financial datasets. A comparison with some existing models is made.

Third, I also incorporate the leverage effect and the heavy-tailed error distribution into the MSV model. A Particle Gibbs Sampling Algorithm is developed for the extended MSV model. Simulation results indicate that our algorithm performs well when a small number of particles are used. Empirical studies of the stock indices are considered. I have found strong evidence of the leverage effect and, more, importantly, heavy-tails in the errors.

Contents

List of Figures

List of Tables

Acknowledgment

I would like to express gratitude to my advisor, Professor Jun Yu, for his patience, support and encouragement. He guides me with the necessary skill and knowledge to complete the thesis. I sincerely appreciate all opportunities and ideas he has provided to me.

I would like to thank my committee members, Professor Yichong Zhang, Professor Daniel Preve, and Professor Tao Zeng, for the insightful comment. I especially own debt to Professor Yichong Zhang for his encouragement and insightful comment during my Ph.D. study.

I also would like to give thanks to all my friends in Professor Yu's reading group of econometrics. In particular, I would like to thank my friend and co-author Yijie Fei for stimulating discussions on this thesis; and thank my senior, Xiaobin Liu, for his encouragement and comment on my thesis.

My sincere thanks also go to Professor Anthony Tay and Thor Qiu Ling for their kindness and help during this five-year study. Finally, I would like to give thanks to my family for their continuous support and encouragement.

1 Introduction

Characterizing the dynamic behavior of volatility of individual asset returns is of great importance to both portfolio allocation and risk management. Starting from the seminal paper by Engle (1982), a wide range of univariate volatility models have been considered in the literature, most of which can be categorized as either GARCH-based models or stochastic volatility (SV) models. Another notable development during the past few decades is the focus on multivariate financial data analysis. It is increasingly recognized that analyzing the asset return individually is not adequate, and the dependence structure among different assets must be taken into account. To this end, a plethora of multivariate extensions to univariate GARCH and SV models have emerged and been developed and applied in practice. See Bauwens et al. (2006) for an extensive review of multivariate GARCH (MGARCH) models and Asai and McAleer (2006) for a review of multivariate SV (MSV) models.

The first MSV model is proposed in Harvey et al. (1994) and is an analogy of the constant conditional correlation (CCC) model in the MGARCH literature. In this basic setup, the volatility of each individual asset is assumed to follow a univariate SV process, while the correlation matrix among all assets is constant over time. This is obviously a rather restrictive and perhaps unrealistic assumption, as it implies that the correlation structure is unchanged over time. Great efforts have been dedicated to the relaxation of the constant correlation assumption ever since. For instance, Yu and Meyer (2006) proposed a model that mirrors the dynamic conditional correlation (DCC) model of Engle (2002) in MGARCH. Another parametrization that is also based on DCC can be found in Asai and McAleer (2009).

All the models mentioned above are built upon a variance-correlation decomposition of the covariance matrix. It is also possible to characterize the dynamic conditional correlation by modeling the covariance matrix per se. Similar to the case in the MGARCH literature, the major challenge here is to make sure that the model can produce a positive definite covariance matrix. One possibility is to model the matrix logarithm of the original covariance matrix, which must exist and is free

of algebraic constraint by design. This idea, as an extension of exponential GARCH in Kawakatsu (2006), is explored in Ishihara et al. (2016). A significant drawback of this approach is that the volatilities and correlations are interwoven with each other, and thus the estimated model is very hard to interpret. Another existing approach, considered in Lopes et al. (2010), takes advantage of the well-known Cholesky decomposition of a symmetric positive definite matrix. The useful property of this method is that there is a clear and meaningful connection between elements of the original covariance matrix and those of the component matrices. A shortcoming of this modeling strategy is that the driving forces underlying volatilities and correlations are not completely separated. Last but not least, the Wishart autoregressive multivariate process provides a flexible modeling tool for MSV as well; see Philipov and Glickman (2006) and Gouriéroux et al. (2009) for details.

In this dissertation, we propose to apply a recently developed new parameterization of the correlation matrix to develop a new MSV model. Such a parameterization, originally proposed in Archakov and Hansen (2018a), has been successfully employed to introduce a new MGARCH model; see Archakov et al. (2020). Archakov et al. (2020) show that this approach can be deemed as a generalization of the well-known Fisher's z-transformation to the higher dimensional case. In this chapter, we seek to use the new parameterization of the correlation matrix to introduce a new MSV model. Under this new modeling design, the underlying latent states that determine the correlation among assets are allowed to have an unrestricted domain, and no algebraic constraint of any kind is necessary. This is because this new parameterization by construction automatically guarantees that the estimated correlation matrix is indeed valid. Meanwhile, the underlying shocks to volatilities and correlations are fully separated in our model. This is an appealing feature, as in practice, these two quantities may be determined by completely distinct factors. Last but not least, our model is invariant to the reordering of assets, and thus no ex-ante ordering is necessary. All these features indicate that our model is very flexible in terms of capturing dynamic patterns of volatilities and correlations, imposing a minimum level of ex-ante restrictions.

2

As is the case for other MSV models, estimation of our model faces challenges such as a large number of parameters and high dimensional latent variables. To accommodate the first challenge, we follow the existing work and apply the Bayesian Markov Chain Monte Carlo (MCMC) method. A Bayesian method is considered because the maximum-likelihood-based approach is unstable or even fails to work in this setting. Departing from much of the literature, where a carefully designed single-move sampler or multi-move Gibbs sampler is used to tackle the latent variables issue, we opt to work with the recently proposed particle-filter-based MCMC (PMCMC) algorithm. Ever since the seminal paper by Andrieu et al. (2010), the research on the theoretical foundation of PMCMC and its applications in many different fields have mushroomed. Though theoretically applicable under a very general setup, the practical performance of this approach for a particular model depends on many factors and requires careful examination. To strike a balance between the satisfactory estimation accuracy and the acceptable computational cost, we choose a method called Particle Gibbs Ancestor Sampling (PGAS). It is a modified version of the Particle Gibbs Sampling (PG) considered in the Andrieu et al. (2010), which dramatically improves the mixing property under a small number of particles. Extensive simulation results are presented to justify the choice and also provide some guidance for empirical applications.

The MSV model is estimated solely based on the daily return data. While this approach of modeling is prevalent in the existing literature, it does not fully utilize the available information. Another well-known potential source of information for return fluctuation is realized volatility (RV) calculated using intraday high-frequency data; see Andersen et al. (2010b) for a recent survey. Inspired by the realized-GARCH models proposed in Hansen et al. (2014) and Noureldin et al. (2012), MSV models based on both return series and realized volatility data, referred to as RSV, have appeared in the literature. RV, though shown to be a consistent estimator of the latent variance of log-return under perfect market condition, is subject to market microstructure noises and may exhibit large bias. On the contrary, the daily return is not seriously contaminated by noises and can be used to eliminate the bias.

Besides, adding RV into the SV model as an extra measurement can also improve the inference efficiency of parameters in the basic SV model. Combining RV with SV/GARCH, therefore, seems to be a promising direction.

In light of these advances in the literature, we also incorporate realized measures into our MSV model. This is done by applying the new transformation to the realized covariance matrix, and the resulting variables provide additional measurements to the latent variables in the original model. As argued in Yamauchi and Omori (2019), this extra information source can help stabilize the parameter estimation. One thing to notice, however, is that this will introduce more parameters into the model and these additional parameters are not separately identified with parameters in the original model. For example, the intercept coefficient in the autoregression equation for latent log-variance will not be identified if we impose that realized measures are biased proxies for unobserved factors. A similar concern applies to other parameters. To circumvent this problem, we propose to do a two-stage estimation. Namely, we first estimate a model without realized measures and then obtain the rest parameters by imposing identification conditions.

When examining the residuals generated by fitting our model to real data, we find that there are a few significant outliers. This indicates that our model, though quite flexible, still leaves some important stylized features unaccounted for. Indeed, in the basic model we consider, no leverage effect and heavy-tailed innovation are allowed. Given the strong evidence in the literature supporting the existence of these features, we extend our model by making less restrictive assumptions on some building blocks. Specifically, we allow the innovations to return and volatility to be correlated. We also assume that the innovations to return are t-distributed instead of Gaussian. As shown by the Q-Q plot of model residuals, this extended model greatly reduces the number of outliers, which suggests an improved fitting of real data.

The rest of the dissertation is organized as follows. In Chapter 2, we provide a selective literature review on the MSV models, as well as realized measures constructed from high-frequency data. In Chapter 3, we introduce the new parametrization of

the correlation matrix. We also present the basic MSV model and discuss how to estimate the model and make the statistical inference of parameters in the model in Chapter 3. In Chapter 4, we discuss how to incorporate the information in realized measures. In particular, we discuss both the joint estimation method and the two-stage estimation method and explain why we choose the two-stage method. Chapter 5 presents generalized MSV models with the leverage effect and the conditional distribution with heavy tails. Chapter 6 concludes.

2 A Literature Review

We first review the literature on the MSV models. Asai et al. (2006) summarize this area of research up to that time, and discussed both the estimation techniques and the model comparison methods. A similar account can be found in Chib et al. (2009). For early studies in the literature, we refer the reader to these review papers. Here, we mainly consider models proposed over the last ten years, paying particular attention to how they ensure the positivity of the covariance matrices and to the choice of inference methods.[1]

2.1 Model Setup

The basic structure of the MSV model is

$$r_t | C_t \sim N(0, C_t),$$

where r_t is a vector of asset returns. We aim at characterizing the dynamics of its variance-covariance matrix C_t. Clearly, C_t must be symmetric and positive definite for all t. Different models rely on different ways to ensure this property. Broadly speaking, we can categorize the MSV models into two groups. In the first group, a model is directly built for C_t. In the second group, a covariance decomposition is first carried out and then each component in the decomposition is modeled separately.

Within the first group of models, three methods have been considered. The first method is based on the matrix exponential. Ishihara et al. (2016) assume that

$$C_t = \exp(H_t/2),$$

and propose to model $vech(H_t)$ as a vector autoregressive process. Due to the definition of the matrix exponential, C_t is guaranteed to be a valid variance-covariance matrix. The major drawback of this model is that the relationship between the latent

---

[1]Note that we do not consider models based on the factor structure in this review, as our new model is based on the strategy of direct modeling of the covariance matrix.

variables in the model and the original volatilities/correlations is highly nonlinear and thus very hard to interpret.

The second method utilizes the well-known Cholesky decomposition. For instance, Lopes et al. (2010) propose to decompose C_t as

$$C_t = A_t H_t A_t',$$

where H_t is a diagonal matrix and A_t is a lower triangular matrix, and then model all the nonzero elements in A_t and H_t as the autoregressive process. Similarly, Shirota et al. (2017) also use this decomposition to set up their MSV model. Although under this parameterization, we can easily link the latent variables to the volatilities/correlations, other problems arise with the Cholesky decomposition. For instance, the resulting matrix depends on the ordering of assets. This dependence is highly undesirable. Meanwhile, the underlying dynamics of the volatilities and that of the correlations are not fully separated.

The third method takes advantage of the Wishart distribution, whose support includes only positive definite matrices. This is considered in Gouriéroux et al. (2009), where a Wishart autoregressive process is used. Specifically, Gouriéroux et al. (2009) assume that

$$C_t = \sum_{i=1}^{m} x_{it} x_{it}',$$

$$x_{it} = A x_{i,t-1} + \epsilon_{it} \text{ and } \epsilon_{it} \sim N(0, \Sigma),$$

where (v, d, A) are unknown parameters. Alternatively, one can also model C_t uses the inverse Wishart as in Philipov and Glickman (2006). In this case, we have

$$C_t^{-1}|v, C_{t-1}^{-1} \sim Wishart\left(v, \frac{1}{v}(A^{1/2})(C_{t-1}^{-1})^d(A^{1/2})'\right),$$

where (m, A, Σ) are unknown parameters. A similar model specification is presented in Jin and Maheu (2013).

The models in the second group treat the volatility of each asset and the corre-

lation matrix separately, based on the following decomposition

$$C_t = V_t^{1/2} R_t V_t^{1/2},$$

where V_t is a diagonal matrix collecting all the variances, and R_t is the correlation matrix. For our purpose, the major difference of designs among this group lies in how R_t is parameterized. Note that the critical issue in this setup is again how to ensure the positivity of the correlation matrix. The first and the simplest model in this fashion is the constant correlation MSV in Harvey et al. (1994), where

$$R_t = R, \text{ for all } t.$$

A similar assumption is made in Chan et al. (2006) and Asai and McAleer (2006). It can be easily seen that in these models, the dynamic movement of the correlations is not allowed. Although the inference under this assumption is simple, it is clearly too restrictive for modeling most financial data.

To allow for time-varying correlations, Asai and McAleer (2009) consider two models motivated by the Dynamic Conditional Correlation (DCC) model of Engle (2002) in the MGARCH literature. The idea is to write the correlation matrix in the following form

$$R_t = \tilde{Q}_t^{-1} Q_t \tilde{Q}_t^{-1},$$

where $\tilde{Q}_t = (diag(vecd(Q_t)))^{1/2}$. By construction, the diagonal elements of R_t must be 1 and the symmetry and the positive definiteness can be achieved if Q_t is symmetric positive definite. Two Wishart-driven models for Q_t that they propose are

$$Q_{t+1} = (1 - \phi)\bar{Q} + \phi Q_t + \Xi_t, \text{ where } \Xi_t \sim Wishart(k, \Lambda),$$

and

$$Q_{t+1}^{-1}|k, Q_t^{-1} \sim Wishart\left(k, \frac{1}{k} Q_t^{-\phi/2} \Lambda Q_t^{-\phi/2}\right),$$

where the unknown parameters are k, ϕ, Λ. Between these two parameterizations,

the authors argued that the second one is preferred.

Inspired by the Dynamic Equicorrelation (DECO) model of Engle and Kelly (2012) in the MGARCH literature, Kurose and Omori (2016) proposed to model R_t as

$$R_t = (1 - \rho_t)I + \rho_t J,$$

where I is an identity matrix, and J is a square matrix with all elements equal to 1. To ensure that ρ_t is within $(-1, 1)$, following Yu and Meyer (2006), they model the Fisher transformation of ρ_t as an autoregressive process. Note that in this model, R_t is positive definite only if ρ_t is large than some lower bound, depending on the number of assets. When more assets are considered, this lower bound approaches 0. Kurose and Omori (2020) further extend this model to the multiple-block case and include other features.

More recently, Yamauchi and Omori (2019) propose to model the pairwise correlations by the Fisher transformation. Their parameterization is

$$R_t = \{\rho_{ij,t}\}, \text{ where } \rho_{ij,t} = \frac{\exp(g_{ij,t}) - 1}{\exp(g_{ij,t}) + 1},$$

and $g_{ij,t}$ is assumed to follow a random walk. Since this element-wise operation does not guarantee the validity of R_t as a correlation matrix, they further derive algebraic bounds for $\rho_{ij,t}$ that ensure the positive definiteness of R_t. Note that the bounds for one particular $\rho_{ij,t}$ are conditional on all other elements in R_t. Therefore, it is well suited for the single-move Gibbs sampling technique, but hard to be extended to another estimation method.

2.2 Estimation Method

Unlike observation-driven models like univariate and multivariate GARCH, which can be estimated straightforwardly by the maximum likelihood method, stochastic volatility models are particularly challenging in terms of estimation and inference. This is due to the high-dimensional latent variables involved in the models, as well as a large number of parameters. To deal with these complications, historically, dozens

of methods were proposed and implemented to make inferences for SV models. For a detailed survey on this topic, see Broto and Ruiz (2004).

The earliest and perhaps the most straightforward method for estimating SV models is the method of moments, which is considered in Taylor (1986a). Other moment-based procedures include the generalized method of moments as in Melino and Turnbull (1990) and Andersen and Sørensen (1996), and the simulated method of moments proposed in Duffie and Singleton (1993). Though very easy to implement, these methods have less satisfactory finite sample properties. Moreover, they do not provide estimates of underlying latent processes. Later, a quasi-maximum-likelihood estimator (QMLE) is proposed in Harvey et al. (1994), which relies on a log-linearization of the original model. This transformation results in a linear state space form with log chi-squared errors, and the Kalman filter is subsequently used to compute the likelihood. However, QMLE does not rely on the exact likelihood of data, and the approximation of log chi-squared errors by Gaussian density could be somewhat inappropriate. A better approximation to the exact likelihood can be achieved either by using a mixture of normals or through numerical simulation. The former is proposed in Kim et al. (1998). Monte Carlo Likelihood approach proposed in Sandmann and Koopman (1998) provides an example of the latter.

A majority of recent papers on stochastic volatility work under the Bayesian framework and base their inference on the Markov Chain Monte Carlo technique. This is particularly the case for multivariate models, as they usually involve a large number of parameters, and optimizing over a high-dimensional parameter space is well known to be complicated. For a Bayesian method, this problem becomes computing high-dimensional integral, which can be efficiently done by MCMC sampling. The critical step of this procedure is to draw a sample of latent processes given a particular set of static parameters. The first method proposed in the literature is the single-move sampler in Jacquier et al. (1994), where each latent variable is drawn one at a time, given all the other ones. This approach is well known to be inefficient, as it will generate seriously autocorrelated samples from the Markov chain, suggesting a vast amount of random draws is required to achieve a satisfac-

tory accuracy of estimation. This is particularly true for empirically relevant cases of strongly persistent volatility. Yu and Meyer (2006) discusses how to implement the simple-move sampler under a popular Bayesian software known as WinBUGS. In light of the inefficiency in mixing, Kim et al. (1998) provides an alternative MCMC method, which is highly efficient. Their method also starts with a log-linearization of the model, after which they introduce extra auxiliary variables to reach a linear state-space form. The simulation smoother of De Jong and Shephard (1995) is then applied to draw the whole latent process simultaneously. Although this method is very popular in the univariate SV literature, it is not easy to extend to MSV models with complicated correlation structures. Another significant improvement of the single-move sampler is proposed in Pitt and Shephard (1999). These authors suggest blocking to improve the speed of convergence for simulators of non-Gaussian state-space models. Within each block, a second-order Gaussian approximation is used to obtain a good proposal density. Since sampling the whole sequence of latent variables in one block leads to a high rejection rate, they also discuss how to choose a balanced block size. This method is later modified by, e.g., Watanabe and Omori (2004),Omori and Watanabe (2008), and attracts a lot of applications.

2.3   Realized Variance and Realized Covariance Matrix

Let us first discuss the realized variance (RV) of a univariate asset return. To illustrate the basic idea behind RV, consider the price dynamics driven by the Brownian motion as follows,

$$ds(t) = \alpha dt + \sigma(t) dW(t),$$

where α and $\sigma(t) = \sigma > 0$ represent constant mean and volatility, respectively. Suppose one have n equally-spaced discrete observations in interval $[0, K]$. Since the drift term α is of smaller order on a short interval, one can estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{K} \sum_{j=1}^{nK} \left[s\left(\frac{j}{n}\right) - s\left(\frac{j-1}{n}\right) \right]^2.$$

11

It is easy to show that $\hat{\sigma}^2$ is consistent and obtain its in-fill asymptotic distribution as $\sqrt{nK}(\hat{\sigma}^2 - \sigma^2) \to^d N(0, 2\sigma^4)$. Similar results exist when the volatility process is stochastic rather than time-invariant. Specifically, realized variance defined as

$$RV(t - K, t) = \sum_{j=1}^{nK} \left[s\left(t - K + \frac{j}{n}\right) - s\left(t - K + \frac{j-1}{n}\right) \right]^2,$$

on interval $[t - K, t]$ will be a consistent estimator for the integrated variance $IV(t - K, t) = \int_{t-K}^{t} \sigma(\tau)^2 d\tau$ and asymptotic normality still applies as n goes to infinity. Now consider a p-dimensional asset price vector which follows diffusion process defined by

$$dS(t) = \alpha(t)dt + \Sigma(t)dW(t),$$

where $W(t)$ is now a p-dimensional independent standard Brownian motion. To estimate integrated convariance $ICOV(t - K, t) = \int_{t-K}^{t} \Sigma(\tau)\Sigma(\tau)' d\tau$ of these assets on the interval $[t - K, t]$, one can analogously construct realized covariance as follows

$$RCOV(t - K, t) = \sum_{j=1}^{nK} \left[S\left(t - K + \frac{j}{n}\right) - S\left(t - K + \frac{j-1}{n}\right) \right]$$
$$\left[S\left(t - K + \frac{j}{n}\right) - S\left(t - K + \frac{j-1}{n}\right) \right]'.$$

For detailed discussions, see Andersen et al. (2001a), Barndorff-Nielsen and Shephard (2002), and Hurn et al. (2020).

Though theoretically, we can obtain an estimator for integrated variance/covariance which is arbitrarily close to the true value by increasing sampling frequency, there are a few practical challenges that complicate the application of this asymptotic argument. These include measurement errors induced by discretization, as well as various types of market microstructure noise like price discreteness, bid-ask spread, and asynchronous trading. In general, such contamination causes bias in the RV and RCOV. Hence, it is critical to choose a sampling frequency that strikes a balance between estimation accuracy and the adverse effects of microstructure noise. A popular and simple method in practice is to sample sparsely, such as every 1 minute,

5 minutes, or 30 minutes. A partial list of more sophisticated solutions in the literature includes multi-scale realized variance by Zhang et al. (2005) and Zhang (2006); realized kernel of Barndorff-Nielsen et al. (2009); pre-averaged realized variance by Podolskij and Vetter (2009) and Jacod et al. (2009); realized range-based variance of Christensen and Podolskij (2007).

2.4 Multivariate Stochastic Volatility and Realized Measures

During the last two decades of the 20th century, modeling asset volatility was achieved by using daily return data only. Leading examples are the GARCH-type models (Engle (1982); Bollerslev (1986)) and the SV-type models (Taylor (1986a)). These models have been widely studied in academia and applied by practitioners. More recently, an increasing number of papers turn to work with intra-daily high-frequency data and use the methodology introduced in the last subsection to investigate the characteristics of volatility. As both data sources potentially provide useful information for second moments of asset returns, it is natural to combine them when making statistical inferences and predictions.

Within the GARCH framework, the effort was initially made by Engle (2000), where realized measures are included as exogenous variables. This model, however, is incomplete in the sense that realized measures are not explicitly modeled. Complete models have been proposed soon after. Three leading models are MEM by Engle and Gallo (2006), HEAVY model by Shephard and Sheppard (2010), and Realized GARCH by Hansen et al. (2012). Multivariate extensions are also available now, see e.g. Noureldin et al. (2012), Hansen et al. (2014). Since these models are all observation-driven and no latent variables are involved, they can be easily estimated by the standard maximum likelihood method. Prediction of future volatility is straightforward as well.

In the SV literature, fewer studies exist on joint modeling of daily and intra-daily returns. Existing studies include Takahashi et al. (2009) and Koopman and Scharth

(2012), which both consider the univariate model. The former was extended in several directions by Venter and de Jongh (2014), while Zheng and Song (2014) extend the latter paper by considering the Box-Cox transformation. Recent contributions to multivariate modeling include Shirota et al. (2017), Kurose and Omori (2020), and Yamauchi and Omori (2019). Models proposed in these papers incorporate a lot of stylized features such as the leverage effect, endogeneity, and block structure. To ensure the positive definiteness of the covariance matrix, these studies use techniques such as the Cholesky decomposition, the matrix exponential, and the pairwise Fisher transformation. See also Jin and Maheu (2013) and Jin and Maheu (2016), where the Wishart autoregression is used to model the stochastic covariance matrix. Due to the high-dimensional latent variables involved, estimation and inference for such models are usually computationally intensive and rely heavily on simulation-based approaches, especially MCMC.

2.5   Leverage Effect and Heavy-tailedness

It is well known that the basic SV model cannot capture all important stylized facts of asset returns. Another important stylized fact in much financial time series is the asymmetric effect of return on volatility. In the literature, this effect is often called the leverage effect. According to Black (1976), when there is a negative shock in return, the firm's financial leverage increases, and hence the future expected volatility also increases. However, when there is a positive shock in return, the firm's financial leverage decreases, and hence the future expected volatility also decreases. Therefore, the negative shock in return generates a different impact on the expected volatility from the positive shock of the same magnitude. The empirical evidence of the leverage effect is provided by Chirstie (1982). In the univariate SV literature, Harvey and Shephard (1996) propose a model with the leverage effect and estimate the model by quasi-maximum likelihood while Meyer and Yu (2000) estimate the same model using the MCMC method. Jacquier et al. (2004) estimate the SV model with the leverage effect using MCMC. Yu (2005) notes that there is a difference between the specifications in Harvey and Shephard (1996) and Jacquier

et al. (2004). He compares the two specifications and shows that specification in Jacquier et al. (2004) is inferior to that in Harvey and Shephard (1996). Omori et al. (2007) shows the basic approach in Kim et al. (1998) can be extended to the model with leverage effect. In the context of the MSV model, the leverage effect can be extended to the cross-leverage effect (i.e., a nonzero correlation between the return of the ith asset at time t and the log-volatility of the jth asset at time $t + 1$). Chan et al. (2006) and Asai and McAleer (2006) proposed alternative MSV models with the leverage effect. Ishihara and Omori (2012) proposes an efficient Bayesian method to estimate the MSV model with the cross-leverage effect. More recent contributions in the literature include Ishihara et al. (2016), and Yamauchi and Omori (2019).

Although heavy tails in the unconditional return distribution are possible in the SV model with the conditional Gaussian error distribution, empirical evidence often suggests that the tails are not heavy enough compared with data. One way to generate heavy tails that match with those of the data is to use the Student-t distribution. In the context of univariate SV models, the heavy-tailed error distribution has been used in Kim et al. (1998), Meyer and Yu (2000), Liesenfeld and Jung (2000), Berg et al. (2004), Jacquier et al. (2004), Chib et al. (2002), and Omori et al. (2007). In the context of MSV model, the heavy-tailed error distribution has been used in Yu and Meyer (2006), Ishihara and Omori (2012).

3 A Multivariate Stochastic Volatility Model with the Generalized Fisher Transformation

3.1 The Generalized Fisher z-Transformation and MSV-GFT Model

In this section, we first review the new parametrization of the correlation matrix proposed in Archakov and Hansen (2018a). This novel parametrization can be considered as a high-dimensional generalization of the widely-used Fisher z-transformation in bivariate correlation modeling. Based on this new tool, a flexible MSV model is proposed.

3.1.1 Parametrization of Correlation Matrix

When the correlation coefficient between two random variables, say ρ, is to be modeled, an essential constraint is that its value must be within the interval $(-1, 1)$. To avoid complexity brought by this constraint in modeling, one can instead model Fisher's z-transformation of ρ, defined as $F(\rho) = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$. We know that $F(\cdot)$ is a one-to-one mapping and for any $F(\rho) \in \mathscr{R}$, there exists a unique corresponding $\rho \in (-1, 1)$. Therefore, one can impose any structure on $F(\rho)$ and transform it back to obtain ρ without worrying about the validity of the resulting correlation coefficient. This idea was first introduced to the multivariate stochastic volatility literature in Yu and Meyer (2006). Unfortunately, it is acknowledged by the authors that this approach "is not easy to be generalized into higher dimension situations". In particular, a pair-wise transformation applied to each entry in a high-dimensional correlation matrix, though seems to be natural, is not a valid choice as it fails to guarantee the positive definiteness of the resulting correlation matrix.

Fisher's z-transformation has many nice properties. It is desirable to obtain a valid high-dimensional version of this transformation. The proposal made in Archakov and Hansen (2018a) is to apply the matrix logarithm to the original cor-

relation matrix and then model all the lower off-diagonal elements in that new matrix. To fix the idea, suppose we have a p-dimensional correlation matrix C and let $G = \log C = \sum_{k=1}^{\infty} \frac{(-1)^k (C-I)^k}{k}$. Note that the convergence of the infinite summation and thus the existence of G is ensured by the fact that C, as a correlation matrix, is symmetric and positive definite. Further, denote $q = vecl(G)$ as the $\frac{p(p-1)}{2}$-dimensional vector containing all lower off-diagonal entries of G. In summary, the new parametrization of the original correlation matrix is now defined by the mapping $q = vecl(\log C)$. One of key theoretical contributions of Archakov and Hansen (2018a) is that this mapping is shown to be one-to-one. That is to say, given any $\frac{p(p-1)}{2}$-dimensional vector q, there exists a unique valid p-dimensional correlation matrix C. Although the inverse mapping from q to C does not admit a closed-form analytical solution, C can easily be reconstructed from q using an iterative algorithm.[2]

It is straightforward to show that when $p = 2$, the above-defined transformation reduces to Fisher's z-transformation. As a generalization, the new parametrization inherits a few advantages of the z-transformation and enjoy some additional desirable properties. First and foremost, it is very flexible in the sense that when modeling q, we do not need to impose any algebraic constraint. This suggests that we can consider any reasonable dynamics for q without worrying about the positive-definiteness of the resulting correlation matrix, which is attractive in the stochastic volatility model. Second, compared with original elements in correlation matrix C, the sampling distribution of elements in transformed vector q usually appears to be closer to a Gaussian distribution. Hence, it is reasonable to model q through a Gaussian autoregressive process. Third, this transformation is invariant to the reordering of variables, in contrast to those approaches based on the Cholesky decomposition. Fourth, although elements of q depend on C in a nonlinear way, some interesting properties carry over to $G = \log(C)$, notably the equicorrelation and block-equicorrelation structure, see Archakov et al. (2020).

---

[2]The algorithm has been implemented in Matlab, Python, Ox, and the corresponding computer code can be found in the online appendix of Archakov and Hansen (2018b). We thank the authors for making it available to the public.

For the sake of notational simplicity, in the rest of the paper, we refer to the mapping $vecl(\log(\cdot))$ as $F(\cdot)$ and the corresponding inverse mapping as $F^{-1}(\cdot)$.

3.1.2 A Multivariate Stochastic Volatility Model with Generalized Fisher Transformation

We specify our new MSV model in this section. Let $r_t = (r_{1t}, ..., r_{pt})'$ denote the $p \times 1$ vector of asset returns and $h_t = (h_{1t}, ..., h_{pt})'$ the corresponding vector of latent log-volatilities at time t. We denote the vector of latent variables at time t that underlie the correlations by $q_t = (q_{1t}, ..., q_{dt})'$, where $d = \frac{p \times (p-1)}{2}$ by construction. This vector q_t is connected to correlation matrix R_t through the transformation detailed in Section 3.1.1. Then, the basic model, which we refer to as MSV-GFT, is given by

$$r_t = V_t^{1/2} \epsilon_t, \text{ where } \epsilon_t \sim N(0, R_t), \ t = 1, ..., T, \tag{3.1}$$

$$V_t = \exp(H_t), \text{ and } h_t = diag(H_t), \ t = 1, ..., T, \tag{3.2}$$

$$q_t = F(R_t), \ t = 1, ..., T, \tag{3.3}$$

$$h_{t+1} = \mu_h + \Phi_h(h_t - \mu_h) + \eta_{ht}, \text{ where } \eta_{ht} \sim N(0, \Sigma_h), \ t = 1, ..., T-1, \tag{3.4}$$

$$q_{t+1} = \mu_q + \Phi_q(q_t - \mu_q) + \eta_{qt}, \text{ where } \eta_{qt} \sim N(0, \Sigma_q), \ t = 1, ..., T-1, \tag{3.5}$$

$$h_0 \sim N(\mu_h, (I_p - \Phi_h^2)^{-1}\Sigma_h), \text{ and } q_0 \sim N(\mu_q, (I_d - \Phi_q^2)^{-1}\Sigma_q), \tag{3.6}$$

where $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})'$, $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hpt})'$, $\eta_{qt} = (\eta_{q1t}, ..., \eta_{qdt})'$, $\mu_h = (\mu_{h1}, ..., \mu_{hp})'$ and $\mu_q = (\mu_{q1}, ..., \mu_{qd})'$. We assume that both Φ_h and Φ_q are diagonal with the diagonal elements being $\phi_h = (\phi_{h1}, ..., \phi_{hp})$ and $\phi_q = (\phi_{q1}, ..., \phi_{qd})$, respectively. Σ_h and Σ_q are also supposed to be diagonal with the diagonal elements being $\sigma_h^2 = (\sigma_{h1}^2, ..., \sigma_{hp}^2)$ and $\sigma_q^2 = (\sigma_{q1}^2, ..., \sigma_{qd}^2)$. Here h_t is a p-dimensional latent process that determines the volatilities via the exponential transformation and q_t is a d-dimensional latent process that determines the correlation coefficients via the F transformation. All the latent variables (elements in h_t and q_t) are assumed to be generated by independent AR(1) models.

Note that in the MSV-GFT model, persistence in the elements of q_t is allowed to vary. This is in sharp contrast to models based on Engle's DCC framework or the Wishart autoregression, where the persistence of all the correlation sequences is assumed to be the same. While the 'equi-persistence makes the model specification more parsimonious, the empirical validity of this assumption has to be verified. To empirically examine the relevance of our generalization, we also consider a restricted model in this chapter. The restricted model is obtained from the general model by assuming that the sequences in the elements of q_t share the same persistence and variance, that is, $\phi_{q1} = ... = \phi_{qd}$ and $\sigma_{q1}^2 = ... = \sigma_{qd}^2$. As can be easily seen, this structure is similar to the equicorrelation model proposed by Engle and Kelly (2012), with a minor difference that we allow the variation in the mean of each correlation sequence.

3.2   Particle Filter and Markov chain Monte Carlo

Due to the difficulty of evaluating likelihood, the literature on MSV models carries out statistical inference under a Bayesian framework. Jacquier et al. (1994) introduced the single-move Gibbs sampler, in which all the latent variables are sampled using the full conditional distributions one by one. Meyer and Yu (2000) explained how to implement the single-move Gibbs sampler in a popular Bayesian software WinBUGS. Kim et al. (1998) showed that this approach could be quite inefficient and produce highly autocorrelated MCMC draws. To improve efficiency, the multi-move sampler proposed in Shephard and Pitt (1997) is employed by many papers. See also the discussion in Watanabe and Omori (2004) and Omori and Watanabe (2008). In this chapter, we take advantage of a recently proposed technique known as PMCMC, which builds an efficient, high-dimensional MCMC kernel. Specifically, we use an improved Particle Gibbs sampler that enjoys the good mixing property even with a small number of particles. Details of our estimation procedure are discussed in this section.

3.2.1 Review of Particle Filter

Before explaining PMCMC, we first briefly introduce the concept of particle filter. Consider a general non-linear state-space model given by

$$y_t | x_t = x \sim f_\theta(\cdot | x), \tag{3.7}$$

$$x_{t+1} | x_t = x \sim g_\theta(\cdot | x), \text{ and } x_1 \sim \mu_\theta(\cdot), \tag{3.8}$$

where y_t is the observable variable, x_t is the latent variable, and θ contains all the parameters of interest. In this subsection, we assume for the moment that θ is given and our target is to infer the latent variables $(x_1, ... x_t)$ using the observations $(y_1, ..., y_t)$. Specifically, we wish to obtain an estimator of the conditional distribution $p_\theta(x_1, ..., x_t | y_1, ..., y_t)$. A closed-form analytical solution for this problem is usually not available, except in very special cases such as in the Gaussian linear models, where the Kalman filter is applicable. Approximation must be relied upon in general. Particle filter, also known as sequential Monte Carlo (SMC) in the literature, is exactly the tool to apply in such a case. The only requirements for the validity of particle filter are that (i) the measurement density $f_\theta(\cdot | x)$ can be numerically evaluated; (ii) one can simulate from the transition density $g_\theta(\cdot | x)$.

The methodology of particle filter combines importance sampling and Monte Carlo simulations to approximate the target distribution. The key idea is to represent the distribution by a set of random samples with the corresponding weights and calculate the quantity of interest based on these samples and weights. To fix the idea, let $\{x_{1:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$ be a random measure, where $\{x_{0:t}^{(i)}, i = 1, ..., N\}$ is a set of support points and $\{w_t^{(i)}, i = 1, ..., N\}$ are the associated weights. Here, we use $x_{1:t} = \{x_j, j = 1, ..., t\}$ to denote the set of all states up to time t. Each point is called a particle, and N is the number of particles used. The approximate distribution can then be written as

$$\hat{p}_\theta(dx_{1:t} | y_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{x_{1:t}^{(i)}}(dx_{1:t}), \tag{3.9}$$

where $y_{1:t}$ is similarly defined and $\delta(\cdot)$ is the Dirac function. $\hat{p}_\theta$ is a discrete weighted

approximation to the target distribution p_θ. Apparently, the accuracy of the approximation can be improved as an increasing number of particles are included. Doing so, however, also dramatically raises the computational burden.

To obtain the weights, we resort to importance sampling. That is to say, we sample N times from a candidate distribution, say $q_\theta(x_{1:t}|y_{1:t})$, and assign the weight

$$w_t^{(i)} \propto p_\theta(x_{1:t}^{(i)}|y_{1:t})/q_\theta(x_{1:t}^{(i)}|y_{1:t})$$

to each sample drawn. In practice, it is hard, if not impossible, to pick up a proper importance density for the joint distribution of $x_{1:t}$ conditional on the data when the sample size is large. Hence, this approach usually proceeds in a sequential fashion. Specifically, the importance density is chosen to admit the factorization such that

$$q_\theta(x_{1:t}|y_{1:t}) = q_\theta(x_t|x_{t-1}, y_t)q_\theta(x_{1:t-1}|y_{1:t-1}).$$

For any existing weighted sample $\{x_{1:t-1}^{(i)}, w_{1:t-1}^{(i)}\}$ that follows from $p_\theta(x_{1:t-1}|y_{1:t-1})$, we augment it with the new state $x_t^{(i)}$ randomly drawn from $q_\theta(x_t|x_{t-1}, y_t)$. The joint sample, $(x_{t-1}^{(i)}, x_t^{(i)})$ is then a realization from the targeted joint importance density. The corresponding weight for i^{th} sample can easily be updated through

$$\tilde{w}_t^{(i)} \propto w_{t-1}^{(i)} \frac{f_\theta(y_t|x_t^{(i)})g_\theta(x_t^{(i)}|x_{t-1}^{(i)})}{q_\theta(x_t^{(i)}|x_{t-1}^{(i)}, y_t)},$$

and normalized to be $w_t^{(i)} = \frac{1}{N}\sum_{i=1}^{N}\tilde{w}_t^{(i)}$. An unavoidable problem of this procedure, known as degeneracy, is that after a few iterations, only one particle has a non-negligible weight, which means a large computational cost is spent on particles with almost no contribution. To alleviate this problem, a resampling step is necessary.

An important by-product of this filtering strategy is an approximation to $p_\theta(y_{1:t}|y_{1:t-1})$, which has a simple formula

$$\hat{p}_\theta(y_{1:t}|y_{1:t-1}) = \frac{1}{N}\sum_{i=1}^{N} w_t^{(i)}.$$

The joint likelihood can then be easily obtained as

$$\hat{p}_\theta(y_{1:T}) = \prod_{t=2}^{T} \hat{p}_\theta(y_{1:t}|y_{1:t-1}) \cdot \hat{p}_\theta(y_1).$$

Unlike the Kalman filter, where the exact likelihood is available through recursive evaluations of closed-form functions, the likelihood computed using particle filter only estimates the true likelihood. Thus, it is subject to the approximation error arising from random sampling.

Despite its general applicability, when implementing the particle filter for a particular model, many subtle issues must be considered. These include how to choose a proper importance density $q_\theta(x_t|x_{t-1}, y_t)$, how many particles to use, and whether a resampling step should be added. For a thorough discussion, see Arulampalam et al. (2002) and Johansen and Doucet (2008).

3.2.2 Particle Gibbs Sampler and Ancestor Sampling

In practice, model parameters are unknown. In the Bayesian framework with MCMC sampling, model parameters must be drawn together with the latent variables. To sample from the joint density $p(\theta, x_{1:T}|y_{1:T})$, typically we proceed by running a Gibbs sampler, which means drawing alternately from the two conditional densities, namely $p(\theta|x_{1:T}, y_{1:T})$ and $p(x_{1:T}|y_{1:T}, \theta)$. Usually, the former is easy to sample, either by imposing conjugate priors or through the Metropolis-Hasting algorithm. The latter, on the other hand, can be handled by the particle filter approach introduced above. This is the basic idea of the particle Gibbs (PG) sampler proposed in Andrieu et al. (2010).

One subtlety of this algorithm, as indicated in Andrieu et al. (2010), is that to ensure the targeted joint distribution is indeed the invariant distribution of a Markov chain, we have to modify the sequential Monte Carlo step. Specifically, one particle trajectory must be specified a prior, and it serves as a reference trajectory. Therefore, after a complete pass of the particle filter is run, a trajectory $x_{1:T}^{(i)}$ is randomly picked with probability proportional to the corresponding weight $w_T^{(i)}$.

22

For the next MCMC iteration, when running particle filter, we only draw $N-1$ particles, and the Nth particle is fixed at the chosen one from the last iteration. The intuition of such modification is that this path can guide the simulated particles to move within a relevant region of the state space. For a formal justification, see Theorem 5 of Andrieu et al. (2010).

A particle-filter-based MCMC procedure has a few desirable properties, which make it preferred compared with traditional methods. First and foremost, a significant improvement is in terms of efficiency. Traditional single-move sampler is well known to be quite inefficient, as it usually leads to highly autocorrelated samples across MCMC iterations. Such strong dependency implies that one has to draw a vast number of samples to achieve satisfactory accuracy. As suggested by the simulation studies below, the PG method we use significantly reduces the sample autocorrelation and thereby is much more efficient computationally. Another alternative that could also alleviate the inefficiency in the single-move sampler is various multi-move approaches. Those methods, however, in most cases, require the derivation of the second-order approximation, which could be tedious and difficult for multivariate non-linear models, including the one considered in this chapter. Designing a PG sampler, on the contrary, requires a minimal modification across different models, as long as they could be cast into a state-space form.

Second, one can gain a lot from adopting sequential Monte Carlo when performing model comparison. As will be discussed in Section 3.2.4, two popular approaches have been used in practice to compare competing Bayesian models. The first one is based on the Bayes factor and the second one is based on the Deviance Information Criterion (DIC). If either single-move or multi-move sampler is used, both the Bayes factor and DIC are more difficult to compute from the algorithms based on the single-move sampler and the multi-move sampler than from the proposed PG sampler. The reason is that in the single-move sampler and the multi-move sampler, the latent variables are treated as parameters due to the use of the data augmentation technique of Tanner and Wong (1987). As a result, the likelihood function at period t is defined as $p(y_t|x_t, \theta)$. Whereas, in the proposed PG sampler, the likeli-

hood function at period t is defined as $p(y_t|\theta)$. The computation of both the Bayes factor and DIC requires $p(y_t|\theta)$, not $p(y_t|x_t, \theta)$. While the single-move sampler and the multi-move sampler can generate MCMC draws of $p(y_t|x_t, \theta)$ as a byproduct, x_t has to be integrated out from $p(y_t|x_t, \theta)$ to obtain MCMC draws of $p(y_t|\theta)$. This step is computationally expensive although the particle filtering technique can be used here. On the other hand, the PG sampler obtains an estimated $p(y_t|\theta)$ as a by-product, facilitating the calculation of the Bayes factor and DIC.

Though theoretically correct, PG has been shown to perform quite poorly when the underlying SMC sampler suffers from path degeneracy. As observed in Lindsten et al. (2014) and Chopin and Singh (2015), the mixing of the Markov kernel induced by PG is rather slow under those circumstances. What makes things worse is that for the high-dimensional problem, such as the one we consider in this study, path degeneracy is inevitable. To overcome this severe drawback, Lindsten et al. (2014) proposes a new method, which includes an additional step called ancestor sampling. While this is a small modification, the new particle Gibbs with ancestor sampling (PGAS) enjoys fast mixing of the Markov kernel even when only a seemingly small number of particles are used in the underlying SMC. Informally, in the original PG, when degeneracy occurs, the particle system collapses toward the chosen reference trajectory. Whereas, in the PGAS, it degenerates toward something entirely different. As a consequence, the update rates of latent variables are much higher with the additional ancestor sampling step. Therefore, the mixing is much faster. Lindsten et al. (2014) also show that for a state-space model, PGAS is probabilistically equivalent to the particle Gibbs sampler with a backward smoothing step under certain conditions.

For our purpose, a fast mixing under a small number of particles is highly desirable, as our likelihood function contains a component that has no closed-form solution and thus must be computed numerically. Although the cost for one-time computation is relatively low, it soon becomes infeasible when a vast number of particles are included in the system. Indeed, for MCMC with S iterations, if the sample size is T and N particles are used, $F^{-1}(\cdot)$ must be evaluated $S \times T \times N$

times. As S and T are usually quite large in the empirical application, we can gain a lot in terms of computational efficiency by choosing the PGAS approach. Due to the same consideration, the particle Metropolis-Hasting approach is not chosen either as it requires an accurate calculation of $f(y_t|\theta)$ and a large number of particles. In summary, we think that PGAS could be a suitable estimation tool given our model setup. Its performance will be further examined through extensive simulation studies reported later.

3.2.3 Bayesian Analysis of MSV-GFT

After introducing the critical component of our estimation strategy, we can now present our particle-filter-based MCMC sampling algorithm for a Bayesian analysis of the MSV-GFT model. As any other Bayesian methods, we first need to specify the prior distributions of all the parameters $\theta \equiv (\mu_h, \mu_q, \phi_h, \phi_q, \sigma_h^2, \sigma_q^2)$. In this regard, our specification follows those adopted in the literature, especially in Kim et al. (1998). For the prior distribution of μ_h and μ_q, we assume independent multivariate normal distributions. The persistence parameters ϕ_h and ϕ_q are assumed to have Beta priors. The prior distribution of σ_h and σ_q are chosen to be inverse gamma. In summary, we choose the following prior distributions:

- $\mu_{hi} \sim N(m_{\mu 0}, s_{\mu 0}^2)$ and $\mu_{qj} \sim N(m_{\mu 0}, s_{\mu 0}^2)$;

- $\frac{\phi_{hi}+1}{2} \sim Beta(a,b)$ and $\frac{\phi_{qj}+1}{2} \sim Beta(a,b)$;

- $\sigma_{hi}^2 \sim IG(\frac{n_{m0}}{2}, \frac{d_{m0}}{2})$ and $\sigma_{hi}^2 \sim IG(\frac{n_{m0}}{2}, \frac{d_{m0}}{2})$,

for $i = 1, ..., p$ and $j = 1, ..., d$ and $m_{\mu 0}, s_{\mu 0}^2, a, b, n_{m0}, d_{m0}$ are hyperparameters.

Let $r = (r1', ..., r_T')$, $h = (h_1', ..., h_T')$ and $q = (q_1', ..., q_T')$. To carry out the inference, we implement a Gibbs sampler with four blocks. In the following, we use $\theta_{/\alpha}$ to denote the parameters θ excluding α. Then, the algorithm proceeds as:

1. Initialize h, q and θ.

2. Draw $h, q|r, \theta$.

3. Draw $\mu_h, \mu_q | r, h, q, \theta_{/(\mu_h, \mu_q)}$.

4. Draw $\phi_h, \phi_q | r, h, q, \theta_{/(\phi_h, \phi_q)}$.

5. Draw $\sigma_h^2, \sigma_q^2 | r, h, q, \theta_{/(\sigma_h^2, \sigma_q^2)}$.

Iteration over steps 2-5 consists of a complete sweep of MCMC sampler. The joint posterior distribution of our model can be written as

$$
\begin{aligned}
p(\theta, h, q | r) &\propto p(r|\theta, h, q)p(\theta, h, q) \\
&= f(r|h, q)g_\theta(h)g_\theta(q)\pi(\theta) \\
&= f(r_1|h_1, q_1)g_\theta(h_1)g_\theta(q_1) \prod_{t=2}^{T} \left[f(r_t|h_t, q_t)g_\theta(h_t|h_{t-1})g_\theta(q_t|q_{t-1}) \right] \pi(\theta) \\
&= \prod_{t=1}^{T} \left[\left(\sum_{i=1}^{p} h_{it} \right) |R_t|^{-1/2} \exp\left[-\frac{1}{2} r_t'(V_t^{1/2} R_t V_t^{1/2})^{-1} r_t \right] \right] \\
&\quad \times \prod_{t=2}^{T} \prod_{i=1}^{p} \left[(\sigma_{hi}^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_{hi}^2} (h_{it+1} - \mu_{hi} - \phi_{hi}(h_{it} - \mu_{hi}))^2 \right) \right] \\
&\quad \times \prod_{t=2}^{T} \prod_{j=1}^{d} \left[(\sigma_{qj}^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_{qj}^2} (q_{jt+1} - \mu_{qj} - \phi_{qj}(q_{jt} - \mu_{qj}))^2 \right) \right] \\
&\quad \times \prod_{i=1}^{p} \left(\frac{\sigma_{hi}^2}{1 - \phi_{hi}^2} \right)^{-1/2} \exp\left(-\frac{(h_{i1} - \mu_{h1})^2}{2\sigma_{hi}^2/(1 - \phi_{hi}^2)} \right) \\
&\quad \times \prod_{j=1}^{d} \left(\frac{\sigma_{qj}^2}{1 - \phi_{qj}^2} \right)^{-1/2} \exp\left(-\frac{(q_{j1} - \mu_{q1})^2}{2\sigma_{qj}^2/(1 - \phi_{qj}^2)} \right) \times \pi(\theta).
\end{aligned}
$$

$$(3.10)$$

We apply the PGAS introduced in the last subsection to sample the latent variables h and q given all the observations r and one particular set of parameter values. The detailed description of the algorithm is presented in Appendix A.1. On the other hand, from the joint posterior density, it is straightforward to sample each parameter in θ given one particular realization of latent processes h and q. Specifically, we can do the following:

1. We can directly sample from the full conditional posterior of μ_{hi} and μ_{qi} from

a normal distribution. For $i = 1, ..., p$ and $j = 1, ..., d$,

$$\mu_{hi}|r, h, q, \theta_{/\mu_{hi}} \sim N(\tilde{m}_{h\mu}, \tilde{s}^2_{h\mu}) \text{ and } \mu_{qj}|r, h, q, \theta_{/\mu_{qj}} \sim N(\tilde{m}_{q\mu}, \tilde{s}^2_{q\mu}) \quad (3.11)$$

where

$$\tilde{m}_{h\mu} = \tilde{s}^2_{h\mu} \left\{ \frac{1 - \phi^2_{hi}}{\sigma^2_{hi}} h_{i1} + \frac{1 - \phi_{hi}}{\sigma^2_{hi}} \sum_{t=1}^{T-1}(h_{it+1} - \phi_{hi}h_{it}) \right\},$$

$$\tilde{m}_{q\mu} = \tilde{s}^2_{q\mu} \left\{ \frac{1 - \phi^2_{qj}}{\sigma^2_{qj}} q_{j1} + \frac{1 - \phi_{qj}}{\sigma^2_{qj}} \sum_{t=1}^{T-1}(q_{jt+1} - \phi_{qj}q_{jt}) \right\},$$

and

$$\tilde{s}^2_{h\mu} = \sigma^2_{hi} \left[(T - 1)(1 - \phi_{hi})^2 + (1 - \phi_{hi})^2 \right]^{-1},$$

$$\tilde{s}^2_{q\mu} = \sigma^2_{qj} \left[(T - 1)(1 - \phi_{qj})^2 + (1 - \phi_{qj})^2 \right]^{-1}.$$

2. To draw random samples from the full conditional posterior density of ϕ_{hi} and ϕ_{qi}, one can resort to the Metropolis-Hasting sampler. Since

$$\log p(\phi_{hi}|y, h, q, \theta_{/\phi_{hi}}) \propto \log p(h_i|\phi_{hi}, \theta_{/\phi_{hi}}) + \log \pi(\phi_{hi})$$

$$= \log \pi(\phi_{hi}) - \frac{(h_{i1} - \mu_{hi})^2(1 - \phi^2_{hi})}{2\sigma^2_{hi}} + \frac{1}{2}\log(1 + \phi^2_{hi})$$

$$- \frac{\sum_{t=1}^{T-1} \left[(h_{it+1} - \mu_{hi}) - \phi_{hi}(h_{it} - \mu_{hi}) \right]^2}{2\sigma^2_{hi}},$$

$$(3.12)$$

we draw ϕ^*_{hi} from the proposal normal density $N(\hat{\phi}_{hi}, V_{\phi_{hi}})$, where

$$\hat{\phi}_{hi} = \left[\sum_{t=1}^{T-1}(h_{it+1} - \mu_{hi})(h_{it} - \mu_{hi}) \right] \bigg/ \left[\sum_{t=1}^{T-1}(h_{it} - \mu_{hi})^2 \right],$$

is the ordinary least square estimator of ϕ_{hi} given h_i and

$$V_{\phi_{hi}} = \sigma^2_{hi} \left[\sum_{t=1}^{T-1}(h_{it} - \mu_{hi})^2 \right]^{-1}.$$

Then, the draw is accepted with probability $\min \left[1, \exp\left\{ g(\phi^*_{hi})/g(\phi^{(i-1)}_{hi}) \right\} \right]$,

where $\phi_{hi}^{(i-1)}$ is the sample from last MCMC iteration and

$$g(\phi_{hi}) = \log \pi(\phi_{hi}) - \frac{(h_{i1} - \mu_{hi})^2(1 - \phi_{hi}^2)}{2\sigma_{hi}^2} + \frac{1}{2}\log(1 + \phi_{hi}^2).$$

ϕ_{qi} can be treated in the same fashion.

3. Similar to the case for μ, due to the conjugacy, draws of σ_{hi}^2 can come from an inverse gamma distribution. For $i = 1, ..., p$ and $j = 1, ..., d$,

$$\sigma_{hi}^2 | y, h, q, \theta_{/\sigma_{hi}^2} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{d}_{hm}}{2}\right) \text{ and } \sigma_{qj}^2 | y, h, q, \theta_{/\sigma_{qj}^2} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{d}_{qm}}{2}\right)$$

$$(3.13)$$

where $\tilde{n}_m = n_{m0} + T$ and

$$\tilde{d}_{hm} = d_{m0} + (h_{i1} - \mu_{hi})^2(1 - \phi_{hi}^2) + \sum_{t=1}^{T-1} \left[(h_{it+1} - \mu_{hi}) - \phi_{hi}(h_{it} - \mu_{hi})\right]^2,$$

$$\tilde{d}_{qm} = d_{m0} + (q_{j1} - \mu_{qj})^2(1 - \phi_{qj}^2) + \sum_{t=1}^{T-1} \left[(q_{jt+1} - \mu_{qj}) - \phi_{qj}(q_{jt} - \mu_{qj})\right]^2.$$

3.2.4 Model Comparison

To investigate the performance of the proposed model against alternative model specifications, we compare competing models based on DIC of Spiegelhalter et al. (2002) and the Bayes factor of Kass and Raftery (1995).

DIC is a Bayesian version of AIC. As shown in Li et al. (2017), the justification of DIC can be made in a similar way to that of AIC. As AIC, DIC consists of two terms, that is,

$$\mathrm{D}IC = E_{\theta|r}[D(\theta)] + P_D \tag{3.14}$$

where $D(\theta) = -2\log p(r|\theta)$ and $P_D = E_{\theta|r}[D(\theta)] - D(E_{\theta|r}[\theta])$. The first term measures the goodness of fit of the model while the second term measures the complexity of the model. Due to the use of the second term which penalizes the increasing size

of a model, DIC can deal with the problem of over-fitting. The smaller the DIC value, the better the model.

To compute DIC based on MCMC output, the log-likelihood component in $D(\theta)$ can be approximated by $\log(\hat{p}(r|\theta))$. Note that $\hat{p}(r|\theta)$ is generated from PGAS. By the law of large numbers for ergodic processes, $\frac{1}{B}\sum_{i=1}^{B} D(\theta^i) \xrightarrow{p} E_{\theta|r}[D(\theta)]$ as $B \to \infty$ where B is the number of MCMC iterations and θ^i is i^{th} MCMC draw from the posterior distribution, $p(\theta|r)$. The second term P_D depends on $D(E_{\theta|r}[\theta])$, which can be approximated by $D(\bar{\theta})$, where $\bar{\theta}$ is the posterior mean. Given that $\hat{p}(r|\theta)$ is generated from PGAS, $D(\bar{\theta})$ is easy to compute. As PGAS estimates the log-likelihood $\log p(r|\bar{\theta})$, the numeric standard error affects the precision of DIC. Following Ishihara and Omori (2012), we repeat PGAS with 10000 particle numbers for ten times. At each time, we obtain the log-likelihood value at $\bar{\theta}$. We then use the average of these log-likelihood values to estimate $\log p(r|\bar{\theta})$.

It is attempting to calculate DIC from $D(\theta) = -2\log p(r|h,q,\theta)$ and $P_D = E_{\theta,h,q|r}[D(\theta)] - D(E_{\theta,h,q|r}[\theta])$ for models with latent variables because $\log p(r|h,q,\theta)$ is numerically more tractable than $\log p(r|\theta)$ for most models with latent variables. In $\log p(r|h,q,\theta)$ all the latent variables, including h and q, are treated as parameters. This treatment makes the calculation of DIC based on the single-move and the multi-move algorithms easy to implement. This method of calculating DIC is proposed in Spiegelhalter et al. (2002) and implemented in WinBUGS when there are latent variables in a model. Unfortunately, as pointed out in **?**, this method lacks of theoretical justification because when the latent variables are incidental parameters which cannot be consistently estimated. As $\hat{p}(r|\theta)$ is generated by PGAS but not by the single-move and the multi-move algorithms, PGAS makes the calculation of DIC, the version based on $\log p(r|\theta)$, straightforward.

The Bayes factor is an alternative way to compare the competing models. Like DIC, the Bayes factor can also deal with over-fitting. Assume that there are two competing models, M_0 and M_1, which may be nested or not nested. The idea of the Bayes factor comes from the posterior odds defined by $p(M_0|r)/p(M_1|r)$. By the

Bayes theorem, we have

$$\frac{p(M_0|r)}{p(M_1|r)} = \frac{p(r|M_0)}{p(r|M_1)} \times \frac{p(M_0)}{p(M_1)} := BF_{01} \times \frac{p(M_0)}{p(M_1)},$$

where $p(M_0), p(M_1$ represent the prior model probabilities, $BF_{01} = \frac{p(r|M_0)}{p(r|M_1)}$, the Bayes factor, is the ratio of the two marginal likelihoods, $p(r|M_0)$ and $p(r|M_1)$. If we assume the equal prior model probabilities as is typically done, then the posterior odds is the same as the Bayes factor.

From the discussion above, to calculate the Bayes factor, one needs to calculate the two marginal likelihood values of two competing models. To calculate the log Bayes factor, one needs to calculate the difference between the two log marginal likelihood values. In general marginal likelihood conducts integrations over the entire parameter space, that is, for any model M_i,

$$p(r|M_i) = \int p(r, \theta|M_i)d\theta = \int p(r|\theta, M_i)p(\theta|M_i)d\theta.$$

When the parameter space is of high dimension and the integral is not analytically available, the computational cost in evaluating marginal likelihood can be very high. In the literature, several MCMC-based approaches are available to compute the marginal likelihood numerically from MCMC output, including Chib (1995), Chib and Jeliazkov (2001), Friel and Pettitt (2008), and Li et al. (2020).

To avoid numeric integrations, for any model M_i, Chib (1995) suggests evaluating the log marginal likelihood as,

$$\log p(r|M_i) = \log p(r|\bar{\theta}M_i) + \log p(\bar{\theta}|M_i) - \log p(\bar{\theta}|r, M_i), \qquad (3.15)$$

where $\log(p(r|\bar{\theta}, M_i))$ is the log likelihood of model i, $\log p(\bar{\theta}, M_i)$ is the log prior density for parameters in model i, $\log p(\bar{\theta}|r, M_i)$ is the log posterior density of model i, and $\bar{\theta}$ is the posterior mean of parameters in model i.

In the right-hand side of Equation (3.15), the first component $\log p(r|\bar{\theta}, M_i)$ is approximated by our proposed method. The second component is directly computed

by the prior distribution in Section 3.2.3. The posterior density is only known up to a normality constant. Follows as Kim et al. (1998), we approximate it by using the multivariate kernel density estimate. If a single-move or multi-move method is used, one would obtain $\log p(r|\bar{\theta}, h, q, M_i)$ by a by-product. Therefore, additional marginalization is required to integrated out h and q from $\log p(r|\bar{\theta}, h, q, M_i)$.

Unlike the maximum likelihood which tends to take a larger value for a model with more parameters and can lead to over-fitting, the marginal likelihood can handle the problem of over-fitting. However, unlike DIC, the penalty attached to over-fitting is implicit in the marginal likelihood. In a recent study, Fong and Holmes (2020) shows that the marginal likelihood is formally equivalent to a cross-validation method which is designed to address the problem of over-fitting.

After DIC and the log marginal likelihood are obtained for competing models (either nested or non-nested), their values can be directly compared across models. Let K be the difference between the log marginal likelihoods or the DIC values of two competing model.[3] When $K > 0$, it means that M_0 is more strongly supported by the data than M_1. Jeffreys (1998) gives a scale for interpretation of K. If $K < 0$, there is negative evidence for M_0. If $K \in (0, 1.15)$, there is barely worth mentioning about the difference between the two models. If $K \in (1.15, 2.30)$, there is substantial evidence to support M_1. If $K \in (2.30, 3.45)$, there is strong evidence to support M_1. If $K \in (3.45, 4.60)$, there is very strong evidence to support M_1. If $K > 4.60$, there is decisive evidence to support M_1.

3.2.5 Simulation Studies

To investigate the performance of our estimation procedure, we conduct some simulation exercises in this section. Our simulation design is frequentist as we fix the parameters at their true value and generate data from the same data generating process for 100 times. However, we use the posterior mean as a point estimator for all the parameters in the model. Since we know the true value of those parameters,

---

[3]For log marginal likelihoods, the difference is the value of M_1 minus the value of M_0. For DIC, the difference is the value of M_0 minus the value of M_1.

Table 1: Posterior statistics of μ with simulated data

	N	True Value	μ_{h1} 0.3	μ_{h2} 0.3	μ_{h3} 0.3	μ_{q1} 0.7	μ_{q2} 0.7	μ_{q3} 0.7
T=500	50	Mean	0.311	0.317	0.305	0.632	0.639	0.631
		Std	0.094	0.083	0.080	0.063	0.054	0.059
		IF	11.5	9.8	9.5	17.3	16.0	14.3
	100	Mean	0.305	0.313	0.301	0.646	0.654	0.645
		Std	0.093	0.083	0.082	0.064	0.055	0.061
		IF	8.5	7.6	7.8	16.4	15.1	12.5
	200	Mean	0.304	0.311	0.299	0.657	0.665	0.656
		Std	0.092	0.083	0.082	0.064	0.055	0.061
		IF	6.4	5.8	5.5	12.2	11.6	12.0
T=1000	50	Mean	0.316	0.315	0.322	0.646	0.651	0.648
		Std	0.055	0.051	0.048	0.037	0.038	0.034
		IF	11.5	10.8	10.3	19.6	19.1	19.5
	100	Mean	0.310	0.310	0.317	0.661	0.666	0.664
		Std	0.053	0.052	0.049	0.036	0.038	0.034
		IF	9.0	7.6	6.7	17.2	16.9	14.7
	200	Mean	0.307	0.308	0.315	0.671	0.676	0.674
		Std	0.053	0.051	0.049	0.037	0.038	0.035
		IF	6.4	6.7	5.3	14.4	13.5	15.4
T=2000	50	Mean	0.316	0.316	0.316	0.651	0.657	0.655
		Std	0.036	0.033	0.034	0.028	0.025	0.024
		IF	10.1	9.3	8.6	21.2	25.1	20.8
	100	Mean	0.312	0.311	0.312	0.666	0.672	0.671
		Std	0.036	0.033	0.035	0.028	0.025	0.024
		IF	6.8	6.4	5.9	16.9	19.6	20.2
	200	Mean	0.308	0.308	0.309	0.675	0.682	0.681
		Std	0.037	0.032	0.035	0.028	0.026	0.024
		IF	6.3	4.7	5.4	15.7	15.4	15.5

1. T is the number of observations for each asset and N is the number of particles used in PGAS.
2. Mean is the average posterior mean across replications.
3. Std is the standard error of the posterior mean across replications.
4. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

Table 2: Posterior statistics of ϕ with simulated data

	N	True Value	ϕ_{h1} 0.9	ϕ_{h2} 0.9	ϕ_{h3} 0.9	ϕ_{q1} 0.8	ϕ_{q2} 0.8	ϕ_{q3} 0.8
$T=500$	50	Mean	0.894	0.885	0.887	0.784	0.791	0.790
		Std	0.033	0.043	0.046	0.051	0.058	0.06
		IF	56.4	54.6	55.6	56.9	59.5	58.7
	100	Mean	0.898	0.891	0.892	0.794	0.801	0.800
		Std	0.034	0.041	0.045	0.055	0.060	0.063
		IF	69.1	61.4	68.6	61.3	65.9	64.4
	200	Mean	0.902	0.893	0.893	0.798	0.807	0.803
		Std	0.033	0.043	0.047	0.059	0.061	0.062
		IF	61.3	69.3	71.1	60.7	65.2	72.6
$T=1000$	50	Mean	0.899	0.889	0.895	0.785	0.787	0.781
		Std	0.032	0.036	0.029	0.063	0.051	0.055
		IF	91.5	92.6	90.9	94.5	93.3	91.8
	100	Mean	0.900	0.891	0.898	0.793	0.795	0.791
		Std	0.032	0.032	0.028	0.060	0.051	0.057
		IF	103.86	87.1	95.1	101.0	92.5	91.7
	200	Mean	0.900	0.894	0.899	0.799	0.800	0.798
		Std	0.033	0.032	0.029	0.059	0.056	0.056
		IF	102.9	102.0	92.7	103.2	107.3	110.2
$T=2000$	50	Mean	0.900	0.897	0.899	0.793	0.785	0.787
		Std	0.022	0.022	0.023	0.047	0.040	0.044
		IF	112.4	105.8	107.0	128.2	125.0	120.9
	100	Mean	0.901	0.898	0.900	0.798	0.792	0.792
		Std	0.023	0.024	0.022	0.051	0.042	0.042
		IF	114.0	106.1	100.9	135.7	144.4	136.3
	200	Mean	0.900	0.898	0.900	0.801	0.795	0.796
		Std	0.023	0.023	0.024	0.049	0.041	0.044
		IF	103.6	99.2	102.0	134.9	131.5	141.8

1. T is the number of observations for each asset and N is the number of particles used in PGAS.
2. Mean is the average posterior mean across replications.
3. Std is the standard error of the posterior mean across replications.
4. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

Table 3: Posterior statistics of σ^2 with simulated data

			σ^2_{h1}	σ^2_{h2}	σ^2_{h3}	σ^2_{q1}	σ^2_{q2}	σ^2_{q3}
	N	True Value	0.05	0.05	0.05	0.05	0.05	0.05
T=500	50	Mean	0.026	0.030	0.028	0.027	0.029	0.029
		Std	0.013	0.013	0.013	0.012	0.012	0.012
		IF	173.6	150.7	148.0	124.2	124.1	122.8
	100	Mean	0.032	0.036	0.034	0.032	0.034	0.035
		Std	0.017	0.016	0.016	0.015	0.015	0.018
		IF	170.7	134.4	143.3	124.3	133.0	125.5
	200	Mean	0.036	0.041	0.038	0.037	0.039	0.040
		Std	0.019	0.020	0.019	0.018	0.017	0.020
		IF	138.0	136.8	134.8	121.8	126.2	138.2
T=1000	50	Mean	0.030	0.030	0.029	0.026	0.028	0.026
		Std	0.015	0.011	0.010	0.009	0.010	0.010
		IF	198.0	195.1	198.5	185.6	184.5	188.3
	100	Mean	0.036	0.037	0.035	0.032	0.034	0.032
		Std	0.016	0.012	0.012	0.011	0.012	0.013
		IF	182.3	157.5	170.8	180.43	169.5	163.0
	200	Mean	0.041	0.040	0.039	0.037	0.038	0.037
		Std	0.018	0.014	0.012	0.013	0.014	0.015
		IF	170.1	164.7	153.1	169.5	175.9	188.5
T=2000	50	Mean	0.030	0.031	0.031	0.025	0.026	0.024
		Std	0.010	0.009	0.009	0.008	0.007	0.006
		IF	198.9	196.7	193.0	240.4	239.4	235.0
	100	Mean	0.036	0.038	0.037	0.032	0.033	0.031
		Std	0.011	0.011	0.010	0.010	0.010	0.008
		IF	183.2	173.3	167.5	216.5	232.4	230.0
	200	Mean	0.041	0.042	0.041	0.037	0.039	0.037
		Std	0.012	0.011	0.011	0.012	0.012	0.010
		IF	161.8	157.3	157.8	205.7	200.4	219.6

1. T is the number of observations for each asset and N is the number of particles used in PGAS.
2. Mean is the average posterior mean across replications.
3. Std is the standard error of the posterior mean across replications.
4. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

34

we are thus able to calculate bias (defined as the difference between the true value and the average value of posterior means) and the standard deviation.[4]

To evaluate the sampling efficiency, following Kim et al. (1998), we calculate the average inefficiency factor (IF), which is defined as the variance of the sample mean from MCMC sampling divided by that from a hypothetical sampler which draws independent samples. The variance of the MCMC sample mean is the square of the numerical standard error estimated by

$$NSE = 1 + \frac{2B_M}{B_M - 1} \sum_{i=1}^{B_M} K\left(\frac{i}{B_M}\right) \hat{\rho}(i),$$

where $\hat{\rho}(i)$ is estimated autocorrelation at lag i, B_M is the bandwidth and $K(\cdot)$ is the Parzen kernel. We choose the bandwidth B_M to be 1000. A smaller IF indicates a better mixing of the Markov chain and thereby a higher sampling efficiency.

The number of assets considered for simulation is three. There are 18 parameters whose true values are given by:

1. $\mu_{h1} = \mu_{h2} = \mu_{h3} = 0.3$ and $\mu_{q1} = \mu_{q2} = \mu_{q3} = 0.7$,

2. $\phi_{h1} = \phi_{h2} = \phi_{h3} = 0.9$ and $\phi_{q1} = \phi_{q2} = \phi_{q3} = 0.8$,

3. $\sigma_{h1}^2 = \sigma_{h2}^2 = \sigma_{h3}^2 = 0.05$ and $\sigma_{q1}^2 = \sigma_{q2}^2 = \sigma_{q3}^2 = 0.05$.

All the simulation results reported in this section is based on 5000 MCMC iterations, among which the first 1000 samples are discarded as burn-in period.[5] We consider three different sample sizes, namely $T = 500$, $T = 1000$ and $T = 2000$, as well as three numbers of particles, namely $N = 50$, $N = 100$ and $N = 200$. It is worthwhile to mention that, the simulated data used across different particle numbers for given sample size are the same, while it changes when the sample size increases.

The estimation results for the mean parameters μs are reported in Table 1. It can be seen that even for a small sample size such as 500 and a relatively small number

---

[4]Note that here, the standard deviation refers to the variation across replications, rather than the numerical standard error of MCMC sampler introduced below.

[5]Plot of autocorrelation function suggests that the MCMC sampling has converged after at most 1000 iterations.

of particles such as 50, the posterior means for both h and q are reasonable close to their respective true values while there is an upward bias for μ_h and a downward bias for μ_q. Given a particular number of particles used, as sample size increases, bias in the estimator for μ_q increases, but this is not the case for μ_h. Nevertheless, as expected, the standard deviations for both parameters decrease as T increases. On the other hand, an increasing number of particles can reduce bias substantially. For example, when the sample size is 2000, bias in the posterior mean of h reduces from 0.016 to 0.008 if 200 particles are used instead of 50. A similar improvement applies to μ_q. However, an increasing number of particles has no effect on the standard deviation.

Table 2 presents the simulation results related to the persistence parameters ϕs. These parameters can be estimated accurately, even with 500 observations and 50 particles. The estimates have very small biased and low standard deviations. When 200 particles are used, bias almost completely vanishes.

Table 3 presents the simulation results related to parameters σ^2s. A substantial downward bias is observed for the variance estimator of both hs and qs when 50 particles are used. This bias is insensitive to the number of observations. Fortunately, it can be improved by using more particles. Indeed, when N=200, bias becomes much smaller, although it seems that a larger number of particles are necessary to fully cancel this bias.

As for the IF, in general, it does not vary much as we change either the sample size or the number of particles. Consistent with earlier studies, the IF is the lowest for μs and the highest for σ^2s. Compared with the traditional single-move or multi-move Gibbs sampler (for example, see Kim et al. (1998)), our new PGAS sampler enjoys a much better mixing property. In summary, the simulation results confirm that our chosen approach works well for the model considered in our study. In light of the excellent performance, 200 particles are used for all of our empirical applications.

At last, we report the filtered h and q, together with the 95% credible interval and their true values, in Figure 1 and 2, respectively. These figures show that sampling

of the latent variables based on the proposed particle filter is reliable.

3.3 Empirical Analysis

In this section, we consider two empirical applications of our models. The first one is based on weekly exchange rate returns of three European currencies, while the second one is based on the daily returns of three Asian equity indexes.

In each application, we estimate the following four models:

1. MSV-GFT model: The general model that we propose in Section 3.

2. MSV-GFT model with equi-persistence: the same as the general MSV-GFT model except for the additional assumption of 'equi-persistence' on the correlation coefficients. It specify that each sequence in h and q follows as same AR(1) process with different mean.

$$
\begin{aligned}
r_t &= V_t^{1/2}\epsilon_t, \ \epsilon_t \sim N(0, R_t) \\
V_t &= \exp(H_t), \ h_t = diag(H_t), \ q_t = F(R_t) \\
h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht}, , \ \eta_{ht} \sim N(0, \sigma_h^2) \\
q_{t+1} &= \mu_q + \phi_q(q_t - \mu_q) + \eta_{qt}, , \ \eta_{qt} \sim N(0, \sigma_q^2)
\end{aligned}
\tag{3.16}
$$

where ϕ_h, ϕ_q, σ_h^2 and σ_q^2 are all scalar instead of vector.

3. MSV-CC: Model with constant correlation matrix over time, similar to the one considered in Harvey et al. (1994).

$$
\begin{aligned}
r_t &= V_t^{1/2}\epsilon_t, \ \epsilon_t \sim N(0, R_t) \\
V_t &= \exp(H_t), \ h_t = diag(H_t), \ q_t = F(R_t) \\
h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht}, , \ \eta_{ht} \sim N(0, \sigma_h^2) \\
q_{t+1} &= \mu_q
\end{aligned}
\tag{3.17}
$$

4. MSV-DCC: Model proposed in Asai and McAleer (2009), where a DCC structure with a Wishart transition dynamics is used to characterize the movement

Figure 1: True and filtered hs
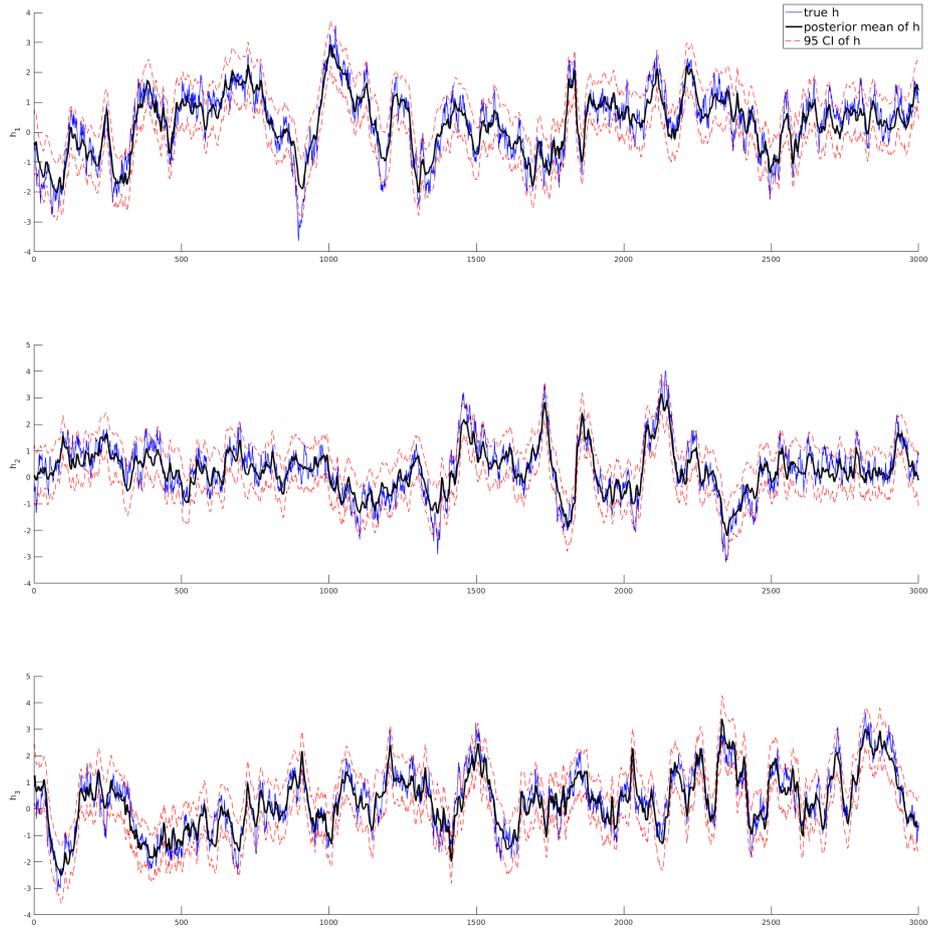
Note: Compare the true hs (blue solid) with their posterior means (black solid) and 95% credible intervals (red dash).

38

Figure 2: True and filtered qs
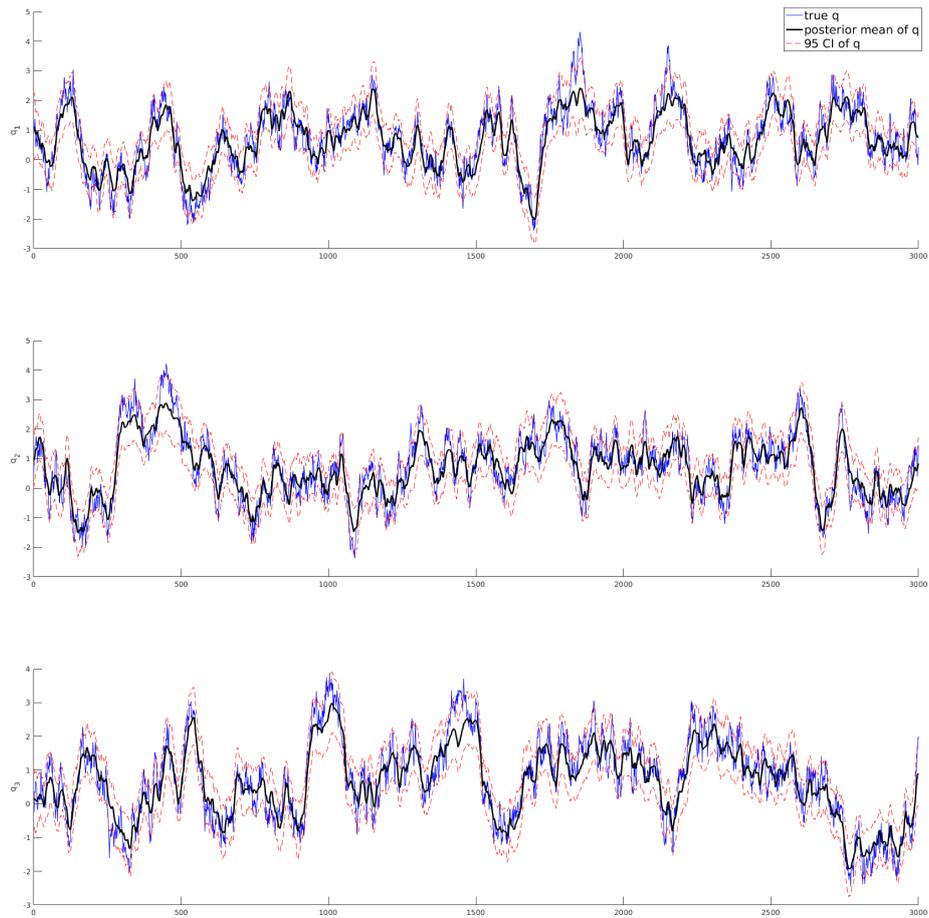
Note: Compare the true qs (blue solid) with the posterior means of q (black solid) and 95% credible intervals (red dash).

of the correlation matrix.

$$r_t = V_t^{1/2}\epsilon_t, \ \epsilon_t \sim N(0, R_t),$$

$$V_t = \exp(H_t), \ h_t = diag(H_t),$$

$$h_{t+1} = \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht}, \ \eta_h \sim N(0, \sigma_h^2), \qquad (3.18)$$

$$R_t = \tilde{Q}_t^{-1} Q_t \tilde{Q}_t^{-1},$$

$$Q_{t+1}^{-1} | k, Q_t^{-1} \sim Wishart\left(k, \frac{1}{k} Q_t^{-\phi/2} \Lambda Q_t^{-\phi/2}\right),$$

$$\Lambda = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix},$$

where unknown parameters are $(\mu_h, \phi_h, \sigma_h^2, k, \phi, \Lambda)$.

In all models, the parametrization of the volatility movement is the same. Hence, we focus on alternative specifications about the dynamics of the correlations. To estimate and compare these models in a unified framework, we treat all four models as nonlinear state-space models and use the proposed PGAS algorithm to estimate them. The details of the estimation procedure and the computation of the log marginal likelihood are presented in the appendix.

3.3.1 Weekly Foreign Exchange Rates

In the first empirical application, the data contains 1406 weekly mean-corrected log-returns of Euro, Pound sterling, and Swiss franc exchange rates, all against the US dollar, from January 13, 1993 to December 25, 2019. The three series are expected to be correlated, as the underlying economies are closely connected. The three time series are plotted in Figure 3. For the MSV-GFT model, the filtered h sequences of all three exchange rate and the filtered q sequences for all three pairs are plotted in Figure 4 and Figure 5, respectively.

Table 4 reports the posterior statistics of parameters related to volatility for all four competing models, including the posterior mean, the posterior standard deviation, and the 95% credible interval. As expected, all h sequences have a very

high level of persistence, with the autoregressive root close to 1.

Table 5 reports the posterior statistics of parameters related to correlation for all four competing models. In the second model, since the persistence levels and the standard deviations are restricted to be the same among all q sequences, we have only one ϕ_q and one σ_q^2 to estimate. In the third model, no ϕ_q and one σ_q^2 is in presence as all correlations are assumed to be constant over time. An important finding from Table 5 is that, when the three q sequences are allowed to have separate dynamics, q_1 has a very low level of persistence and hence, is highly stationary. Moreover, the posterior mean of ϕ_q is $(0.212, 0.911, 0.812)$ in this model. This finding is in sharp contrast to that in the MSV-GFT model with equi-persistence where only one parameter is assumed to govern all the latent variables. In the latter case, the posterior mean of ϕ_q is 0.957. To see whether this additional flexibility leads to any statistical improvement, we compare the log marginal likelihoods and DIC values of all four models. The results are reported in the last row of Table 5. The fully flexible model has a much higher value for the log marginal likelihood and much lower value for DIC than all the competing models. For example, the difference between the log marginal likelihood value in the fully flexible model and each of the competing models is 190.9, 123.3 and 64.6, indicating the decisive evidence to support the fully flexible model. We can then conclude that adopting the generalized Fisher transformation is indeed beneficial for modeling the multivariate exchange rate returns.

Figure 3: Time series of exchange rate returns

Table 4: Posterior statistics of parameters in the h sequences for competing models based on the exchange rate data

		MSV-GFT	MSV-GFT (Equi-persist)	MSV-CC	MSV-DCC
μ_{h1}	Mean	-1.701	-1.68	-1.872	-1.764
	SD	0.120	0.108	0.182	0.098
	95%CI	[-1.963,-1.470]	[-1.904,-1.468]	[-2.231,-1.507]	[-1.964,-1.575]
	IF	6.5224	12.154	14.924	17.084
μ_{h2}	Mean	-1.804	-1.774	-1.864	-1.893
	SD	0.101	0.099	0.142	0.097
	95%CI	[-2.005,-1.599]	[-1.973,-1.582]	[-2.132,-1.565]	[-2.034,-1.655]
	IF	5.5755	12.67	15.559	15.577
μ_{h3}	Mean	-1.504	-1.481	-1.688	-1.578
	SD	0.106	0.093	0.131	0.083
	95%CI	[-1.733,-1.317]	[-1.672,-1.311]	[-1.939,-1.406]	[-1.752,-1.422]
	IF	6.4733	13.139	17.795	17.456
ϕ_{h1}	Mean	0.980	0.977	0.98	0.972
	SD	0.016	0.016	0.01	0.015
	95%CI	[0.958,0.994]	[0.952,0.993]	[0.962,0.992]	[0.939,0.990]
	IF	55.044	44.04	47.832	49.084
ϕ_{h2}	Mean	0.962	0.960	0.966	0.970
	SD	0.019	0.031	0.015	0.036
	95%CI	[0.921,0.987]	[0.849,0.989]	[0.929,0.987]	[0.843,0.990]
	IF	132.17	90.76	111.05	71.137
ϕ_{h3}	Mean	0.970	0.957	0.960	0.962
	SD	0.027	0.037	0.014	0.036
	95%CI	[0.927,0.991]	[0.888,0.988]	[0.928,0.982]	[0.847,0.987]
	IF	83.019	131.2	101.49	92.167
σ_{h1}^2	Mean	0.005	0.004	0.01	0.005
	SD	0.005	0.004	0.006	0.005
	95%CI	[0.002,0.008]	[0.002,0.008]	[0.005,0.020]	[0.002,0.012]
	IF	233.94	204.19	140.9	212.09
σ_{h2}^2	Mean	0.013	0.013	0.017	0.006
	SD	0.008	0.016	0.009	0.016
	95%CI	[0.004,0.028]	[0.004,0.080]	[0.008,0.037]	[0.002,0.066]
	IF	228.46	246.53	207.7	195.5
σ_{h3}^2	Mean	0.006	0.008	0.022	0.005
	SD	0.007	0.008	0.009	0.012
	95%CI	[0.002,0.021]	[0.003,0.017]	[0.008,0.037]	[0.002,0.032]
	IF	198.83	251.07	197.85	380.74

1. Mean is the posterior mean based on 20000 MCMC samples after a 2000 burn-in period.
2. SD is the numerical standard errors of the posterior means.
3. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws.
4. IF is the inefficiency factor.

Figure 4: Filtered h sequences in MSV-GFT for the exchange rate returns
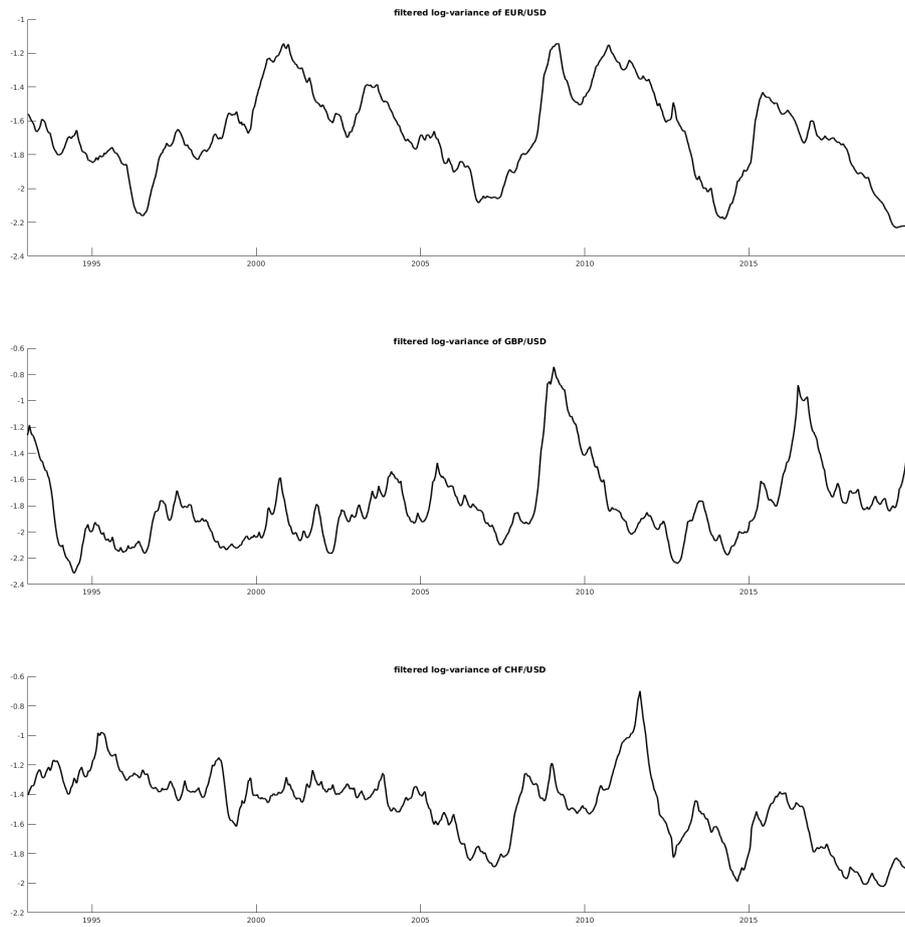
3.3.2 Daily Stock Market Indices

Data in our second empirical application contains 2237 daily mean-corrected log-returns of three stock market indices, namely the Hong Kong Hang Seng Index, the Nikkei 225 Index of Tokyo Stock Exchange, and the SSE Composite Index of Shanghai Stock Exchange. The sample period is from January 5, 2005 to December 30, 2014 that covers the global financial crisis period. This exercise is interesting, as it can provide some insights on the co-movement of these three arguably most important stock market indices in Asia. The three series are plotted in Figure 6. For
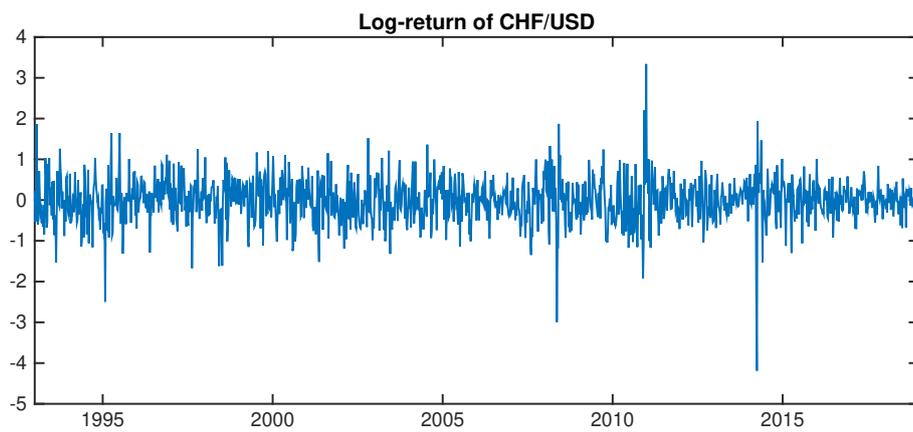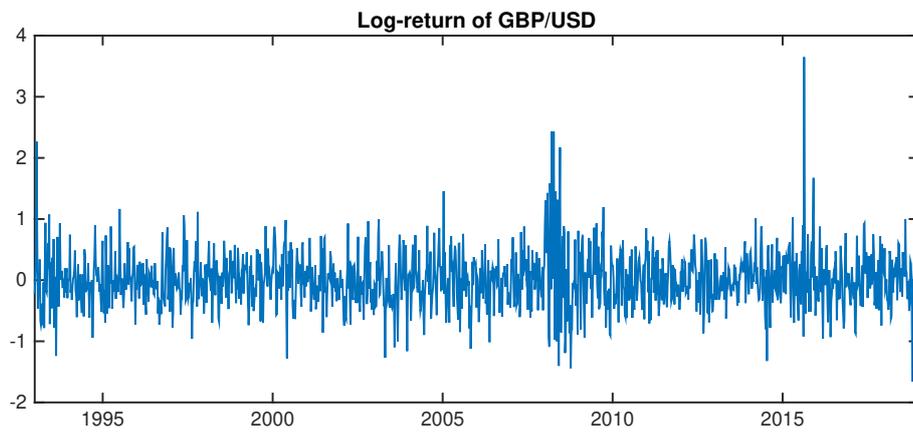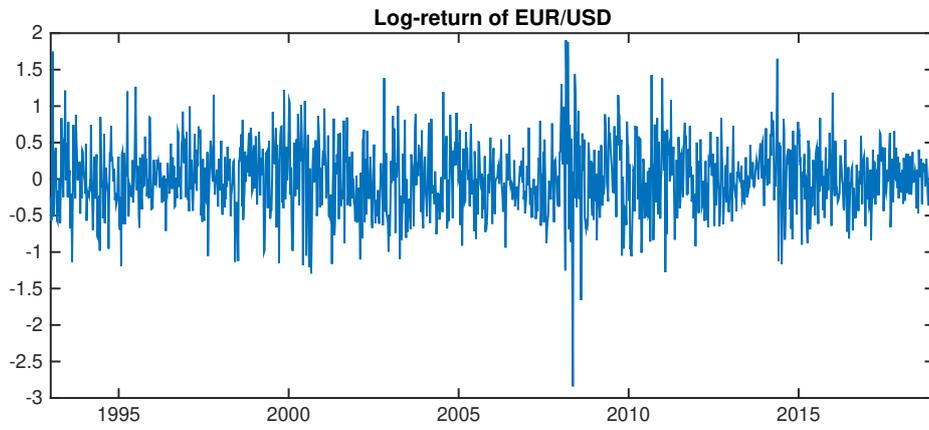
Table 5: Posterior statistics of parameters in the q sequences for competing models based on the exchange rate data

		MSV-GFT	MSV-GFT (Equi-persist)	MSV-CC		MSV-DCC
μ_{q1}	Mean	0.705	0.615	0.697	k	9.690
	SD	0.029	0.126	0.127		0.145
	95%CI	[0.647,0.762]	[0.327,0.812]	[0.438,0.946]		[9.430,10.064]
	IF	26.781	13.741	49.809		91.265
μ_{q2}	Mean	1.511	1.392	1.103	d	0.005
	SD	0.085	0.135	0.123		0.014
	95%CI	[1.337,1.674]	[1.078,1.600]	[0.871,1.354]		[-0.015,0.026]
	IF	7.9712	23.481	60.228		671.79
μ_{q3}	Mean	0.444	0.379	0.359	a_{11}	5.639
	SD	0.047	0.126	0.121		0.487
	95%CI	[0.351,0.533]	[0.114,0.576]	[0.128,0.589]		[4.900,6.506]
	IF	10.287	13.69	50.536		68.964
ϕ_{q1}	Mean	0.212	0.957		a_{21}	-1.232
	SD	0.107	0.019			0.137
	95%CI	[0.014,0.417]	[0.916,0.988]			[-1.484,-1.001]
	IF	76.346	86.206			62.115
ϕ_{q2}	Mean	0.911			a_{22}	-4.312
	SD	0.022				0.425
	95%CI	[0.862,0.948]				[-5.053,-3.644]
	IF	51.812				66.618
ϕ_{q3}	Mean	0.812			a_{31}	1.751
	SD	0.055				0.085
	95%CI	[0.685,0.898]				[1.612,1.912]
	IF	97.187				66.67
σ_{q1}^2	Mean	0.116	0.018		a_{32}	0.098
	SD	0.018	0.007			0.097
	95%CI	[0.083,0.155]	[0.010,0.032]			[-0.086,0.298]
	IF	59.407	138.66			48.733
σ_{q2}^2	Mean	0.061			a_{33}	4.779
	SD	0.014				0.397
	95%CI	[0.039,0.091]				[4.135,5.459]
	IF	77.28				63.357
σ_{q3}^2	Mean	0.055				
	SD	0.016				
	95%CI	[0.031,0.095]				
	IF	134.73				
log marg like		-1285.8	-1476.7	-1409.1		-1350.4
DIC		2275.2	3427.8	2711.4		3110.4

1. Mean is the posterior mean based on 20000 MCMC draws after a 2000 burn-in period.
2. SD is the numerical standard error of the posterior mean.
3. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws.
4. IF is the inefficiency factor.

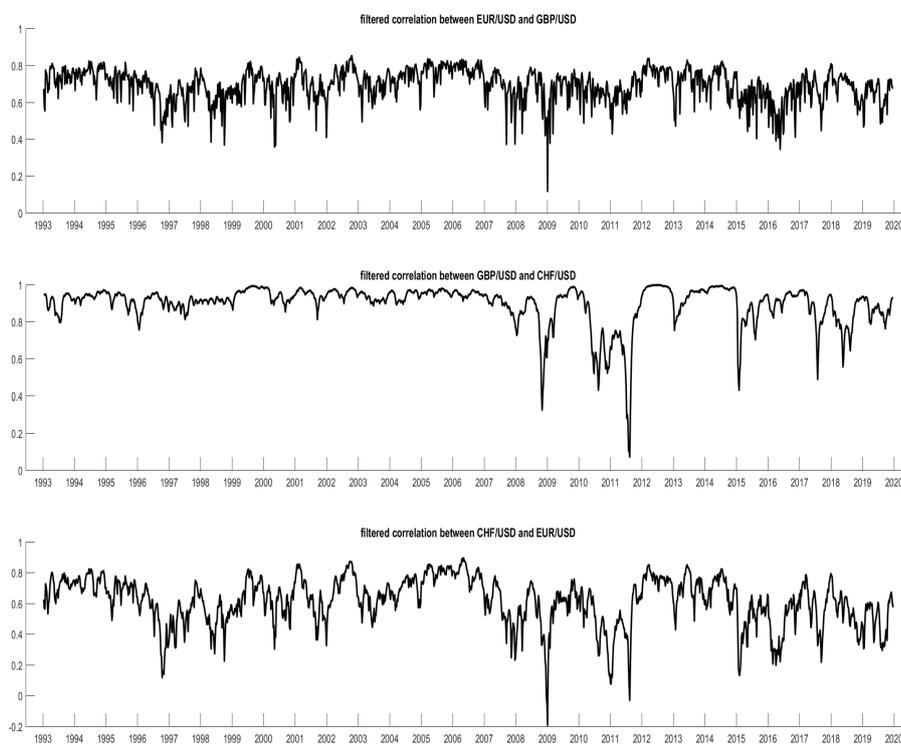Figure 5: Filtered q sequences in MSV-GFT for the exchange rate returns

the MSV-GFT model, the filtered h sequences of all three returns and the filtered q sequences for all three pairs are plotted in Figure 7 and Figure 8, respectively.

The general conclusion from this application is similar to that in the exchange rate data. For example, as shown in Table 6, all the h sequences are strongly persistent with the autoregressive root being very close to one. As shown presented in Table 7, in the flexible model, the posterior means of ϕ_q are $(0.519, 0.654, 0.664)$ while the equi-persistence model suggests a higher persistence level (0.859). Due to this, the improvement in the log marginal likelihood value and DIC values are not as large as that in the empirical application based on the exchange rate data. However, the fully flexible model still has a much higher value for the log marginal likelihood and much lower value for DIC than all the competing models. For example, the difference between the log marginal likelihood value in the fully flexible model and each of the competing models is 16, 26 and 19, indicating the decisive evidence to support the fully flexible model. We can then conclude that adopting the generalized Fisher transformation is indeed beneficial for modeling the multivariate equity returns.

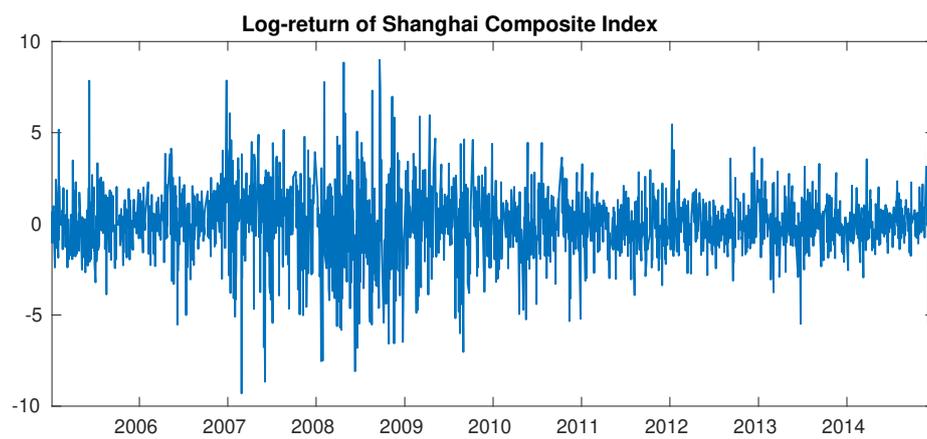Figure 6: Time series of stock index returns

Log-return of Hang Seng Index

Log-return of Nikkei 225 Index

Log-return of Shanghai Composite Index

Table 6: Posterior statistics of parameters in the h sequences for competing models based on the stock index data

		MSV-GFT	MSV-GFT (Equi-persist)	MSV-CC	MSV-DCC
μ_{h1}	Mean	0.370	0.343	0.397	0.375
	SD	0.246	0.257	0.247	0.232
	95%CI	[-0.144,0.851]	[-0.180,0.845]	[-0.096,0.889]	[-0.085,0.849]
	IF	4.8598	10.674	10.467	10.238
μ_{h2}	Mean	0.525	0.502	0.539	0.525
	SD	0.175	0.175	0.171	0.161
	95%CI	[0.186,0.872]	[0.147,0.836]	[0.195,0.887]	[0.203,0.839]
	IF	4.8977	10.783	10.59	10.063
μ_{h3}	Mean	0.758	0.738	0.767	0.757
	SD	0.203	0.216	0.215	0.195
	95%CI	[0.357,1.156]	[0.287,1.155]	[0.335,1.194]	[0.347,1.144]
	IF	4.8082	9.435	9.6687	9.9254
ϕ_{h1}	Mean	0.991	0.990	0.987	0.989
	SD	0.004	0.005	0.005	0.005
	95%CI	[0.982,0.995]	[0.981,0.95]	[0.977,0.994]	[0.979,0.995]
	IF	37.696	58.003	49.239	88.546
ϕ_{h2}	Mean	0.981	0.980	0.978	0.979
	SD	0.007	0.006	0.007	0.008
	95%CI	[0.968,0.991]	[0.965,0.990]	[0.962,0.989]	[0.961,0.991]
	IF	61.049	64.753	56.269	79.122
ϕ_{h3}	Mean	0.983	0.984	0.983	0.983
	SD	0.007	0.006	0.006	0.007
	95%CI	[0.968,0.993]	[0.970,0.994]	[0.968,0.993]	[0.967,0.994]
	IF	66.925	89.743	109.45	133.1
σ_{h1}^2	Mean	0.011	0.014	0.016	0.012
	SD	0.006	0.006	0.008	0.007
	95%CI	[0.006,0.021]	[0.008,0.023]	[0.009,0.027]	[0.006,0.026]
	IF	127.23	129.22	135.69	253.17
σ_{h2}^2	Mean	0.020	0.023	0.024	0.020
	SD	0.006	0.006	0.007	0.008
	95%CI	[0.013,0.030]	[0.013,0.035]	[0.016,0.040]	[0.010,0.036]
	IF	135.53	164.95	125.29	159.86
σ_{h3}^2	Mean	0.022	0.020	0.022	0.019
	SD	0.009	0.007	0.008	0.009
	95%CI	[0.011,0.035]	[0.011,0.034]	[0.011,0.039]	[0.008,0.034]
	IF	150.91	185.25	216.22	244.47

1. Mean is the posterior mean based on 20000 MCMC draws after a 2000 burn-in period.
2. SD is the numerical standard error of the posterior means.
3. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws.
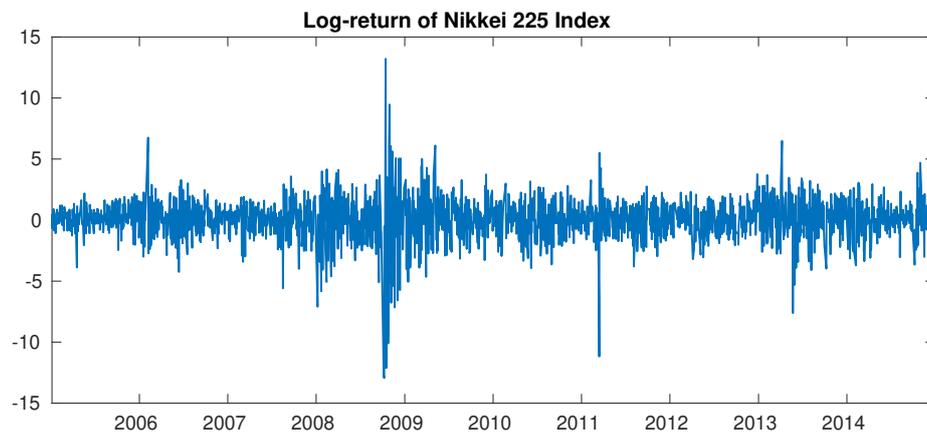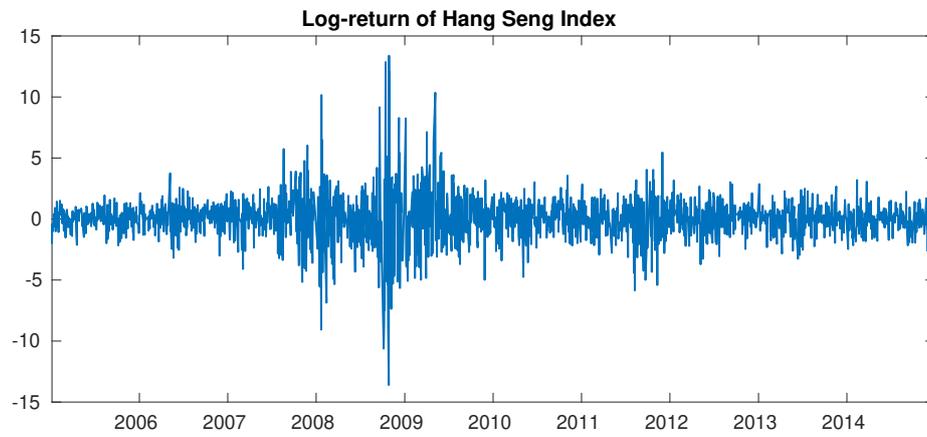4. IF is the inefficiency factor.

Table 7: Posterior statistics of parameters in the q sequences for competing models based on the stock index data

		MSV-GFT	MSV-GFT (Equi-persist)	MSV-CC		MSV-DCC
μ_{q1}	Mean	0.715	0.631	0.674	k	39.634
	SD	0.028	0.065	0.071		0.174
	95%CI	[0.663,0.771]	[0.476,0.736]	[0.540,0.827]		[39.261,39.968]
	IF	23.577	157.3	38.843		200.85
μ_{q2}	Mean	0.542	0.5	0.535	d	0.000
	SD	0.027	0.052	0.067		0.015
	95%CI	[0.489,0.595]	[0.368,0.578]	[0.408,0.658]		[-0.017,0.015]
	IF	18.057	98.375	34.34		607.12
μ_{q3}	Mean	0.123	0.109	0.123	a_{11}	1.915
	SD	0.024	0.047	0.065		0.115
	95%CI	[0.077,0.171]	[-0.003,0.179]	[-0.009,0.248]		[1.745,2.063]
	IF	30.631	42.512	37.416		148.66
ϕ_{q1}	Mean	0.519	0.859		a_{21}	-0.962
	SD	0.115	0.141			0.098
	95%CI	[0.277,0.733]	[0.524,0.994]			[-1.077,-0.829]
	IF	182.65	447.4			159.58
ϕ_{q2}	Mean	0.654			a_{22}	-0.685
	SD	0.104				0.075
	95%CI	[0.429,0.836]				[-0.793,-0.583]
	IF	204.24				121.1
ϕ_{q3}	Mean	0.664			a_{31}	1.561
	SD	0.090				0.070
	95%CI	[0.479,0.820]				[1.451,1.654]
	IF	185.73				160.63
σ_{q1}^2	Mean	0.069	0.007		a_{32}	0.058
	SD	0.020	0.010			0.039
	95%CI	[0.037,0.114]	[0.001,0.038]			[-0.012,0.133]
	IF	208.97	461.33			145.64
σ_{q2}^2	Mean	0.029			a_{33}	1.323
	SD	0.010				0.044
	95%CI	[0.015,0.050]				[1.256,1.399]
	IF	276.64				134.45
σ_{q3}^2	Mean	0.017				
	SD	0.008				
	95%CI	[0.008,0.038]				
	IF	318.23				
log marg like		-11038	-11054	-11064		-11067
DIC		22017	22076	22145		22533

1. Mean is the posterior mean based on 20000 MCMC draws after a 2000 burn-in period.
2. SD is the numerical standard error of the posterior means.
3. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws.
4. IF is the inefficiency factor.

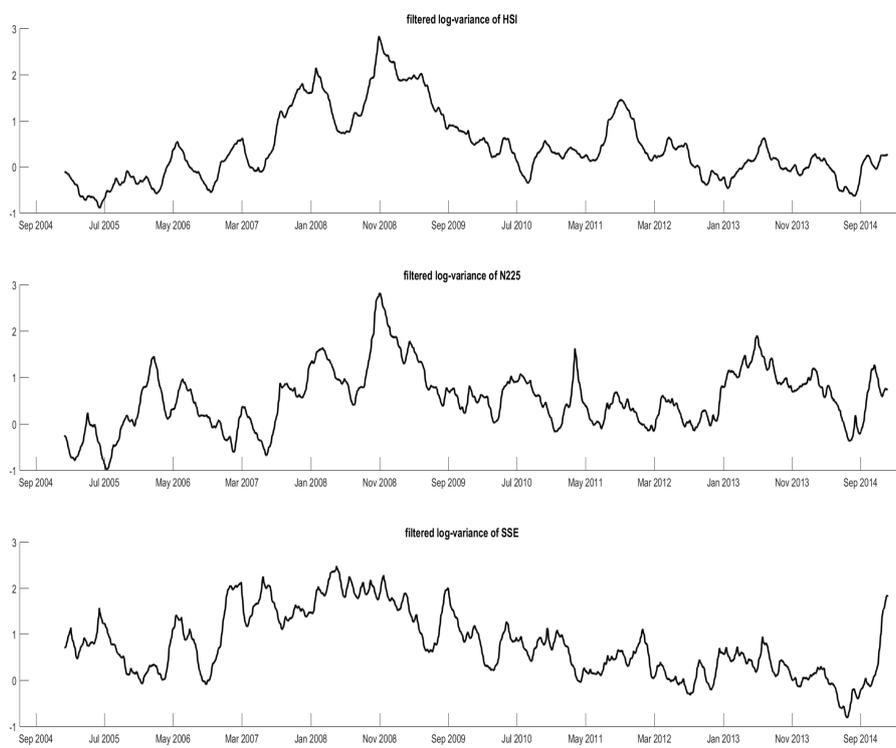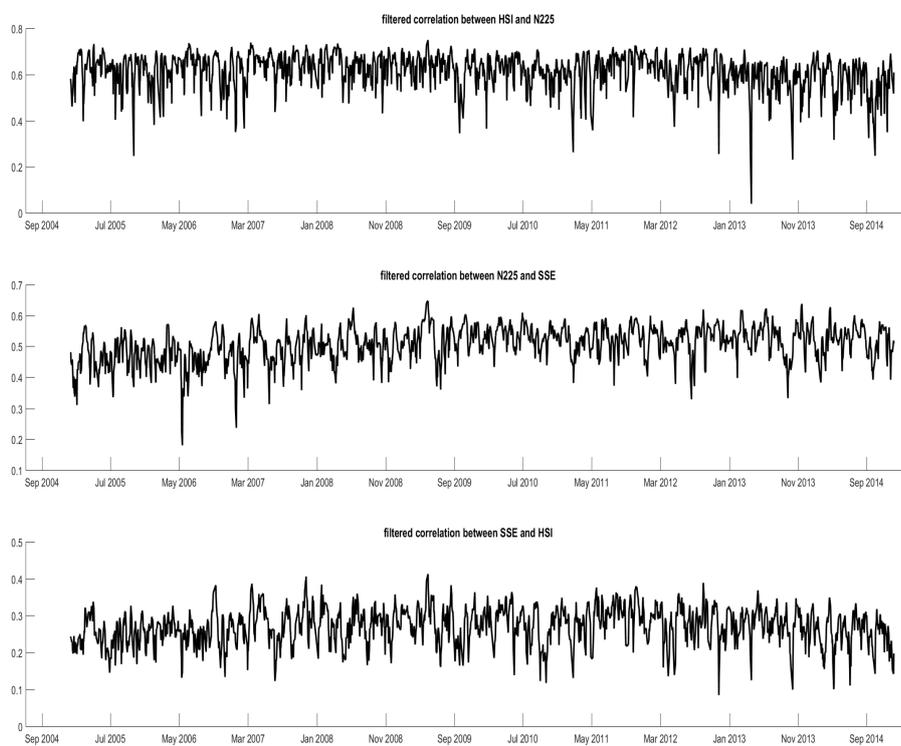Figure 7: Filtered h sequences in MSV-GFT for the stock index

Figure 8: Filtered q sequences in MSV-GFT for the stock index

4 MSV Model with Realized Measures

4.1 Model with Realized Measures

Given a large number of parameters in the MSV models given in Chapter 3, only using daily returns may not be enough to produce stable inference. Thanks to the increasing availability of intraday high-frequency data, we can now construct informative realized approximations to unobservable volatilities and covariances. When realized measures for the latent process h_t and q_t are available, we can incorporate them into a MSV model and greatly improve parameter estimation efficiency as well as the fit of the model; see Hansen et al. (2012) and Hurn et al. (2020).

Specifically, we assume that the researchers have access to the $p \times p$ realized covariance matrices $RCOV_t$ computed from intraday high-frequency returns. From the matrix, we first extract the $p \times 1$ vector of the diagonal elements $\tilde{v}_t = diag(RCOV_t)$ as the approximation to the latent variances. Then a realized measure of the unobservable correlation matrix is obtained through

$$Q_t = v_t^{-1/2} \cdot RCOV_t \cdot v_t^{-1/2},$$

where v_t is a diagonal matrix with the elements of $\tilde{v}_t$ on the main diagonal. Since the latent processes in the MSV model considered in Chapter 3 are the transformation of the original variances and correlations, we also apply the same transformation to the corresponding realized measures. Specifically, we define

$$
\begin{aligned}
h_t^r &= (h_{1t}^r, ..., h_{pt}^r)' = (\log \tilde{v}_{1t}, ..., \log \tilde{v}_{pt})', \\
q_t^r &= (q_{1t}^r, ..., q_{dt}^r)' = F(Q_t),
\end{aligned}
\tag{4.1}
$$

where superscripts denote realized measures.

The relationship between the latent variables and their realized counterparts is

modeled as

$$h_t^r = \psi_h + h_t + \xi_{ht}, \text{ where } \xi_{ht} \sim N(0, \Sigma_h^r), \ t = 1, ..., T - 1, \qquad (4.2)$$

$$q_t^r = \psi_q + q_t + \xi_{qt}, \text{ where } \xi_{qt} \sim N(0, \Sigma_q^r), \ t = 1, ..., T - 1, \qquad (4.3)$$

where $\psi_h = (\psi_{h1}, ..., \psi_{hp})'$ and $\psi_q = (\psi_{q1}, ..., \psi_{qd})'$ capture potential approximation biases in the realized measures, $\xi_h = (\xi_{h1}, ..., \xi_{hp})'$ and $\xi_q = (\xi_{q1}, ..., \xi_{qd})'$ are innovations which are normally distributed. These innovations are assumed to be independent of each other so that Σ_h^r and Σ_q^r are both diagonal matrices, with the diagonal entries being $\eta_h^2 = (\eta_{h1}^2, ... \eta_{hp}^2)$ and $\eta_q^2 = (\eta_{q1}^2, ..., \eta_{qd}^2)$.

It can be seen that extra measurement equations have been added to the MSV models. These measurement equations are based on the transformation of the realized measure and the same transformation applied to the latent covariance matrix. In the literature of realized volatility, it has been shown that the realized volatility converges to the integrated volatility as well as the log realized volatility converges to the log integrated volatility. However, Barndorff-Nielsen and Shephard (2002) has argued that the approximation to the log integrated volatility by the log realized volatility is better than that to the integrated volatility by the realized volatility. The property has been used in Hansen and Huang (2016) to introduce a realized EGARCH model and in Phillips and Yu (2009) to construct a two-stage method to estimate continuous time models. The empirical justification of the measurement equation will be shown in Section 4.4.

4.2 Bayesian Analysis of RMSV-GFT

If we combine equations (3.1)-(3.6) with equations (4.2)-(4.3), we have the so-called realized MSV-GFT (RMSV-GFT) model. The parameters of the model are $\theta \equiv (\mu_h, \mu_q, \phi_h, \phi_q, \sigma_h^2, \sigma_q^2, \psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)$.

4.2.1   Joint Estimation of RMSV-GFT

Let $r = (r1', ..., r_T')'$, $h = (h_1', ..., h_T')'$, $q = (q_1', ..., q_T')'$, $h^r = (h_1^{r'}, ..., h_T^{r'})'$ and $q^r = (q_1^{r'}, ..., q_T^{r'})'$. To carry out the inference, we implement a Gibbs sampler with four blocks. In the following, we use $\theta_{/\alpha}$ to denote θ excluding α. Then, the algorithm proceeds as

1. Initialize h, q and θ.

2. Draw $h, q | r, h^r, q^r, \theta$.

3. Draw $\mu_h, \mu_q | r, h, q, \theta_{/(\mu_h, \mu_q)}$.

4. Draw $\phi_h, \phi_q | r, h, q, \theta_{/(\phi_h, \phi_q)}$.

5. Draw $\sigma_h^2, \sigma_q^2 | r, h, q, \theta_{/(\sigma_h^2, \sigma_q^2)}$.

6. Draw $\Sigma_h^r, \Sigma_q^r | h, q, h^r, q^r, \theta_{/(\Sigma_h^r, \Sigma_q^r)}$.

7. Draw $\psi_h, \psi_q | h, q, h^r, q^r, \theta_{/(\psi_h, \psi_q)}$.

Iterating over steps 2-7 consists of a complete sweep of MCMC sampler.

We apply the PGAS introduced in the last chapter to sample from the latent variables h and q given r, (h^r, q^r), and a particular set of parameter values. Steps 3 to 5 sample from the conditional posterior distributions in (3.11), (3.12) and (3.13). We assume (ψ_h, Σ_h^r) and (ψ_q, Σ_q^r) follow the following normal-inverse-gamma priors

- $\psi_{hi} \sim N(m_{\psi 0}, s_{\psi 0}^2)$ and $\psi_{qj} \sim N(m_{\psi 0}, s_{\psi 0}^2)$;

- $\eta_{hi}^2 \sim IG\left(\frac{n_{m0}}{2}, \frac{\eta_{m0}}{2}\right)$ and $\eta_{qj}^2 \sim IG\left(\frac{n_{m0}}{2}, \frac{\eta_{m0}}{2}\right)$,

Steps 6 and 7 draw from the conditional posterior distributions of ψ_{hi} and η_{hi}^2:

$$\psi_{hi} | h, h^r, \theta_{/\psi_{hi}} \sim N(\tilde{m}_{hi\psi}, \tilde{s}_{hi\psi}^2) \text{ and } \eta_{hi}^2 | h, h^r, \theta_{/\eta_{hi}^2} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{\eta}_{hi}}{2}\right), \qquad (4.4)$$

where

$$\tilde{m}_{hi\psi} = \tilde{s}_{hi\psi}^2 \left(\frac{m_{\psi 0}}{T} + \sum_{t=1}^{T}(h_{it}^r - h_{it})\right), \ \tilde{s}_{hi\psi}^2 = \left(\frac{1}{s_{\psi 0}^2} + \frac{T}{\eta_{hi}^2}\right)^{-1},$$

and

$$\tilde{n}_m = n_{m0} + T, \ \tilde{\eta}_{hi} = \frac{\eta_{m0}}{2} + \frac{\sum_{t=1}^{T}(h_{it}^r - h_{it})^2}{2}.$$

We draw (ψ_{qj}, η_{qj}^2) following the similar procedure.

4.2.2 Two-stage Estimation of RMSV-GFT

As equations (4.2) and (4.3) two additional independent measurement equations, the conditional likelihood can be written as

$$p(r, h^r, q^r | h, q, \theta) = p(r | h, q, \theta_{/(\psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)}) + p(h^r, q^r | h, q, \psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r). \quad (4.5)$$

We draw θ, h, q from the full posterior distribution

$$p(\theta, h, q | r, h^r, q^r) \propto \left[p(r | h, q, \theta_{/(\psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)}) + p(h^r, q^r | h, q, \psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r) \right] p(h, q, \theta). \quad (4.6)$$

Note that if we only use the measurement equations 4.2 and 4.3 and the state equations for h_t and q_t, the system is in a linear Gaussian state space form and the Kalman filter can provide the straightforward likelihood-based inference. Unfortunately, in the same spirit as in Koopman and Scharth (2012), the setting precludes us to identify all parameters in the RMSV-GFT model. In particular, (μ_h, μ_q) and (ψ_h, ψ_q) are not separately identified. To explain the lack of identification, let x_t denote either h_{it} or q_{jt} and x_t^r denote the realized measures of x_t. Combining the realized measurement equations and the dynamics of the latent variables, we can easily show that x_t^r admits following ARMA(1,1) representation

$$x_t^r = (\mu + \psi) + \phi(x_{t-1}^r - \mu - \psi) + \eta_t - \phi\eta_{t-1} + \xi_t. \quad (4.7)$$

From this representation, it is clear that μ and ψ cannot be separately identified. To deal with the problem of lack of identification, the return equation is also used, as in Koopman and Scharth (2012) and Takahashi et al. (2009).

However, with a small or moderate number of MCMC iterations, the joint esti-

mation of all parameters in the model cannot produce satisfactory results as there are too many parameters in the joint estimation. A large number of MCMC iterations is required to produce good estimates. In fact, the number of MCMC iterations depends on the ratio of the likelihood of the intra-day return and that of the realized measure, which is given by

$$\frac{p(r|h, q, \theta_{/(\psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)})}{p(h^r, q^r|h, q, \psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)}.$$

The smaller the likelihood ratio, the larger number of MCMC iteration required. Note that the dimension of the numerator and the denominator is p and $\frac{p(p+1)}{2}$, respectively. As p increases, the difference between the two dimensions increases, and hence, it is expected that the likelihood ratio decreases. As a result, a worse mixing of MCMC samplers is expected. So the joint estimation is not feasible in empirical applications.

To overcome this problem of slow convergence, we propose to use a two-stage estimation procedure. In the literature of SV models with realized measures, this issue has been treated in various ways. Takahashi et al. (2009), for instance, ignore this problem and estimate all the parameters jointly. Yamauchi and Omori (2019) also fails to take into account the restrictions related to log-variances, while circumventing the constraints on correlations by imposing driftless random walks. Koopman and Scharth (2012) propose a two-step estimation method, which uses realized measurements only in the first step and relies on the Kalman filter. They impose a zero bias in the first step as an identification condition and then estimate it in the second step.

Keeping this constraint in mind, in this paper, we propose to use a two-stage procedure, which estimates the bias parameters in the first stage and then keep them fixed in the second stage. The detail is as follows.

- In the first stage, we estimate MSV-GFT without realized measures and obtain the posterior mean of μ, denote as $\hat{\mu}$. The procedure for this step is same as

in Section 3.2.3. Then estimate ψ_h and ψ_q by

$$\hat{\psi}_{h_i} = \bar{h}_i^r - \hat{\mu}_{h_i}, i = 1, ..., p,$$
$$\hat{\psi}_{q_j} = \bar{q}_j^r - \hat{\mu}_{q_j}, j = 1, ..., qd,$$

(4.8)

where $\bar{h}^r$ and $\bar{q}^r$ is the sample mean of h^r and q^r, respectively.

- In the second stage, we re-estimate the model using the joint likelihood by plug-in $\hat{\psi}$. That is, we sample from

$$p(\theta_{/(\psi_h, \psi_q)}, h, q | r, h^r, q^r, \hat{\psi}_h, \hat{\psi}_q)$$

(4.9)

by using steps 1-6 in Section 4.2.1. After sampling from (4.9), we estimate θ, h, q by their posterior means.

Note that the inclusion of realized measures into the likelihood function influences the weights computed during the particle filter. Meanwhile, in the second stage, we fix the ψ to be the one computed in the first stage and only re-sample μ. In the end, ψ is updated by subtracting the new μ from average realized measures. By doing this two-stage estimation, we circumvent the slow convergence problem and are able to obtain a much less biased estimator of ψ and μ, as shown in the simulation studies reported below.

We use the joint estimation and the two-stage estimation methods to fit the RMSV-GFT model based on simulated data. The autocorrelation functions (ACFs) for μ_h and μ_q are displayed in Figure 9. It can be seen that the MCMC draws for μ_h and μ_q mix much better in the two-stage method than in the joint method.

4.3 Simulation Studies

We now investigate the performance of the proposed method for estimating the RMSV-GFT model that incorporates the realized measures. The true values of

Figure 9: Comparison of ACFs of MCMC Draws for μ_h and μ_q
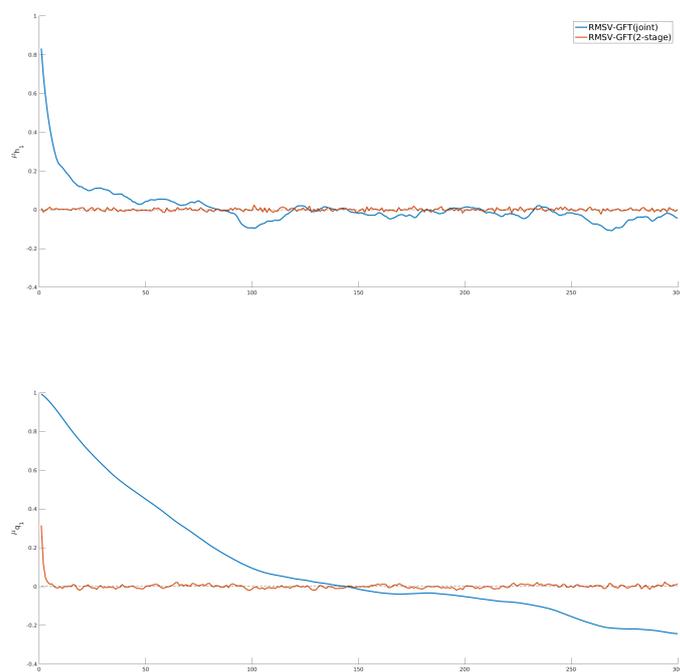
Note: The 2 ACFs of MCMC draws for μ_h and μ_q are displayed when the joint and two-stage methods are used to estimate the RMSV-GFT model. The top figure is for μ_{h_1} and the bottom figure is for μ_{q_1}.

parameters are the same as in the previous subsection, with the following additional ones,

$$\psi_{h1} = \psi_{h2} = \psi_{h3} = -0.5, \psi_{q1} = \psi_{q2} = \psi_{q3} = -0.3,$$

$$\eta_{h1}^2 = \eta_{h2}^2 = \eta_{h3}^2 = 0.2, \eta_{q1}^2 = \eta_{q2}^2 = \eta_{q3}^2 = 0.2.$$

The sample size considered in the simulation study is 500. Again, all the simulation results reported are based on 5000 MCMC iterations, with the first 1000 samples discarded as the burn-in period.

The results are reported in Table 8 where both the joint method and the two-stage method are used. It can be seen that when the joint estimation method is used, the posterior means of both μ and ϕ are biased. Specifically, the posterior means of μ_h and μ_q are biased downward. The posterior means of ψ_h and ψ_q are upward biased with a similar magnitude. Similarly, the posterior means of σ^2s are 0.01 and 0.015 lower than the true values for the log-variance and the transformed correlations, respectively. The posterior means of η^2s are upward biased.

On the other hand, our two-stage estimation method works very well, with a very small bias for all the parameters. Unbiased parameter estimation is critical to the filtering of the latent processes, which influences the out-of-sample prediction performance. To this end, we can examine the in-sample performance of the filtered $\{h_t\}$ and $\{q_t\}$ as the true latent processes are available for comparison in the simulation studies. Indeed, the average mean squared error of all six unobservables is 0.339 when the model is estimated jointly and is 0.327 when the two-stage method is used. In summary, our inference procedure produces accurate estimators for the parameters and the latent volatility/correlation.

As we re-sample (μ_h, μ_q) in Equation (4.9), the posterior means of the parameters are different from those in the MSV-GFT model. To further validate the stability of the two-stage method, we consider a three-stage method using the same simulated data. In the three-stage method, After the second stage estimation, we re-estimate (ψ_h, ψ_q) and sample from Equation (4.9) with the new estimators of (ψ_h, ψ_q).

Table 9 reports the simulation results for μ in all three stages. Firstly, we observe

Table 8: Posterior statistics when the RMSV-GFT model is fitted to the simulated data

	μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}	ψ_{h1}	ψ_{h2}	ψ_{h3}	ψ_{q1}	ψ_{q2}	ψ_{q3}
Joint Estimation												
Mean	0.272	0.281	0.268	0.656	0.662	0.651	-0.474	-0.478	-0.475	-0.265	-0.269	-0.265
Std	0.088	0.084	0.081	0.063	0.049	0.057	0.073	0.068	0.069	0.05	0.037	0.052
IF	9.867	10.172	10.002	10.111	9.832	10.08	26.383	27.638	27.716	15.618	15.625	15.642
2-Stage Estimation												
Mean	0.299	0.307	0.295	0.65	0.657	0.648	-0.502	-0.505	-0.504	-0.255	-0.261	-0.259
Std	0.093	0.087	0.085	0.066	0.054	0.061	0.081	0.074	0.074	0.054	0.044	0.059
IF	1.272	1.307	1.021	2.106	1.878	1.936						
True Value	0.3	0.3	0.3	0.7	0.7	0.7	-0.5	-0.5	-0.5	-0.3	-0.3	-0.3

	ϕ_{h1}	ϕ_{h2}	ϕ_{h3}	ϕ_{q1}	ϕ_{q2}	ϕ_{q3}
Joint estimation						
Mean	0.91	0.907	0.908	0.831	0.83	0.825
Std	0.021	0.023	0.023	0.037	0.039	0.034
IF	12.715	13.312	15.732	24.868	24.926	24.419
2-Stage Estimation						
Mean	0.89	0.883	0.886	0.817	0.816	0.809
Std	0.021	0.025	0.024	0.038	0.042	0.036
IF	17.244	18.724	20.989	25.839	25.104	27.277
True Value	0.9	0.9	0.9	0.8	0.8	0.8

	σ^2_{h1}	σ^2_{h2}	σ^2_{h3}	σ^2_{q1}	σ^2_{q2}	σ^2_{q3}	η^2_{h1}	η^2_{h2}	η^2_{h3}	η^2_{q1}	η^2_{q2}	η^2_{q3}
Joint Estimation												
Mean	0.039	0.04	0.04	0.034	0.035	0.036	0.22	0.215	0.22	0.223	0.222	0.222
Std	0.009	0.008	0.01	0.009	0.009	0.01	0.018	0.015	0.018	0.017	0.02	0.017
IF	26.6	28.104	30.259	41.99	43.879	42.951	7.822	8.08	9.008	11.967	13.02	12.949
2-Stage Estimation												
Mean	0.048	0.051	0.05	0.034	0.036	0.037	0.211	0.198	0.207	0.228	0.226	0.224
Std	0.012	0.013	0.014	0.01	0.011	0.012	0.029	0.029	0.031	0.021	0.027	0.025
IF	40.446	44.711	43.358	52.335	52.938	53.244	14.36	16.557	13.835	27.365	28.418	29.568
True Value	0.05	0.05	0.05	0.05	0.05	0.05	0.2	0.2	0.2	0.2	0.2	0.2

1. Mean is the average posterior mean across replications.

2. Std is the standard error of the posterior mean across replications.

3. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

Table 9: Posterior statistics of μ under three alternative methods

		μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}
	True Value	0.3	0.3	0.3	0.7	0.7	0.7
First Stage	Mean	0.304	0.311	0.299	0.657	0.665	0.656
	Std	0.092	0.083	0.082	0.064	0.055	0.061
	IF	6.4	5.8	5.5	12.2	11.6	12.0
Second Stage	Mean	0.299	0.307	0.295	0.65	0.657	0.648
	Std	0.088	0.084	0.081	0.063	0.049	0.057
	IF	1.272	1.307	1.021	2.106	1.878	1.936
Third Stage	Mean	0.298	0.307	0.295	0.65	0.657	0.648
	Std	0.093	0.087	0.085	0.065	0.054	0.062
	IF	1.235	1.276	1.037	2.102	2.027	1.959

1. Mean is the average posterior mean across replications.
2. Std is the standard error of the posterior mean across replications.
3. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

that the difference between the true parameter values and their posterior means is smaller in the second stage than in the first stage, although the posterior standard deviations are similar. Secondly, incorporating realized measures in the second stage greatly reduces the inefficiency factor, suggesting that the RMSV-GFT model results in better mixed chains than the MSV-GFT model. Thirdly, when comparing with the two-stage method and the three-stage method, we find the results from the two methods are nearly identical, suggesting that the two-stage method is stable and reliable. Since the two-stage method performs better than the joint method in simulation and is computationally more efficient than the three-stage method, we only report the posterior results based on the two-stage method in the empirical applications.

Figure 10: Finite Sample Distributions of 3 Stages
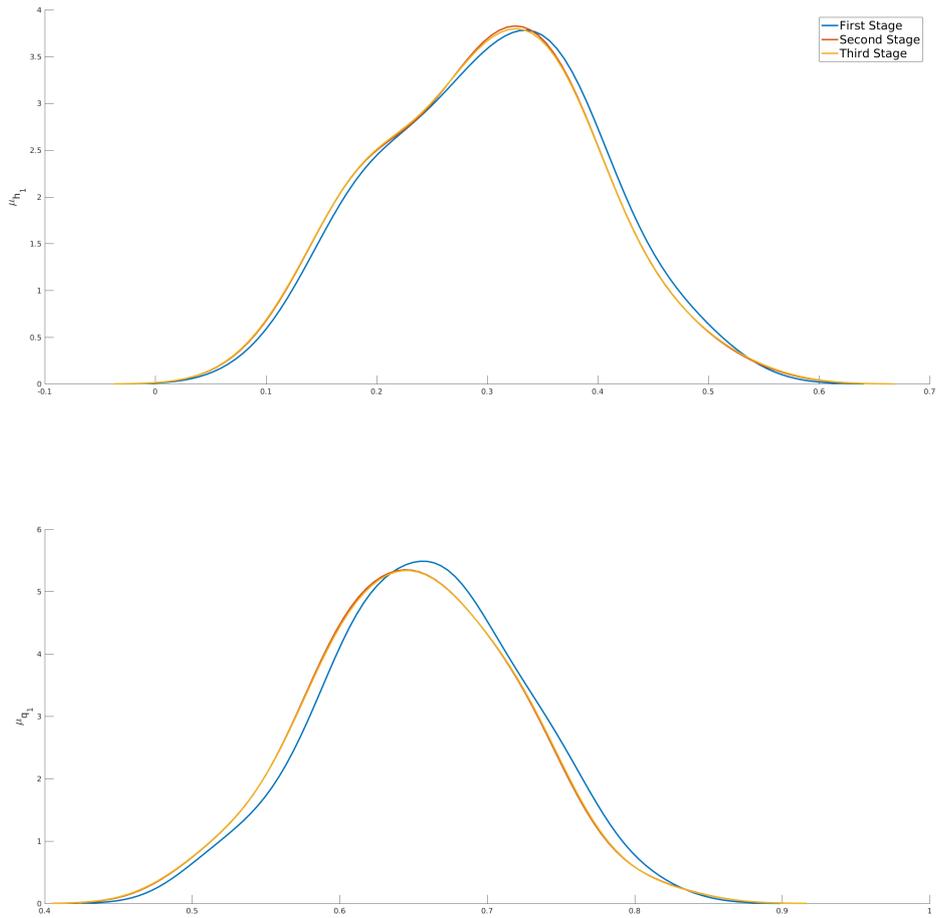
Note: The finite sample distribution is based on simulation data. The top figure is distribution for $\hat{\mu}_{h_1}$ and bottom figure is distribution for $\hat{\mu}_{q_1}$.

4.4 Empirical Application

4.4.1 Data

In this empirical application, we use daily stock returns and corresponding realized measures. We consider the daily log-returns of the following three stocks ($p = 3$): Bank of America (BAC), JP Morgan (JPM) and American Express (AXP). The full sample period is from February 1, 2001 to December 31, 2009. The total sample size is $T = 2232$. The daily log returns are the demeaned close-to-close returns. The realized covariance is constructed by using 5-minute returns with subsampling and the Parzen weight function. The detailed construction of these realized measures is discussed in Noureldin et al. (2012). The daily return data and the corresponding realized covariance are downloaded from the data library of the Oxford Man Institute website.[6] The three time series are plotted in Figure 11. We transfer the realized variance and realized correlation by using equation (4.1).

Our model assumes that the counterparts of these realized measures follow independent Gaussian autoregressions specified in (4.2) and (4.3). The distribution of the log realized volatility had been carefully and extensively discussed in the literature; see Andersen et al. (2001a) and Andersen et al. (2001b). One way to check the validity of the model specification is to examine the distribution properties of the realized measures. In this paper, we report the Q-Q plot of q^r in Figure 12 and the scatter plot of each two q^rs in Figure 13. These plots show the assumption of the independent normal distributed error terms seems to be supportive by the data.

4.4.2 Model Estimation

Let $\hat{\theta}, \hat{h}, \hat{q}$ denote the posterior mean of the parameters and latent variables. Let $\hat{V}_t$ and $\hat{R}_t$ denote the corresponding estimates of the variance and the correlation, which are defined as

$$\hat{V}_t = I_p * \left(\mathbb{1}_p \otimes \exp(\hat{h}_t) \right),$$
$$\hat{R}_t = F(\hat{q}_t)^{-1}. \tag{4.10}$$

Figure 11: Time series plot of stock returns

equity return of BAC

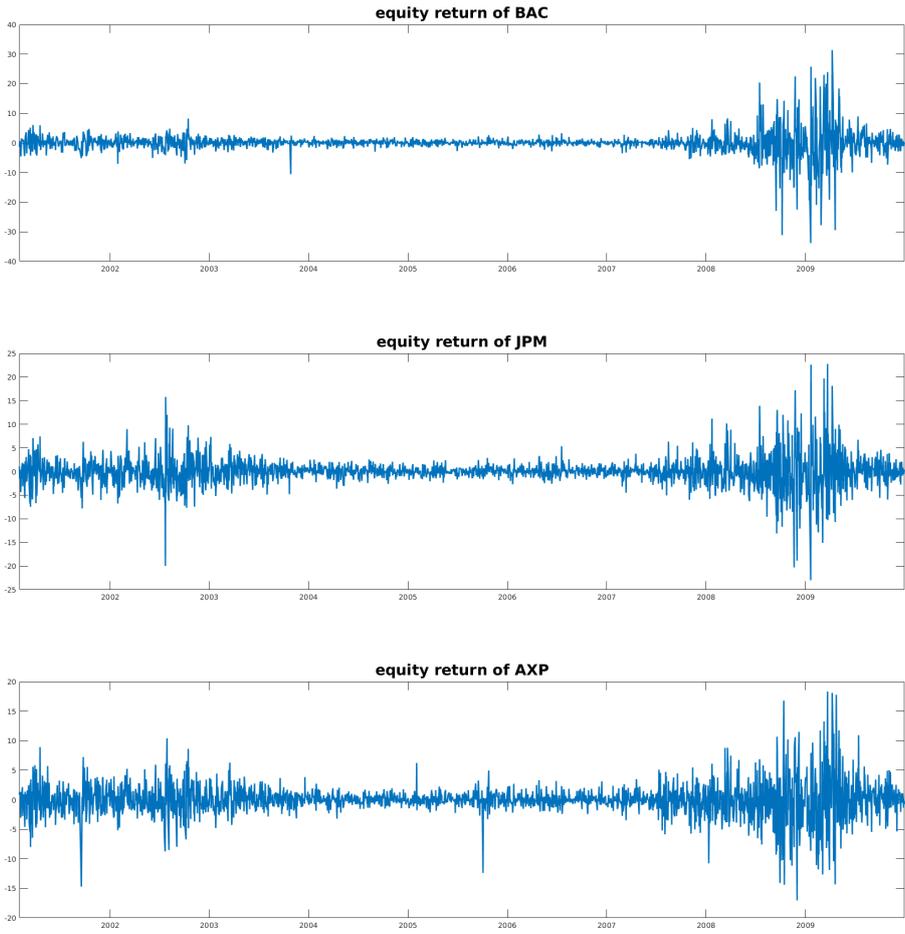equity return of JPM
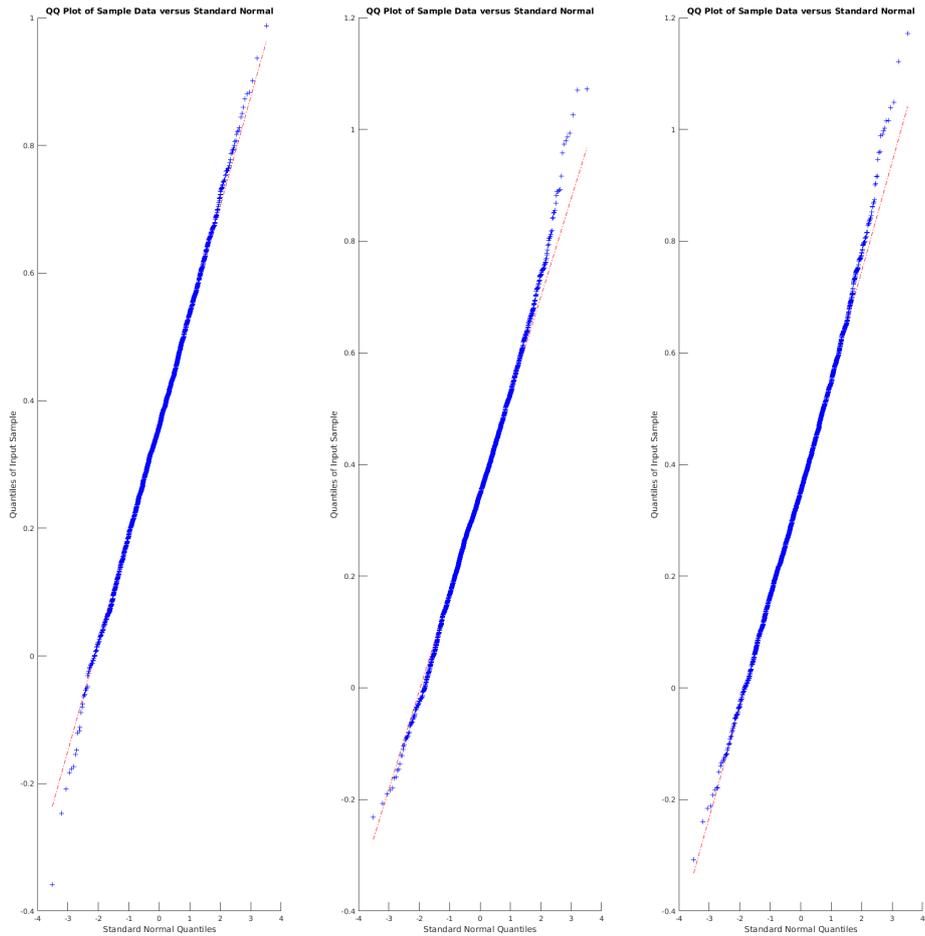
equity return of AXP

Figure 12: Q-Q plots of parameterized realized correlation

Note: The Q-Q plots for parameterized realized correlitons q^r, are constructed using daily realized correlation of BAC,JPM and AXP for the period since 2 Jan 2001, until 31 Dec 2009.
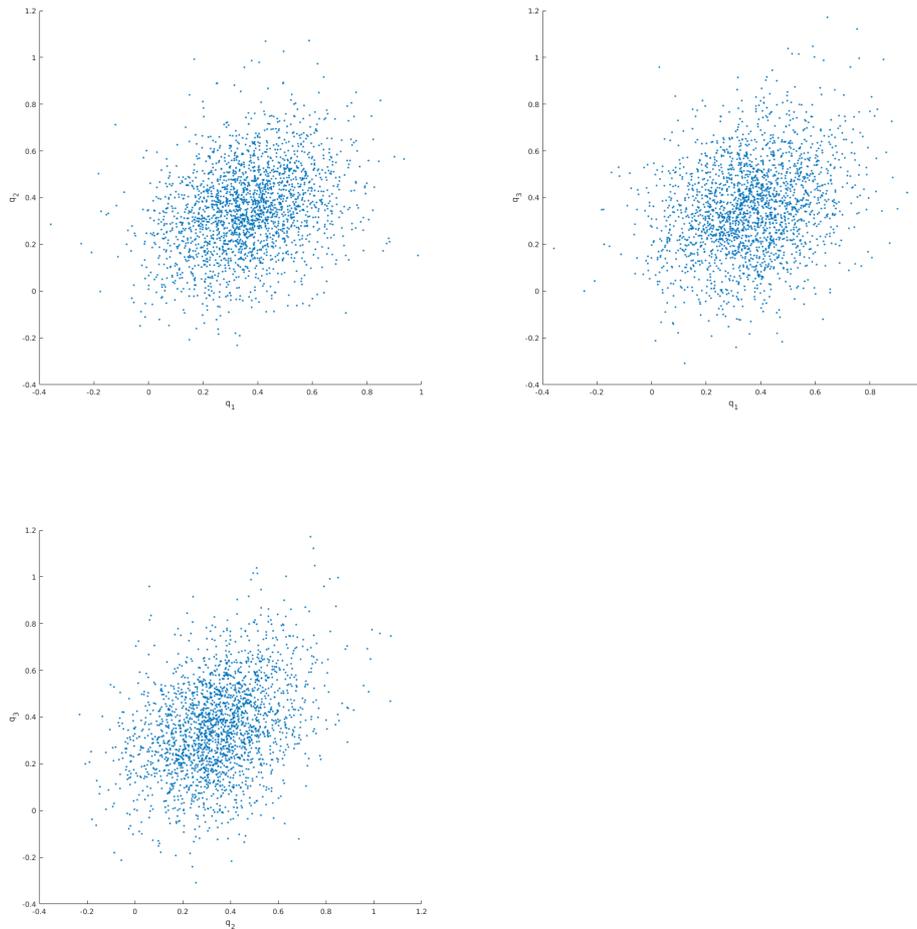
Figure 13: Scatter plot of parameterized realized correlation

Note: The scatter plots for parameterized realized correlitons q^r, are constructed using daily realized correlation between BAC and JPM, BAC and AXP,JPM and BAC for the period since 2 Jan 2001, until 31 Dec 2009.

Table 10: Empirical Posterior parameter estimation without realized measures

	μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}
mean	0.885	1.21	1.163	0.863	0.633	0.721
std	0.414	0.339	0.354	0.027	0.028	0.028
IF	8.476	9.397	8.998	26.738	34.805	26.225
	ϕ_{h1}	ϕ_{h2}	ϕ_{h3}	ϕ_{q1}	ϕ_{q2}	ϕ_{q3}
mean	0.993	0.993	0.992	0.586	0.567	0.585
std	0.002	0.002	0.002	0.132	0.089	0.099
IF	26.584	55.228	31.551	257.92	158.21	195.12
	σ_{h1}^2	σ_{h2}^2	σ_{h3}^2	σ_{q1}^2	σ_{q2}^2	σ_{q3}^2
mean	0.018	0.012	0.016	0.048	0.066	0.07
std	0.006	0.006	0.006	0.016	0.017	0.02
IF	80.872	109.16	83.834	257.76	250.05	213.49

1. Mean is the average posterior mean based on 18000 MCMC samples after a 2000 burn-in period.
2. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

Table 11: Empirical Posterior parameter estimation with realized measures

	μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}
mean	0.868	1.18	1.139	0.861	0.632	0.723
std	0.293	0.18	0.174	0.007	0.009	0.014
IF	8.714	8.915	8.59	10.549	9.537	9.762
	ϕ_{h1}	ϕ_{h2}	ϕ_{h3}	ϕ_{q1}	ϕ_{q2}	ϕ_{q3}
mean	0.988	0.979	0.981	0.773	0.842	0.924
std	0.004	0.007	0.007	0.047	0.038	0.031
IF	14.177	18.475	21.2	83.014	58.096	41.363
	σ_{h1}^2	σ_{h2}^2	σ_{h3}^2	σ_{q1}^2	σ_{q2}^2	σ_{q3}^2
mean	0.028	0.028	0.022	0.004	0.003	0.002
std	0.004	0.005	0.004	0.001	0.001	0.001
IF	55.265	69.463	84.922	89.295	54.492	39.511
	ψ_{h1}	ψ_{h2}	ψ_{h3}	ψ_{q1}	ψ_{q2}	ψ_{q3}
mean	-0.165	-1.103	-0.704	-0.501	-0.284	-0.364
	η_{h1}^2	η_{h2}^2	η_{h3}^2	η_{q1}^2	η_{q2}^2	η_{q3}^2
mean	0.4	0.207	0.211	0.022	0.024	0.028

1. Mean is the average posterior mean based on 18000 MCMC samples after a 2000 burn-in period.
2. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

We estimate MSV-GFT and RMSV-GFT based on the full sample data. The empirical results are reported in Tables 10 and 11 for MSV-GFT and RMSV-GFT, respectively. The posterior quantities of (μ_h, μ_q) are similar in the two tables. For example, the posterior means of ϕ_h are close to 1. This finding is well-known in the literature. However, the posterior means of q are very different. In MSV-GFT, the posterior means of ϕ_q are $(0.586, 0.567, 0.585)$ which are close to the persistence levels of q^r, which are $(0.33, 0.358, 0.405)$. By adding the realized measures, the posterior means of ϕ_q in RMSV-GFT is higher than those in MSV-GFT. Moreover, the posterior means of ψ_h and ψ_q are negative which are consistent with that has been reported in the literature; see Koopman and Scharth (2012) and Yamauchi and Omori (2019). Furthermore, the posterior means of η_h^2 and η_q^2 are larger than those of σ_h^2 and σ_q^2. Furthermore, the chains mix better in RMSV-GFT than in MSV-GFT as the inefficiency factors are smaller in RMSV-GFT than those in MSV-GFT. Figure 16 shows the Q-Q plot of the residuals in the RMSV-GFT model, $\hat{V}_t^{-\frac{1}{2}}\hat{R}_t^{-\frac{1}{2}}r_t$. It suggests that the tail distribution is not close to the normal distributed assumption. We will address this inadequacy in Section 5.

To validate the estimation results, we compare the in-sample fit of the realized measures with the realized measures. Using the measurement equations in (4.2) and (4.3), we conduct the in-sample fit of the realized variance $\hat{v}_t$ and correlation $\hat{Q}_t$ as following

$$
\begin{aligned}
\hat{v}_t &= I_p * \left(\mathbb{1}_p \otimes \exp(\hat{h}_t + \hat{\psi}_h) \right), \\
\hat{Q}_t &= F(\hat{q}_t + \hat{\psi}_q)^{-1},
\end{aligned}
\tag{4.11}
$$

where I_p is a p-dimensional identity matrix, $\mathbb{1}_p$ is a p-dimensional column vector of ones.

We compare $\hat{v}_t$ and $\hat{Q}_t$ with v_t and Q_t. Note that v_t is a diagonal matrix and Q_t is a symmetric matrix with diagonal elements equal to 1. Hence, we only focus the comparison on the diagonal elements in v_t and the off-diagonal elements in Q_t. Figure 14 and Figure 15 show the in-sample comparisons. For the realized variance, the two models provide a reasonably similar fit except during the financial crisis period. Nevertheless, MSV-GFT provides a less accurate fit than RMSV-GFT.

However, for the realized correlation, RSMV-GFT provides a more reasonable fit than MSV-GFT.

4.4.3 Out-of-Sample Forecasting Performance

In the univariate context, comparing the out-of-sample performance of the alternative model is relatively easy. For example, one can use the mean squared error (MSE) to quantify the distance between the predictive volatility and a proxy of the true volatility. This task becomes more difficult in the multivariate context.

In this section we compare the out-of-sample performance of alternative MSV models by using both the statistical and economic methods. For the statistical evaluation, since the true volatility is unobserved, we use the realized measure as the proxy of the true volatility. Under the RMSV-GFT model, we forecast the realized measure using equation (4.2) and (4.3) with the plug-in forecast of volatility. We then compare the forecast of realized measure with the realized measure using the MSE and the quasi-likelihood loss function(Qlike). MSE and Qlike are two common loss function to measure the distance between forecasts and the ex-post values.

$$
\begin{aligned}
\textbf{MSE} : & L(\hat{v}, v) = (\hat{v} - v)^2, \\
\textbf{Qlike} : & L(\hat{v}, v) = \frac{\hat{v}}{v} - log(\frac{\hat{v}}{v}) - 1,
\end{aligned}
\tag{4.12}
$$

where $\hat{v}$ is the forecast and v is the ex-post value.

For the economic evaluation, we apply the forecast of volatility (covariance) in a mean-variance framework and focus on the out-of-sample performance of the portfolio. Engle and Colacito (2006) argues that the correct covariance produces a smaller expected portfolio variance. Therefore, we construct the portfolio based on the forecast of covariance and evaluate the performance of alternative models by comparing the variance of the portfolio.

4.4.3.1 Forecasting Procedure

In this section, we consider the performance of MSV-GFT and RMSV-GFT in terms their ability in one-step-ahead out-of-sample forecasts of h and q. The period over which I try to obtain forecasts is between February 7, 2008 and December 31, 2009. There 480 trading days in the forecasting period. To estimate the MSV-GFT and RMSV-GFT models, we use the most recent 1,000 observations at each point in time. As the new observation becomes available, we roll the window forward but keep the window size fixed at 1,000 when estimating the two models. The general procedure is illustrated as follows.

1. Given data $\{r_{t-1000:t-1}, h^r_{t-1000:t-1}, q^r_{t-1000:t-1}\}$, we draw N samples from $p(\theta, h, q \mid r_{t-1000:t-1}, h^r_{t-1000:t-1}, q^r_{t-1000:t-1})$ by using the procedure proposed in Section 4.2.2 in RMSV-GFT or in 3.2.3 for MSV-GFT. We draw 5,000 MCMC iterations from the posterior distributions and estimate the parameters using the posterior means of $(\theta, h, q \mid r_{t-1000:t-1}, h^r_{t-1000:t-1}, q^r_{t-1000:t-1})$ denoted by $\tilde{\theta}, \tilde{h}_{t-1}$ and $\tilde{q}_{t-1}$.

2. Forecast h_t and q_t by using

$$
\begin{aligned}
\hat{h}_{t|t-1} &= \tilde{\mu}_h + \tilde{\phi}_h(\tilde{h}_{t-1} - \tilde{\mu}_h), \\
\hat{q}_{t|t-1} &= \tilde{\mu}_q + \tilde{\phi}_h(\tilde{q}_{t-1} - \tilde{\mu}_q).
\end{aligned}
\tag{4.13}
$$

The out-of-sample forecast of the variance and correlation are obtained as

$$
\begin{aligned}
\hat{V}_{t|t-1} &= I_p * \left(\mathbb{1}_p \otimes \exp(\hat{h}_{t|t-1}) \right), \\
\hat{R}_{t|t-1} &= F^{-1}(\hat{q}_{t|t-1}).
\end{aligned}
\tag{4.14}
$$

Therefore, the out-of-sample forecast of the covariance is obtained as

$$
\hat{\Sigma}_{t|t-1} = \hat{V}^{\frac{1}{2}}_{t|t-1} \hat{R}_{t|t-1} \hat{V}^{\frac{1}{2}}_{t|t-1}.
\tag{4.15}
$$

3. Iterate steps 1 and 2 for t from February 7, 2008 to December 31, 2009. We

71

end up with 480 one-step-ahead out-of-sample forecasts of h and q.

4.4.3.2 Forecasting Realized Measure

In terms of volatility forecasting, extensive studies in the literature focus on forecasting realized volatility. See, for example, Corsi et al. (2012), Koopman and Scharth (2012) forecast realized volatility by using the HAR model and the SV model, respectively. In this Chapter, we forecast both realized volatility and realized correlation with MSV-GFT and RMSV-GFT. The forecasting procedure is as follows.

$$
\begin{aligned}
\hat{v}_{t|t-1} &= I_p * \left(\mathbb{1}_p \otimes \exp(\tilde{\psi}_h + \hat{h}_{t|t-1}) \right), \\
\hat{Q}_{t|t-1} &= F^{-1}(\tilde{\psi}_q + \hat{q}_{t|t-1}),
\end{aligned}
\tag{4.16}
$$

where $\hat{v}_{t|t-1}$ is the forecast of the realized variance and $\hat{Q}_{t|t-1}$ is the forecast of the realized correlation. To evaluate the out-of-sample performance, we consider both the univariate setting and the multivariate setting. For comparison in the univariate setting, Patton (2011) establish the robustness of MSE and Qlike. Unfortunately, Qlike is not possible for realized correlation as it takes a value between -1 and 1. When the forecast correlation takes a negative value, Qlike is not well-defined. So we report MSE, Qlike for the forecasts of realized variance and only report MSE for the forecasts of realized correlation. To check the robustness of MSE and Qlike, we can report the Mean Absolute Error (MAE). MSE, Qlike and MAE are computed as follows.

$$
\begin{aligned}
MSE_v &= \frac{1}{480} \sum_{t=1763}^{2242} (v_t - \hat{v}_{t|t-1})^2, \quad MSE_Q = \frac{1}{480} \sum_{t=1763}^{2242} (Q_t - \hat{Q}_{t|t-1})^2, \\
Qlike_v &= \frac{1}{480} \sum_{t=1763}^{2242} \left(\frac{\hat{v}_{t|t-1}}{v_t} - log(\frac{\hat{v}_{t|t-1}}{v_t}) - 1 \right), \\
MAE_v &= \frac{1}{480} \sum_{t=1763}^{2242} |v_t - \hat{v}_{t|t-1}|, \quad MAE_Q = \frac{1}{480} \sum_{t=1763}^{2242} |Q_t - \hat{Q}_{t|t-1}|.
\end{aligned}
\tag{4.17}
$$

To evaluate out-of-sample performance of alternative models for forecasting the realized covariance, we calculate the multivariate Qlike of Patton and Sheppard

Table 12: Out-of-Sample Predicted RM

	MSE		Qlike		MAE	
	MSV-GFT	RMSV-GFT	MSV-GFT	RMSV-GFT	MSV-GFT	RMSV-GFT
$\exp(h_1^r)$	398.153	269.284	0.609	0.307	11.2695	6.592
$\exp(h_2^r)$	36.5326	21.0515	0.480	0.178	3.4162	1.5199
$\exp(h_3^r)$	37.2511	25.0701	0.251	0.150	3.391	2.0018
$Q(1,2)$	0.0386	0.0308			0.1441	0.1263
$Q(1,3)$	0.0472	0.0379			0.1498	0.1282
$Q(2,3)$	0.0755	0.0526			0.1829	0.1479
	MSV-GFT				RMSV-GFT	
Qlike	1.5143				0.7235	

1.$\exp(h_i^r)$ is the ith diagonal element in the realized variance.
2.Q(i,j) denotes the (i,j) components in the realized correlation matrix, while the diagonal elements are all 1.

(2009) which is follows as

$$Qlike = \frac{1}{480} \sum_{t=1763}^{2242} \left(tr(\widehat{RCOV}_{t|t-1}^{-1} RCOV_t) - \log |\widehat{RCOV}_{t|t-1}^{-1} RCOV_t| - p \right), \quad (4.18)$$

where $\widehat{RCOV}_{t|t-1}^{-1} = \hat{v}_{t|t-1}^{\frac{1}{2}} \hat{Q}_{t|t-1} \hat{v}_{t|t-1}^{\frac{1}{2}}$ is the realized covariance forecast.

In the univariate setting, Table 12 reports MSE, Qlike and MAE for MSV-GFT and RMSV-GFT, respectively. According to MSE and Qlike, by incorporating the information from the realized measure, RMSV-GFT provides more accurate forecasts than MSV-GFT. According to MAE, RMSV-GFT continues to provide more accurate forecasts than MSV-GFT. In the univariate setting, Table 12 reports the multivariate Qlike in the last row. According to multivariate Qlike, RMSV-GFT provides more accurate forecasts of realized covariance than that of MSV-GFT.

Figure 17 plots the out-of-sample forecasts for the realized measures. For the realized variance, it shows MSV-GFT is less precise from Jan 2009 to Dec 2010. And MSV-GFT provides more downward outliers for the forecasts of realized correlation than RMSV-GFT.

4.4.3.3 Global Minimum Variance Portfolio Analysis

In this section, we compare alternative models in terms of their economic value. In particular, we use the forecasted covariance to construct a Global Minimum

Variance (GMV) portfolio in each candidate model. According to Markowitz (1952), the GMV portfolio is optimal in the sense that it has the smallest variance among all portfolios on the efficient frontier. In our application, at period $t - 1$, we construct the GMV portfolio with optimal weights $w_t = (w_{1t}, ..., w_{pt})$, which minimize the variance of the portfolio return over the next period. The optimal weights are a unique solution to

$$\min_{w_t \in \mathbb{R}^p} w_t' \hat{\Sigma}_{t|t-1} w_t, \quad \text{s.t. } w_t' \mathbb{1}_p = 1, \tag{4.19}$$

where $\hat{\Sigma}_{t|t-1}$ is the forecast of the covariance in (4.15), and $\mathbb{1}_p$ is a p-by-1 vector of ones.

In general, negative weights are allowed so that short-sells are possible. The well-known solution of w_t is

$$w_t^\star = \frac{\hat{\Sigma}_{t|t-1}^{-1} \mathbb{1}_p}{\mathbb{1}_p' \hat{\Sigma}_{t|t-1}^{-1} \mathbb{1}_p}, \tag{4.20}$$

and the optimal portfolio return at time t is

$$R_t^p = w_t^{\star\prime} r_t. \tag{4.21}$$

We construct the GMV portfolio using MSV-GFT and RMSV-GFT, respectively. As a benchmark, an equal-weighted portfolio is also constructed and its return is

$$R_{ew}^p = \frac{1}{p} \mathbb{1}_p' r_t.$$

Figure 18 shows the out-of-sample optimal weights of the MSV-GFT, RMSV-GFT and equal-weighted portfolios. According to Chopra and Ziemba (1993), the estimation error in the expected return may not lead to an optimal portfolio. To facilitate comparison, we assume all stocks have equal expected returns and focus on the variance of the portfolios. The variance of the portfolios is measured by on average squared return, which is $\frac{1}{480} \sum_{t=1743}^{2242} (R_t^p)^2$. Since the average squared return is sensitive to outlier, we also compute average absolute return, $\frac{1}{480} \sum_{t=1743}^{2242} |R_t^p|$ as a robustness check.

Table 13 reports the out-of-sample variance under MSV-GFT, RMSV-GFT and

Table 13: Out-of-Sample Portfolio

	period	Equal Weight	MSV-GFT	RMSV-GFT
Squared Return	2008-2010	26.0852	19.1104	**19.0969**
	2008-2009	25.8988	**20.1271**	20.3104
	2009-2010	25.914	17.832	**17.7604**
Absolute Return	2008-2010	3.4972	3.2023	**3.1815**
	2008-2009	3.7123	3.4155	**3.3773**
	2009-2010	3.2822	2.9892	**2.9858**

1.Bold figures indicates the optimal values

2.We report the result for $100 \times R_t^p$ return for precision.

3.Squared return is $\frac{1}{480} \sum_{t=1743}^{2242} (R_t^p)^2$ and Absolute Return is defined as $\frac{1}{480} \sum_{t=1743}^{2242} |R_t^p|$.

equal-weight. The equal-weighted portfolio has a larger variance than the MSV-GFT and RMSV-GFT portfolios. In the full out-of-sample period, the RMSV-GFT model produces the smallest variance of the portfolio. We also report the out-of-sample variance for each year in the out-of-sample period. Table 13 shows that in 2008, MSV-GFT constructs the smallest variance portfolio while the average absolute return is larger than RMSV-GFT. In 2009, the portfolio based on RMSV-GFT dominates the other two portfolios.

In summary, we can conclude that incorporating realized measures in the MSV model leads to a better portfolio.

Figure 14: MSV-GFT: In-sample comparison with realized measure
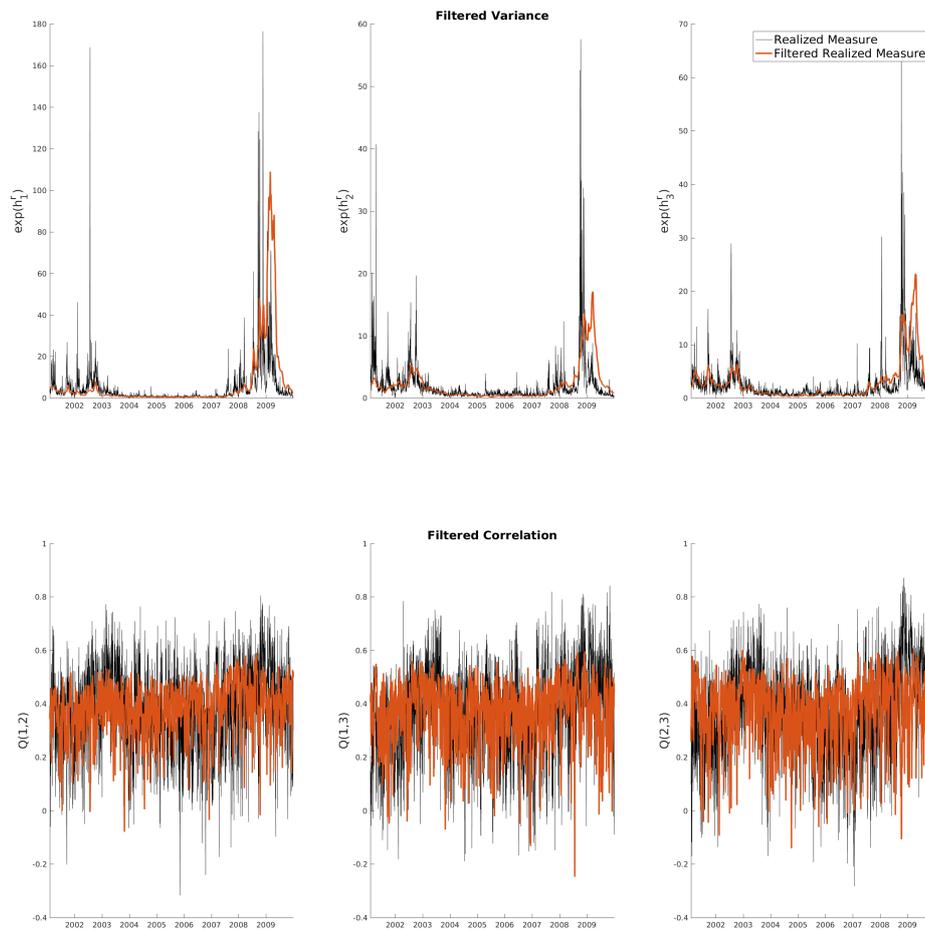
Note: The top three panels report in-sample fit of realized variance for BAC, JPM and AXP respectively; The bottom three panels report in-sample fit of the realized correlation for BAC and JPM, BAC and AXP, JPM and AXP respectively. The time period is from January 2, 2001 to December 31, 2009.

Figure 15: RMSV-GFT: In-sample comparison with realized measure
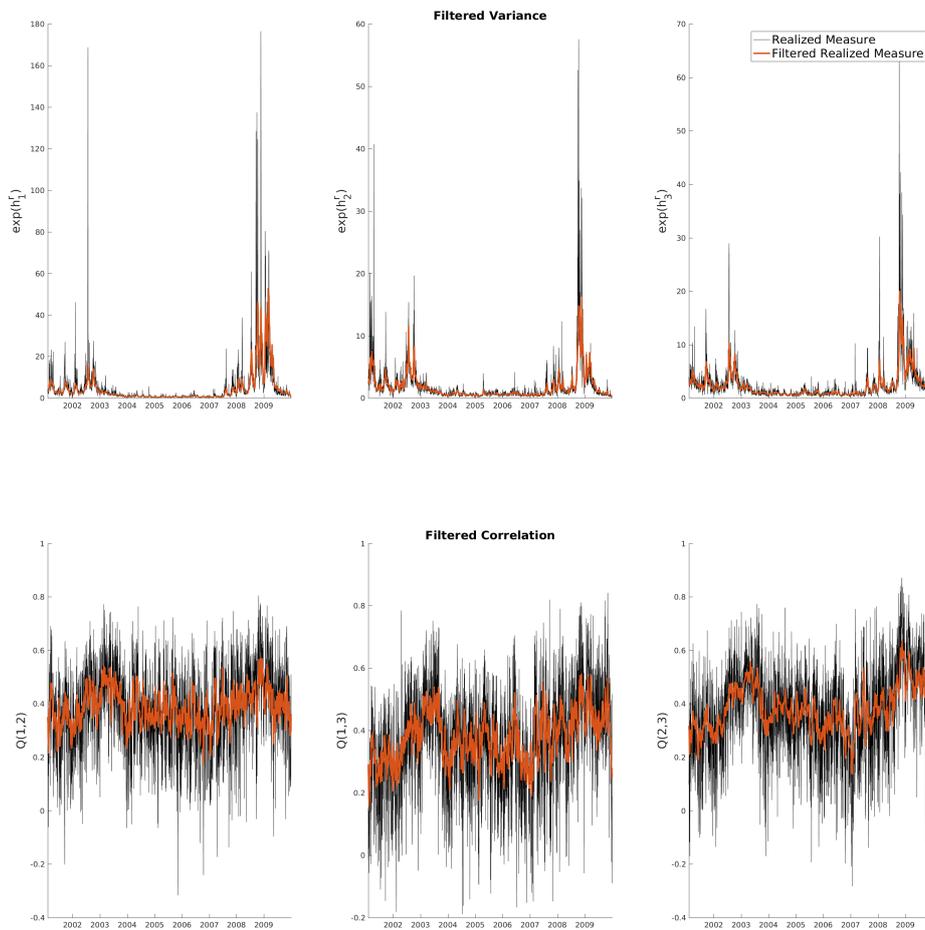
Note: The top three panels report in-sample fit of the realized variance for BAC, JPM and AXP respectively; The bottom three panels report in-sample fit of the realized correlation for BAC and JPM, BAC and AXP, JPM and AXP respectively. The time period is from January 2, 2001 to December 31, 2009.

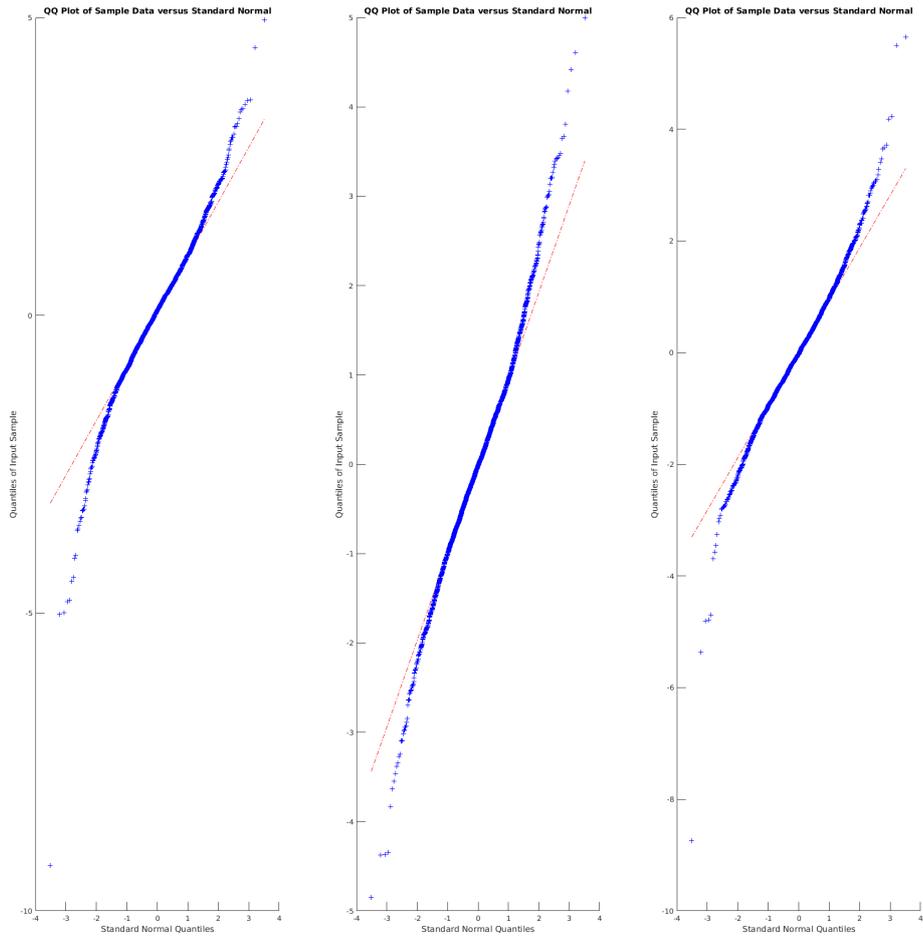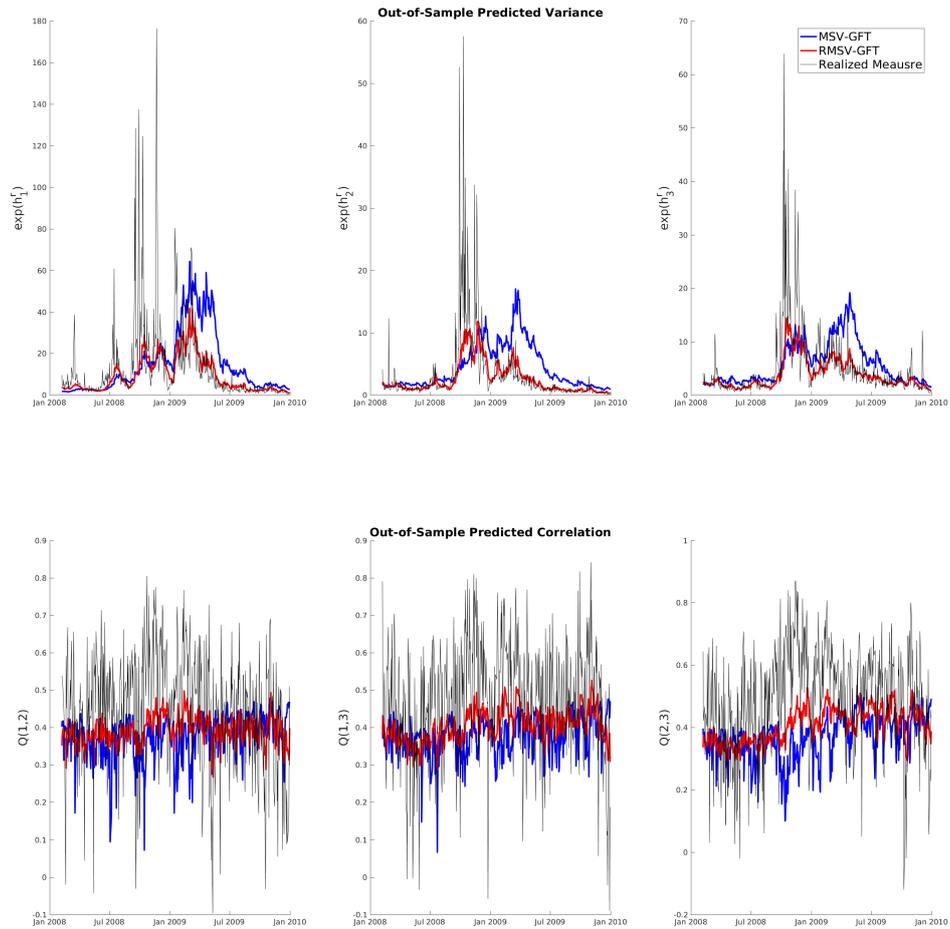Figure 16: Q-Q plot of the residuals in RMSV-GFT

Figure 17: Out-of-sample prediction of realized measures

Note: The top three panels report out-of-sample fit of realized variance for BAC, JPM and AXP respectively; The bottom three panels report out-of-sample fit of realized correlation for BAC and JPM, BAC and AXP, JPM and AXP respectively. The time period RMSV-GFT

Figure 18: Time series plot of the portfolio weights for the GMV portfolio
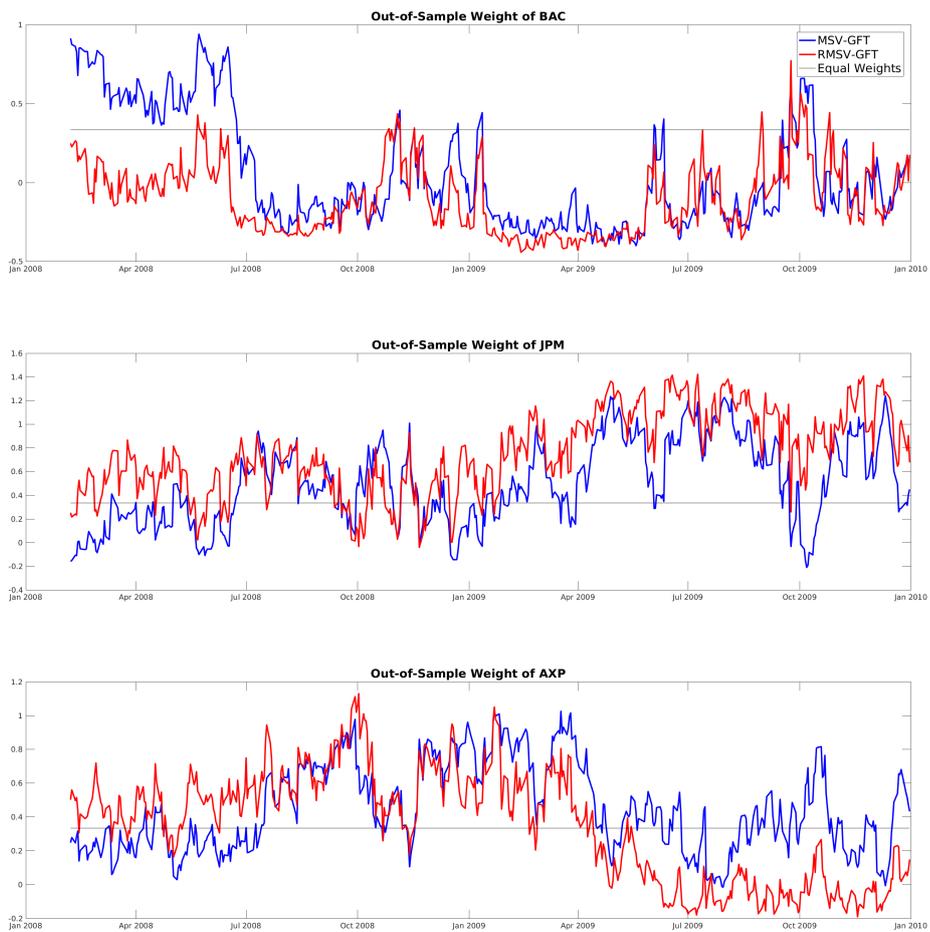
5 MSV-GFT Model with Leverage Effect and Heavy-Tailed Error

We are motivated by Figure 16, which gives the Q-Q plot of residuals from the estimated MSV model without the leverage effect or heavy-tailed error distribution. Clearly, the Q-Q plot suggests that the estimated MSV model is inadequate in two aspects. First, there are large movements in the market that cannot be explained by the estimated model. Second, the Q-Q plot is not symmetric. In this chapter, we consider extending the basic model (MSV-GFT) by incorporating the leverage-effect and heavy-tailed error distribution.

5.1 Model Setup

Following Harvey et al. (1994), we assume that $\epsilon_t = \lambda_t^{\frac{1}{2}} z_t$, where z_t follows as the p-dimensional standard normal distribution, $\lambda_t = diag\{\lambda_{1t}, ... \lambda_{pt}\}$, and

$$\lambda_{it}|\nu_i \overset{iid}{\sim} \nu_i/\chi^2(\nu_i),, \text{ where } i = 1, ..., p. \tag{5.1}$$

Under this parameterization, $\epsilon_{it} \overset{iid}{\sim} t_{\nu_i}$. The tail thickness in the distribution of ϵ_{it} is controlled by ν_i.

To allow for an asymmetric effect on return on expected volatility, we assume

$$\begin{pmatrix} z_t \\ \eta_{ht} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_p & \Sigma_h^{\frac{1}{2}}\Omega \\ \Omega' \left(\Sigma_h^{1/2}\right)' & \Sigma_h \end{pmatrix} \right), \tag{5.2}$$

where Ω is a $p \times p$ matrix. Let ρ_{ij} denote the $(i, j)^{th}$ element of Ω, that is, $\rho_{ij} = corr(z_{it}, \eta_{h_j t})$. When $p = 1$, if $\Omega = \rho = corr(z_{1t}, \eta_{h_1 t}) < 0$, our model is the same as that in the univariate literature (Yu (2005)). If $\Omega = 0$, our model is the same as the original model. When the self-leverage and cross-leverage effects exist for all assets, we expect all the elements Ω be negative.

Combining Equations (3.1)-(3.6), (5.1) and (5.2), we have the new MSV model

with leverage effects and t distributed error term (MSVLt-GFT). In particular, the MSVLt-GFT model is given by

$$r_t = V_t^{1/2}\epsilon_t, \ \epsilon_t = \lambda_t^{\frac{1}{2}} z_t, \ z_t \sim N(0, R_t), \ t = 1, ..., T,$$

$$\lambda_{it}|\nu \sim v_i/\chi^2(v_i), \ i = 1, ..., p, \ t = 1, ..., T,$$

$$V_t = \exp(H_t), \text{ and } h_t = diag(H_t), \ t = 1, ..., T,$$

$$q_t = F(R_t), \ t = 1, ..., T, \tag{5.3}$$

$$h_{t+1} = \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht}, \ \eta_{ht} \sim N(0, \Sigma_h), \ t = 1, ..., T-1,$$

$$q_{t+1} = \mu_q + \phi_q(q_t - \mu_q) + \eta_{qt}, \ \eta_{qt} \sim N(0, \Sigma_q), \ t = 1, ..., T-1,$$

$$h_0 \sim N(\mu_h, (I_p - \phi_h^2)^{-1}\Sigma_h), \text{ and } q_0 \sim N(\mu_q, (I_d - \phi_q^2)^{-1}\Sigma_q),$$

and

$$\begin{pmatrix} z_t \\ \eta_{ht} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_p & \Sigma_h^{\frac{1}{2}}\Omega \\ \Omega'\left(\Sigma_h^{1/2}\right)' & \Sigma_h \end{pmatrix} \right),$$

The parameters of the model are $\theta \equiv (\mu_h, \mu_q, \phi_h, \phi_q, \Sigma_h, \Sigma_q, \Omega, \nu)$.

5.2 Bayesian Analysis of MSVLt-GFT

5.2.1 Estimation Procedure

Let $r = (r_1', \ldots, r_T')$, $\lambda = (\lambda_1', \ldots, \lambda_T')$, $h = (h_1', \ldots h_T')$ and $q = (q_1', \ldots q_T')$. To conduct the inference, we implement a Gibbs sampler with four blocks. In the following, we use θ/α to denote the parameters θ excluding α. Then, the algorithm proceeds as follows.

1. Initialize h, q and θ.

2. Draw $h, q|r, m, \lambda, \theta$.

3. Draw $\mu_h, \mu_q|h, q, \theta_{/(\mu_h, \mu_q)}$.

4. Draw $\phi_h, \phi_q|h, q, \theta_{/(\phi_h, \phi_q)}$.

5. Draw $\Sigma_q, \Sigma_h|h, q, \theta_{/(\Sigma_h, \Sigma_q)}$.

6. Draw $\lambda, \nu | r, h, q$.

7. Draw $\Omega | r, h, q, \theta_{/\Omega}$.

Iterating over Steps 2-7 consists of a complete sweep of MCMC sampler. In Step 2, we replace return r with $\lambda^{-\frac{1}{2}} r$ and apply the PGAS introduced in the Section 3.2.2 to sample from the latent variables h and q given $\lambda^{-\frac{1}{2}} r$ and one particular set of parameter values. Steps 3-5 sample from the conditional posterior distribution in (3.11), (3.12) and (3.13). We will describe Steps 6 and 7 in the following section.

5.2.2 Sampling Leverage Effects

The joint distribution of (r_t, h_{t+1}) given h_t is

$$
\left(\begin{array}{c|c} r_t \\ h_{t+1} \end{array} \; h_t \right) \sim N \left(\left(\begin{array}{c} 0 \\ \mu_h + \phi_h(h_t - \mu_h) \end{array} \right), \left(\begin{array}{cc} \lambda_t V_t^{\frac{1}{2}} R_t V_t^{\frac{1}{2}} & \lambda_t^{\frac{1}{2}} V_t^{\frac{1}{2}} R_t^{\frac{1}{2}} \Sigma_h^{\frac{1}{2}} \Omega \\ \Omega' \left(\Sigma_h^{1/2} \right)' \left(R_t^{1/2} \right)' V_t^{1/2} \lambda_t^{1/2} & \Sigma_h \end{array} \right) \right).
$$

(5.4)

The conditional distribution of $h_{t+1} | h_t, r_t, \theta$ is

$$
h_{t+1} | h_t, r_t, q_t, \theta \sim N(\mu_h + \phi_h(h_t - \mu_h)) + \Omega \Sigma_h^{\frac{1}{2}} z_t, \Psi)
$$

(5.5)

where $z_t = R_t^{-\frac{1}{2}} V_t^{-\frac{1}{2}} \lambda_t^{\frac{1}{2}} r_t$ and $\Psi = \Sigma_h - \Omega \Sigma_h \Omega'$ is a positive definite matrix. Following the literature, we assume that the prior of (Ω, Ψ) is from the Normal-Inverse-Wishart family:

$$
vec(\Omega) | \Psi \sim N(vec(b), \Psi \otimes H_0),
$$
$$
\Psi \sim IW(\tilde{\nu}, \tilde{\Psi}).
$$

(5.6)

We sample Ω from

$$
vec(\Omega) | r, h, q, \theta_{/\Omega} \sim N(vec(B), \Psi \otimes H_1),
$$

(5.7)

83

where

$$B = H_1 \left(\sum_{t=1}^{T-1} \Sigma_h^{\frac{1}{2}} z_t \eta_t' + H_0^{-1} b \right),$$

$$H_1 = \left(\sum_{t=1}^{T-1} z_t \Sigma_h z_t' + H_0^{-1} \right)^{-1}, \tag{5.8}$$

with $\eta_t = h_{t+1} - \mu_h - \phi_h (h_t - \mu_h)$ and $z_t = R_t^{-\frac{1}{2}} V_t^{-\frac{1}{2}} \lambda_t^{\frac{1}{2}} r_t$.

5.2.3 Sampling λ and ν

To implement MCMC for λ and ν, we assume the conjugate prior for $\lambda_{it}|\nu_i \sim IG(\nu_i/2, 2/\nu_i)$ for $t = 1, ..., T$, $i = 1, ..., p$ and $\nu_i \sim U(3, 40)$ for $i = 1, ..., p$. Moreover, we denote the prior of ν as $\pi(\nu)$.

Under the assumption of the heavy-tailed distribution, conditional on V_t and R_t, r_t follows a multivariate t-distribution with mean 0 and variance $V_t^{\frac{1}{2}} R_t V_t^{\frac{1}{2}}$ with ν degrees of freedom. We construct $V_t^{-\frac{1}{2}} R_t^{-\frac{1}{2}} r_t$ so that multiple degrees of freedom are supported in the multivariate t distribution. The joint distribution of (λ, ν) is:

$$p(\lambda_t, \nu | r_t, h_t, q_t) \propto p(r_t | \lambda_t, h_t, q_t, \nu) p(\lambda_t | \nu) p(\nu)$$

$$\propto p(\nu) \left\{ \frac{\left(\frac{\nu}{2} \right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2} \right)} \right\}^T \prod_{t=1}^{T} \lambda_t^{-\frac{p+\nu}{2} - 1} \tag{5.9}$$

$$\times \exp\left[-\frac{1}{2} \sum_{t=1}^{T} \left\{ \frac{\nu}{\lambda_t} + r_t (V_t^{\frac{1}{2}} R_t V_t^{\frac{1}{2}})^{-1} r_t \right\} \right].$$

Sampling λ. Given the conjugate prior, the conditional posterior of $\lambda_t = (\lambda_{1t}, ..., \lambda_{pt})$ is

$$p(\lambda_t | r_t, h_t, q_t, \nu) \propto \lambda_t^{\frac{\nu+p}{2} - 1} \exp\left\{ -\frac{1}{2\lambda_t} r_t' (V_t^{\frac{1}{2}} R_t V_t^{\frac{1}{2}})^{-1} r_t \right\}, \tag{5.10}$$

which follows as $IG\left(\frac{v+p}{2}, \frac{r_t' (V_t^{\frac{1}{2}} R_t V_t^{\frac{1}{2}})^{-1} r_t}{2} \right)$.

Sampling ν. The conditional posterior distribution of $\nu = (\nu_1, .., \nu_p)$ is

$$p(\nu|r,h,q,\lambda) \propto p(\lambda|\nu)p(\nu) \propto p(\nu) \left\{ \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \right\}^T \left(\prod_{t=1}^n \lambda_t \right)^{-\frac{\nu}{2}} \exp\left\{ -\frac{1}{2}\sum_{t=1}^T \frac{\nu}{\lambda_t} \right\}.$$
(5.11)

To draw random samples from the full conditional posterior density of ν, one can resort to the random walk Metropolis-Hasting sampler. We sample $\nu^\star$ from $N(\nu^{(i-1)}, I_p)$, then accept it with probability $min\left[1, \frac{p(\nu^\star|r,h,q,\lambda)}{p(\nu^{i-1}|r,h,q,\lambda)}\right]$, where ν^{i-1} is the sample from the most recent MCMC iteration.

5.3   Simulation Studies

To investigate the performance of our estimation procedure, we conduct a simulation study in this section. Our simulation design is frequentist as we fix the parameters at their true value and generate data from the same data generating process for 100 times. However, we use the posterior mean as a point estimator for all the parameters in the model. Since we know the true value of those parameters, we are able to calculate bias (defined as the difference between the true value and the average value of the posterior means across 100 replications) and the standard deviation (across 100 replications).

The number of assets considered for simulation is three. There are 24 parameters whose true values are given by:

1. $\mu_{h1} = \mu_{h2} = \mu_{h3} = 0.3$ and $\mu_{q1} = \mu_{q2} = \mu_{q3} = 0.7$,

2. $\phi_{h1} = \phi_{h2} = \phi_{h3} = 0.9$ and $\phi_{q1} = \phi_{q2} = \phi_{q3} = 0.8$,

3. $\sigma_{h1}^2 = \sigma_{h2}^2 = \sigma_{h3}^2 = 0.05$ and $\sigma_{q1}^2 = \sigma_{q2}^2 = \sigma_{q3}^2 = 0.05$,

4. $\Omega_{11} = \Omega_{22} = \Omega_{33} = -0.5$,

5. $\nu_1 = \nu_2 = \nu_3 = 20$.

All the simulation results reported in this section are based on 5000 MCMC iterations, among which the first 1000 samples are discarded as a burn-in period. We consider sample size $T = 1000$ and number of particles $N = 200$.

Table 14: Simulation results for MSV-GFT with leverage effects and heavy-tailed errors

	μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}
true value	0.3	0.3	0.3	0.7	0.7	0.7
mean	0.177	0.179	0.170	0.655	0.656	0.653
std	0.121	0.129	0.114	0.034	0.036	0.036
IF	70.415	54.570	62.127	10.487	11.227	9.212
	ϕ_{h1}	ϕ_{h2}	ϕ_{h3}	ϕ_{q1}	ϕ_{q2}	ϕ_{q3}
true value	0.9	0.9	0.9	0.8	0.8	0.8
mean	0.898	0.891	0.893	0.787	0.783	0.792
std	0.036	0.034	0.044	0.036	0.060	0.049
IF	71.729	75.685	70.782	95.130	98.740	95.485
	σ_{h1}^2	σ_{h2}^2	σ_{h3}^2	σ_{q1}^2	σ_{q2}^2	σ_{q3}^2
true value	0.05	0.05	0.05	0.05	0.05	0.05
mean	0.046	0.048	0.047	0.037	0.037	0.035
std	0.013	0.009	0.013	0.010	0.010	0.011
IF	140.810	140.800	140.690	154.610	160.540	162.440
	Ω_{11}	Ω_{22}	Ω_{33}	ν_1	ν_2	ν_3
true value	-0.5	-0.5	-0.5	20	20	20
mean	-0.350	-0.34	-0.344	12.868	12.528	11.964
std	0.041	0.036	0.038	7.351	7.213	7.196
IF	22.943	18.823	21.961	235.140	206.800	226.490

1. Mean is the average posterior mean across replications.
2. Std is the standard error of the posterior mean across replications.
3. IF is the average inefficiency factor across replications calculated as suggested in Kim et al. (1998).

Table 14 reports simulation results including the posterior mean, posterior standard deviation, as well as the inefficiency factor for all parameters. For the leverage effects, Ω, we find upward bias in the proposed estimators. For ν, there is a downward bias in the proposed estimators. However, the bias decreases when the sample size increases. For parameters related to correlation, $(\mu_q, \phi_q, \sigma_q^2)$, the performance of the proposed method is good. For parameters related to volatility, the performance of the proposed method is good for ϕ_h and σ_h^2 while there is a noticeable downward bias for μ_h. This bias is caused by the downward bias in estimating ν. A smaller ν causes a larger λ. Hence, $\lambda_t^{-\frac{1}{2}} r_t$ is smaller, leading to a downward bias in μ_h. On the other hand, compared with the basic MSV-GFT model, we find that the IF of estimators of μ_h is larger than that reported in Section 3.2.5. This finding suggests that adding the heavy-tailed error distribution to our MSV model affects the convergent speed of MCMC sampler of μ_h.

5.4 Empirical Application

5.4.1 Data

In this section, we consider the empirical application based on the daily returns of three Asian equity indices. The data contains 2237 daily mean-corrected log-returns of three stock market indices, namely the Hong Kong Hang Seng Index, the Nikkei 225 Index of Tokyo Stock Exchange, and the SSE Composite Index of Shanghai Stock Exchange. The sample period is from January 5, 2005 to December 30, 2014 that covers the global financial crisis period. This exercise is interesting, as it can provide some insights on the co-movement of these three arguably most important stock market indices in Asia. The three series are plotted in Figure 19.

5.4.2 Estimation

We estimate and compare four models with dynamic correlation defined as follows.

1. MSV-GFT. The model was proposed in Section 3. It is given by

$$
\begin{aligned}
r_t &= V_t^{1/2}\epsilon_t,\ \epsilon_t \sim N(0, R_t) \\
V_t &= \exp(H_t),\ h_t = diag(H_t),\ q_t = F(R_t) \\
h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht} \\
q_{t+1} &= \mu_q + \phi_q(q_t - \mu_q) + \eta_{qt}
\end{aligned}
\tag{5.12}
$$

2. MSVL-GFT: The MSV-GFT model with leverage effects only. It is given by

$$
\begin{aligned}
r_t &= V_t^{1/2}\epsilon_t,\ \epsilon_t \sim N(0, R_t) \\
V_t &= \exp(H_t),\ h_t = diag(H_t),\ q_t = F(R_t) \\
h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht} \\
q_{t+1} &= \mu_q + \phi_q(q_t - \mu_q) + \eta_{qt}
\end{aligned}
\tag{5.13}
$$

Figure 19: Time series of stock index returns
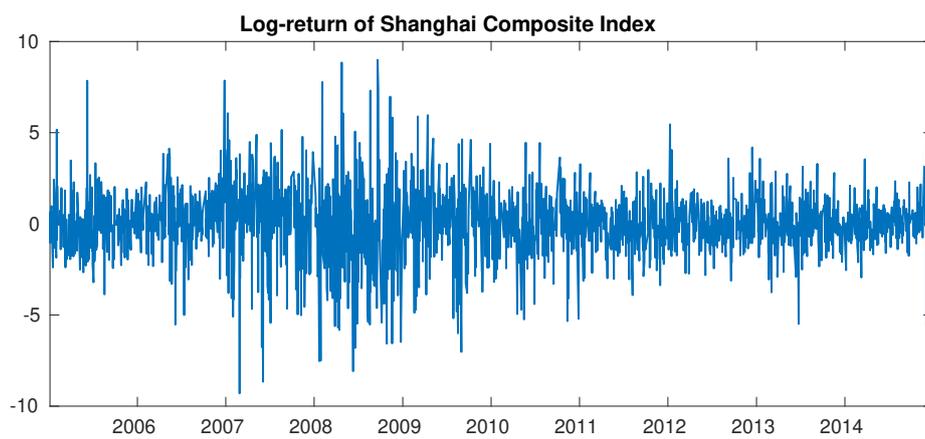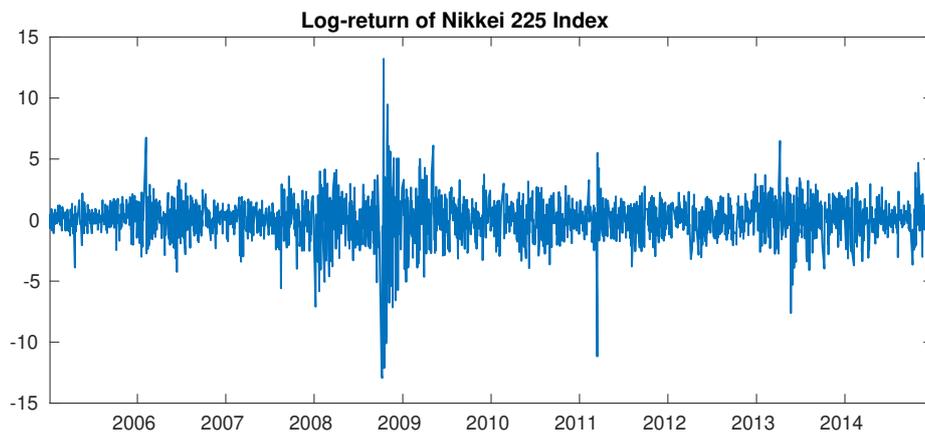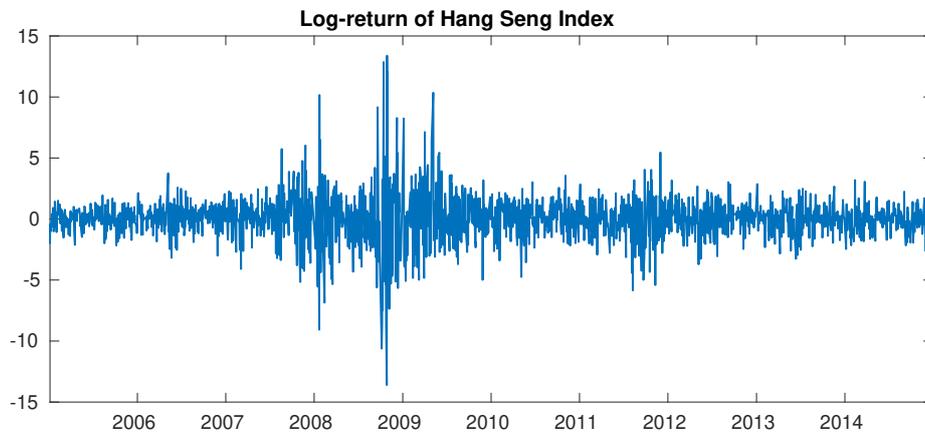
$$\begin{pmatrix} \epsilon_t \\ \eta_{ht} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_p & \Sigma_h^{\frac{1}{2}}\Omega \\ \Omega'\Sigma_h^{'\frac{1}{2}} & \Sigma_h \end{pmatrix} \right),$$

3. MSVt-GFT: The MSV-GFT model with heavy-tailed errors only. It is given by

$$\begin{aligned} r_t &= \lambda_t^{\frac{1}{2}} V_t^{1/2} \epsilon_t, \ \epsilon_t \sim N(0, R_t), \\ V_t &= \exp(H_t), \ h_t = diag(H_t), \ q_t = F(R_t), \ \lambda_t|\nu \ \sim v/\chi^2(v), \\ h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht}, \\ q_{t+1} &= \mu_q + \phi_q(q_t - \mu_q) + \eta_{qt}. \end{aligned} \tag{5.14}$$

4. MSVLt-GFT: The model is given in Equation (5.2) and (5.3).

For each model, we draw 20000 MCMC samples and discard first 5000 as a burn-in period. For the MSVLt-GFT model, we plot the posterior mean of log-volatility and correlation in Figure 20 and Figure 21, respectively. Figure 22 shows the Q-Q plot of the residuals ϵ_t in the estimated MSVLt-GFT model. The Q-Q plot suggests that the residuals follow the normal distribution very well. We reports the posterior means, standard deviations, the 95% credible intervals, inefficiency factors in Table 15, 16 and 17.

Table 15 reports the posterior statistics related to volatility. The estimators of ϕ_h and σ_h^2 are all reasonable. Moreover, the estimators of μ in the heavy-tailed model are smaller than that in MSV-GFT and MSVL-GFT. This is also reasonable as more large movements in the stock markets can be explained by the t distribution in the MSVLt-GFT model than that by the MSV-GFT and MSVL-GFT models. As expected, the MSVLt-GFT model leads to a lower and smoother h series than that in the basic model. Figure 23 compares the posterior mean of the h sequence from MSV-GFT and MSVt-GFT and confirms the expectation.

Table 16 reports the posterior statistics related to the correlation. The posterior means, posterior standard deviations are similar to each other. That is consistent with the expectation; that is, the leverage effects and heavy-tailed error term should not affect correlation.

Figure 20: Compare filtered h of MSV-GFT and MSVt-GFT

filtered log-volatility of HSI

filtered log-volatility of N225

filtered log-volatility of SSE

Figure 21: Compare filtered h of MSV-GFT and MSVt-GFT

filtered correlation between HSI and N225

filtered correlation between N225 and SSE

filtered correlation between SSE and HSI
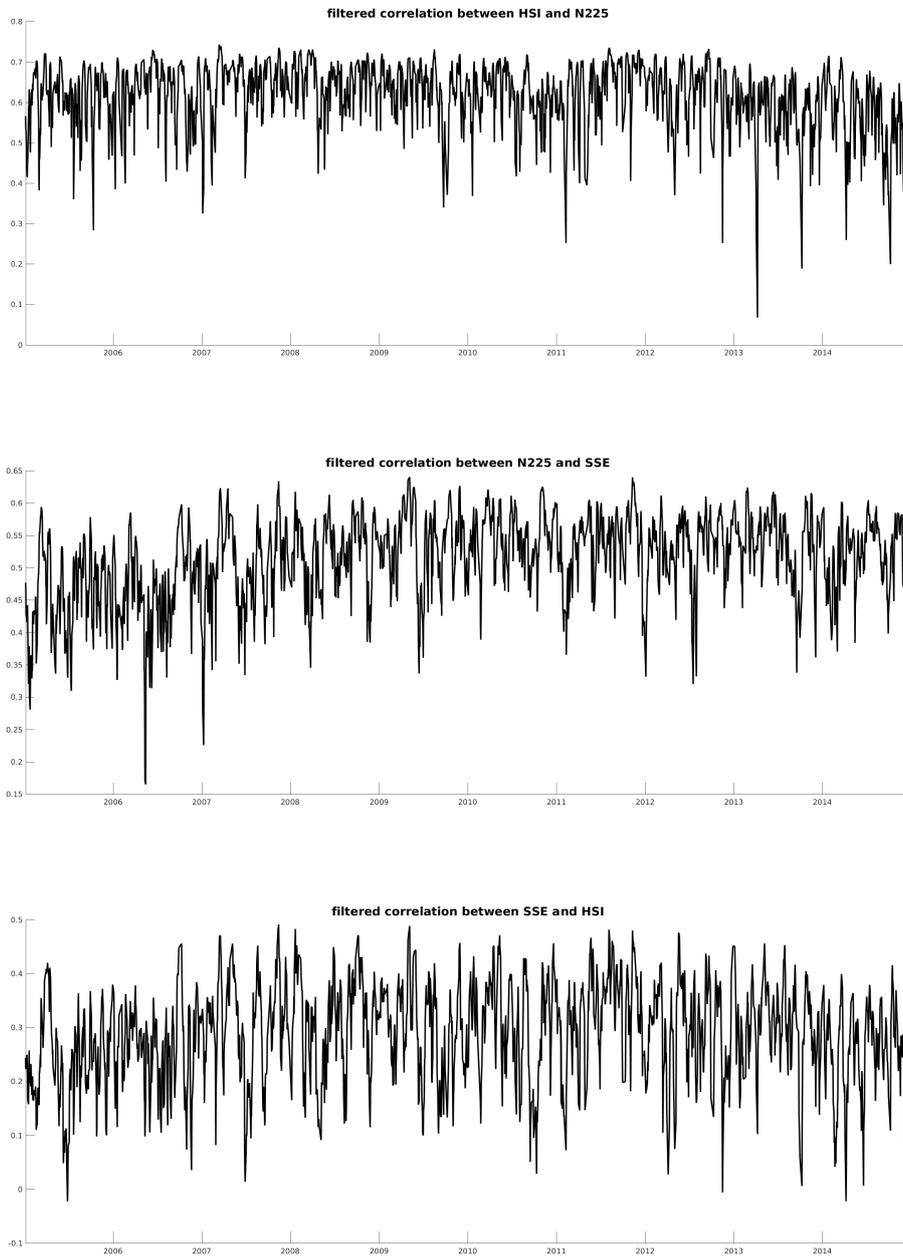
Figure 22: Q-Q plot of the residuals in MSVLt-GFT
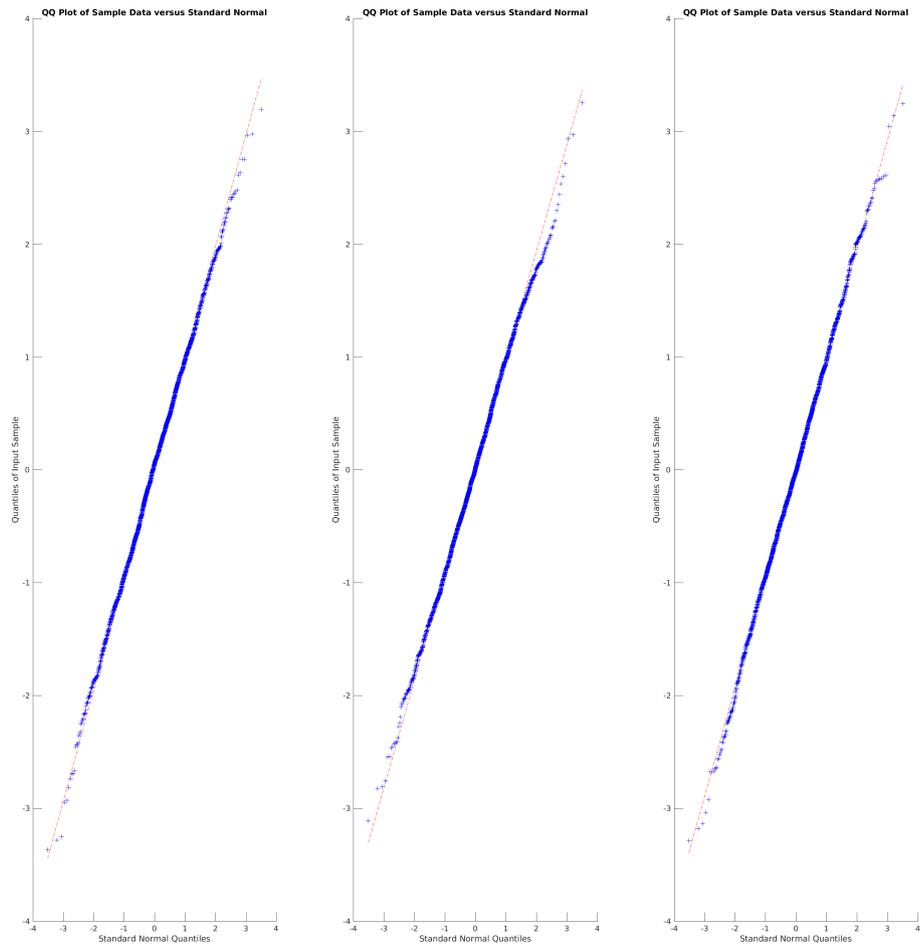
Figure 23: Compare filtered h of MSV-GFT and MSVt-GFT

Table 17 reports the posterior statistics related to the leverage effects and parameters in the error distribution. Several findings emerge. Firstly, all leverage effects are negative, which is consistent with the expectation. However, the leverage effect of SSE is not very strong. Secondly, in the literature, Ishihara and Omori (2012) have proposed an MSV model with the error term following a multivariate t-distribution with the same degrees of freedom. However, we find evidence that the error terms of the three stock indices have quite different degrees of freedom parameter. In particular, the error terms for HSI and SSE have much smaller νs than N225.

We can compare the four alternative specifications using the deviance information criterion (DIC) and log marginal likelihood. The details about the DIC and log marginal likelihood can be found in Section 3.2.4. The last two rows of Table 17 reports the DIC and log-marginal likelihood. The smaller the DIC, the better the model. The difference between the two log marginal likelihood values gives the log Bayes factor. The larger the log marginal likelihood, the better the model. The best model to describe the stock indices is the MSVLt-GFT model, followed by MSVt-GFT, then by MSVL-GFT, and finally by MSV-GFT. Comparing only the MSVt-GFT and MSVL-GFT models, we find that MSVt-GFT performs much better than MSVL-GFT. Hence, as far as this dataset is concerned, it is more important to model heavy-tailed errors than to model the leverage effects.

Table 15: Posterior statistics of parameters in the h sequences for competing models based on the stock index

		MSV-GFT	MSVL-GFT	MSVt-GFT	MSVLt-GFT
μ_{h1}	Mean	0.373	0.382	0.101	0.108
	SD	0.256	0.277	0.297	0.293
	95% CI	[-0.146,0.886]	[-0.184,0.933]	[-0.497,0.684]	[-0.482,0.690]
	IF	0.917	0.893	7.494	6.854
μ_{h2}	Mean	0.523	0.524	0.473	0.454
	SD	0.171	0.177	0.193	0.175
	95% CI	[0.182,0.863]	[0.169,0.879]	[0.085,0.848]	[0.104,0.792]
	IF	1.170	1.192	2.223	2.775
μ_{h3}	Mean	0.753	0.755	0.356	0.357
	SD	0.205	0.221	0.259	0.258
	95% CI	[0.345,1.161]	[0.306,1.188]	[-0.164,0.865]	[-0.163,0.854]
	IF	0.618	1.435	8.808	5.965
ϕ_{h1}	Mean	0.991	0.989	0.992	0.991
	SD	0.003	0.003	0.002	0.003
	95% CI	[0.984,0.995]	[0.982,0.995]	[0.986,0.995]	[0.985,0.995]
	IF	26.915	42.531	32.147	43.485
ϕ_{h2}	Mean	0.980	0.976	0.981	0.975
	SD	0.006	0.007	0.006	0.008
	95% CI	[0.967,0.991]	[0.962,0.988]	[0.968,0.991]	[0.958,0.988]
	IF	44.186	71.412	46.472	60.339
ϕ_{h3}	Mean	0.983	0.984	0.992	0.992
	SD	0.005	0.006	0.003	0.003
	95% CI	[0.971,0.993]	[0.970,0.993]	[0.984,0.995]	[0.985,0.995]
	IF	52.526	113.065	118.505	81.791
σ_{h1}^2	Mean	0.012	0.016	0.011	0.013
	SD	0.003	0.004	0.003	0.003
	95% CI	[0.007,0.018]	[0.010,0.024]	[0.007,0.019]	[0.008,0.021]
	IF	98.436	145.461	132.573	159.462
σ_{h2}^2	Mean	0.021	0.032	0.025	0.035
	SD	0.005	0.007	0.006	0.009
	95% CI	[0.013,0.033]	[0.020,0.048]	[0.015,0.039]	[0.021,0.056]
	IF	101.608	151.118	106.359	112.252
σ_{h3}^2	Mean	0.021	0.022	0.009	0.009
	SD	0.005	0.007	0.003	0.003
	95% CI	[0.012,0.033]	[0.011,0.039]	[0.005,0.018]	[0.005,0.017]
	IF	130.360	195.886	302.221	205.034

1. Mean is the posterior mean based on 20000 MCMC draws after a 5000 burn-in period.
2. SD is the numerical standard error of the posterior means.
3. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws.
4. IF is the inefficiency factor.

Table 16: Posterior statistics of parameters in the q sequences for competing models based on the stock index

		MSV-GFT	MSVL-GFT	MSVt-GFT	MSVLt-GFT
μ_{q1}	Mean	0.715	0.727	0.717	0.711
	SD	0.028	0.029	0.028	0.028
	95% CI	[0.659,0.769]	[0.670,0.784]	[0.661,0.771]	[0.656,0.765]
	IF	13.960	30.399	20.105	16.935
μ_{q2}	Mean	0.542	0.557	0.544	0.546
	SD	0.026	0.027	0.027	0.028
	95% CI	[0.490,0.594]	[0.503,0.611]	[0.490,0.598]	[0.491,0.601]
	IF	11.287	13.140	14.872	18.844
μ_{q3}	Mean	0.125	0.130	0.142	0.144
	SD	0.024	0.025	0.027	0.027
	95% CI	[0.077,0.173]	[0.080,0.178]	[0.089,0.193]	[0.090,0.197]
	IF	21.576	25.492	9.119	10.134
ϕ_{q1}	Mean	0.556	0.575	0.519	0.545
	SD	0.107	0.107	0.112	0.114
	95% CI	[0.328,0.751]	[0.349,0.757]	[0.284,0.717]	[0.296,0.740]
	IF	146.187	231.513	223.231	162.251
ϕ_{q2}	Mean	0.660	0.725	0.643	0.614
	SD	0.132	0.110	0.148	0.155
	95% CI	[0.365,0.860]	[0.465,0.886]	[0.336,0.887]	[0.290,0.883]
	IF	198.705	266.335	320.489	325.425
ϕ_{q3}	Mean	0.617	0.694	0.681	0.672
	SD	0.127	0.108	0.099	0.089
	95% CI	[0.335,0.827]	[0.424,0.854]	[0.468,0.841]	[0.483,0.822]
	IF	169.348	193.125	175.052	151.433
σ_{q1}^2	Mean	0.064	0.083	0.083	0.075
	SD	0.020	0.026	0.022	0.020
	95% CI	[0.032,0.108]	[0.044,0.147]	[0.043,0.129]	[0.041,0.116]
	IF	162.667	274.750	223.332	184.875
σ_{q2}^2	Mean	0.028	0.030	0.039	0.041
	SD	0.011	0.012	0.016	0.017
	95% CI	[0.011,0.053]	[0.013,0.056]	[0.012,0.074]	[0.012,0.080]
	IF	285.125	330.669	379.679	328.221
σ_{q3}^2	Mean	0.020	0.023	0.038	0.039
	SD	0.008	0.011	0.015	0.013
	95% CI	[0.007,0.040]	[0.010,0.053]	[0.015,0.072]	[0.019,0.065]
	IF	293.989	335.082	248.723	221.188

1. Mean is the posterior mean based on 20000 MCMC draws after a 5000 burn-in period.
2. SD is the numerical standard error of the posterior means.
3. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws.
4. IF is the inefficiency factor.

Table 17: Posterior statistics of parameters in the leverage and heavy-tail sequences for competing models based on the stock index

		MSV-GFT	MSVL-GFT	MSVt-GFT	MSVLt-GFT
Ω_{11}	Mean		-0.239		-0.196
	SD		0.052		0.057
	95% CI		[-0.338,-0.136]		[-0.306,-0.083]
	IF		30.238		31.924
Ω_{22}	Mean		-0.261		-0.272
	SD		0.050		0.048
	95% CI		[-0.358,-0.163]		[-0.365,-0.177]
	IF		41.986		39.015
Ω_{33}	Mean		-0.054		-0.007
	SD		0.051		0.058
	95% CI		[-0.151,0.052]		[-0.118,0.109]
	IF		34.847		34.734
ν_1	Mean			6.099	6.466
	SD			1.510	1.804
	95% CI			[4.027,10.030]	[3.951,11.185]
	IF			87.960	107.398
ν_2	Mean			30.187	28.245
	SD			6.889	6.961
	95% CI			[14.648,39.392]	[13.980,39.143]
	IF			195.756	209.572
ν_3	Mean			5.304	5.330
	SD			0.901	0.892
	95% CI			[3.926,7.480]	[3.911,7.444]
	IF			93.892	61.991
log marg like		-11038	-10897	-10144	-10142
DIC		22017	21880	20238	20161

1. Mean is the posterior mean based on 20000 MCMC draws after a 5000 burn-in period.
2. SD is the numerical standard error of the posterior means.
3. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws.
4. IF is the inefficiency factor.

6   Conclusions

In this dissertation, we propose a novel approach to model volatilities and correlations in the multivariate context.

In Chapter 3, we propose the basic multivariate stochastic volatility model with dynamic correlation (MSV-GFT). Our approach is to use a generalized version of Fisher's z-transformation to characterize the dynamics in volatilities and correlations in a highly flexible manner. The most important feature in our models is that they automatically generate a positive definite correlation matrix, and the driving forces underlying volatilities and correlations are fully separated. Different from much of the existing literature, when making inference for our model, we apply the state-of-the-art particle-filter-based MCMC technique. We have considered a simulation based on different sample size($T = 500, 1000, 2000$) and different number of particles ($N = 50, 100, 200$). The simulation result shows the finite sample performance is stable. In the empirical application of exchange rate and stock indices, we found that the flexible correlation setting can capture more information than some existing models. For model comparison, DIC and log marginal likelihood confirmed the basic model is better than the competing models.

In Chapter 4, we accommodate a few essential stylized facts commonly observed in real data and consider the Multivariate stochastic volatility model with dynamic correlation and realized measures (RMSV-GFT). We consider the bias from the micro-structure noises and construct two measurement equations of realized measures to capture the bias. However, the new setting causes the problem of under-identification. We estimate the model by particle-filter-based MCMC technique and propose a two-stage estimation to address the identification problem. The simulation results suggest that the two-stage estimation method works well in finite samples. We apply the proposed model and estimation method to the daily return of three equities with their realized measure. The empirical result shows RMSV-GFT out-performs the MSV-GFT in terms of both realized-measure forecast and portfolio optimization.

In Chapter 5, we extend the basic model (MSV-GFT) by incorporating the leverage effects and heavy-tailed error distribution. We develop the particle-filter-based MCMC technique to conduct Bayesian analysis of the extended models. The simulation study shows the estimation technique works well in the finite sample. In the empirical application, we estimate four models using stock indices data. We use DIC and log marginal-likelihood to compare the model specifications. The model with both the leverage effect and heavy-tailed error distribution outperforms the other three specifications, suggesting that both the leverage effect and heavy-tailed error distribution are important for the data.

References

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2010a): "Parametric and nonparametric volatility measurement," in Handbook of Financial Econometrics: Tools and Techniques, Elsevier, 67–137.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001a): "The distribution of realized stock return volatility," Journal of Financial Economics, 61, 43–76.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001b): "The distribution of realized exchange rate volatility," Journal of the American statistical association, 96, 42–55.

ANDERSEN, T. G., T. BOLLERSLEV, P. FREDERIKSEN, AND M. ØRREGAARD NIELSEN (2010b): "Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns," Journal of Applied Econometrics, 25, 233–261.

ANDERSEN, T. G. AND B. E. SØRENSEN (1996): "GMM estimation of a stochastic volatility model: A Monte Carlo study," Journal of Business & Economic Statistics, 14, 328–352.

ANDRIEU, C., A. DOUCET, AND R. HOLENSTEIN (2010): "Particle markov chain monte carlo methods," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72, 269–342.

ARCHAKOV, I. AND P. R. HANSEN (2018a): "A new parametrization of correlation matrices," Working paper, 1–32.

——— (2018b): "Web Appendix for "A new parametrization of correlation matrices"," Working paper, 1–44.

ARCHAKOV, I., P. R. HANSEN, AND A. LUNDE (2020): "A Multivariate Realized GARCH Model," Working paper, 1–42.

ARULAMPALAM, M. S., S. MASKELL, N. GORDON, AND T. CLAPP (2002): "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Transactions on Signal Processing, 50, 174–188.

ASAI, M. AND M. MCALEER (2006): "Asymmetric multivariate stochastic volatility," Econometric Reviews, 25, 453–473.

——— (2009): "The structure of dynamic correlations in multivariate stochastic volatility models," Journal of Econometrics, 150, 182–192.

ASAI, M., M. MCALEER, AND J. YU (2006): "Multivariate stochastic volatility: a review," Econometric Reviews, 25, 145–175.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2009): "Realized kernels in practice: Trades and quotes," .

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 253–280.

BAUWENS, L., S. LAURENT, AND J. V. ROMBOUTS (2006): "Multivariate GARCH models: a survey," Journal of Applied Econometrics, 21, 79–109.

BERG, A., R. MEYER, AND J. YU (2004): "Deviance information criterion for comparing stochastic volatility models," Journal of Business & Economic Statistics, 22, 107–120.

BLACK, F. (1976): "Studies of stock market volatility changes," 1976 Proceedings of the American Statistical Association Bisiness and Economic Statistics Section.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," Journal of Econometrics, 31, 307–327.

BROTO, C. AND E. RUIZ (2004): "Estimation methods for stochastic volatility models: a survey," Journal of Economic Surveys, 18, 613–649.

CHAN, D., R. KOHN, AND C. KIRBY (2006): "Multivariate stochastic volatility models with correlated errors," Econometric Reviews, 25, 245–274.

CHIB, S. (1995): "Marginal likelihood from the Gibbs output," Journal of the American Statistical Association, 90, 1313–1321.

CHIB, S. AND I. JELIAZKOV (2001): "Marginal likelihood from the Metropolis–Hastings output," Journal of the American Statistical Association, 96, 270–281.

CHIB, S., F. NARDARI, AND N. SHEPHARD (2002): "Markov chain Monte Carlo methods for stochastic volatility models," Journal of Econometrics, 108, 281–316.

CHIB, S., Y. OMORI, AND M. ASAI (2009): "Multivariate stochastic volatility," in Handbook of Financial Time Series, Springer, 365–400.

CHIRSTIE, A. (1982): "The stochastic behavior of common stock variances," Journal of Financial Economics, 10, 407–432.

CHOPIN, N. AND S. S. SINGH (2015): "On particle Gibbs sampling," Bernoulli, 21, 1855–1883.

CHOPRA, V. K. AND W. T. ZIEMBA (1993): "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," The Journal of Portfolio Management, 19, 6–11.

CHRISTENSEN, K. AND M. PODOLSKIJ (2007): "Realized range-based estimation of integrated variance," Journal of Econometrics, 141, 323–349.

CORSI, F., F. AUDRINO, AND R. REN (2012): HAR Modeling for Realized Volatility Forecasting, 363 – 382.

DE JONG, P. AND N. SHEPHARD (1995): "The simulation smoother for time series models," Biometrika, 82, 339–350.

DUFFIE, D. AND K. J. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," Econometrica, 61, 929–952.

ENGLE, R. (2002): "New frontiers for ARCH models," Journal of Applied Econometrics, 17, 425–446.

ENGLE, R. AND R. COLACITO (2006): "Testing and valuing dynamic correlations for asset allocation," Journal of Business & Economic Statistics, 24, 238–253.

ENGLE, R. AND B. KELLY (2012): "Dynamic equicorrelation," Journal of Business & Economic Statistics, 30, 212–228.

ENGLE, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," Econometrica, 987–1007.

——— (2000): "The econometrics of ultra-high-frequency data," Econometrica, 68, 1–22.

ENGLE, R. F. AND G. M. GALLO (2006): "A multiple indicators model for volatility using intra-daily data," Journal of Econometrics, 131, 3–27.

FONG, E. AND C. HOLMES (2020): "On the marginal likelihood and cross-validation," Biometrika, 107, 489–496.

FRIEL, N. AND A. N. PETTITT (2008): "Marginal likelihood estimation via power posteriors," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70, 589–607.

GALLANT, A. R., D. HSIEH, AND G. TAUCHEN (1997): "Estimation of stochastic volatility models with diagnostics," Journal of Econometrics, 81, 159–192.

GOURIÉROUX, C., J. JASIAK, AND R. SUFANA (2009): "The Wishart autoregressive process of multivariate stochastic volatility," Journal of Econometrics, 150, 167–181.

HANSEN, P. R. AND Z. HUANG (2016): "Exponential GARCH modeling with realized measures of volatility," Journal of Business & Economic Statistics, 34, 269–287.

HANSEN, P. R., Z. HUANG, AND H. H. SHEK (2012): "Realized GARCH: a joint model for returns and realized measures of volatility," Journal of Applied Econometrics, 27, 877–906.

HANSEN, P. R., A. LUNDE, AND V. VOEV (2014): "Realized beta GARCH: A multivariate GARCH model with realized measures of volatility," Journal of Applied Econometrics, 29, 774–799.

HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate stochastic variance models," The Review of Economic Studies, 61, 247–264.

HARVEY, A. C. AND N. SHEPHARD (1996): "Estimation of an asymmetric stochastic volatility model for asset returns," Journal of Business & Economic Statistics, 14, 429–434.

HURN, S., V. MARTIN, P. PHILLIPS, AND J. YU (2020): Financial Econometric Modeling, Oxford University Press.

ISHIHARA, T. AND Y. OMORI (2012): "Efficient Bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors," Computational Statistics & Data Analysis, 56, 3674–3689.

ISHIHARA, T., Y. OMORI, AND M. ASAI (2016): "Matrix exponential stochastic volatility with cross leverage," Computational Statistics & Data Analysis, 100, 331–350.

JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): "Microstructure noise in the continuous case: the pre-averaging approach," Stochastic Processes and their Applications, 119, 2249–2276.

JACQUIER, E., N. G. POLSON, P. ROSSI, ET AL. (1999): Stochastic volatility: Univariate and multivariate extensions, CIRANO.

JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (1994): "Bayesian analysis of stochastic volatility models," Journal of Business & Economic Statistics, 12, 371–417.

——— (2004): "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," Journal of Econometrics, 122, 185–212.

JEFFREYS, H. (1998): The Theory of Probability, OUP Oxford.

JIN, X. AND J. M. MAHEU (2013): "Modeling realized covariances and returns," Journal of Financial Econometrics, 11, 335–369.

——— (2016): "Bayesian semiparametric modeling of realized covariance matrices," Journal of Econometrics, 192, 19–39.

JOHANSEN, A. M. AND A. DOUCET (2008): "A note on auxiliary particle filters," Statistics & Probability Letters, 78, 1498–1504.

KASS, R. E. AND A. E. RAFTERY (1995): "Bayes factors," Journal of the American statistical Association, 90, 773–795.

KAWAKATSU, H. (2006): "Matrix exponential GARCH," Journal of Econometrics, 134, 95–128.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic volatility: likelihood inference and comparison with ARCH models," The Review of Economic Studies, 65, 361–393.

KOOPMAN, S. J. AND M. SCHARTH (2012): "The analysis of stochastic volatility in the presence of daily realized measures," Journal of Financial Econometrics, 11, 76–115.

KUROSE, Y. AND Y. OMORI (2016): "Dynamic equicorrelation stochastic volatility," Computational Statistics & Data Analysis, 100, 795–813.

——— (2020): "Multiple-block dynamic equicorrelations with realized measures, leverage and endogeneity," Econometrics and Statistics, 13, 46–68.

LI, Y., N. WANG, AND Y. JUN (2020a): "Improved marginal likelihood estimation via power posteriors and importance sampling," SMU Economics and Statistics Working Paper Series, Paper No. 16-2019.

LI, Y., J. YU, AND T. ZENG (2020b): "Deviance information criterion for latent variable models and misspecified models," Journal of Econometrics, 216, 450–493.

LI, Y., T. ZENG, AND J. YU (2017): "Deviance Information Criterion for Model Selection: Justification and Variation," Working Paper.

LIESENFELD, R. AND R. C. JUNG (2000): "Stochastic volatility models: conditional normality versus heavy-tailed distributions," Journal of Applied Econometrics, 15, 137–160.

LINDSTEN, F., M. I. JORDAN, AND T. B. SCHÖN (2014): "Particle Gibbs with ancestor sampling," The Journal of Machine Learning Research, 15, 2145–2184.

LINDSTEN, F. AND T. B. SCHÖN (2013): "Backward simulation methods for Monte Carlo statistical inference," Foundations and Trends® in Machine Learning, 6, 1–143.

LOPES, H. F., R. MCCULLOCH, AND R. TSAY (2010): "Cholesky stochastic volatility," Unpublished Technical Report, University of Chicago, Booth Business School, 2.

MARKOWITZ, H. (1952): "Portfolio analysis," Journal of Finance, 8, 77–91.

MELINO, A. AND S. M. TURNBULL (1990): "Pricing foreign currency options with stochastic volatility," Journal of Econometrics, 45, 239–265.

MEYER, R. AND J. YU (2000): "BUGS for a Bayesian analysis of stochastic volatility models," The Econometrics Journal, 3, 198–215.

NOURELDIN, D., N. SHEPHARD, AND K. (2012): "Multivariate high-frequency-based volatility (HEAVY) models," Journal of Applied Econometrics, 27, 907–933.

OMORI, Y., S. CHIB, N. SHEPHARD, AND J. NAKAJIMA (2007): "Stochastic volatility with leverage: Fast and efficient likelihood inference," Journal of Econometrics, 140, 425–449.

OMORI, Y. AND T. WATANABE (2008): "Block sampler and posterior mode estimation for asymmetric stochastic volatility models," Computational Statistics & Data Analysis, 52, 2892–2910.

PATTON, A. J. (2011): "Volatility forecast comparison using imperfect volatility proxies," Journal of Econometrics, 160, 246–256.

PATTON, A. J. AND K. SHEPPARD (2009): "Evaluating volatility and correlation forecasts," in Handbook of Financial Time Series, Springer, 801–838.

PHILIPOV, A. AND M. E. GLICKMAN (2006): "Multivariate stochastic volatility via Wishart processes," Journal of Business & Economic Statistics, 24, 313–328.

PHILLIPS, P. C. AND J. YU (2009): "A two-stage realized volatility approach to estimation of diffusion processes with discrete data," Journal of Econometrics, 150, 139–150.

PITT, M. K. AND N. SHEPHARD (1999): "Filtering via simulation: Auxiliary particle filters," Journal of the American Statistical Association, 94, 590–599.

PODOLSKIJ, M. AND M. VETTER (2009): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," Bernoulli, 15, 634–658.

SANDMANN, G. AND S. J. KOOPMAN (1998): "Estimation of stochastic volatility models via Monte Carlo maximum likelihood," Journal of Econometrics, 87, 271–301.

SHEPHARD, N. AND M. K. PITT (1997): "Likelihood analysis of non-Gaussian measurement time series," Biometrika, 84, 653–667.

SHEPHARD, N. AND K. SHEPPARD (2010): "Realising the future: forecasting with high-frequency-based volatility (HEAVY) models," Journal of Applied Econometrics, 25, 197–231.

SHIROTA, S., Y. OMORI, H. F. LOPES, AND H. PIAO (2017): "Cholesky realized stochastic volatility model," Econometrics and Statistics, 3, 34–59.

SPIEGELHALTER, D. J., N. G. BEST, B. P. CARLIN, AND A. VAN DER LINDE (2002): "Bayesian measures of model complexity and fit," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 583–639.

TAKAHASHI, M., Y. OMORI, AND T. WATANABE (2009): "Estimating stochastic volatility models using daily returns and realized volatility simultaneously," Computational Statistics & Data Analysis, 53, 2404–2426.

TANNER, M. A. AND W. H. WONG (1987): "The calculation of posterior distributions by data augmentation," Journal of the American statistical Association, 82, 528–540.

TAYLOR, S. J. (1986a): Modelling Financial Time Series, Wiley.

——— (1986b): Modelling financial time series, World Scientific.

VENTER, J. H. AND P. J. DE JONGH (2014): "Extended stochastic volatility models incorporating realised measures," Computational Statistics and Data Analysis, 76, 687–707.

WATANABE, T. AND Y. OMORI (2004): "A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997)," Biometrika, 246–248.

YAMAUCHI, Y. AND Y. OMORI (2019): "Multivariate stochastic volatility model with realized volatilities and pairwise realized correlations," Journal of Business & Economic Statistics, 1–17.

YU, J. (2005): "On leverage in a stochastic volatility model," Journal of Econometrics, 127, 165–178.

YU, J. AND R. MEYER (2006): "Multivariate stochastic volatility models: Bayesian estimation and model comparison," Econometric Reviews, 25, 361–384.

ZHANG, L. (2006): "Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach," Bernoulli, 12, 1019–1043.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," Journal of the American Statistical Association, 100, 1394–1411.

ZHENG, T. AND T. SONG (2014): "A realized stochastic volatility model with Box–Cox transformation," Journal of Business & Economic Statistics, 32, 593–605.

A   Appendix to Chapter 3

A.1   Details of PGAS algorithm

Recall that a state space model is in the form

$$y_t|x_t = x \sim f_\theta(\cdot|x), \tag{A.1}$$

$$x_{t+1}|x_t = x \sim g_\theta(\cdot|x), \text{ and } x_1 \sim \mu_\theta(\cdot), \tag{A.2}$$

The output of a PGAS algorithm is a random draw from the joint smoothing distribution $p_\theta(x_{1:T}|y_{1:T})$, conditional on one particular set of parameter values. In the following, we will omit parameters in all densities with an understanding that they depend on a same θ. The input of this algorithm, except for θ, is a reference trajectory of $x_{1:T}$, which is a sample from last MCMC iteration. Let's denote that reference trajectory by $x'_{1:T}$. Then, the algorithm proceeds as following:

- 1. Draw $x_1^{(i)}$ from $q_1(x_1|y_1)$, for $i = 1, 2, ..., N - 1$.

- 2. Set $x_1^{(N)} = x'_1$.

- 3. Set $w_1^{(i)} = f(y_1|x_1^{(i)})/q_1(x_1^{(i)}|y_1)$, for $i = 1, 2, ..., N$.

- 4. For $t = 2$ to T, do the following:

  - (a). Generate $\{\tilde{x}_{1:t-1}^{(i)}\}_{i=1}^{N-1}$ by sampling with replacement $N - 1$ times from $\{x_{1:t-1}^{(i)}\}_{i=1}^{N}$ with probabilities proportional to the importance weights $\{w_{t-1}^{(i)}\}_{i=1}^{N}$.

  - (b). Draw J from $\{1, 2, ..., N\}$ with probabilities proportional to $w_{t-1}^{(i)}g(x'_t|x_{t-1}^{(i)})$ and then set $\tilde{x}_{1:t-1}^{(N)} = x_{1:t-1}^{(J)}$.

  - (c). Simulate $x_t^{(i)}$ from $q_t(x_t|\tilde{x}_{t-1}^{(i)}, y_t)$, for $i = 1, 2, ..., N - 1$.

  - (d). Set $x_t^{(N)} = x'_t$.

  - (e). Set $x_{i:t}^{(i)} = (\tilde{x}_{1:t-1}^{(i)}, x_t^{(i)})$

– (f). Set weight to be $w_t^{(i)} = f(y_t|x_t^{(i)})g(x_t^{(i)}|\tilde{x}_{t-1}^{(i)})/q_t(x_t^{(i)}|\tilde{x}_{t-1}^{(i)}, y_t)$, for $i = 1, 2, ..., N$.

- 5. Draw k from $\{1, 2, ..., N\}$ with probabilities proportional to $w_T^{(i)}$ and return $x_{1:T}^* = x_{1:T}^{(k)}$.

Note that this procedure is very similar to the original conditional sequential Monte Carlo-based particle Gibbs sampler. The major modification is in the Step 4(b), where a new index is drawn and thus the Nth trajectory may not be the reference one from last iteration. In conditional PG, on the contrary, we fix the last particle to follow the input trejectory $x_{1:T}'$.

It is worth mentioning that the probability of drawing J depends on $g(x_t'|x_{t-1}^{(i)})$ and x_t' is drawn in the last iteration conditional on all observations $y_{1:T}$. Therefore, this step makes the algorithm more like a smoothing instead of filtering.

A.2 Additional Figures

This section includes some additional figures that are related to the empirical application with weekly exchange rate data we consider in the main text. Corresponding figures for stock return data are similar and thus omitted for saving the space. Specifically, we present the following figures:

- Figure 24: Kernel-smoothed posterior distributions of volatility-related parameters.

- Figure 25: Kernel-smoothed posterior distributions of correlation-related parameters.

- Figure 26: Mixing of volatility-related parameters represented by the decay of autocorrelation.

- Figure 27: Mixing of correlation-related parameters represented by the decay of autocorrelation.

- Figure 28: MCMC iterations of volatility-related parameters.

- Figure 29: MCMC iterations of correlation-related parameters.

Figure 24: Posterior distributions of volatility-related parameters for weekly exchange rate data
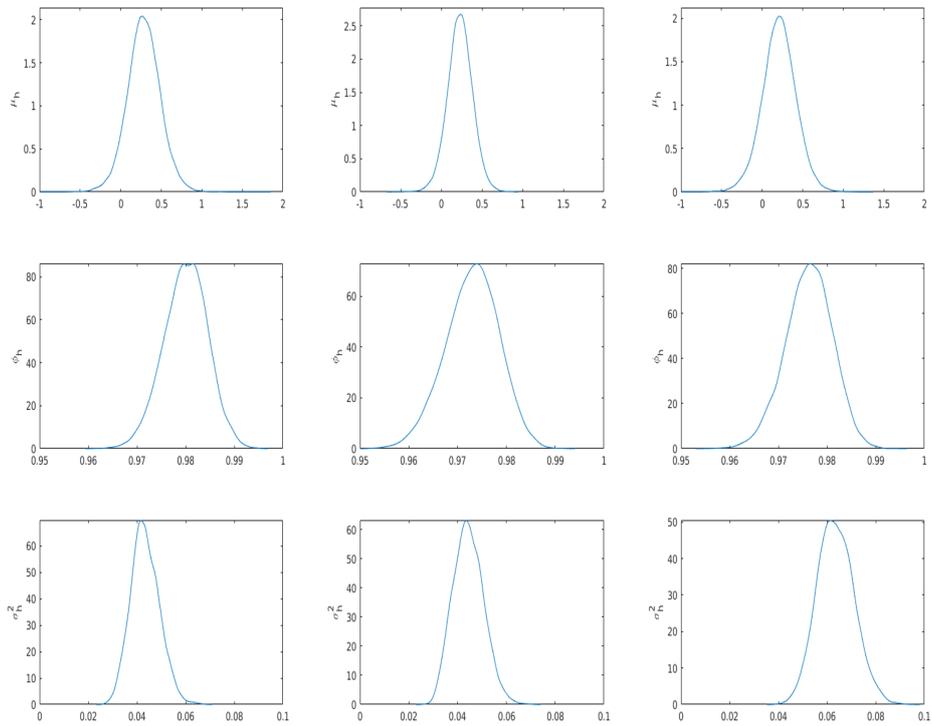
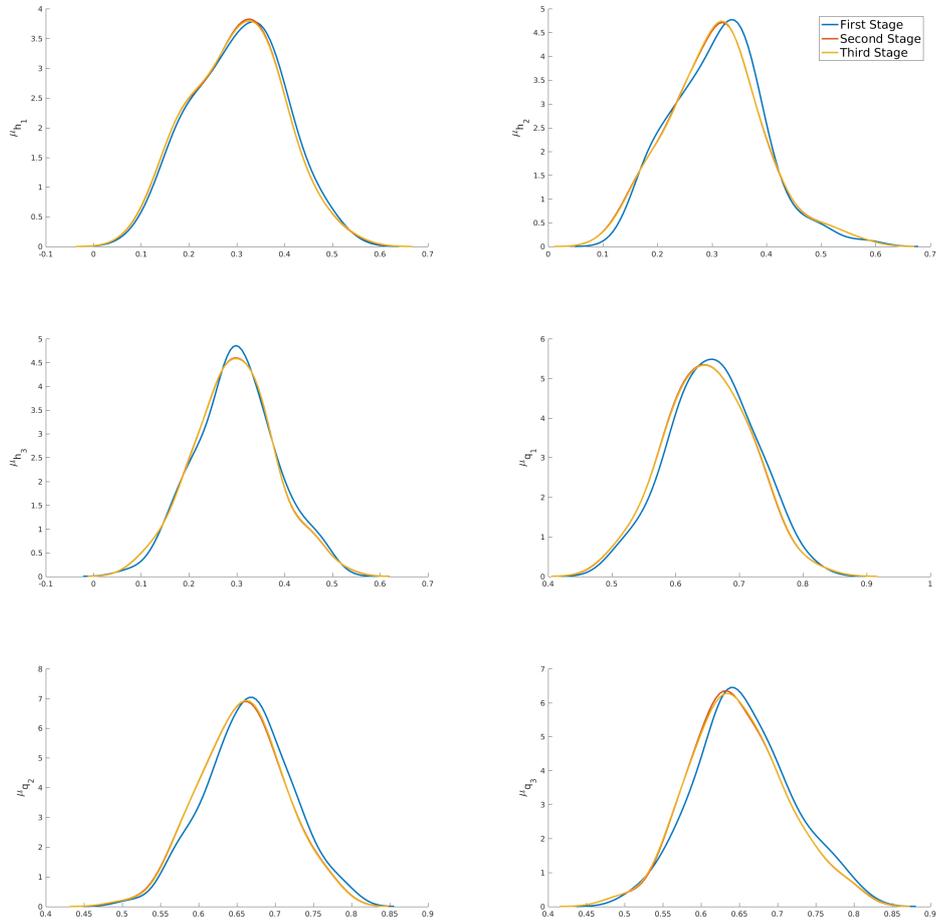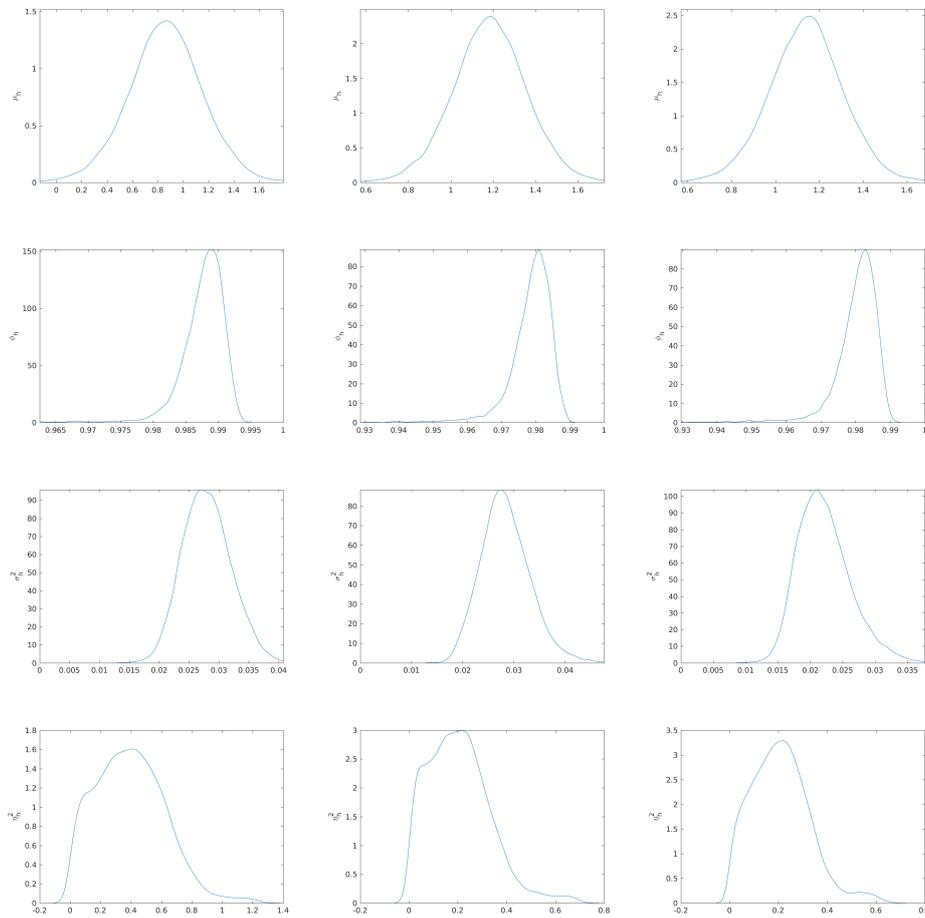Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

Figure 25: Posterior distributions of correlation-related parameters for weekly exchange rate data
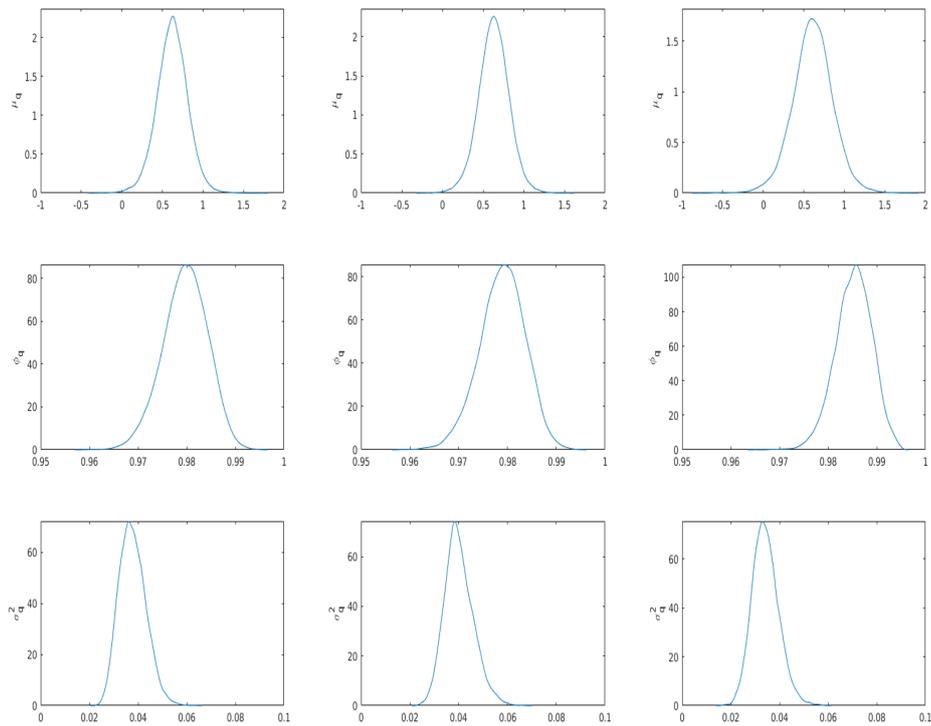
Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

Figure 26: Mixing of volatility-related parameters for weekly exchange rate data
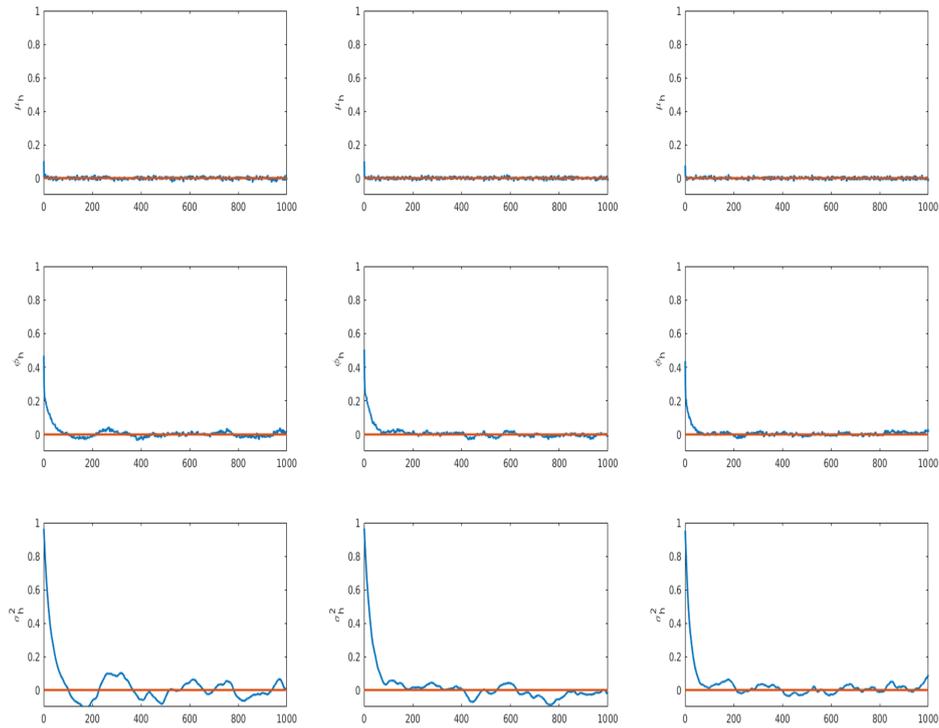
Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. Blue curves show the decay of sample autocorrelations of MCMC sampler.

Figure 27: Mixing of correlation-related parameters for weekly exchange rate data
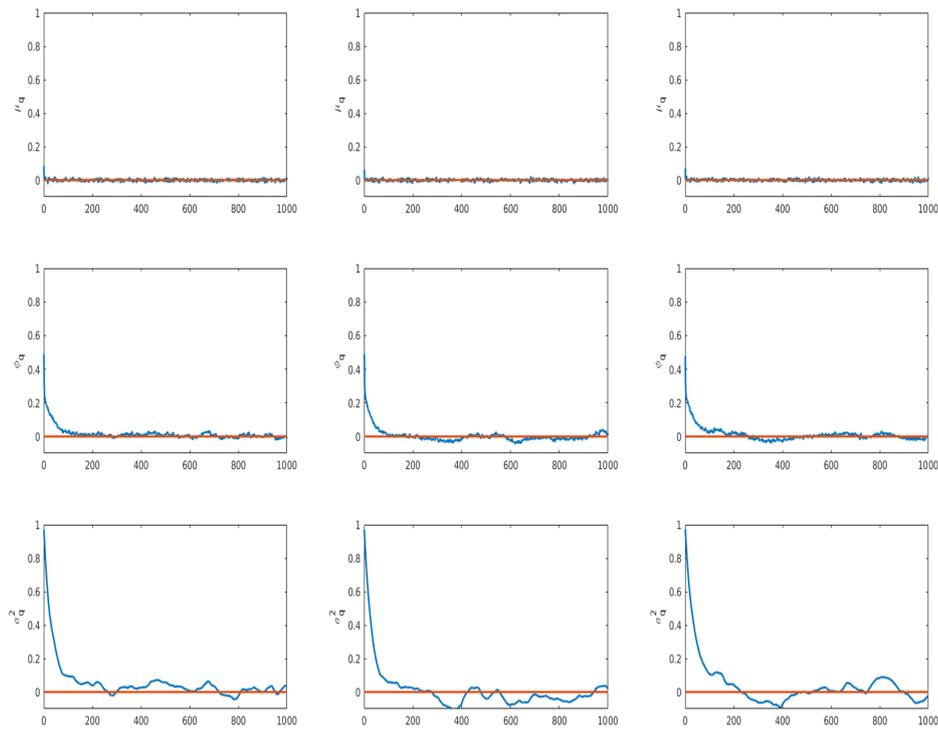
Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. Blue curves show the decay of sample autocorrelations of MCMC sampler.

Figure 28: MCMC iterations of volatility-related parameters for weekly exchange rate data
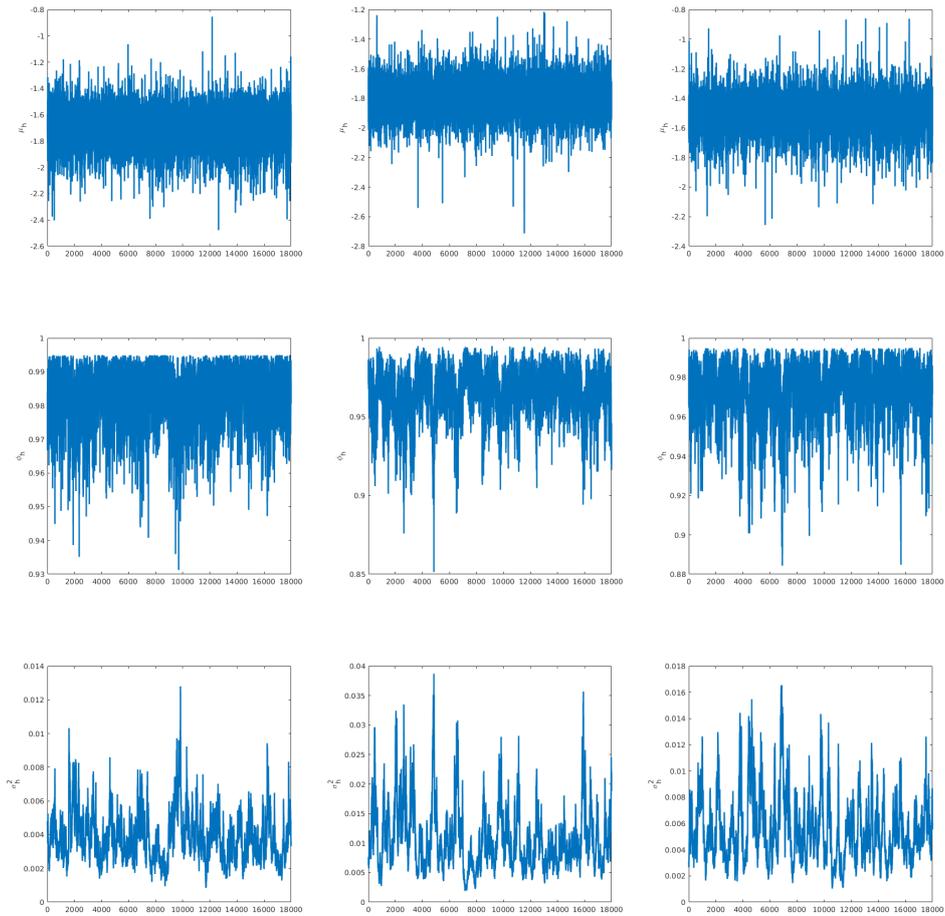
Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's.

Figure 29: MCMC iterations of correlation-related parameters for weekly exchange rate data
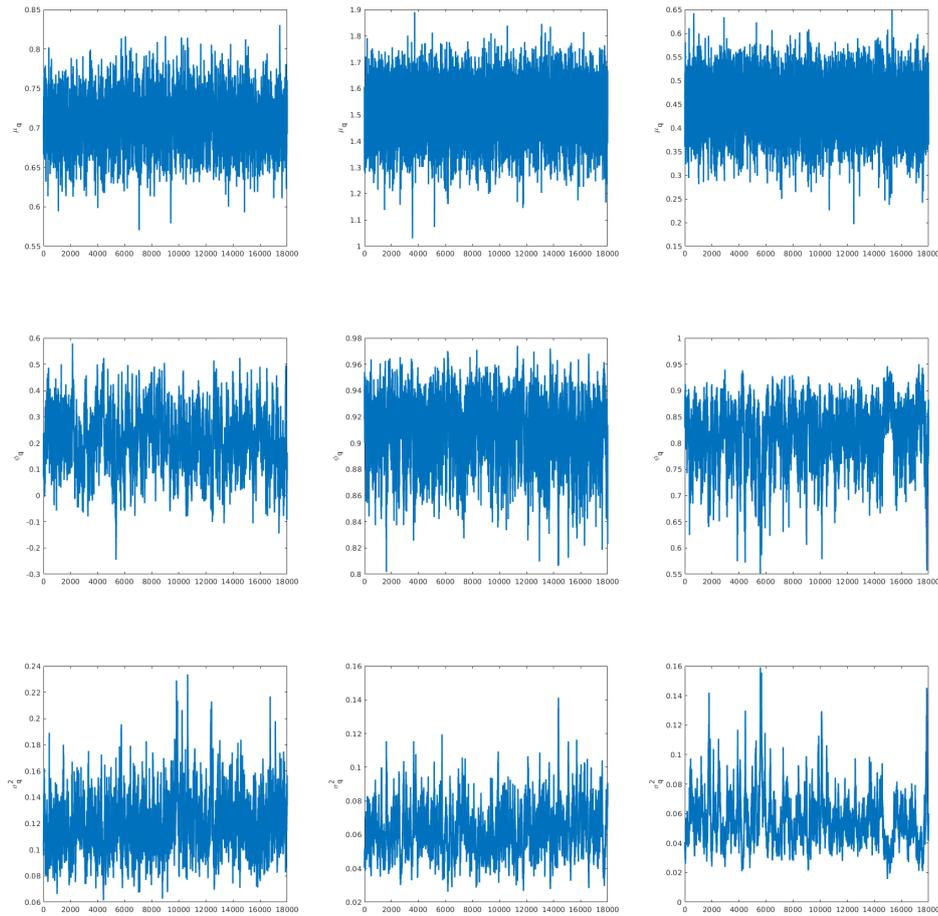
Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's.

B Appendix to Chapter 4

B.1 Additional Figures

This section includes some additional figures of that are related to the empirical application with stock return data we consider in the main text.

- Figure 30: Finite Sample Distribution of $\hat{\mu}_h$ and $\hat{\mu}_q$ across 3 Stages

- Figure 31: Kernel-smoothed posterior distributions of volatility-related parameters.

- Figure 32: Kernel-smoothed posterior distributions of correlation-related parameters.

- Figure 33: Mixing of volatility-related parameters represented by the decay of autocorrelation.

- Figure 34: Mixing of correlation-related parameters represented by the decay of autocorrelation.

- Figure 35: MCMC iterations of volatility-related parameters.

- Figure 36: MCMC iterations of correlation-related parameters.

Figure 30: Finite Sample Distribution of $\hat{\mu}_h$ and $\hat{\mu}_q$ across 3 Stages

Note: The finite sample distribution based on simulation data.

Figure 31: Posterior distributions of volatility-related parameters for daily stock return data

The first row includes μ's, the second row includes ϕ's, the third row includes σ^2's and forth row includes η^2. All densities are kernel-smoothed.

Figure 32: Posterior distributions of correlation-related parameters for daily stock return data
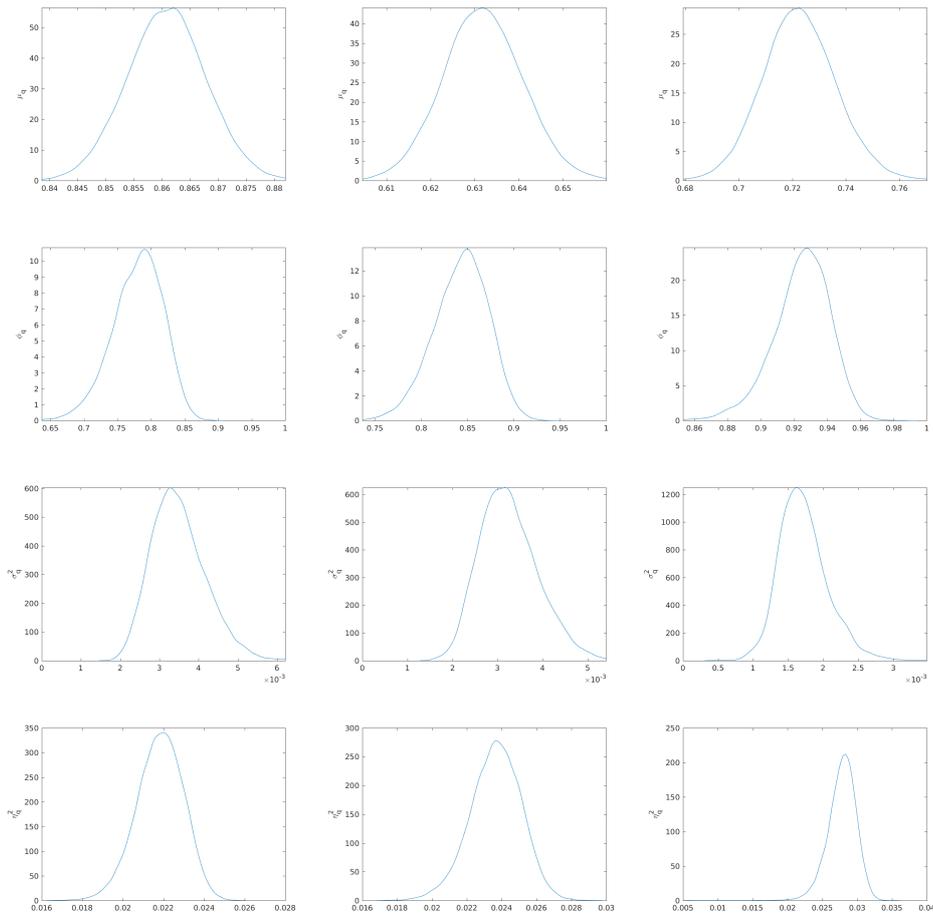
Note: The first row includes μ's, the second row includes ϕ's, the third row includes σ^2's and forth row includes η^2. All densities are kernel-smoothed.

Figure 33: Mixing of volatility-related parameters for daily stock return data
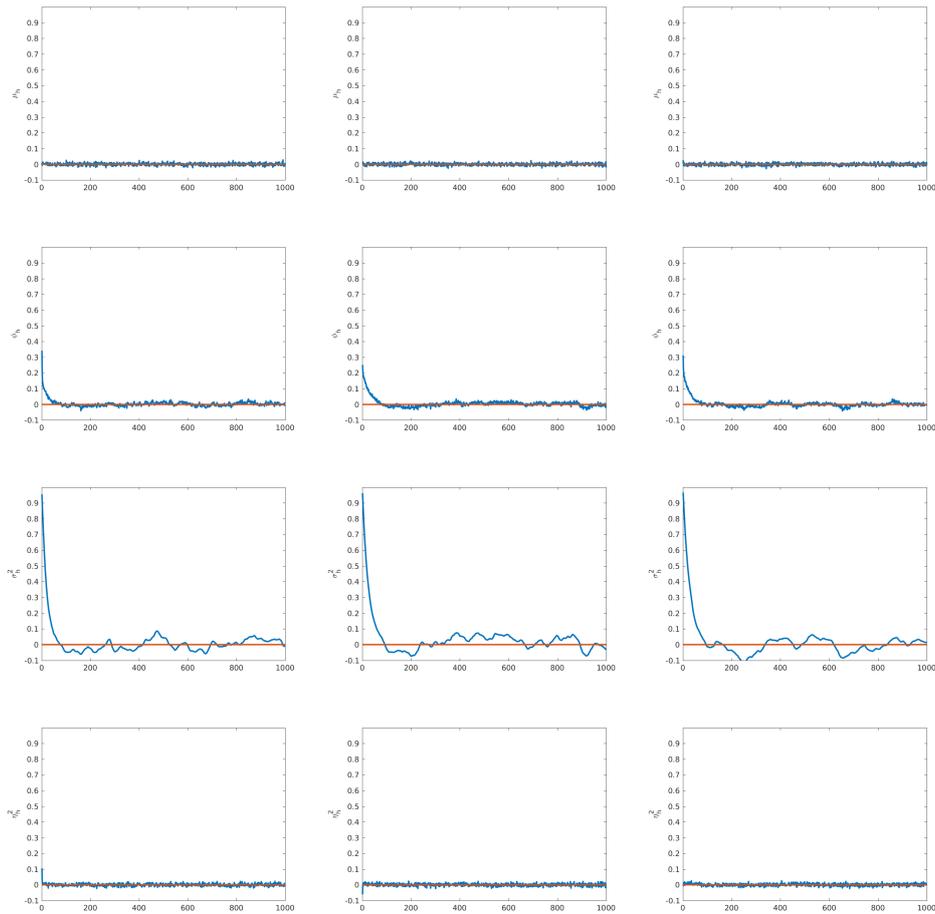
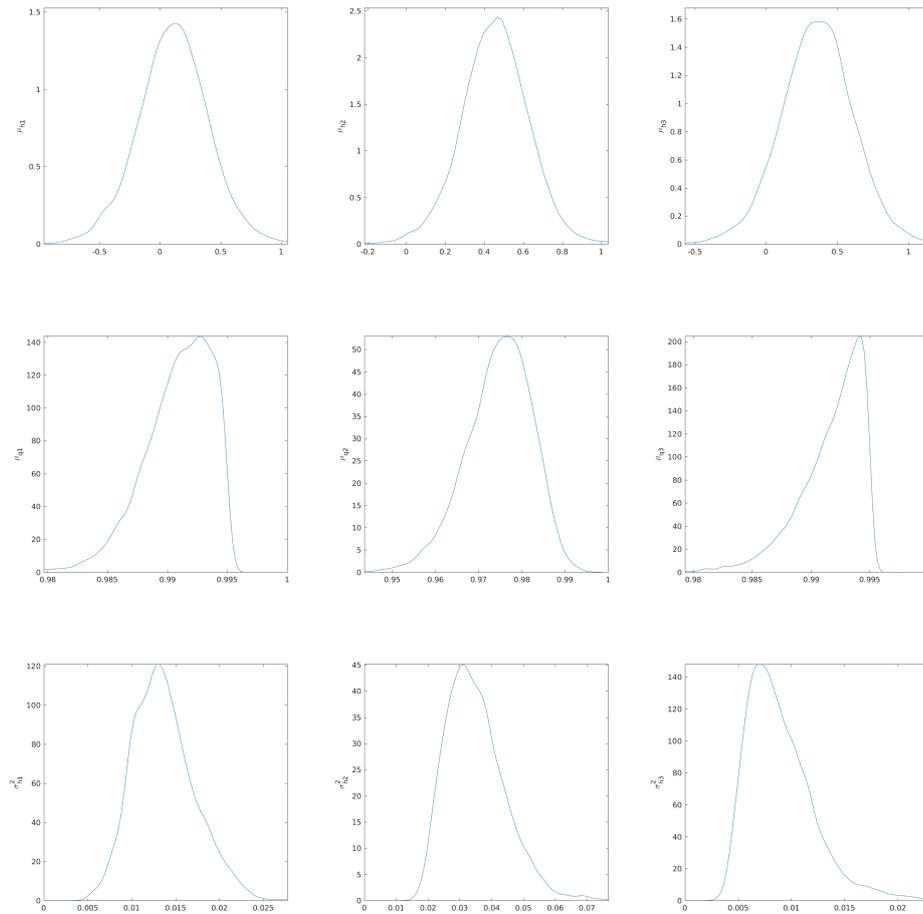Note: The first row includes μ's, the second row includes ϕ's,the third row includes σ^2's and forth row includes η^2. Blue curves show the decay of sample autocorrelations of MCMC sampler.

Figure 34: Mixing of correlation-related parameters for daily stock return data
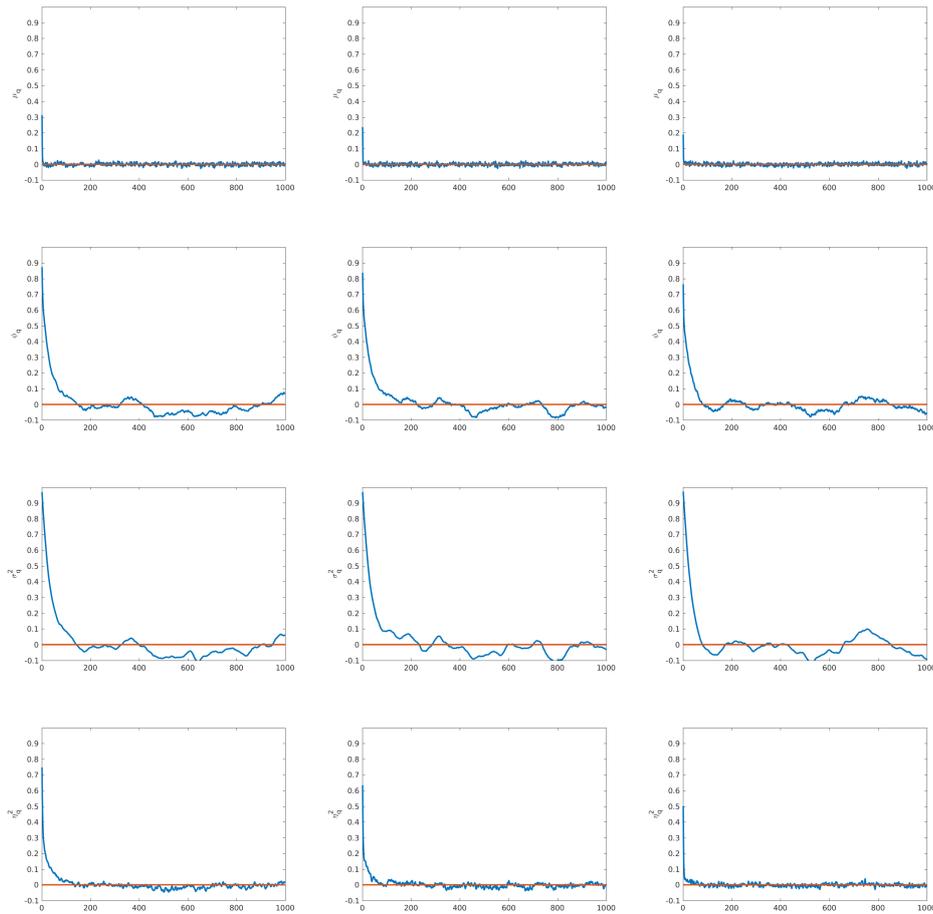
Note: The first row includes μ's, the second row includes ϕ's, the third row includes σ^2's and forth row includes η^2. Blue curves show the decay of sample autocorrelations of MCMC sampler.

Figure 35: MCMC iterations of volatility-related parameters for daily stock return data
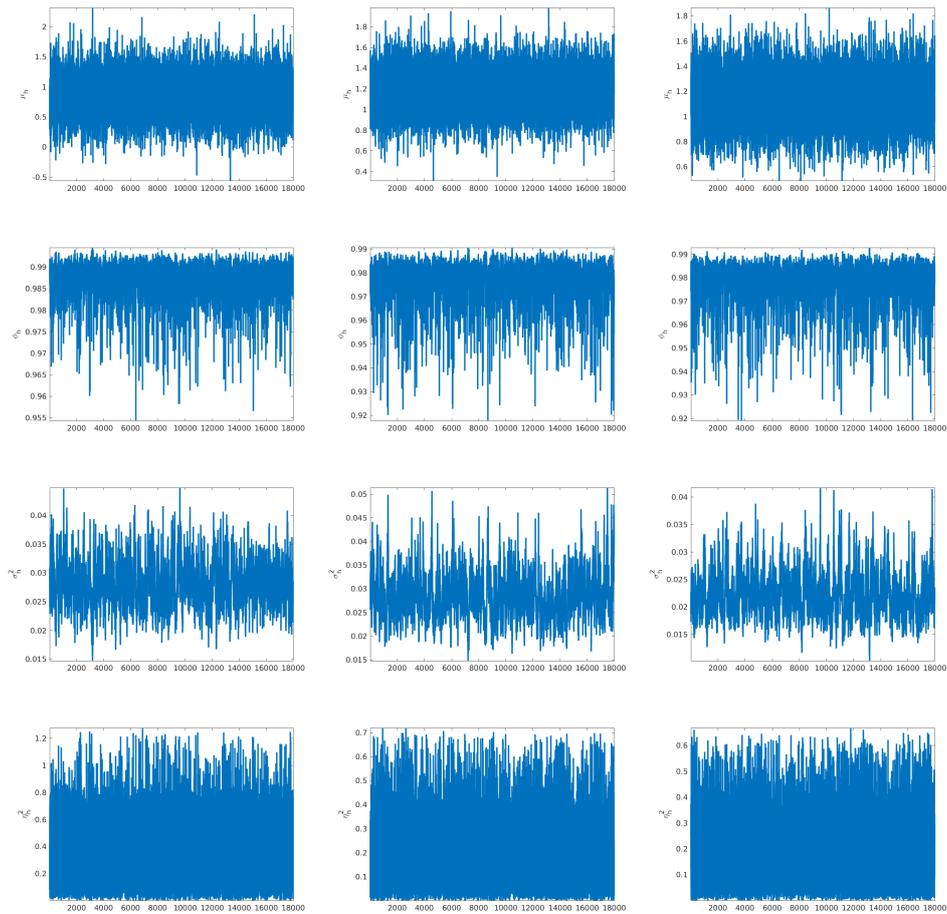
Note: The first row includes μ's, the second row includes ϕ's, the third row includes σ^2's and forth row includes η^2.

Figure 36: MCMC iterations of correlation-related parameters for daily stock return data
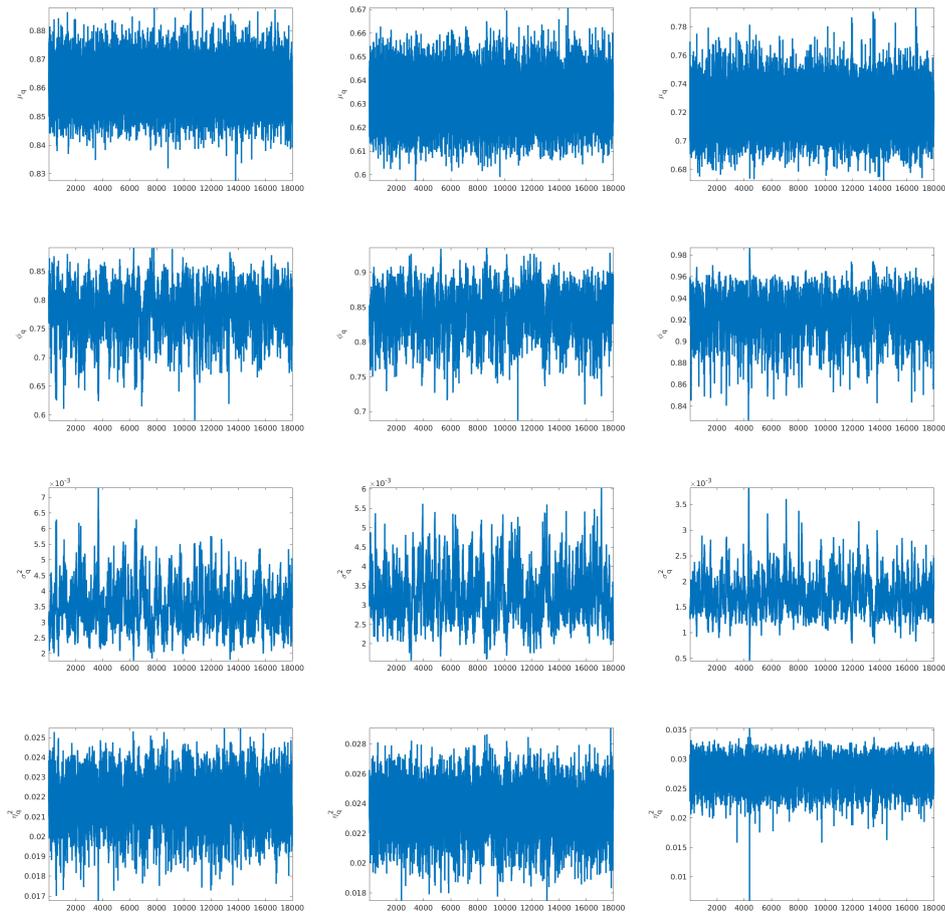
Note: TThe first row includes μ's, the second row includes ϕ's, the third row includes σ^2's and forth row includes η^2.

C Appendix to Chapter 5

C.1 Additional Figures

This appendix includes some additional figures that supports model estimation in empirical application. The figures below shows the convergence property of MSVLt-GFT model. Corresponding figures for MSV-GFT, MSVL-GFT and MSVt-GFT are similar and thus omitted for saving the space. Specifically, we present the following figures:

- Figure 37: Kernel-smoothed posterior distributions of volatility-related parameters.

- Figure 38: Kernel-smoothed posterior distributions of correlation-related parameters.

- Figure 39: Kernel-smoothed posterior distributions of leverage and heavy tail-related parameters.

- Figure 40: Mixing of volatility-related parameters represented by the decay of autocorrelation.

- Figure 41: Mixing of correlation-related parameters represented by the decay of autocorrelation.

- Figure 42: Mixing of leverage and heavy tail-related parameters represented by the decay of autocorrelation.

- Figure 43: MCMC iterations of volatility-related parameters.

- Figure 44: MCMC iterations of correlation-related parameters.

- Figure 45: MCMC iterations of leverage and heavy tail-related parameters.

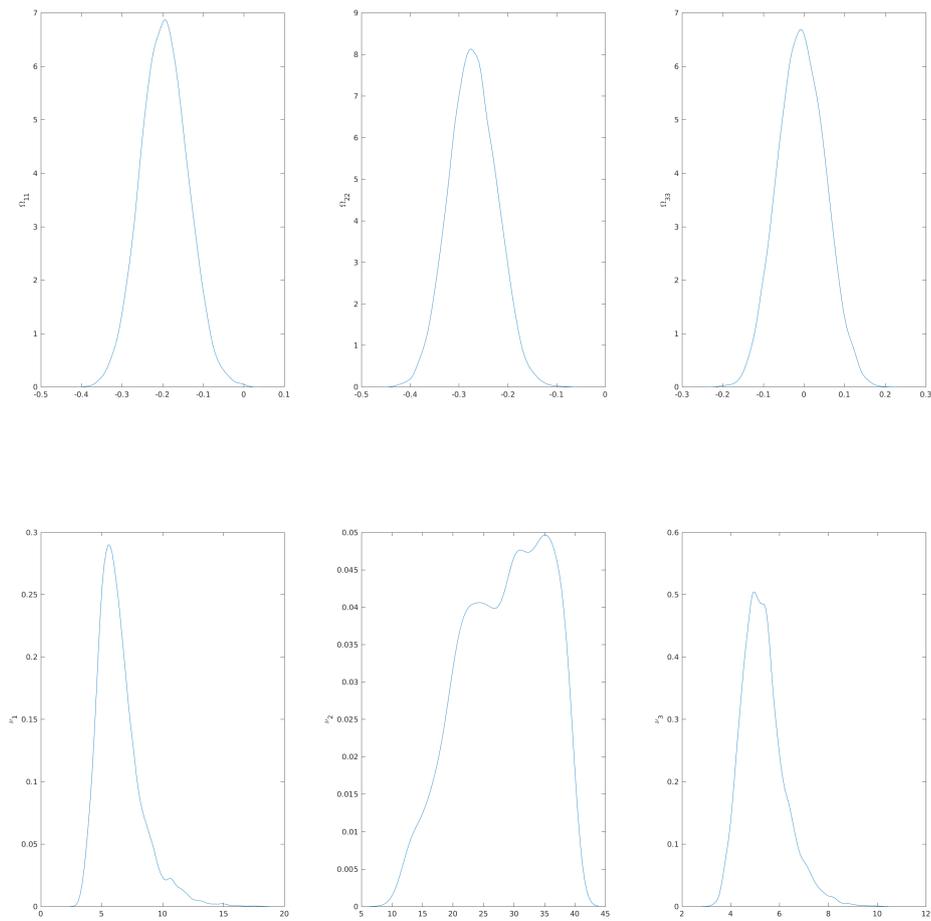Figure 37: Posterior distributions of volatility-related parameters for stock index data

Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

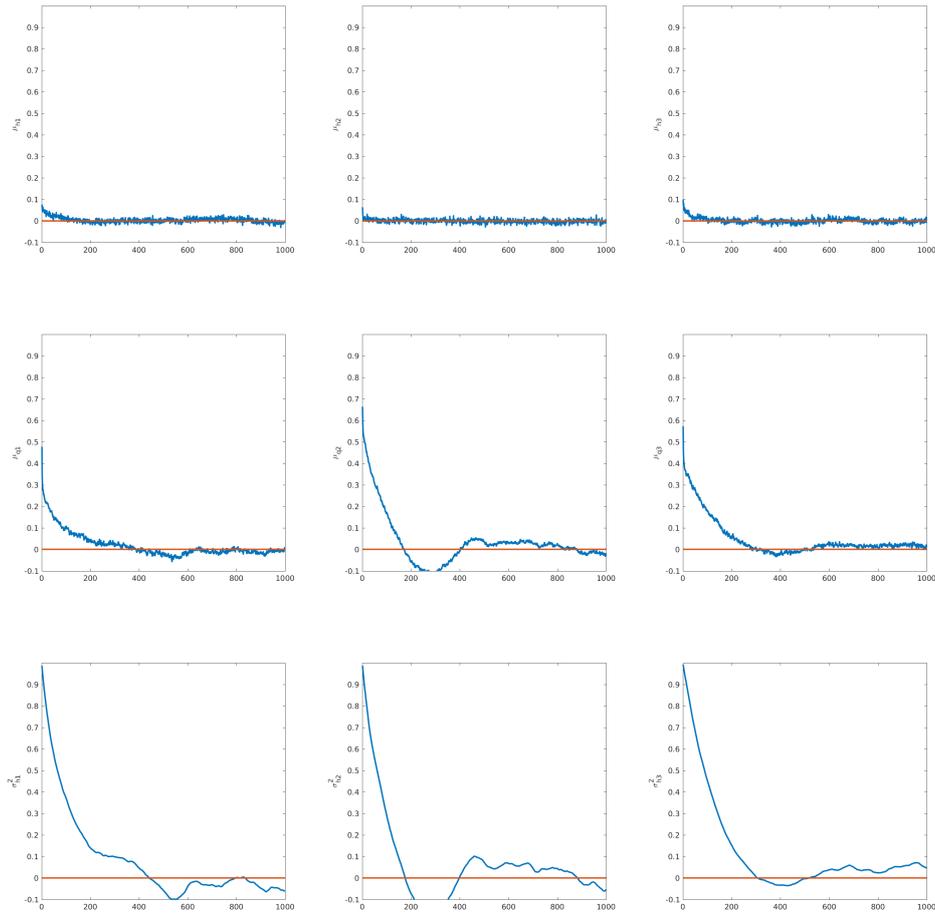Figure 38: Posterior distributions of correlation-related parameters for stock index data

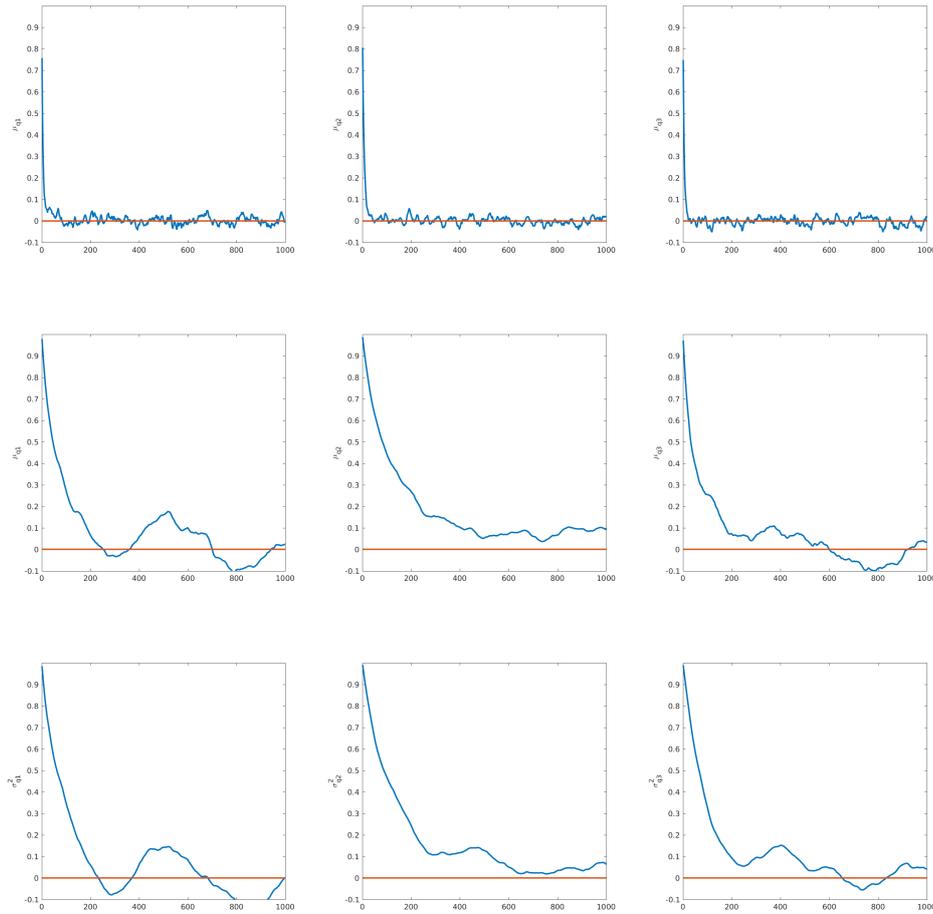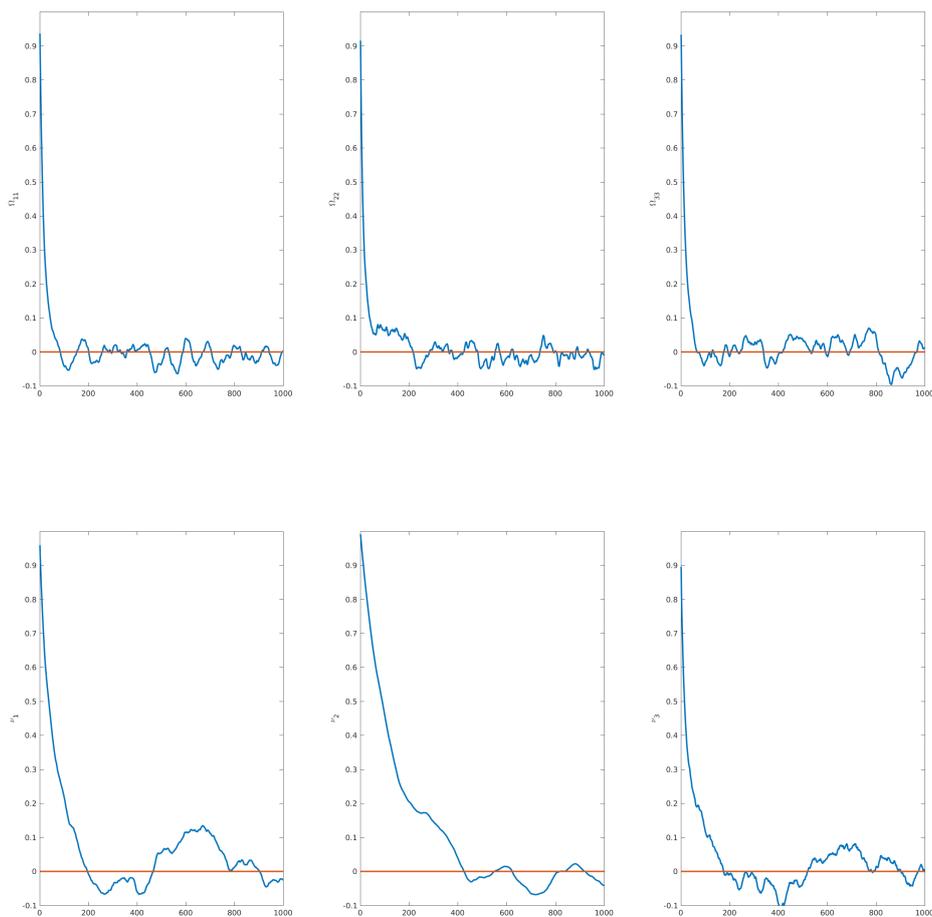Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

Figure 39: Posterior distributions of leverage and heavy tail-related parameters for stock index data

Note: The first row includes Ω's and the second row includes ν's.

Figure 40: Mixing of volatility-related parameters for stock index data

Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

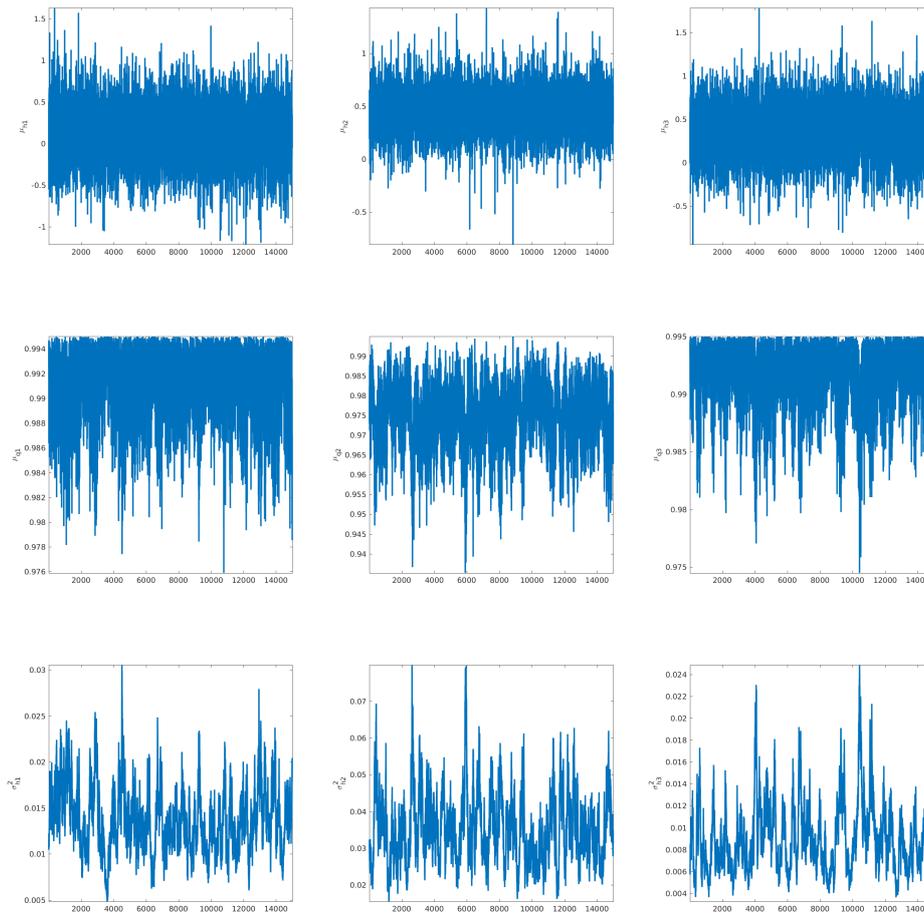Figure 41: Mixing of correlation-related parameters for stock index data

Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

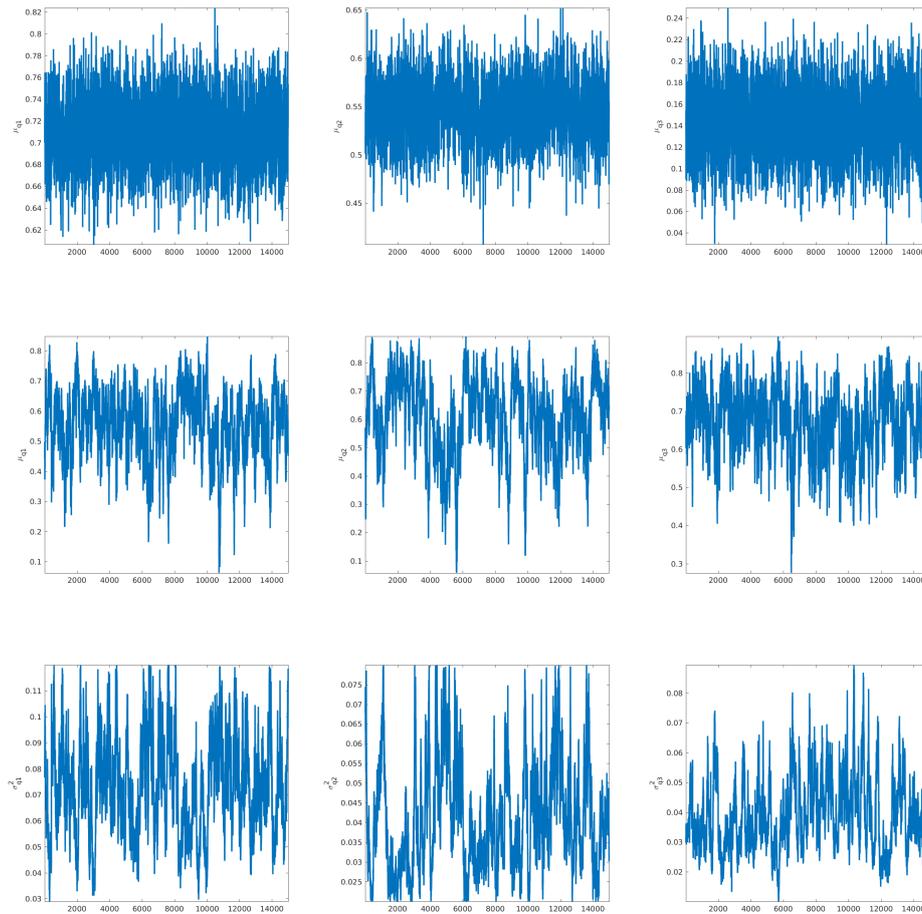Figure 42: Mixing of leverage and heavy tail-related parameters for stock index data

Note: The first row includes Ω's and the second row includes ν's.

133

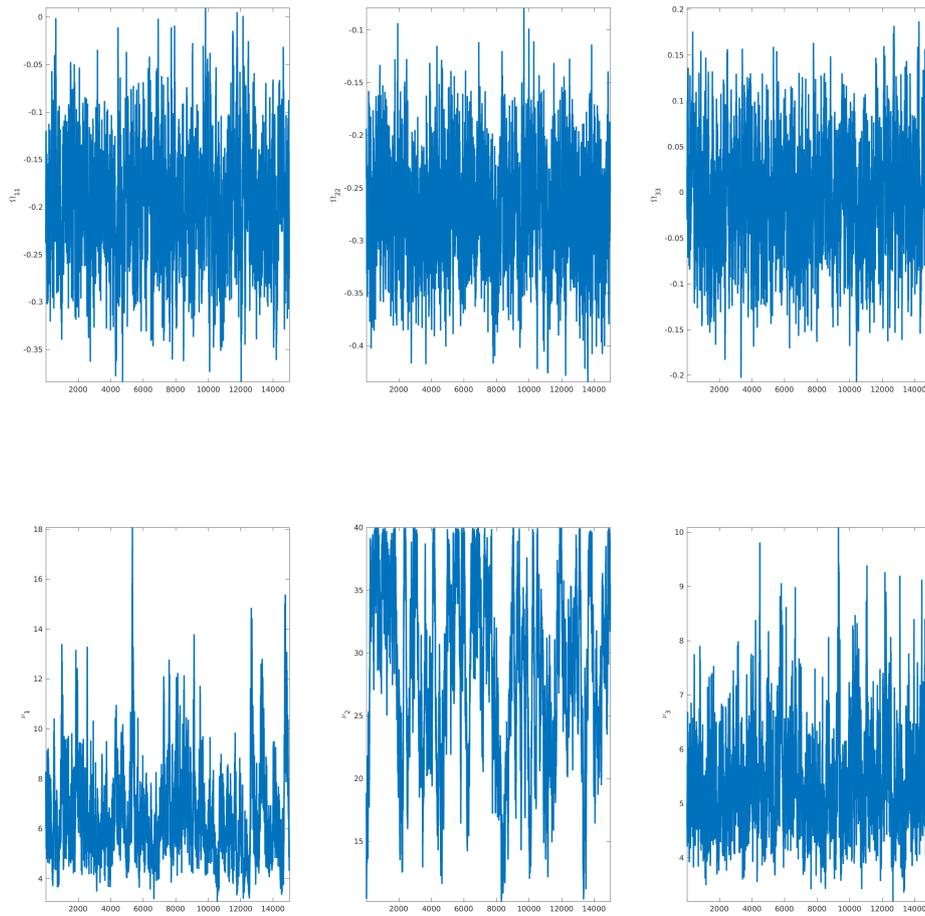Figure 43: MCMC iterations of volatility-related parameters for stock index data

Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

Figure 44: MCMC iterations of correlation-related parameters for stock index data

Note: The first row includes μ's, the second row includes ϕ's and the third row includes σ^2's. All densities are kernel-smoothed.

Figure 45: MCMC iterations of leverage and heavy tail-related parameters for stock index data

Note: The first row includes Ω's and the second row includes ν's.