7-2023

Multivariate stochastic volatility models based on generalized fisher transformation

Han CHEN

Yijie FEI

Jun YU
Singapore Management University, yujun@smu.edu.sg

SMU ECONOMICS &
STATISTICS

Multivariate Stochastic Volatility Models based on Generalized Fisher Transformation

Han Chen, Yijie Fei, Jun Yu,

July 2023

Paper No. 09-2023

Multivariate Stochastic Volatility Models based on Generalized Fisher Transformation*

Han Chen
Hunan University

Yijie Fei
Hunan University

Jun Yu
Singapore Management University

July 6, 2023

Abstract

Modeling multivariate stochastic volatility (MSV) can be challenging, particularly when both variances and covariances are time-varying. In this paper, we address these challenges by introducing a new MSV model based on the generalized Fisher transformation of Archakov and Hansen (2021). Our model is highly flexible and ensures that the variance-covariance matrix is always positive-definite. Moreover, our approach separates the driving factors of volatilities and correlations. To conduct Bayesian analysis of the model, we use a Particle Gibbs Ancestor Sampling (PGAS) method, which facilitates Bayesian model comparison. We also extend our MSV model to cover the leverage effect in volatilities and the threshold effect in correlations. Our simulation studies demonstrate that the proposed method performs well for the MSV model. Furthermore, empirical studies based on exchange-rate returns and equity returns show that our MSV model outperforms alternative specifications in both in-sample and out-of-sample performances. Overall, our novel MSV model and inferential method offer a more reliable and flexible framework for modeling time-varying variances and covariances of financial assets.

*We would like to thank Peter Hansen for useful discussions and comments. Han Chen, Email:han.chen.2015@phdecons.smu.edu.sg and Yijie Fei, Email:yijiefei@hnu.edu.cn, College of Finance and Statistics, Hunan University, 109 Shijiachong Road, Changsha, China. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, Singapore, 178903. Email: yujun@smu.edu.sg.

Keywords: Multivariate stochastic volatility; Dynamic correlation; Leverage effect; Particle filter; Markov chain Monte Carlo;

JEL Codes: G10, C53, C12, C32, C58

1 Introduction

The characterization of the dynamic behavior of return volatility is crucial for asset pricing, portfolio allocation, and risk management. Univariate volatility models have been extensively studied in the literature since the seminal paper by Engle (1982). These models can be broadly categorized into two types: GARCH-based and stochastic volatility (SV) models. In recent decades, there has been a growing focus on multivariate financial data analysis. It is now widely recognized that analyzing asset returns individually is insufficient, and the dependence structure among assets must be taken into account. To address this, a plethora of multivariate extensions to univariate GARCH and SV models have been proposed and applied in practice. Multivariate GARCH (MGARCH) models have been extensively reviewed in Bauwens et al. (2006), while multivariate SV (MSV) models have been reviewed in Asai et al. (2006). These multivariate models enable us to capture the co-movements of volatilities and correlations among multiple assets, providing a more accurate representation of the underlying dependence structure. The development of multivariate models has significantly improved our understanding of asset return dynamics and has important practical implications for financial risk management and investment strategies.

The first multivariate stochastic volatility (MSV) model, proposed by Harvey et al. (1994), is an extension of the constant conditional correlation (CCC) model in multivariate GARCH (MGARCH). In this basic setup, each asset's volatility is modeled by a univariate stochastic volatility process, while the correlation matrix among all assets remains constant over time. However, this assumption is rather restrictive. Subsequent efforts have been devoted to relaxing this assumption in the literature of MSV. For instance, Yu and Meyer (2006) proposed a model that mirrors the dynamic conditional correlation (DCC) model of Engle (2002) in MGARCH. The DCC-based model allows for time-varying correlation among assets while still assuming that each asset's volatility follows a univariate stochastic volatility process. Another parametrization based on DCC can be found in Asai and McAleer (2009b). Other studies have proposed even more flexible models that allow for both time-varying volatilities and correlations among assets.

In this paper, we propose a new MSV model that builds upon a recently developed parameterization of the correlation matrix. This parameterization, first introduced in Archakov and Hansen (2021), is a generalization of the well-known Fisher z-transformation from the bivariate case to the multivariate case. It has been successfully used in other models, such as the multivariate realized GARCH model of Archakov et al. (2020) and the dynamic conditional score model of Hafner and Wang (2021). Recent empirical evidence in Bucci et al. (2022) shows that this parameterization provides more accurate forecasts of the realized covariance matrix than other existing methods.

Our new MSV model allows the underlying latent variables that determine the correlations among assets to have an unrestricted domain because the correlation matrix is always valid by construction. In addition, the shocks to the volatility dynamics and the correlation dynamics are fully separated in our model. This is an appealing feature, as in practice, these two types of shocks may be determined by completely distinct factors. Finally, our model is invariant to the reordering of assets, which eliminates the need for an ex-ante ordering of assets. All of these features indicate that our model is highly flexible and imposes a minimal level of ex-ante restrictions.

The importance of accommodating asymmetric effects in both GARCH and SV literature has long been recognized. For equity returns, it has been emphasized that bad news has a greater impact on future volatility than good news, known as the leverage effect. This has been incorporated in several existing MSV models, such as Asai and McAleer (2006) and Asai and McAleer (2009a). Recently, it has been observed that the degree of co-movement among stocks responds differently to good and bad news, with realized correlations between stock returns tending to be higher when stock prices are falling, and lower when there is a surge in stock prices. This phenomenon has been studied in Gorgi and Koopman (2021) using a novel threshold beta observation-driven model and in Cappiello et al. (2006) using a generalized asymmetric DCC-GARCH model.[1]

To allow for asymmetric effects across multiple assets in an MSV model, Ishihara et al. (2016) and Asai et al. (2022) propose to consider the lower-diagonal elements of the matrix logarithm of the covariance matrix and assume that the return vector is correlated with modeled variables, which they call cross-leverage. However, this approach is inconsistent with the original idea of leverage effect since their latent variables are generated by complex nonlinear transformations of the covariance matrix and correspond to both volatility and correlation. As argued in Asai et al. (2006), leverage should refer only to the negative correlations between the current return and future volatility.

[1]Asymmetry in realized correlations and DCC-GARCH models has also been studied in Audrino and Corsi (2010) and Audrino and Trojani (2011) using a tree-structured threshold approach.

To address this issue, in this paper, we propose an extension of our baseline MSV model that explicitly and independently accounts for asymmetric effects in volatility and correlation. Specifically, we assume that the volatility variable of a stock is correlated with its own return, consistent with the leverage effect, and impose a threshold structure for correlation variables. Following Gorgi and Koopman (2021), we use lagged S&P 500 returns as the threshold variable for correlations. To the best of our knowledge, the proposed model is the first MSV model that explicitly and independently accounts for asymmetric effects in correlations.

In this paper, we propose a Bayesian approach to estimate our new MSV model. Unlike the majority of the existing Bayesian MSV literature, which relies on standard Markov chain Monte Carlo (MCMC) algorithms, we use a recently developed Particle MCMC (PMCMC) algorithm. PMCMC algorithms have become increasingly popular since the seminal paper by Andrieu et al. (2010) and have been applied to a wide range of fields. While theoretically applicable to a broad class of models, the practical performance of PMCMC algorithms depends on various factors and requires careful examination. In our case, we choose the Particle Gibbs Ancestor Sampling (PGAS) method proposed by Lindsten et al. (2014). This method is a modified version of the Particle Gibbs (PG) Sampler and has been shown to have improved mixing properties even with a small number of particles. We present extensive simulation results to justify our choice of estimation strategy and provide useful guidance for empirical applications. Our approach strikes a balance between estimation accuracy and computational cost, making it an attractive option for practitioners.

Our MSV models, unfortunately, face limitations in scalability, as estimating them with a moderately large number of assets is not feasible. This is to be expected, as our model is highly flexible, allowing for time-varying volatilities, correlations, and covariances, while also ensuring the positive-definiteness of the variance-covariance matrix. Nonetheless, estimating our low-dimensional MSV model can still provide valuable insights into important features of data with a low dimension, which can in turn guide the selection of more restrictive MSV models for high-dimensional data. For instance, our estimation of a three-dimensional MSV model reveals the critical importance of allowing pairwise correlation coefficient sequences to have different levels of persistence. Therefore, a reasonable and more restrictive MSV model for high-dimensional data should also retain this feature.

The rest of the paper is organized as follows. In Section 2, we provide a selective literature review on MSV models, introduce the new parametrization of correlation matrix and present our new model. In Section 3, we introduce the estimation and inferential method based on PGAS algorithm. Simulation and empirical studies are provided in Section 4 and Section 5, respectively. Section 6 concludes. Additional technical details can be found in the Appendix.

Throughout the paper, we let $diag(A)$ denote the column vector formed by the diagonal elements of a square matrix A or the diagonal matrix whose diagonal elements are elements in A if A is a column vector;[2] $vech(A)$ denote the $p(p+1)/2 \times 1$ column-vector obtained by vectorizing only the lower triangular part of a p-dimensional matrix A (including the diagonal elements); $vecl(A)$ denote the $p(p-1)/2 \times 1$ column-vector containing all lower off-diagonal elements of A (excluding the diagonal elements); I_p denote a p-dimensional identity matrix, $\mathbb{1}_p$ denote a p-dimensional column vector of ones. We use $I(x)$ to denote the indicator function such that $I(x) = 1$ if $x \geq 0$ and $I(x) = 0$ otherwise.

2 New Multivariate Stochastic Volatility Model

In this section, we will provide a brief overview of existing MSV models, with a focus on the parametrization of the covariance matrix. We will then introduce the generalized Fisher transformation (GFT), a novel technique proposed in Archakov and Hansen (2021), and review its application to MSV models. Finally, we will propose a new MSV model that utilizes the GFT.

2.1 Review of existing MSV models

The existing literature on multivariate stochastic volatility (MSV) models is vast, with numerous studies proposing different modeling approaches and estimation techniques. Two comprehensive reviews of this area of research up to a certain point in time are Asai et al. (2006) and Chib et al. (2009). These reviews discuss various estimation techniques and methods for model comparison and provide an overview of the earlier studies in the literature. In this section, we focus on the models proposed over the last decade, which have aimed to address the challenge of ensuring the positive-definiteness of the variance-covariance matrix. We provide a critical review of these recent models and their parametrizations.[3]

The basic structure of the MSV model is

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

where r_t is a vector of q 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

[2] If A is a square matrix, $diag(diag(A))$ is a diagonal matrix whose diagonal elements are the diagonal elements in A.

[3] We do not review models based on the factor structure in this review, as our new model is based on direct modeling of the variance-covariance matrix.

models rely on different techniques to ensure this positive-definiteness. 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 variance-covariance decomposition is first carried out and then each component in the decomposition is modeled separately.

Within the first group of models, four methods have been considered. The first method is based on the matrix exponential. For example, Ishihara et al. (2016) and Asai et al. (2022) assume that

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

and propose to model $vech(H_t)$ as vector autoregressive (VAR) process. By the definition of the matrix exponential, C_t is guaranteed to be positive-definite. The major drawback of this model is that the relationship between latent variables and the original volatilities/correlations is highly nonlinear and hence, 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, Nakajima (2017), Shirota et al. (2017) and Zaharieva et al. (2020) also use this decomposition to set up their MSV models. As a well-known problem in the VAR literature, in the Cholesky decomposition, order matters. That is, the resulting variance-covariance matrix depends on the ordering of assets. This dependence is highly undesirable as the performance of the model depends on the ordering. Moreover, the dynamics of the volatilities and the correlations are not separated.

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

$$\begin{aligned} C_t &= \sum_{i=1}^{m} x_{it} x_{it}', \\ x_{it} &= A x_{i,t-1} + \epsilon_{it} \text{ and } \epsilon_{it} \sim N(0, \Sigma), \end{aligned}$$

where (m, A, Σ) are unknown parameters. Alternatively, one can also model C_t using 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 (v, d, A) are unknown parameters. Clearly, the dynamics of the volatilities and the correlations are not separated.

Dellaportas et al. (2023) point out that Wishart-based MSV models are not able to scale well with p, the number of assets, as the computational complexity is $O(p^3)$. In view of this drawback, they propose a new MSV model assuming Gaussian latent processes for functions of the eigenvalues and rotation angles of C_t. In particular, spectral decomposition implies that

$$C_t = P_t \Lambda_t P_t',$$

where Λ_t is a diagonal matrix of eigenvalues and P_t is the eigenvector matrices. They further decompose P_t as a product of $p(p-1)/2$ rotation matrices, whose elements are modeled after transformation. This parameterization leads to $O(p^2)$ complexity and thus more scalable. Nevertheless, the modeled variables in this setup are still hard to interpret.

Models in the second group treat the volatilities and the correlation matrix separately and is amenable to easy interpretation of driving factors of volatilities and correlations. Consider 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. By construction, the diagonal elements of R_t are ones and the off-diagonal elements of R_t are pair-wise correlation coefficients. For our purpose, the major difference in model designs among models in this group lies in how R_t is parameterized. The critical issue in this setup is to ensure R_t is a valid correlation matrix, such as the positive-definiteness, symmetry, all the diagonal elements being one, all the off-diagonal elements taking values in $[-1, 1]$. The first and the simplest model in this fashion is the constant correlation MSV in Harvey et al. (1994), where

$$R_t = R, \text{ for all } t. \tag{1}$$

A similar assumption is made in Chan et al. (2006), Asai and McAleer (2006), and Ishihara and Omori (2012). In these models, the dynamic movement of correlations is not allowed. Although the assumption makes statistical inference simple, it is too restrictive for modeling financial time series.

To allow for time-varying correlations, Asai and McAleer (2009b) consider two model

specifications, both motivated by DCC. The idea is to write the correlation matrix as

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

where $\tilde{Q}_t = (diag(diag(Q_t))^{1/2}$. By construction, all diagonal elements of R_t are ones and R_t is a valid correlation matrix as long as Q_t is symmetric positive-definite. The two existing Wishart distribution-based models for Q_t 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. Asai and McAleer (2009b) argue that the second one is preferred.

Inspired by the dynamic equicorrelation (DECO) model of Engle and Kelly (2012), Kurose and Omori (2016) propose 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 being ones. To ensure that ρ_t is within $(-1, 1)$, Kurose and Omori (2016) model the Fisher z-transformation of ρ_t as an autoregressive process. As in the model of Yamauchi and Omori (2020), a lower bound for ρ_t is needed to guarantee the positive-definiteness of R_t. This lower bound approaches zero as the number of assets goes to infinity.

Following Yu and Meyer (2006), Yamauchi and Omori (2020) propose to model the pairwise correlations by the Fisher z-transformation. That is, letting $R_t = \{\rho_{ij,t}\}$ and

$$g_{ij,t} = \frac{1}{2} \log \frac{1+\rho_{ij,t}}{1-\rho_{ij,t}} := F\left(\rho_{ij,t}\right), \tag{2}$$

they assume that $g_{ij,t}$ follows a random walk for any $i \neq j$. By construction, $|\rho_{ij,t}| < 1$. When $p = 2$, that is, only two returns are modeled, this transformation ensures the positive-definiteness of R_t. However, if $p > 2$, this element-wise operation does not guarantee the positive-definiteness of R_t. Yamauchi and Omori (2020) further derive algebraic bounds for $\rho_{ij,t}$ that ensure the positive-definiteness of R_t. The bounds for one particular $\rho_{ij,t}$ are conditional on all other elements in R_t. Therefore, the restriction is well suited for the single-move Gibbs sampling technique, but hard to be implemented by other estimation

methods.

2.2 Generalized Fisher transformation 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 the complexity introduced by this constraint in modeling, one can instead model Fisher z-transformation of ρ, defined by $F(\rho)$ in (2). It is easy to show that $\rho = F^{-1}(g) = \frac{\exp(g)-1}{\exp(g)+1} \in (-1, 1)$ for any $g \in (-\infty, \infty)$. 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 MSV literature in Yu and Meyer (2006) when the number of assets is two. Unfortunately, it is acknowledged by Yu and Meyer (2006) that this approach "is not easy to be generalized into higher dimension situations". In particular, a pairwise 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 ensure the positive-definiteness of the resulting correlation matrix in general.

Clearly, it is desirable to obtain a valid high-dimensional extension to the Fisher z-transformation. This is the exact contribution made in Archakov and Hansen (2021). To fix the idea, let R be a valid p-dimensional correlation matrix and $G = \log R = \sum_{k=1}^{\infty} \frac{(-1)^k (R-I)^k}{k}$. Note that the convergence of the infinite summation and hence, the existence of G are ensured by the fact that R is a correlation matrix. Furthermore, let $q = vecl(G)$. In summary, the Fisher z-transformation of C is defined by the mapping $q = vecl(\log R)$. One of key theoretical contributions of Archakov and Hansen (2021) is to show that that this mapping is one-to-one. Thus, given any $\frac{p(p-1)}{2}$-dimensional vector q, there exists a unique and valid p-dimensional correlation matrix R. Although the inverse mapping from q to R does not have a closed-form expression when $p > 2$, R can be obtained numerically from q using an iterative algorithm as shown in Archakov and Hansen (2021).

When $p = 2$, Archakov and Hansen (2021) show that the above-defined transformation reduces to Fisher z-transformation. The new transformation retains the advantages of Fisher 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. Second, compared with original elements in R, the distribution of elements in q is often closer to Gaussian due to the use of log transformation. Hence, it is reasonable to model elements of q via a Gaussian process. Third, this transformation is invariant to ordering of variables. This is in

sharp contrast to that based on the Cholesky decomposition. Fourth, although elements of q depend on R in a nonlinear way, many interesting properties in R carry over to $G = \log(R)$, including the equicorrelation structure and the block-equicorrelation structure; see Archakov et al. (2020). For the sake of notational simplicity, in the rest of the paper, we refer to the mapping $vecl(\log(\cdot))$ as $F(\cdot)$ and its inverse as $F^{-1}(\cdot)$.

2.3 MSV-GFT model

To introduce our new MSV model, let $r_t = (r_{1t}, ..., r_{pt})'$ denote the $p \times 1$ vector of asset returns and $h_t = (h_{1t}, ..., h_{pt})'$ the vector of latent log-volatilities of these returns at time t. Let $V_t = \exp(diag(h_t))$. Let $q_t = (q_{1t}, ..., q_{dt})'$ denote the vector of latent variables at time t that underlie all the correlation coefficients in R_t, where $d = \frac{p(p-1)}{2}$. In particular, q_t is connected to R_t through the transformation detailed in Section 2.2. Our basic MSV model, which we refer to as MSV-GFT, is given by

$$r_t = V_t^{1/2}\epsilon_t,\ \epsilon_t \sim N(0, R_t), \tag{3a}$$

$$V_t = \exp(diag(h_t)), \tag{3b}$$

$$q_t = F(R_t), \tag{3c}$$

$$h_{t+1} = \mu_h + \Phi_h(h_t - \mu_h) + \eta_{ht},\ \eta_{ht} \sim N(0, \Sigma_h), \tag{3d}$$

$$q_{t+1} = \mu_q + \Phi_q(q_t - \mu_q) + \eta_{qt},\ \eta_{qt} \sim N(0, \Sigma_q), \tag{3e}$$

$$h_0 \sim N\left(\mu_h, (I_p - \Phi_h^2)^{-1}\Sigma_h\right),\ q_0 \sim N\left(\mu_q, (I_d - \Phi_q^2)^{-1}\Sigma_q\right), \tag{3f}$$

where $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})'$, $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hpt})'$, $\eta_{qt} = (\eta_{q1t}, ..., \eta_{qdt})'$, $\mu_h = (\mu_{h1}, ..., \mu_{hp})'$, $\mu_q = (\mu_{q1}, ..., \mu_{qd})'$, $\Phi_h = diag((\phi_{h1}, ..., \phi_{hp})')$, $\Phi_q = diag((\phi_{q1}, ..., \phi_{qd})')$, and $t = 1, ..., T$. It is assumed that ϵ_t, η_{ht} and η_{qt} are independent. This implies that no leverage (neither self-leverage or cross-leverage) effect is allowed. It also implies that the shocks to the volatility dynamics (i.e. η_{ht}) are completely separated from those to the correlation dynamics (i.e. η_{qt}). To reduce the number of parameters, we further assume that $\Sigma_h = diag((\sigma_{h1}^2, ..., \sigma_{hp}^2)')$ and $\Sigma_q = diag((\sigma_{q1}^2, ..., \sigma_{qd}^2)')$.

In MSV-GFT, h_t is a p-dimensional latent variable that determines the volatilities via the exponential transformation and q_t is a d-dimensional latent variable that determines the correlation coefficients via the F transformation. Elements of two types of latent variables are assumed to follow independent AR(1) processes.

It is important to note that in the MSV-GFT model, persistence in elements of $\{q_t\}$ can be heterogeneous across pairs. This is in sharp contrast to models based on the idea

of DCC or the Wishart autoregression, where persistence of all the correlation sequences is assumed to be the same. While the 'equi-persistence' makes the model more parsimonious, the empirical validity of this assumption has to be verified. To empirically examine the importance of heterogeneity in persistence in elements of q_t, we will also consider a restricted model, referred to as MSV-GFT-eq, where we assume $\phi_{q1} = \cdots = \phi_{qd}$.

2.4 MSV-GFT model with asymmetric effects

To incorporate asymmetric effect in volatility into SV models, various specifications have been proposed. So et al. (2002), for example, consider a threshold stochastic volatility model in which log-volatility is assumed to follow a threshold autoregressive process with an exogenous trigger variable. A multivariate extension of this model is proposed in So and Choi (2008). Another popular specification, considered in for example Asai and McAleer (2006), is to impose a negative correlation between the return of a stock and its future volatility. Smith (2009) compares abovementioned two specifications and finds out that negative correlation assumption does a better job of capturing asymmetry. It is also argued in Yu (2005) that this assumption leads to a correct timing and clearer interpretation of leverage effect in volatility. We hence also employ this specification in the current paper. In particular, we assume the leverage effect as following

$$\begin{pmatrix} \epsilon_t \\ \eta_{ht} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} R_t & R_t^{\frac{1}{2}} \Omega \Sigma_h^{\frac{1}{2}} \\ \Sigma_h^{\frac{1}{2}} \Omega R_t^{\frac{1}{2}} & \Sigma_h \end{pmatrix} \right), \tag{3g}$$

where $\Omega = diag(\rho)$ and $\rho = (\rho_1, ..., \rho_p)'$. The model defined by equations (3a)-(3g) will be referred to as MSVL-GFT.

As for asymmetric effect in correlations, we follow Gorgi and Koopman (2021) to impose a threshold autoregressive structure. The transition dynamics of correlation variables is modified to switch between two regimes determined by the sign of lagged market return, in the similar spirit of So et al. (2002). Specifically, we replace the equation (3e) with

$$q_{t+1} = \mu_q^{s_t} + \Phi(q_t - \mu_q^{s_t}) + \eta_{qt}, \tag{3e'}$$

where $\mu_q^{s_t} = \mu_q + \gamma_\mu I(r_t^{sp} < 0)$, $\gamma_\mu = (\gamma_{\mu_1}, ...\gamma_{\mu_d})'$ and r_t^{sp} denotes the return of S&P 500 at time t. It is well known in the literature that there is increased correlation during the bear market; see Longin and Solnik (2001) and Campbell et al. (2002). Hence, γ_μ is expected to be positive. In the rest of paper, we refer to the model defined by equations (3a)-(3d), (3e') and (3f)-(3g) as MSVLA-GFT.

3 Inference of MSV-GFT and MSVLA-GFT Model

Due to the difficulty of evaluating the likelihood function, the literature on MSV models relies on Bayesian methods to carry out statistical inference. To fix some notations, let $r = (r_1', ..., r_T')'$, $r^{sp} = \left(r_0^{sp}, ..., r_{T-1}^{sp}\right)'$, $h = (h_1', ..., h_T')'$, $q = (q_1', ..., q_T')'$, $x = (h', q')' := (x_1', ..., x_T')'$ so that $x_t = (h_t', q_t')'$. Vector x contains all latent variables and vector x_t contains all latent variables at period t. We use $x_{1:t}$ to denote $(x_1', ..., x_t')'$ for any $t = 1, \cdots, T$, θ denote the set of parameters in the model, and $p(r|\theta)$ denote the likelihood function of the model.

3.1 Review of standard Bayesian methods for MSV models

Unlike univariate and multivariate GARCH, which can be estimated straightforwardly by the frequentist maximum likelihood (ML) method, SV models are particularly challenging in terms of likelihood-based estimation and inference. The difficulty mainly arises from the high-dimensional latent variables involved in SV models. To be more specific, to obtain the likelihood function of SV models, one needs to integrate out the latent variables from the joint probability density of the observables and the latent variables, that is,

$$p(r|\theta) = \int p(r, x|\theta)\mathrm{d}z = \int p(r|h, q, \theta)p(h|\theta)p(q|\theta)\mathrm{d}h\mathrm{d}q.$$

Unfortunately, such an integration, being $((p + d) \times T)$-dimensional, does not have the analytical solution.

In the context of MSV models, there is an extra difficulty with the ML method. MSV models involve a large number of parameters (i.e. the dimension of θ is large). The ML method requires numerical maximization of $\log p(r|\theta)$ over θ. This often imposes a numerical challenge.

To deal with these two complications, the literature on MSV studies rely on Bayesian methods to conduct statistical inference. A popular choice is standard MCMC methods which conduct the Bayesian posterior analysis based on $p(r|\theta, x)$ which is more tractable than $p(r|\theta)$. In particular, standard MCMC methods consists of alternately updating the component of x conditional on θ and θ conditional on x. A single-move MCMC method draws a single element of x at a time while a multiple-move MCMC method draws a block of elements of x at a time.

Standard MCMC algorithms have been applied to estimate MSV models in the literature. Yu and Meyer (2006) use the single-move algorithm to estimate several bivariate MSV models. Yamauchi and Omori (2020) use the single-move algorithm to estimate the pairwise-Fisher-transformation-based MSV model. However, the single-move algorithm is

well known to be inefficient, as it generates highly autocorrelated Markov chains, suggesting a vast amount of random draws are required to achieve a satisfactory accuracy of estimation.

To improve efficiency, Ishihara and Omori (2012), Ishihara et al. (2016), Kurose and Omori (2016) have resorted to the multi-move algorithms to estimate different MSV models. These studies are built on some earlier works by De Jong and Shephard (1995), Pitt and Shephard (1999), Kim et al. (1998), and Watanabe and Omori (2004) in the univariate time series context. The multi-move samplers often require the second-order approximation to the target distribution. In general the derivation of the approximation is model dependent, making a generic multi-move algorithm not possible.

When latent variables of various degree of persistence co-exist in a model, the single-move sampler and the multi-move sampler may be combined. This idea is exploited in Asai and McAleer (2009b) where the multi-move sampler is applied to the latent variables that determine volatilities and the single-move sampler is applied to the latent variables that determine the correlations.

While standard MCMC algorithms can draw random samples from $p(\theta, x|r)$, additional efforts are needed to compute the likelihood $p(r|\theta)$ and the marginal likelihood $p(r)$ of the model. The marginal likelihood is an important quantity for model comparison.

3.2 Gibbs sampler based on particle filter

In this paper, instead of using above-mentioned standard MCMC techniques, we apply a PMCMC method known as PG, due to Andrieu et al. (2010), to estimate the proposed MSV model.[4] The intuition is to construct a high-dimensional efficient Markov kernel for latent processes using the particle filter.

3.2.1 Introduction to PG approach

Consider a general state-space model given by

$$r_t|x_t = x, \theta \sim f(\cdot|x, \theta),$$

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

where $f(\cdot|x, \theta)$ is the measurement density, $g(\cdot|x, \theta)$ is the transition probability density and $\mu_\theta(\cdot)$ is the initial density.

[4] Another PMCMC method potentially applicable here is Particle Metropolis-Hasting. See Xu and Jasra (2019) for its application in MSV model with constant correlation matrix and cross-leverage. It is not chosen, however, as it requires an accurate estimation of the likelihood and hence a very large number of particles.

To sample from $p(\theta, x_{1:T}|y_{1:T})$, a Gibbs sampler draws alternately from the two conditional densities, $p(\theta|x_{1:T}, r_{1:T})$ and $p(x_{1:T}|r_{1:T}, \theta)$. PG draws random samples from $p(x_{1:T}|r_{1:T}, \theta)$ based on the particle filter, which is applicable as long as the measurement density $f(\cdot|x, \theta)$ can be numerically evaluated and the transition density $g(\cdot|x, \theta)$ can be simulated.[5]

The 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. Let $\{x_{1:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$ be a random measure, where $\{x_{1:t}^{(i)}, i = 1, ..., N\}$ is a set of support points and $\{w_t^{(i)}, i = 1, ..., N\}$ are the associated weights. Each point is called a particle, and N is the number of particles used. The approximate distribution can then be written as

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

where $r_{1:t}$ is similarly defined and $\delta(\cdot)$ is the Dirac delta function. $\hat{p}_\theta$ is a discrete weighted approximation to the target distribution p_θ. Apparently, the accuracy of the approximation can be improved if an increasing number of particles are included. Doing so, however, also dramatically raises the computational burden.

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

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

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 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 $q_\theta(x_{1:t}|r_{1:t}) = q_\theta(x_t|x_{t-1}, r_t) q_\theta(x_{1:t-1}|r_{1:t-1})$. For any existing weighted sample $\{x_{1:t-1}^{(i)}, w_{1:t-1}^{(i)}\}$ that follows from $p_\theta(x_{1:t-1}|r_{1:t-1})$, we augment it with the new state $x_t^{(i)}$ randomly drawn from $q_\theta(x_t|x_{t-1}, r_t)$. 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

[5]Despite its general applicability, when implementing particle filter for a particular model, many subtle issues must be considered. These include how to choose a proper importance density, 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).

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 $p_\theta(r_{1:t}|r_{1:t-1})$, which has a simple formula $\hat{p}_\theta(r_{1:t}|r_{1:t-1}) = \frac{1}{N}\sum_{i=1}^{N} w_t^{(i)}$. The likelihood can then be easily obtained as $\hat{p}_\theta(r_{1:T}) = \hat{p}_\theta(r_1)\prod_{t=2}^{T}\hat{p}_\theta(r_{1:t}|r_{1:t-1})$.

One subtlety to note is that, to ensure the targeted joint density is indeed the invariant distribution of a Markov chain, we have to modify the particle filter when applying PG. Specifically, one particle trajectory must be specified a priori to serve as a reference. This modified version is known in the literature as conditional particle filter. The intuition is that this particular path can guide the simulated particles to move within a relevant region of the state space. See Theorem 5 of Andrieu et al. (2010) for more details.

3.2.2 Advantages of PG

As a PMCMC method, PG enjoys a few desirable properties compared with standard MCMC methods. First, relative to the single-move sampler, a significant improvement can be achieved in terms of efficiency by PG.

Second, unlike the multi-move samplers that are model dependent, PG requires a minimal modification across different models, as long as they could be cast into a state-space form.

Third, as said before, an important by-product of the filtering strategy is to evaluation of the likelihood $p(r|\theta)$. Once $p(r|\theta)$ is known, the marginal likelihood $p(r)$ can also be calculated easily. 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 on the Deviance Information Criterion (DIC).[6] The computation of the Bayes factor requires $p(r)$ while the computation of DIC requires $p(r|\theta)$. Hence, model comparison is straightforward in PG.

3.2.3 Particle Gibbs with ancestor sampling

As noted in Lindsten et al. (2014) and Chopin and Singh (2015), the mixing of the Markov kernel induced by PG can be rather slow when there is path degeneracy. For the high-dimensional problem, such as the one we consider in this paper, path degeneracy is inevitable.

[6]When comparing two candidate models (nested or non-nested), the log marginal likelihood of the first model minus that of the second model leads to the log Bayes factor (BF); see Kass and Raftery (1995). According to the well-known Jeffreys' table, when the log BF less than 1.6, the first model is barely worth mentioning. When the log BF takes a value in $(1.6, 3.3)$ (or $(3.3, 5.0)$ or $(5.0, 6.6)$ or $(6.6, +\infty)$) , there is substantial (or strong or very strong or decisive) evidence to support the first model. DIC is a Bayesian version of AIC with the aim of favouring models that are likely to make good predictions; see Spiegelhalter et al. (2002) and Li et al. (2020). The smaller DIC, the better the model. Jeffreys' table can be used to interpret the difference between one half of the DIC value of the second model and that of the first model.

To overcome this problem, Lindsten et al. (2014) propose to use an additional step called ancestor sampling in PG. The PGAS algorithm enjoys fast mixing of the Markov kernel even 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.[7] This approach has also been used in Gong and Stoffer (2021) for efficient fitting of stochastic volatility. They show that, for univariate SV model, PGAS algorithm mixes well enough with only 20 particles.

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 practice, we can gain a lot in terms of computational efficiency by using the PGAS algorithm. In summary, we believe that PGAS is a suitable estimation tool given our model setup. Its performance will be further examined in simulation in Section 4.

3.3 Bayesian analysis of MSV-GFT and MSVLA-GFT

3.3.1 Inference of MSV-GFT

We now present the Bayesian analysis of our MSV-GFT model. The first step is to specify the prior distributions of all the parameters $\theta = (\mu_h, \mu_q, \phi_h, \phi_q, \sigma_h^2, \sigma_q^2)'$. In this regard, our specification follows those adopted in Kim et al. (1998). For 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, for $i = 1, ..., p$ and $j = 1, ..., d$, 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 $\sigma_{qj}^2 \sim IG(\frac{n_{m0}}{2}, \frac{d_{m0}}{2})$,

[7] 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.

where $m_{\mu 0}, s^2_{\mu 0}, a, b, n_{m0}, d_{m0}$ are hyperparameters.

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. 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 $\sigma^2_h, \sigma^2_q|r, h, q, \theta_{/(\sigma^2_h,\sigma^2_q)}$.

Iteration over steps 2-5 consists of a complete sweep of MCMC sampler. We apply PGAS introduced in Section 3.2 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. On the other hand, from the joint posterior density, it is straightforward to sample each element in θ given one realization of latent processes h and q. The details are provided in Appendix B.1.

3.3.2 Inference of MSVLA-GFT

We now present the Bayesian analysis of our MSVLA-GFT model. It is an extension of MSV-GFT model with additional $p+d$ parameters, including the leverage effect of volatilities $\Omega = diag((\rho_1, ..., \rho_p)')$ and asymmetric effect of correlations $\gamma_\mu = (\gamma_{\mu_1}, ...\gamma_{\mu_{qd}})'$.We set prior distributions following Yu (2005) and So et al. (2002). We choose the uniform prior on $(-1, 1)$ for the leverage effects while normal priors for the threshold parameters. In summary, for $i = 1, ..., p$ and $j = 1, .., d$, we choose the following prior distribution:

- $\rho_i \sim U(-1, 1)$ and $\gamma_{\mu_j} \sim N(m_{\gamma 0}, s^2_{\gamma 0})$

where $m_{\gamma 0}$ and $s^2_{\gamma 0}$ are hyper-parameters.

Redefining $\theta = (\mu_h, \mu_q, \phi_h, \phi_q, \sigma^2_h, \sigma^2_q, \rho, \gamma_\mu)'$, we implement a Gibbs sampler with six blocks. The algorithm proceeds as:

1. Initialize h, q and θ.
2. Draw $h, q|r, \theta, r^{sp}$.
3. Draw $\mu_h|r, h, q, \theta_{/(\mu_h)}$.
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 $\rho | r, h, \theta_{/(\rho)}$.

7. Draw $\mu_q, \gamma_\mu | q, r^{sp}, \theta_{/(\gamma_\mu, \mu_q)}$.

Iteration over steps 2-7 consists of a complete sweep of MCMC sampler. The algorithm is similar to the one in Section 3.3.1. The differences lie in that we apply PGAS to sample the latent variable given the additional information r^{sp}, and we sample the parameters of asymmetric effects in the step 6 and 7. The details are provided in Appendix B.2.

4 Simulation Studies

To investigate the performance of our estimation procedure, we conduct some simulation exercises in this section. Our simulation design is frequentist in nature as we fix the parameters at their true values and generate data from the same data generating process with 100 repetitions. We use the posterior mean as a point estimator for all the parameters. Since the true values are known, we are thus able to calculate bias (defined as the difference between the true values and the average value of posterior means) and the standard deviation.[8]

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

Our data generating process is MSVLA-GFT model with $p = 3$. There are 24 parameters in the model, 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 $\sigma_{q1}^2 = \sigma_{q2}^2 = \sigma_{q3}^2 = 0.05$.

[8]Here, the standard deviation refers to the variation across replications, rather than the numerical standard error of MCMC sampler introduced below.

4. $\rho_1 = \rho_2 = \rho_3 = -0.5$ and $\gamma_{\mu_1} = \gamma_{\mu_2} = \gamma_{\mu_3} = -0.2$.

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.[9] 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. As three h's and q's in our setup are symmetric, we only report the results for h_1 and q_1. The results for other latent processes are similar and hence omitted.

The estimation results for μ_{h1}, ϕ_{h1}, σ^2_{h1},ρ_1, μ_{q1}, ϕ_{q1}, σ^2_{q1}, γ_{μ_1} including the average values of posterior means, standard deviations and IFs across replications, are reported in Table 1. It can be seen that even for a small sample size such as 500 and a relatively small number of particles such as 50, the posterior means for both μ_{h1} and μ_{q1} are reasonably close to their respective true values, although there is an downward bias for μ_h and a upward bias for μ_{q1}. Nevertheless, as expected, the standard deviations for both μ_{h1} and μ_{q1} 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, the bias for μ_{h1} reduces from 0.009 to 0.005 if 200 particles are used instead of 50. A similar improvement applies to μ_{q1}. However, an increasing number of particles has no effect on the standard deviation.

Meanwhile, the persistence parameters ϕ_{h1} and ϕ_{q1} can be estimated accurately, even with 500 observations and 50 particles. The estimates have very small biases and low standard deviations. When 200 particles are used, the bias almost completely vanishes. Substantial downward biases are observed for σ^2_{h1} and σ^2_{q1} when 50 particles are used. This bias is insensitive to the number of observations. Fortunately, it can be improved by using more particles. When N=200, the bias becomes much smaller, although it seems that a larger number of particles are necessary to completely remove this bias.

In line with Jacquier et al. (2004), we find a downward bias in leverage effect ρ_1. The bias becomes smaller when we increase the number of particles or the sample size T. For threshold effect on correlation, there exists a downward bias when the sample size is small, inducing an upward bias in γ_μ. This bias shrinks when the sample size increases, but it is insensitive to the number of particles.

Finally, the IFs do 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 μ's and the highest for σ^2's. Compared with the traditional single-move or multi-move Gibbs sampler (for example,

[9]Plot of autocorrelation function suggests that the MCMC sampling has already converged after 1000 iterations.

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 good performance, 200 particles are used for the empirical applications reported later in the paper.

As additional evidence to support the proposed method, Figure 1 and 2 plot the filtered hs and qs, together with their 95% credible intervals and true values. In all cases, the filtered values are close to the respective true values.

5 Empirical Studies

As illustrations, we consider two empirical applications of our proposed MSV models in this section. The first application is concerned with exchange rate data and we focus on the in-sample effect of allowing each correlations to have different persistence. In the second application, we consider stock return data and examine both the in-sample fit and the out-of-sample forecasting performance when the asymmetric effects are allowed in the model.

5.1 Weekly foreign exchange rates

In the first empirical study, we use data that contains 1406 weekly mean-corrected log-returns of Euro, Pound sterling, and Swiss franc exchange rates (i.e. $p = 3$), all against the US dollar.[10] The sampling period is from January 13, 1993 to December 25, 2019 and the dynamics of these three sequences are plotted in Figure 3. These series are expected to be correlated, as the underlying economies are closely connected. Indeed, the Panel 1 of Table 2 presents the summary statistics and it can be seen that all series are stationary with sample correlations range from 0.521 to 0.809.

For comparison, four candidate models are considered:

1. MSV-GFT.

2. MSV-GFT-eq.

3. MSV-Chol. This is a model based on Cholesky decomposition proposed by Lopes et al.

[10]The data were obtained from the Sauder School of Business at the University of British Columbia via http://fx.sauder.ubc.ca/data.html.

(2010). The specification is as follows:[11]

$$
\begin{aligned}
&r_t = H_t^{-1} V_t^{\frac{1}{2}} \epsilon_t, \epsilon_t \sim N(0, I_p), \\
&V_t = diag((\exp^{h_{1,t}}, ..., \exp^{h_{p,t}})'), \\
&H_t = \begin{pmatrix} 1 & 0 & \dots & 0 \\ -h^\star_{1,t} & 1 & \dots & 0 \\ -h^\star_{2,t} & -h^\star_{3,t} & 1 & \dots \\ \vdots & \ddots & \ddots & \vdots \\ \dots & \dots & -h^\star_{\frac{p(p-1)}{2},t} & 1 \end{pmatrix}, \\
&h_{i,t+1} = \mu_{hi} + \phi_{hi}(h_{i,t} - \mu_{hi}) + \eta_{h_{i,t}}, \eta_{h_{i,t}} \sim N(0, \sigma^2_{hi}), i = 1, ..., p, \\
&h^\star_{j,t+1} = \mu_{h^\star i} + \phi_{h^\star j}(h^\star_{j,t} - \mu_{h^\star j}) + \eta_{h^\star_{j,t}}, \eta_{h^\star_{j,t}} \sim N(0, \sigma^2_{h^\star_j}), j = 1, ..., \tfrac{p(p-1)}{2},
\end{aligned}
$$

where ϵ_t, $\eta_{h_{i,t}}$ and $\eta_{h^\star_{j,t}}$ are mutually independent for all i, j and t.

4. MSV-DCC. This is the model proposed in Asai and McAleer (2009b), where a DCC-type structure with a Wishart transition dynamics is used to characterize the movement of correlation matrix. It is defined as

$$
\begin{aligned}
r_t &= V_t^{1/2} \epsilon_t, \ \epsilon_t \sim N(0, R_t), \\
V_t &= \exp\left(diag(h_t)\right), \\
h_{t+1} &= \mu_h + \Phi_h(h_t - \mu_h) + \eta_{ht}, \ \eta_{ht} \sim N(0, \Sigma_h), \\
R_t &= \tilde{Q}_t^{-1} Q_t \tilde{Q}_t^{-1}, \\
Q_{t+1}^{-1} | k, Q_t^{-1} &\sim Wishart\left(k, \tfrac{1}{k} Q_t^{-\phi/2} \Lambda Q_t^{-\phi/2}\right), \\
\Lambda &= \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{p1} & \dots & a_{pp} \end{pmatrix},
\end{aligned}
\tag{4}
$$

where $\tilde{Q}_t = (diag(diag(Q_t))^{1/2}$,$\Phi_h = diag((\phi_{h1}, ..., \phi_{hp})')$, $\Sigma_h = diag((\sigma^2_{h1}, ..., \sigma^2_{hp})')$ and k, ϕ, a_{ij} are all scalars for $i = 1, .., p$ and $j = 1, .., p$. It is also assumed that Q_t, η_{ht} and ϵ_t are mutually independent for all t.

It is important to note that all candidate models except for MSV-Chol share the same parametrization of the volatility dynamics. The key difference among these models is in the way how the correlation dynamics is specified. We report the posterior statistics of parameters in all four models, including the posterior means, the posterior standard deviations, and the 95% credible intervals. Also reported are the IFs for all parameters as well as the

[11] When estimating MSV-Chol model, we arbitrarily choose the order Euro, Pound Sterling and Swiss franc.

log marginal likelihood and DIC values for all models. The number of particles is set at 200. We obtain 20000 MCMC samples after a 2000 burn-in period.

Tables 3 reports the posterior statistics of the parameters related to volatilities.[12] The posterior means and posterior standard deviations of all the parameters related to volatility are largely in line with existing literature. In particular, all log volatility sequences have a very high level of persistence, with the autoregressive root close to but smaller than 1 for all assets. The IFs are all reasonably small, suggesting the MCMC draws mix well.

Table 4 reports the posterior statistics of the parameters related to correlations. Figure 4 plots the posterior mean of correlation in MSV-GFT. First and foremost, the posterior means of ϕ_q are $(0.834, 0.946, 0.97)$ in MSV-GFT, suggesting a great deal of heterogeneity in the level of persistency in the elements of q_t. This finding is in sharp contrast to the implication in the MSV-GFT model with equi-persistence where $\phi_{q1} = \phi_{q2} = \phi_{q3}$. By forcing $\phi_{q1} = \phi_{q2} = \phi_{q3}$ in this model, we obtain the posterior mean of 0.969 for ϕ_q, which is very large and close to the posterior mean of ϕ_{q3}, the largest in MSV-GFT. Second, the posterior means of σ^2_{qi} for $i = 1, 2, 3$ in MSV-GFT and of σ^2_{q1} in MSV-GFT with equi-persistence are significantly different from zero, suggesting time-varying correlation coefficients.

In order to assess whether the flexibility in MSV-GFT leads to any improvement in in-sample statistical performance, we compare the log marginal likelihoods and DIC values of all four models. The results are presented in the last two rows of Table 4. Based on the log marginal likelihood and DIC, both MSV-GFT and MSV-Chol outperform MSV-GFT-eq and MSV-DCC, providing compelling evidence in favor of MSV-GFT and MSV-Chol, which aligns with the findings in Bucci et al. (2022).

It is worth noting that although MSV-Chol exhibits a higher log marginal likelihood value than MSV-GFT, MSV-GFT demonstrates the lowest DIC value. Furthermore, despite MSV-DCC and MSV-GFT-eq utilizing a single parameter to control the persistent level of correlations, MSV-GFT-eq displays a higher log marginal likelihood value and a lower DIC value compared to MSV-DCC.

5.2 Daily equity returns

In the second empirical study, we conduct a comparative analysis of the proposed model against several existing models, evaluating their performance in both in-sample and out-of-sample forecasting. The analysis focuses on daily demeaned close-to-close log-returns of three stocks, namely DuPont (DD), Exxon Mobil (XOM), and 3M Company (MMM).[13] Our full sample period is from January 4, 2014 to December 31, 2020 with the sample size $T = 1761$.

[12] Note that $h_{i,t}$ in MSV-Chol model is not the volatility of i^{th} asset when $i > 1$.

[13] The data were obtained from Yahoo Finance via https://finance.yahoo.com/.

The dynamics of these three log-return sequences are plotted in Figure 5 and Panel 2 of Table 2 presents the summary statistics. Note that all three sample pairwise correlations are between 0.5 and 0.6.

For comparison, the following three categories of model specifications are considered

1. MSV models
   (a) MSV-CC defined by (3a)-(3d) and (1).
   (b) MSV-GFT.
   (c) MSV-GFT-eq.
   (d) MSV-DCC.
2. MSVL models
   (a) MSVL-CC. This is MSV-CC model with leverage effect defined in (3g).
   (b) MSVL-GFT.
   (c) MSVL-GFT-eq. This is MSVL-GFT model with equi-persistency restriction.
   (d) MSVL-DCC. This is the model defined by equations (4) with additional assumption (3g).
3. MSVLA-GFT model.

Since it is not clear what is the best way to incorporate the leverage effects in the context of the MSV-Chol specification, we decide not to include any MSV-Chol-based model in the comparative analysis. For the implementation of Bayesian estimation, we draw 20000 MCMC samples and discard first 2000 as the burn-in period.

We first conduct the full sample analysis for stock return data. Figure 6 plots the posterior mean of correlation in MSVLA-GFT. Tables 5 and 6 report the posterior statistics of the parameters related to volatilities and asymmetric effects respectively. It can be seen that the leverage effect ρ is always significantly negative, consistent with the findings in the existing literature. More importantly, in line with our expectation, the asymmetric effects in correlations are all positive, though insignificant at the 5% level in some cases.

Table 7 reports the posterior statistics of the parameters related to correlations as well as the log marginal likelihood and DIC values for all models.[14] The following three conclusions

[14]To save the space, we do not report the in-sample estimation results for MSV-CC and MSVL-CC model in Table 7. The DIC for these two models are 22906 and 22863 respectively. The log marginal likelihood are -11466 and -11442 respectively. It can be seen that MSVL-CC has a better in-sample performance than MSV-CC, and both of them are worse than other candidate models.

can be made from this table. First and foremost, MSVLA-GFT and MSVL-GFT provide the best in-sample fit among all models. While MSVL-GFT has the highest log likelihood value, MSVLA-GFT has the lowest DIC value. This finding indicates that the incorporation of leverage effects and/or asymmetric effects indeed improves the in-sample performance. Second, it is noteworthy that MSVL-GFT-eq performs clearly worse than MSVL-GFT model, suggesting that imposing homogeneity among correlations would significantly reduce the goodness of in-sample fit. Thirdly, MSVL-DCC model is least favorable among all candidates, worse than MSV-GFT model without leverage and MSVL-GFT model with equi-persistency.

We then conduct an out-of-sample forecast comparison. The forecast period is from January 2, 2020 to December 31, 2020, consisting of 252 trading days. A rolling window approach with a fixed in-sample size is adopted to obtain one-step-ahead forecasts. For each out-of-sample trading day (say t), the one-step-ahead forecast of the covariance matrix $C_{t|t-1}$ is generated by the posterior mean of C_t conditional on observations in past four years (1006 days). The forecast of the covariance matrix is used to construct a global minimum variance (GMV) portfolio. According to Markowitz (1952), the GMV portfolio is optimal as it has the smallest variance among all portfolios on the efficient frontier. At period $t-1$, we construct the GMV portfolio with the optimal weights $w_t = (w_{1t}, ..., w_{pt})$,[15] where

$$w_t = \frac{\hat{C}_{t|t-1}^{-1}\mathbb{1}_p}{\mathbb{1}_p'\hat{C}_{t|t-1}^{-1}\mathbb{1}_p}, \tag{5}$$

and the optimal portfolio return at time t is then

$$R_t^p = w_t' r_t. \tag{6}$$

In addition to the MSV models discussed earlier, we also consider a portfolio with equal weights as a benchmark, which is frequently used in practice. To enable a fair comparison across models, we assume that all stocks have equal expected returns and focus solely on the variance of the portfolio. Specifically, we measure the portfolio variance by computing the average squared return over out-of-sample periods. For the sake of presentation, we report the average squared return of each strategy relative to that of equal-weight portfolio. The results are reported in Table 8.

Upon analysis, we note that the equal-weight portfolio exhibits a significantly larger variance across all instances. Additionally, our findings demonstrate that the MSVL models consistently outperform their MSV counterparts. Surprisingly, the MSVL-GFT model generates the smallest average squared return, despite being less flexible compared to the

[15] We assume negative weights are allowed so that short-sells are possible.

MSVLA-GFT model. This observation aligns with the fact that the leverage effect remains significant in-sample, while the threshold effect in correlations can be inconsequential in certain cases. It also suggests that simplicity and parsimony can sometimes offer advantages when constructing the optimal portfolio.

Furthermore, the DCC-based MSV models exhibit the poorest performance, even underperforming the most restrictive models utilizing a constant correlation matrix. In conclusion, we establish that MSV models based on GFT, incorporating the leverage effect and/or the asymmetry effect, provide more reliable outcomes for out-of-sample portfolio construction.

6 Conclusion

We present a new approach to modeling multivariate stochastic volatility in this paper. Our approach uses a generalized version of Fisher's z-transformation to dynamically characterize the correlation structure in a highly flexible manner. One key advantage of our model is that it can automatically generate a positive-definite correlation matrix, while also completely separating the driving forces underlying volatilities and correlations. We go a step further and extend the model to incorporate both the leverage effect in volatility and the threshold effect in correlation.

Unlike many existing studies that rely on standard methods for Bayesian inference, we employ a Gibbs sampler with particle filter to perform the inference for our model. Through simulation, we demonstrate that our approach works well for our model. Overall, our proposed model offers a powerful and flexible tool for capturing the complex dynamics of multivariate stochastic volatility in financial markets.

Our empirical results show that this flexible way of modeling multivariate stochastic volatility lead to better in-sample fit to exchange rate volatilities. Moreover, when incorporating the leverage effect and the threshold effect in the new model specification, we can improve the in-sample and out-of-sample forecasting performances over many existing models.

While our proposed model demonstrates promising properties, there is still room for further improvement. For example, we have not yet incorporated realized measures in our analysis. This additional data source may significantly improve the statistical efficiency of our model.

Another important future direction is to address the challenge of handling a large parameter space when analyzing a large pool of assets simultaneously. As the computational burden of implementing the model increases exponentially with the number of assets, it may become impractical to use the model without modifications. To address this issue, we may

need to make certain assumptions, such as imposing (block) equi-correlation or latent factor structures.

References

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B 72(3), 269–342.

Archakov, I. and P. R. Hansen (2021). A new parametrization of correlation matrices. Econometrica 89(4), 1699–1715.

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(2), 174–188.

Asai, M., C.-L. Chang, and M. McAleer (2022). Realized matrix-exponential stochastic volatility with asymmetry, long memory and higher-moment spillovers. Journal of Econometrics 227(1), 285–304.

Asai, M. and M. McAleer (2006). Asymmetric multivariate stochastic volatility. Econometric Reviews 25(2-3), 453–473.

Asai, M. and M. McAleer (2009a). Multivariate stochastic volatility, leverage and news impact surfaces. The Econometrics Journal 12(2), 292–309.

Asai, M. and M. McAleer (2009b). The structure of dynamic correlations in multivariate stochastic volatility models. Journal of Econometrics 150(2), 182–192.

Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: a review. Econometric Reviews 25(2-3), 145–175.

Audrino, F. and F. Corsi (2010). Modeling tick-by-tick realized correlations. Computational Statistics & Data Analysis 54(11), 2372–2382.

Audrino, F. and F. Trojani (2011). A general multivariate threshold GARCH model with dynamic conditional correlations. Journal of Business & Economic Statistics 29(1), 138–149.

Bucci, A., L. Ippoliti, and P. Valentini (2022). Comparing unconstrained parametrization methods for return covariance matrix prediction. Statistics and Computing 32(5), 1–20.

Campbell, R., K. Koedijk, and P. Kofman (2002). Increased correlation in bear markets. Financial Analysts Journal 58(1), 87–94.

Cappiello, L., R. F. Engle, and K. Sheppard (2006). Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial Econometrics 4(4), 537–572.

Chan, D., R. Kohn, and C. Kirby (2006). Multivariate stochastic volatility models with correlated errors. Econometric Reviews 25(2-3), 245–274.

Chib, S., Y. Omori, and M. Asai (2009). Multivariate stochastic volatility. In Handbook of Financial Time Series, pp. 365–400. Springer.

Chopin, N. and S. S. Singh (2015). On particle gibbs sampling. Bernoulli 21(3), 1855–1883.

De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. Biometrika 82(2), 339–350.

Dellaportas, P., M. K. Titsias, K. Petrova, and A. Plataniotis (2023). Scalable inference for a full multivariate stochastic volatility model. Journal of Econometrics 232(2), 501–520.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20(3), 339–350.

Engle, R. and B. Kelly (2012). Dynamic equicorrelation. Journal of Business & Economic Statistics 30(2), 212–228.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica, 987–1007.

Gong, C. and D. S. Stoffer (2021). A note on efficient fitting of stochastic volatility models. Journal of Time Series Analysis 42(2), 186–200.

Gorgi, P. and S. J. Koopman (2021). Beta observation-driven models with exogenous regressors: A joint analysis of realized correlation and leverage effects. Journal of Econometrics, forthcoming.

Gouriéroux, C., J. Jasiak, and R. Sufana (2009). The wishart autoregressive process of multivariate stochastic volatility. Journal of Econometrics 150(2), 167–181.

Hafner, C. M. and L. Wang (2021). A dynamic conditional score model for the log correlation matrix. Journal of Econometrics, forthcoming.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. The Review of Economic Studies 61(2), 247–264.

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(11), 3674–3689.

Ishihara, T., Y. Omori, and M. Asai (2016). Matrix exponential stochastic volatility with cross leverage. Computational Statistics & Data Analysis 100, 331–350.

Jacquier, E., N. G. Polson, and P. E. Rossi (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122(1), 185–212.

Johansen, A. M. and A. Doucet (2008). A note on auxiliary particle filters. Statistics & Probability Letters 78(12), 1498–1504.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with arch models. The Review of Economic Studies 65(3), 361–393.

Kurose, Y. and Y. Omori (2016). Dynamic equicorrelation stochastic volatility. Computational Statistics & Data Analysis 100, 795–813.

Li, Y., J. Yu, and T. Zeng (2020). Deviance information criterion for latent variable models and misspecified models. Journal of Econometrics 216(2), 450–493.

Lindsten, F., M. I. Jordan, and T. B. Schön (2014). Particle gibbs with ancestor sampling. The Journal of Machine Learning Research 15(1), 2145–2184.

Longin, F. and B. Solnik (2001). Extreme correlation of international equity markets. The Journal of Finance 56(2), 649–676.

Lopes, H. F., R. McCulloch, and R. Tsay (2010). Cholesky stochastic volatility. Unpublished Technical Report, University of Chicago,.

Markowitz, H. (1952). Portfolio analysis. Journal of Finance 8, 77–91.

Nakajima, J. (2017). Bayesian analysis of multivariate stochastic volatility with skew return distribution. Econometric Reviews 36(5), 546–562.

Philipov, A. and M. E. Glickman (2006). Multivariate stochastic volatility via wishart processes. Journal of Business & Economic Statistics 24(3), 313–328.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. Journal of the American Statistical Association 94(446), 590–599.

Shirota, S., Y. Omori, H. F. Lopes, and H. Piao (2017). Cholesky realized stochastic volatility model. Econometrics and Statistics 3, 34–59.

Smith, D. R. (2009). Asymmetry in stochastic volatility models: Threshold or correlation? Studies in Nonlinear Dynamics & Econometrics 13(3).

So, M. K. and C.-Y. Choi (2008). A multivariate threshold stochastic volatility model. Mathematics and Computers in Simulation 79(3), 306–317.

So, M. K., W. K. Li, and K. Lam (2002). A threshold stochastic volatility model. Journal of Forecasting 21(7), 473–500.

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 64(4), 583–639.

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-gaussian time series models: Comments on shephard & pitt (1997). Biometrika 91(1), 246–248.

Xu, Y. and A. Jasra (2019). Particle filters for inference of high-dimensional multivariate stochastic volatility models with cross-leverage effects. Foundations of Data Science 1(1), 61.

Yamauchi, Y. and Y. Omori (2020). Multivariate stochastic volatility model with realized volatilities and pairwise realized correlations. Journal of Business & Economic Statistics 38(4), 839–855.

Yu, J. (2005). On leverage in a stochastic volatility model. Journal of Econometrics 127(2), 165–178.

Yu, J. and R. Meyer (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. Econometric Reviews 25(2-3), 361–384.

Zaharieva, M. D., M. Trede, and B. Wilfling (2020). Bayesian semiparametric multivariate stochastic volatility with application. Econometric Reviews 39(9), 947–970.

Appendices

A Details of PGAS algorithm

Consider a state-space model in the form of equation 3.2.1 and 3.2.1. The output of a PGAS algorithm is a random draw from the joint smoothing distribution $p_\theta(x_{1:T}|r_{1:t})$, 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 trajctory of $x_{1:T}$, which is a sample from last MCMC iteration. Let's denote that reference trajctory by $x'_{1:T}$. Then, the algorithm proceeds as following:

- Draw $x_1^{(i)}$ from $q_1(x_1|y_1)$, for $i = 1, 2, ..., N-1$.
- Set $x_1^{(N)} = x'_1$.
- Set $w_1^{(i)} = f(y_1|x_1^{(i)})/q_1(x_1^{(i)}|y_1)$, for $i = 1, 2, ..., N$.
- For $t = 2$ to T, do the following:
  - 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 $\{w_{t-1}^{(i)}\}_{i=1}^{N}$.
  - Draw J from $\{1, 2, ..., N\}$ with probabilities proportional to $w_{t-1}^{(i)} g(x'_t|x_{t-1}^{(i)})$ and then set $\tilde{x}_{1:t-1}^{(N)} = x_{1:t-1}^{(J)}$.
  - Simulate $x_t^{(i)}$ from $q_t(x_t|\tilde{x}_{t-1}^{(i)}, y_t)$, for $i = 1, 2, ..., N-1$.
  - Set $x_t^{(N)} = x'_t$.
  - Set $x_{i:t}^{(i)} = (\tilde{x}_{1:t-1}^{(i)}, x_t^{(i)})$
  - 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 $i = 1, 2, ..., N$.
- 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 PG sampler. The major modification is in drawing J, where a new index is drawn and thus the N^{th} trajectory may not be the reference one from last iteration. In the conditional PG, on the contrary, we fix the last particle to follow the input trajectory $x'_{1:T}$. It's also 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 $r_{1:t}$. Therefore, this step makes the algorithm more like a smoothing instead of filtering.

B Details of Sampling Model Parameters

B.1 Sampling Parameters of MSV-GFT

The joint posterior distribution can be written as

$$
\begin{aligned}
p(\theta,h,q|r) &\propto p(r|\theta,h,q)p(\theta,h,q) \\
&= f(r|h,q)g_\theta(h)g_\theta(q)\pi(\theta) \\
&= f(r_1|h_1,q_1)g_\theta(h_1)g_\theta(q_1)\prod_{t=2}^{T}\left[f(r_t|h_t,q_t)g_\theta(h_t|h_{t-1})g_\theta(q_t|q_{t-1})\right]\pi(\theta) \\
&= \prod_{t=1}^{T}\left[\left(\sum_{i=1}^{p}h_{it}\right)|R_t|^{-1/2}\exp\left[-\frac{1}{2}r_t'(V_t^{1/2}R_tV_t^{1/2})^{-1}r_t\right]\right] \qquad (7) \\
&\quad\times\prod_{t=2}^{T}\prod_{i=1}^{p}\left[(\sigma_{hi}^2)^{-1/2}\exp\left(-\frac{1}{2\sigma_{hi}^2}\left(h_{it+1}-\mu_{hi}-\phi_{hi}(h_{it}-\mu_{hi})\right)^2\right)\right] \\
&\quad\times\prod_{t=2}^{T}\prod_{j=1}^{d}\left[(\sigma_{qj}^2)^{-1/2}\exp\left(-\frac{1}{2\sigma_{qj}^2}\left(q_{jt+1}-\mu_{qj}-\phi_{qj}(q_{jt}-\mu_{qj})\right)^2\right)\right] \\
&\quad\times\prod_{i=1}^{p}\left(\frac{\sigma_{hi}^2}{1-\phi_{hi}^2}\right)^{-1/2}\exp\left(-\frac{(h_{i1}-\mu_{h1})^2}{2\sigma_{hi}^2/(1-\phi_{hi}^2)}\right) \\
&\quad\times\prod_{j=1}^{d}\left(\frac{\sigma_{qj}^2}{1-\phi_{qj}^2}\right)^{-1/2}\exp\left(-\frac{(q_{j1}-\mu_{q1})^2}{2\sigma_{qj}^2/(1-\phi_{qj}^2)}\right)\pi(\theta).
\end{aligned}
$$

To sample from the posterior distribution of parameters conditional on the realization of latent variables, we can do the following:

1. We can directly sample from the full conditional distribution of μ_{hi} and μ_{qi} which a normal distribution. For $i=1,...,p$ and $j=1,...,d$,

$$
\mu_{hi}|r,h,q,\theta_{/\mu_{hi}} \sim N\left(\tilde{m}_{h\mu},\tilde{s}_{h\mu}^2\right) \text{ and } \mu_{qj}|r,h,q,\theta_{/\mu_{qj}} \sim N\left(\tilde{m}_{q\mu},\tilde{s}_{q\mu}^2\right) \qquad (8)
$$

where

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

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

and

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

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

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

$$\begin{aligned}\log p(\phi_{hi}|y,h,q,\theta_{/\phi_{hi}}) &\propto \log p(h_i|\phi_{hi},\theta_{/\phi_{hi}}) + \log\pi(\phi_{hi}) \\ &= \log\pi(\phi_{hi}) - \frac{(h_{i1}-\mu_{hi})^2(1-\phi^2_{hi})}{2\sigma^2_{hi}} + \frac{1}{2}\log(1+\phi^2_{hi}) \\ &\quad - \frac{\sum_{t=1}^{T-1}\left[(h_{it+1}-\mu_{hi}) - \phi_{hi}(h_{it}-\mu_{hi})\right]^2}{2\sigma^2_{hi}},\end{aligned} \tag{9}$$

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

$$\hat{\phi}_{hi} = \frac{\sum_{t=1}^{T-1}(h_{it+1}-\mu_{hi})(h_{it}-\mu_{hi})}{\sum_{t=1}^{T-1}(h_{it}-\mu_{hi})^2},$$

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

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

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

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

$$\sigma^2_{hi}|y,h,q,\theta_{/\sigma^2_{hi}} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{d}_{hm}}{2}\right) \text{ and } \sigma^2_{qj}|y,h,q,\theta_{/\sigma^2_{qj}} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{d}_{qm}}{2}\right), \tag{10}$$

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

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

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

B.2 Sampling parameters of MSVLA-GFT

In the MSVLA-GFT model, the parameter $\theta = (\mu_h, \mu_q, \phi_h, \phi_q, \sigma_h^2, \sigma_q^2, \sigma_q^2, \rho, \gamma_\mu)'$, and the joint posterior distribution can be written as

$$\begin{aligned}
p(\theta, h, q | r, r^{sp}) &\propto p(r|\theta, h, q, r^{sp}) p(\theta, h, q) \\
&= f(r|h, q, r^{sp}) g_\theta(h) g_\theta(q) \pi(\theta) \\
&= f(r_1|h_1, q_1, r_0^{sp}) g_\theta(h_1) g_\theta(q_1) \prod_{t=2}^{T} \left[f(r_t|h_t, q_t, r^{sp}) g_\theta(h_t|h_{t-1}) g_\theta(q_t|q_{t-1}) \right] \pi(\theta) \\
&= \prod_{t=1}^{T} \left[\left(\sum_{i=1}^{p} h_{it} \right) |R_t|^{-1/2} \exp\left[-\frac{1}{2} r_t' (V_t^{1/2} R_t V_t^{1/2})^{-1} r_t \right] \right] \\
&\quad \times \prod_{t=2}^{T} \prod_{i=1}^{p} \left[(\sigma_{hi}^2 (1 - \rho_i^2))^{-1/2} \exp\left(-\frac{1}{2\sigma_{hi}^2} \left(h_{it+1} - \mu_{hi} - \phi_{hi}(h_{it} - \mu_{hi}) - \rho_i \sigma_{hi} z_{it} \right)^2 \right) \right] \\
&\quad \times \prod_{t=2}^{T} \prod_{j=1}^{d} \left[(\sigma_{qj}^2)^{-1/2} \exp\left(-\frac{1}{2\sigma_{qj}^2} \left(q_{jt+1} - \mu_{qj}^{S_t} - \phi_{qj}(q_{jt} - \mu_{qj}^{S_t}) \right)^2 \right) \right] \\
&\quad \times \prod_{i=1}^{p} \left(\frac{\sigma_{hi}^2}{1 - \phi_{hi}^2} \right)^{-1/2} \exp\left(-\frac{(h_{i1} - \mu_{h1})^2}{2\sigma_{hi}^2/(1 - \phi_{hi}^2)} \right) \\
&\quad \times \prod_{j=1}^{d} \left(\frac{\sigma_{qj}^2}{1 - \phi_{qj}^2} \right)^{-1/2} \exp\left(-\frac{(q_{j1} - \mu_{q1})^2}{2\sigma_{qj}^2/(1 - \phi_{qj}^2)} \right) \pi(\theta),
\end{aligned}$$

where $z_t = R_t^{-\frac{1}{2}} V_t^{-\frac{1}{2}} r_t = (z_{1t}, z_{2t}.., z_{pt})'$ and $\mu_{qj}^{S_t} = \mu_{qj} + \gamma_{\mu j} I(r_t^{sp} < 0)$. To sample from the posterior distribution of parameters conditional on latent variables, we proceed as follows:

1. To draw the leverage effect ρ, note that the conditional distribution of $h_{t+1}|h_t, r_t, \theta$ is

$$h_{t+1}|h_t, r_t, q_t, \theta_{/(\rho)}\theta \sim N(\mu_h + \phi_h(h_t - \mu_h)) + \Omega\Sigma_h^{\frac{1}{2}} z_t, \Sigma_h - \Omega\Sigma_h\Omega')$$

   where $\Omega = diag(\rho)$. The posterior distribution of ρ is

$$f(\rho|h, r, q, \theta_{/(\rho)}) \propto \prod_{t=1}^{T-1} f(h_{t+1}|h_t, r_t, q_t, \theta_{/(\rho)}) I(\rho \in (-1, 1)),$$

   and we conduct a random walk Metropolis-Hasting sampler to draw ρ.

2. In the MSVLA-GFT model, note that q_{jt} follows a threshold model

$$q_{j(t+1)} = (\mu_{qj} + \gamma_{\mu j} I(r_t^{sp} < 0))(1 - \phi_{qj}) + \phi_{qj} q_{jt} + \eta_{qjt},$$

which can be expressed as a linear regression

$$Y = X \begin{pmatrix} \mu_{qj} \\ \gamma_{\mu j} \end{pmatrix} + \frac{\eta_{qjt}}{(1 - \phi_{qj})},$$

where

$$Y = \left(\frac{q_{j2} - \phi_{qj} q_{j1}}{(1 - \phi_{qj})}, ..., \frac{q_{jT} - \phi_{qj} q_{j(T-1)}}{(1 - \phi_{qj})} \right)',$$

and

$$X = \begin{pmatrix} 1 & \dots & 1 \\ I(r_1^{sp} < 0) & \dots & I(r_{T-1}^{sp} < 0). \end{pmatrix}'$$

The joint prior distribution of μ_{qj} and $\gamma_{\mu j}$ is

$$\begin{pmatrix} \mu_{qj} \\ \gamma_{\mu j} \end{pmatrix} \sim N(m_{\mu q}, \Sigma_{\mu q}),$$

where

$$m_{\mu q} = \begin{pmatrix} m_{\mu 0} \\ m_{\gamma 0} \end{pmatrix}, \Sigma_{\mu q} = \begin{pmatrix} s_{\mu 0}^2 & 0 \\ 0 & s_{\gamma 0}^2 \end{pmatrix}.$$

μ_{qj} and $\gamma_{\mu j}$ are directly drawn from the conditional posterior distribution

$$\begin{pmatrix} \mu_{qj} \\ \gamma_{\mu j} \end{pmatrix} | q, r^{sp}, \theta_{/(\gamma_\mu, \mu_q)} \sim N(\tilde{m}_{\mu q}, \tilde{\Sigma}_{\mu q}), \tag{11}$$

where

$$\tilde{\Sigma}_{\mu q}^{-1} = \frac{(1 - \phi_{qj})^2 (X'X)}{\sigma_{qj}^2} + \Sigma_{\mu q}^{-1},$$

and

$$\tilde{m}_{\mu q} = \tilde{\Sigma}_{\mu q} \left(\frac{\sigma_{qj}^2}{(1 - \phi_{qj})^2} X'Y + \Sigma_{\mu q}^{-1} m_{\mu q} \right).$$

Table 1: Simulation results for MSVLA-GFT

			μ_{h1}	ϕ_{h1}	σ^2_{h1}	ρ_1	μ_{q1}	ϕ_{q1}	σ^2_{q1}	γ_{μ_1}
T	N	True Value	0.300	0.900	0.050	-0.500	0.700	0.900	0.050	-0.200
500	50	Mean	0.296	0.880	0.054	-0.265	0.739	0.848	0.066	-0.276
		Std	0.150	0.062	0.031	0.111	0.550	0.068	0.032	1.106
		IF	7.511	82.147	138.723	30.730	23.033	83.494	116.718	25.524
	100	Mean	0.301	0.883	0.051	-0.319	0.742	0.845	0.067	-0.281
		Std	0.149	0.059	0.029	0.121	0.579	0.070	0.033	1.155
		IF	8.089	75.534	137.035	31.287	21.742	81.319	113.849	24.367
	200	Mean	0.310	0.883	0.047	-0.363	0.742	0.846	0.067	-0.281
		Std	0.151	0.063	0.028	0.134	0.673	0.070	0.033	1.303
		IF	9.037	72.478	132.307	33.293	21.176	82.786	115.358	23.623
1000	50	Mean	0.289	0.873	0.062	-0.268	0.730	0.882	0.058	-0.246
		Std	0.093	0.048	0.026	0.076	0.235	0.034	0.020	0.463
		IF	8.624	113.888	167.414	22.627	20.230	75.074	114.397	22.951
	100	Mean	0.295	0.871	0.058	-0.318	0.729	0.882	0.058	-0.246
		Std	0.095	0.052	0.025	0.085	0.237	0.034	0.019	0.467
		IF	10.534	97.360	145.374	29.718	18.837	73.355	110.717	21.484
	200	Mean	0.301	0.872	0.053	-0.363	0.728	0.882	0.057	-0.242
		Std	0.096	0.052	0.024	0.094	0.235	0.034	0.019	0.463
		IF	11.339	95.164	154.379	41.072	19.194	69.665	104.743	21.627
2000	50	Mean	0.286	0.867	0.071	-0.286	0.703	0.888	0.055	-0.201
		Std	0.061	0.035	0.021	0.052	0.162	0.023	0.013	0.315
		IF	5.445	106.104	152.163	19.675	17.726	74.089	112.350	20.258
	100	Mean	0.291	0.873	0.066	-0.337	0.704	0.887	0.055	-0.203
		Std	0.061	0.033	0.020	0.056	0.162	0.023	0.013	0.314
		IF	5.061	105.372	148.509	29.565	17.294	77.869	116.653	19.578
	200	Mean	0.295	0.875	0.062	-0.377	0.704	0.888	0.055	-0.202
		Std	0.061	0.033	0.019	0.061	0.162	0.023	0.013	0.316
		IF	6.187	113.495	168.291	40.770	16.932	71.260	107.229	19.232

Notes: T is the number of observations for each asset. N is the number of particles used in PGAS. Mean is the average value of posterior means across replications. Std is the standard error of the posterior means across replications. IF is the average inefficiency factor across replications.

Table 2: Descriptive statistics for the empirical applications

Panel 1: Exchange Rate Data			
	EUR/USD	GBP/USD	CHF/USD
Mean	0.002	0.005	-0.013
Variance	0.218	0.203	0.284
Skewness	0.014	0.844	-0.370
Kurtosis	4.447	7.993	8.084
Min	-2.841	-1.654	-4.188
Max	1.904	3.649	3.340
JB	0.001	0.001	0.001
	1.000		
Sample Correlation	0.655	1.000	
	0.809	0.521	1.000
Panel 2: Stock Return Data			
	DD	XOM	MMM
Mean	0.000	-0.001	0.000
Variance	0.000	0.000	0.000
Skewness	0.143	-0.153	-0.887
Kurtosis	12.542	13.747	16.047
Min	-0.142	-0.130	-0.139
Max	0.164	0.119	0.119
JB	0.001	0.001	0.001
	1.000		
Sample Correlation	0.588	1.000	
	0.576	0.553	1.000

Notes: The first panel reports the summary statistics for the weekly log-returns of three exchange rates, namely Euro, Pound sterling and Swiss frac, all against the U.S. dollar. The sample period is from January 13, 1993 to December 25, 2019. The second panel reports the summary statistics for the daily log-returns of three equities, namely DuPont (DD), Exxon Mobil (XOM) and 3M company (MMM). The sample period is from January 13, 1993 to December 25, 2019. JB denotes the p-value of the Jarque-Bera normality test.

Table 3: Estimation results related to volatilities in candidate models for the exchange rate data

		MSV-GFT/MSV-GFT-eq/MSV-DCC	MSV-chol
μ_{h1}	Mean	-1.715	-1.722
	Std	0.198	0.194
	95%CI	[-2.131,-1.325]	[-2.120,-1.348]
	IF	0.633	1.428
μ_{h2}	Mean	-1.807	-2.567
	Std	0.121	0.144
	95%CI	[-2.043,-1.562]	[-2.838,-2.268]
	IF	2.800	22.014
μ_{h3}	Mean	-1.520	-3.803
	Std	0.139	0.225
	95%CI	[-1.801,-1.249]	[-4.247,-3.355]
	IF	1.259	14.618
ϕ_{h1}	Mean	0.980	0.980
	Std	0.009	0.009
	95%CI	[0.959,0.993]	[0.960,0.993]
	IF	122.683	80.229
ϕ_{h2}	Mean	0.953	0.950
	Std	0.020	0.020
	95%CI	[0.905,0.982]	[0.904,0.980]
	IF	143.909	157.320
ϕ_{h3}	Mean	0.955	0.928
	Std	0.016	0.019
	95%CI	[0.918,0.980]	[0.886,0.959]
	IF	94.723	101.658
σ^2_{h1}	Mean	0.015	0.015
	Std	0.006	0.006
	95%CI	[0.007,0.032]	[0.007,0.028]
	IF	215.786	176.734
σ^2_{h2}	Mean	0.033	0.048
	Std	0.015	0.020
	95%CI	[0.013,0.073]	[0.021,0.097]
	IF	185.510	212.597
σ^2_{h3}	Mean	0.041	0.285
	Std	0.013	0.076
	95%CI	[0.022,0.073]	[0.172,0.473]
	IF	156.210	182.924

Notes: Mean is the posterior mean. SD is the numerical standard errors of the posterior means. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws. IF is the inefficiency factor. The volatility-related parameters for MSV-GFT, MSV-GFT-eq and MSV-DCC are the same as they are all estimated by first fitting separate univariate SV models.

Table 4: Estimation results related to correlation and the log marginal likelihood and DIC values in candidate models for the exchange rate data

		MSV-GFT	MSV-GFT-eq		MSV-DCC		MSV-chol
Mean	μ_{q1}	0.706	0.708	k	0.780	$\mu_{h^\star 1}$	0.619
Std		0.026	0.041		0.024		0.022
95%CI		[0.655,0.757]	[0.624,0.790]		[0.736,0.822]		[0.576,0.663]
IF		13.426	2.558		650.111		12.462
Mean	μ_{q2}	1.475	1.453	d	15.190	$\mu_{h^\star 2}$	0.973
Std		0.094	0.135		1.121		0.020
95%CI		[1.285,1.654]	[1.184,1.715]		[13.299,17.360]		[0.934,1.013]
IF		1.268	0.751		631.814		10.714
Mean	μ_{q3}	0.475	0.478	a_{11}	1.116	$\mu_{h^\star 3}$	-0.008
Std		0.083	0.072		0.030		0.021
95%CI		[0.310,0.639]	[0.334,0.618]		[1.066,1.180]		[-0.050,0.033]
IF		1.060	0.273		538.066		15.551
Mean	ϕ_{q1}	0.834	0.969	a_{21}	-0.132	$\phi_{h^\star 1}$	0.270
Std		0.098	0.006		0.019		0.113
95%CI		[0.576,0.958]	[0.957,0.980]		[-0.171,-0.099]		[0.046,0.484]
IF		156.889	65.441		511.181		57.978
Mean	ϕ_{q2}	0.946		a_{22}	-0.319	$\phi_{h^\star 2}$	0.355
Std		0.014			0.042		0.100
95%CI		[0.916,0.970]			[-0.403,-0.244]		[0.166,0.548]
IF		41.715			605.758		61.688
Mean	ϕ_{q3}	0.970		a_{31}	1.027	$\phi_{h^\star 3}$	0.634
Std		0.012			0.012		0.106
95%CI		[0.941,0.989]			[1.004,1.052]		[0.401,0.817]
IF		130.851			190.699		107.911
Mean	σ^2_{q1}	0.004	0.002	a_{32}	-0.066	$\sigma^2_{h^\star 1}$	0.086
Std		0.002	0.001		0.011		0.015
95%CI		[0.001,0.010]	[0.001,0.003]		[-0.088,-0.046]		[0.058,0.117]
IF		343.509	163.838		232.215		51.744
Mean	σ^2_{q2}	0.030	0.023	a_{33}	1.103	$\sigma^2_{h^\star 2}$	0.062
Std		0.007	0.005		0.026		0.009
95%CI		[0.018,0.045]	[0.015,0.033]		[1.057,1.159]		[0.045,0.080]
IF		86.546	70.013		519.844		56.377
Mean	σ^2_{q3}	0.006	0.006			$\sigma^2_{h^\star 3}$	0.023
Std		0.003	0.002				0.006
95%CI		[0.003,0.013]	[0.003,0.010]				[0.012,0.035]
IF		209.150	127.976				94.640
DIC		2252.6	2378.1		2385.5		2258.3
log ML		-1136.2	-1192.9		-1205.5		-1126.3

Notes: Mean is the posterior mean. SD is the numerical standard error of the posterior mean. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws. IF is the inefficiency factor.

Table 5: Estimation results related to volatility in candidate models for the stock return data

		MSV models	MSVL/MSVLA models
μ_{h1}	Mean	0.756	0.749
	Std	0.119	0.112
	95%CI	[0.539,0.984]	[0.536,0.975]
	IF	9.992	10.383
μ_{h2}	Mean	0.367	0.372
	Std	0.270	0.273
	95%CI	[-0.147,0.904]	[-0.173,0.922]
	IF	9.173	10.142
μ_{h3}	Mean	-0.019	-0.014
	Std	0.119	0.119
	95%CI	[-0.232,0.204]	[-0.253,0.221]
	IF	11.421	10.429
ϕ_{h1}	Mean	0.930	0.933
	Std	0.017	0.014
	95%CI	[0.895,0.960]	[0.905,0.961]
	IF	154.326	85.370
ϕ_{h2}	Mean	0.984	0.985
	Std	0.005	0.005
	95%CI	[0.972,0.993]	[0.975,0.993]
	IF	56.142	61.866
ϕ_{h3}	Mean	0.909	0.923
	Std	0.022	0.016
	95%CI	[0.862,0.948]	[0.890,0.952]
	IF	146.975	97.037
σ^2_{h1}	Mean	0.117	0.111
	Std	0.030	0.022
	95%CI	[0.067,0.177]	[0.068,0.152]
	IF	201.737	137.497
σ^2_{h2}	Mean	0.033	0.030
	Std	0.009	0.006
	95%CI	[0.021,0.052]	[0.019,0.043]
	IF	106.621	153.344
σ^2_{h3}	Mean	0.189	0.159
	Std	0.049	0.032
	95%CI	[0.105,0.305]	[0.100,0.223]
	IF	181.958	135.422

Notes: Mean is the posterior mean. SD is the numerical standard errors of the posterior means. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws. IF is the inefficiency factor.

Table 6: Estimation results of leverage effect and asymmetric effect in candidate models for the stock return data

		MSVL models	MSVLA-GFT model
ρ_1	Mean	-0.293	-0.293
	Std	0.046	0.046
	95%CI	[-0.381,-0.202]	[-0.381,-0.202]
	IF	58.136	58.136
ρ_2	Mean	-0.211	-0.211
	Std	0.050	0.050
	95%CI	[-0.299,-0.108]	[-0.299,-0.108]
	IF	54.575	54.575
ρ_3	Mean	-0.171	-0.171
	Std	0.047	0.047
	95%CI	[-0.265,-0.075]	[-0.265,-0.075]
	IF	42.360	42.360
γ_{μ_1}	Mean		0.054
	Std		0.104
	95%CI		[-0.145,0.262]
	IF		42.599
γ_{μ_2}	Mean		0.489
	Std		0.268
	95%CI		[0.094,1.163]
	IF		209.135
γ_{μ_3}	Mean		0.401
	Std		0.294
	95%CI		[-0.036,1.111]
	IF		137.946

Notes: Mean is the posterior mean. SD is the numerical standard errors of the posterior means. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws. IF is the inefficiency factor.

Table 7: Estimation results related to correlation and the log marginal likelihood and DIC values in candidate models for the stock return data

		MSV-GFT	MSVL-GFT	MSVL-GFT-eq	MSVLA-GFT		MSVL-DCC
μ_{q1}	Mean	0.539	0.536	0.496	0.528	k	0.359
	Std	0.032	0.030	0.057	0.057		0.040
	95%CI	[0.474,0.601]	[0.478,0.596]	[0.383,0.610]	[0.414,0.637]		[0.286,0.432]
	IF	25.282	19.504	10.331	36.151		653.576
μ_{q2}	Mean	0.593	0.582	0.565	0.377	d	8.585
	Std	0.040	0.040	0.068	0.134		0.531
	95%CI	[0.512,0.667]	[0.503,0.659]	[0.429,0.703]	[0.058,0.576]		[7.528,9.605]
	IF	29.604	12.311	9.687	221.668		477.359
μ_{q3}	Mean	0.468	0.464	0.444	0.280	a_{11}	1.245
	Std	0.035	0.033	0.058	0.145		0.039
	95%CI	[0.396,0.535]	[0.399,0.529]	[0.326,0.559]	[-0.060,0.503]		[1.181,1.329]
	IF	20.985	13.329	10.415	142.104		498.432
ϕ_{q1}	Mean	0.786	0.772	0.976	0.556	a_{21}	-0.312
	Std	0.070	0.064	0.005	0.107		0.031
	95%CI	[0.622,0.913]	[0.614,0.862]	[0.964,0.984]	[0.329,0.769]		[-0.380,-0.261]
	IF	236.836	175.359	187.858	136.179		340.597
ϕ_{q2}	Mean	0.887	0.911		0.809	a_{22}	-0.367
	Std	0.062	0.030		0.104		0.036
	95%CI	[0.706,0.956]	[0.834,0.955]		[0.571,0.944]		[-0.436,-0.298]
	IF	397.421	190.867		443.146		426.523
ϕ_{q3}	Mean	0.875	0.877		0.841	a_{31}	1.188
	Std	0.068	0.036		0.104		0.032
	95%CI	[0.712,0.955]	[0.796,0.934]		[0.520,0.933]		[1.133,1.254]
	IF	478.995	139.680		393.311		474.705
σ^2_{q1}	Mean	0.051	0.054	0.004	0.116	a_{32}	-0.256
	Std	0.021	0.018	0.001	0.029		0.027
	95%CI	[0.016,0.102]	[0.031,0.100]	[0.002,0.007]	[0.055,0.174]		[-0.308,-0.204]
	IF	254.208	205.683	264.051	166.473		371.845
σ^2_{q2}	Mean	0.026	0.019	0.005	0.048	a_{33}	1.219
	Std	0.019	0.007	0.002	0.031		0.036
	95%CI	[0.009,0.083]	[0.009,0.038]	[0.003,0.011]	[0.010,0.122]		[1.153,1.291]
	IF	421.839	232.903	263.898	477.251		494.367
σ^2_{q3}	Mean	0.020	0.020	0.004	0.026		
	Std	0.015	0.007	0.001	0.021		
	95%CI	[0.006,0.059]	[0.010,0.034]	[0.002,0.007]	[0.008,0.094]		
	IF	498.020	196.282	184.989	405.992		
DIC		21290	21256	21298	21242		22261
log ML		-10672	-10647	-10675	-10646		-11108

Notes: Mean is the posterior mean. SD is the numerical standard error of the posterior mean. 95% CI is constructed using the 2.5th and 97.5th percentiles of the MCMC draws. IF is the inefficiency factor.

Table 8: Model comparison based on out-of-sample portfolio construction

	MSV models			
	MSV-CC	MSV-GFT	MSV-GFT-eq	MSV-DCC
Squared Returns	0.897	0.888	0.896	0.902
	MSVL models			
	MSVL-CC	MSVL-GFT	MSVL-GFT-eq	MSVL-DCC
Squared Returns	0.867	**0.852**	0.854	0.869
	MSVLA models			
	MSVLA-GFT			
Squared Returns	0.854			

Notes: Boldface indicates the smallest value. The performance of average squared returns is relative to the equal-weight portfolio.

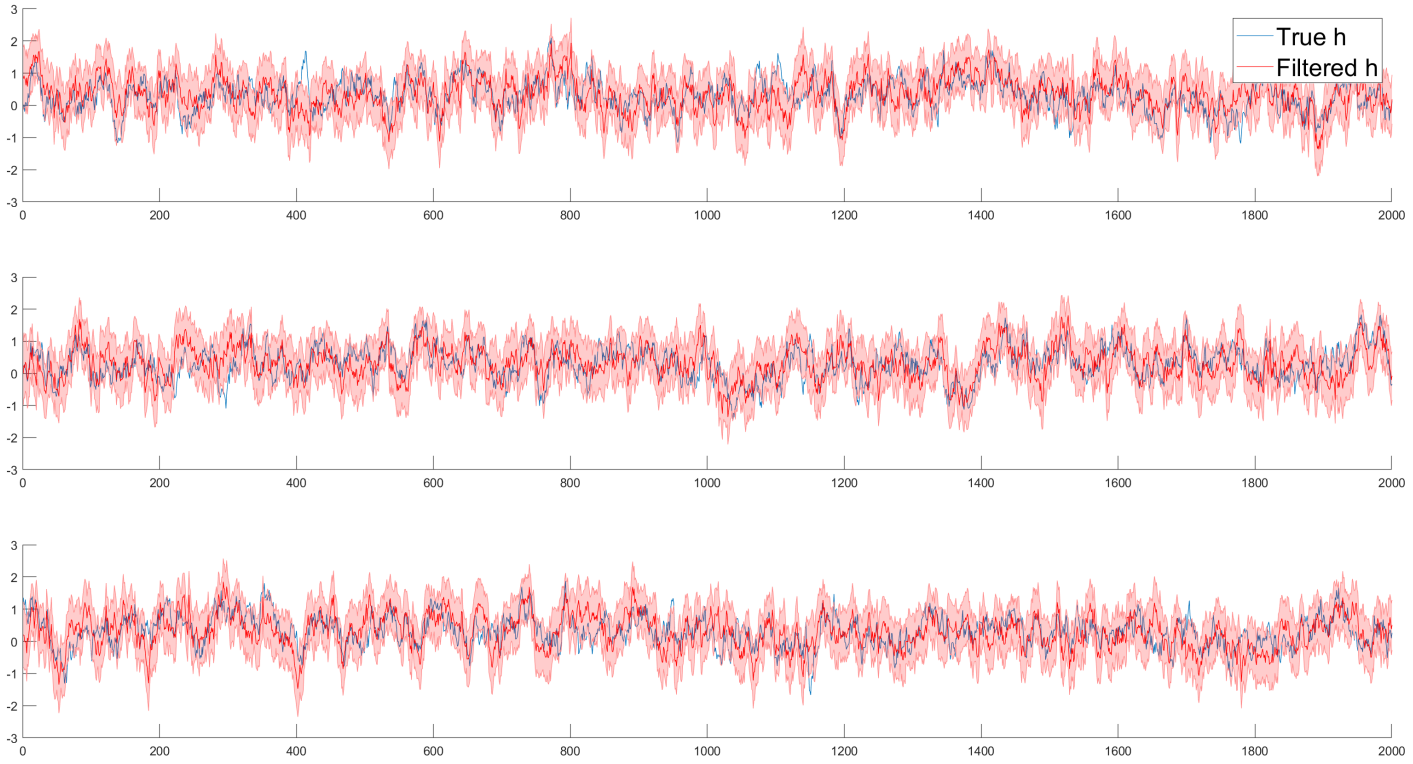

Figure 1: True and filtered log-variances. This figure plots the true h (blue solid), the filtered h (black solid) and the 95% credible intervals of the filtered h (red shaded area).

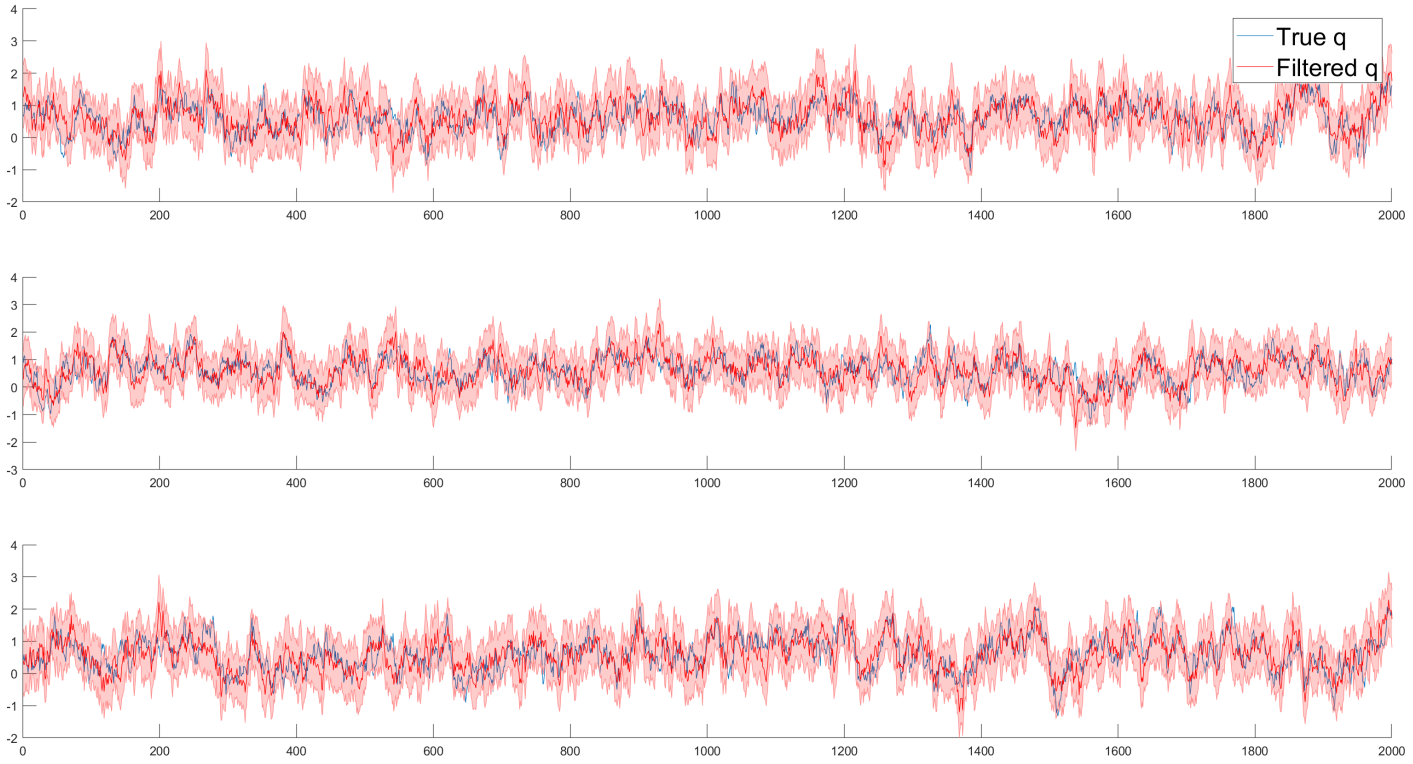

Figure 2: True and filtered transformed-correlation-cofficients. This figure plots the true q (blue solid), the filtered q (black solid) and the 95% credible intervals of the filtered q (red shaded area).

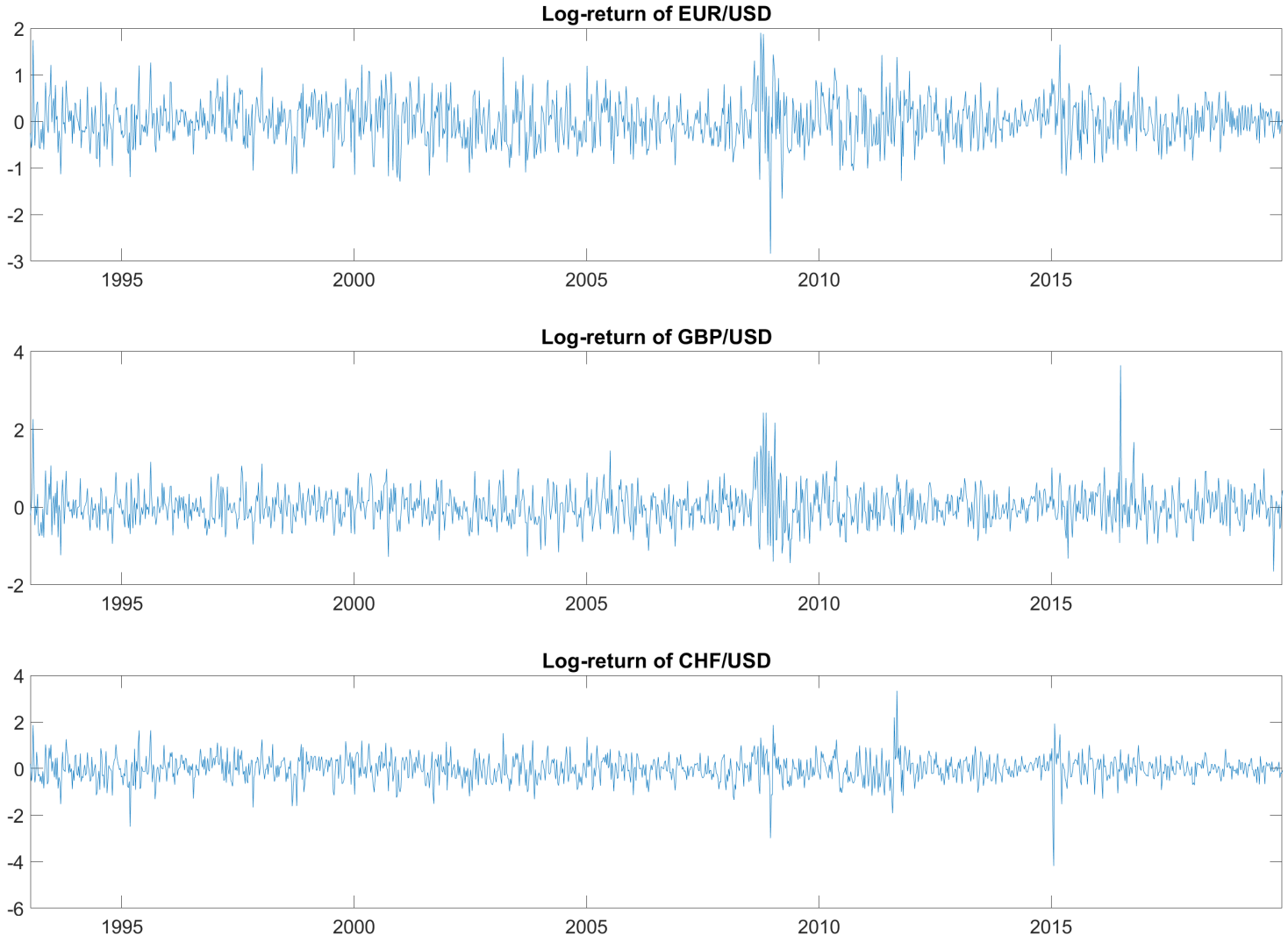

Figure 3: Time series plots of exchange rates data. This figure plots the time series dynamics of the weekly log-returns of three exchange rates, namely Euro, Pound sterling and Swiss frac, all against the US dollar. The sample period is from January 13, 1993 to December 25, 2019.

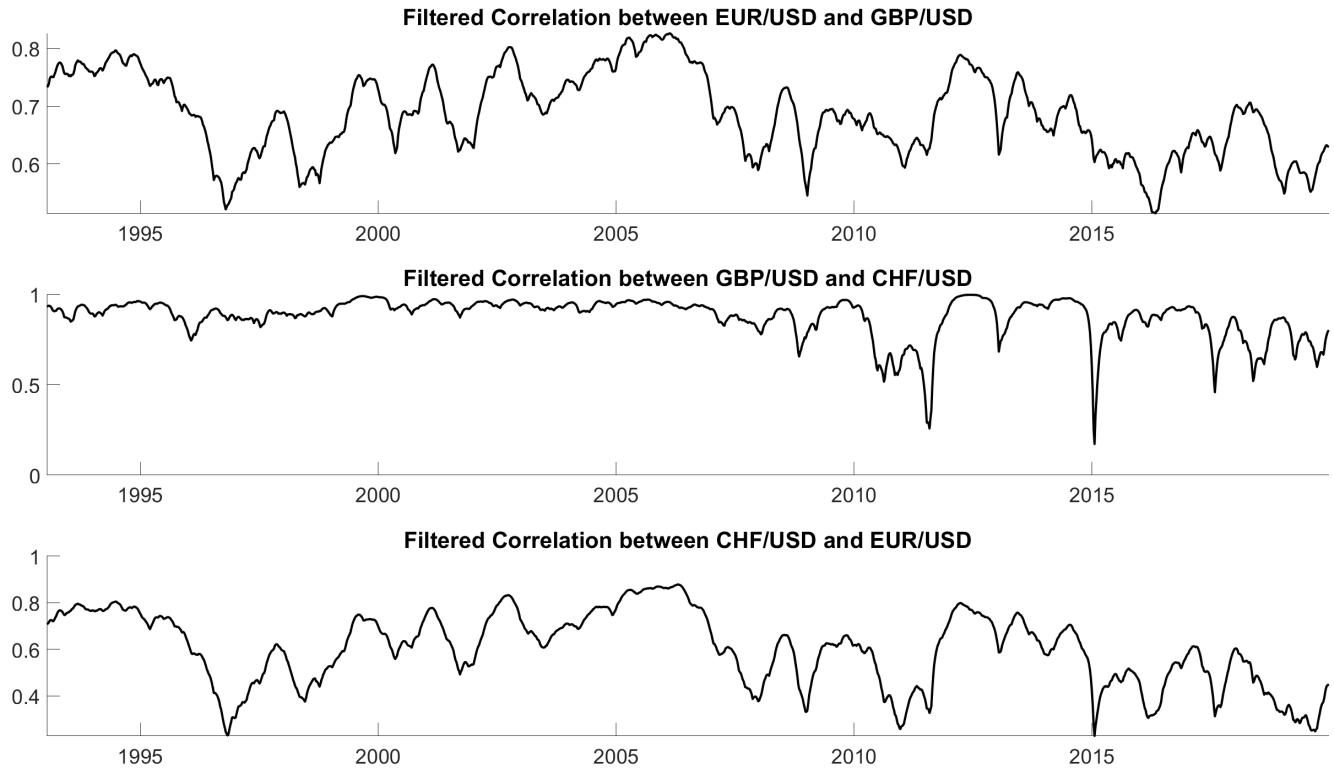

Figure 4: Filtered pairwise correlations in exchange rate application. This figure plots the posterior mean of correlation sequences filtered based on MSV-GFT model, using the weekly log-returns of three exchange rates, namely Euro, Pound sterling and Swiss frac, all against the US dollar.

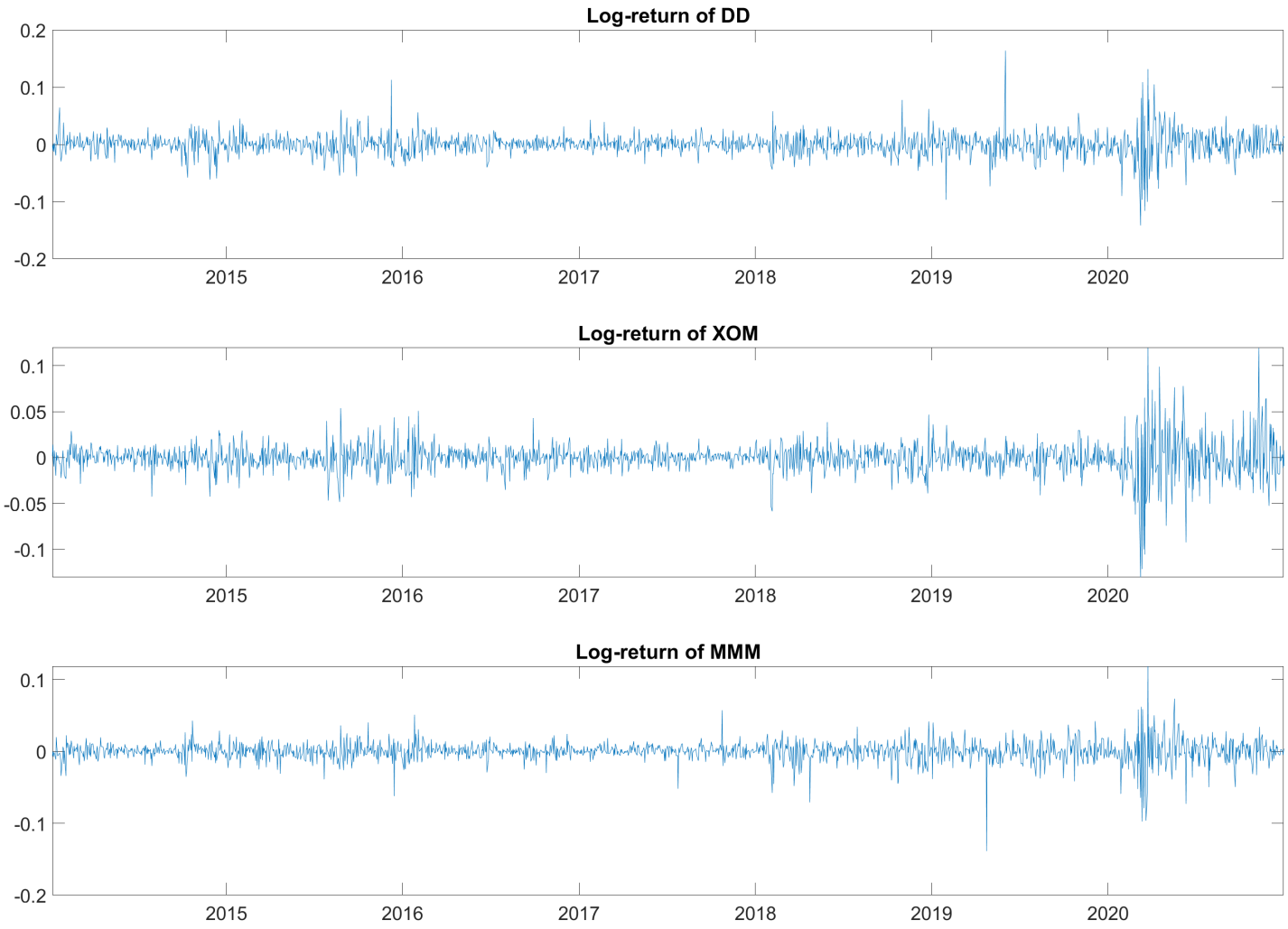

Figure 5: Time series plots of stock returns data. This figure plots the time series dynamics of the daily log-returns of three equities, namely DuPont (DD), Exxon Mobil (XOM) and 3M company (MMM). The sample period is from January 4, 2014 to December 31, 2020.

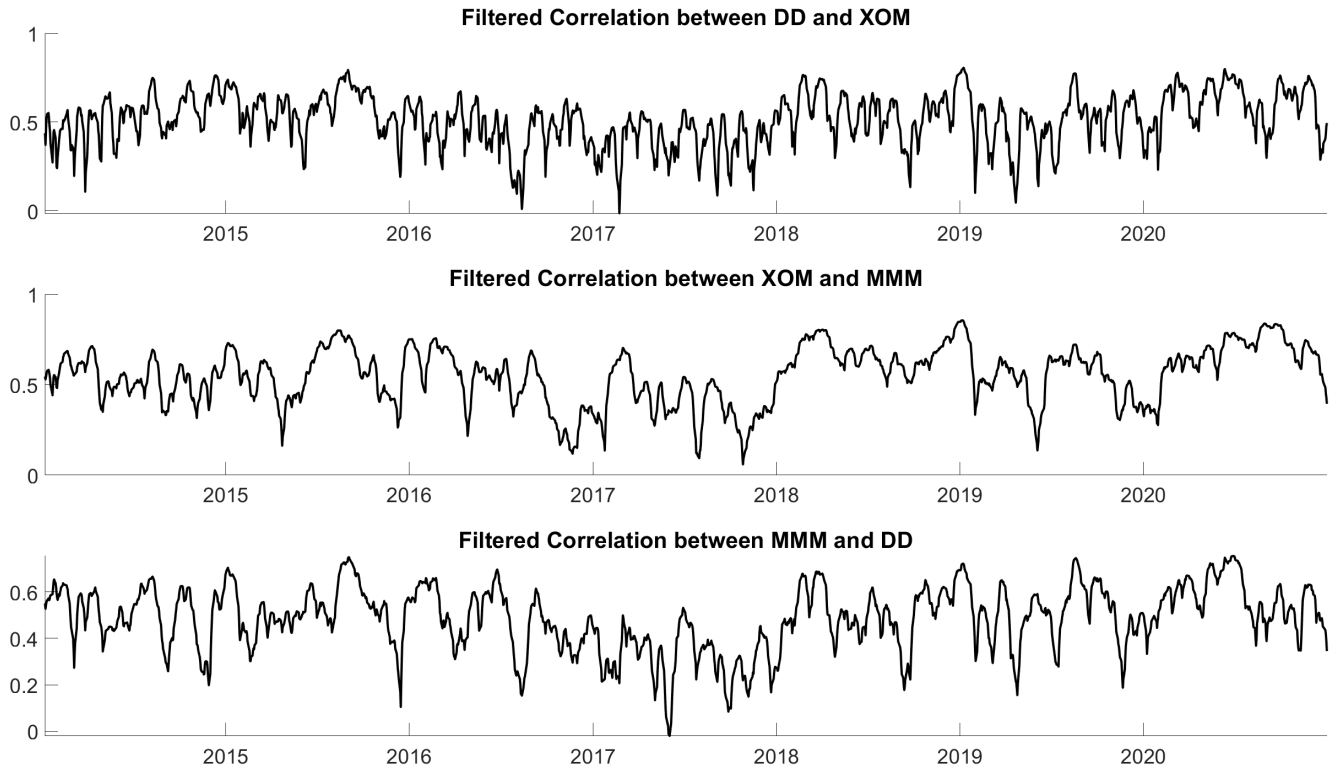

Figure 6: Filtered pairwise correlations in stock return application. This figure plots the posterior mean of correlation sequences filtered based on MSVLA-GFT model, using the daily log-returns of three stocks equities, namely DuPont (DD), Exxon Mobil (XOM) and 3M company (MMM).