Multivariate Stochastic Volatility Models based on Generalized Fisher Transformation*

Han Chen
Singapore Management University

Yijie Fei
Singapore Management University

Jun Yu
Singapore Management University

November 8, 2020

Abstract

Modeling multivariate stochastic volatility (MSV) faces a few challenges when both variances and covariances are time-varying. The main challenges come from the assurance of positive-definiteness of the variance-covariance matrix and the statistical analysis of MSV models. In this paper, we introduce a new MSV model which is based on the generalized Fisher transformation of Archakov and Hansen (2018). This specification is highly flexible where the variance-covariance matrix is always positive-definiteness. Moreover, driving factors of volatilities and correlations are separated in the model. A Particle Gibbs Ancestor Sampling (PGAS) method is proposed to conduct Bayesian analysis of the model. The method facilitates the Bayesian model comparison. The MSV model and the estimation method are extended to include realized measures. Simulation studies show that the proposed method works well for the MSV model. Empirical studies based on exchange-rate and equity returns suggest that the proposed MSV model provides superior in-sample and out-of-sample performance over alternative specifications.

*We would like to thank Peter Hansen for useful discussions and comments. Han Chen, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903. Email: han.chen.2015@phdecons.smu.edu.sg. Yijie Fei, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903. Email: yijie.fei.2015@phdecons.smu.edu.sg. 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: Stochastic volatility; Dynamic correlation; Multivariate asset returns; Particle Filter; Markov Chain Monte Carlo; Realized Measures

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

1 Introduction

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

The first MSV model, proposed in Harvey et al. (1994), is an analogy of the constant conditional correlation (CCC) model in the MGARCH literature. In this basic setup, the volatility of each individual asset is assumed to follow a univariate SV process, while the correlation matrix among all assets is constant over time. This is a rather restrictive assumption. Great efforts have been dedicated 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. Another parametrization that is also based on DCC can be found in Asai and McAleer (2009).

In this paper, we introduce a new MSV model based on a recently developed new parameterization of the correlation matrix. Such a parameterization, originally proposed in Archakov and Hansen (2018), can be deemed as a generalization of the well-known Fisher z-transformation from the bivariate case to the multivariate case. It has been successfully employed to introduce a new MGARCH model by Archakov et al. (2020). In our new MSV model, the underlying latent variables that determine the correlations among assets are allowed to have an unrestricted domain because the correlation matrix, by construction, is always valid. Meanwhile, the underlying 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. Last but not least, our model is invariant to the reordering of assets, and thus no ex-ante ordering is necessary. All these features indicate that our model is very flexible, imposing a minimum level of ex-ante restrictions.

To estimate the proposed MSV model, following the MSV literature, we adopt the Bayesian Markov chain Monte Carlo (MCMC) method. Departing from much of the Bayesian MSV literature, where carefully designed single-move samplers or multi-move Gibbs samplers, all based on the data-augmented likelihood function, have been used for model estimation, we opt to work with the recently proposed particle-filter-based MCMC (PMCMC) algorithm. Ever since the seminal paper by Andrieu et al. (2010), the research on the theoretical foundation of PMCMC and its applications in many different fields have mushroomed. Though theoretically applicable under a very general setup, the practical performance of this approach for a particular model depends on many factors and requires careful examination. To strike a balance between the satisfactory estimation accuracy and the acceptable computational cost, we choose a method called Particle Gibbs Ancestor Sampling (PGAS). This is a modified version of the Particle Gibbs (PG) Sampler considered in Andrieu et al. (2010), which dramatically improves the mixing property under a small number of particles. Extensive simulation results are presented to justify the choice of the estimation strategy and also provide useful guidance for empirical applications.

While traditional MSV models are estimated solely based on the daily return data, they do not fully utilize the available information. Another source of information for return fluctuation is realized volatility (RV) calculated using intra-daily high-frequency data; see Andersen et al. (2010) for a recent survey. Inspired by the realized-GARCH models of Hansen et al. (2014) and Noureldin et al. (2012), SV models based on both return series and RV, referred to as RSV, have appeared in the literature. In this paper, we also incorporate realized measures into our MSV model. This is done by applying the new transformation to the realized covariance matrix, which provides additional measurements to the latent variables. As argued in Yamauchi and Omori (2019), this extra information can help stabilize the parameter estimation. However, this added information incurs an identification problem. In particular, the mean of the latent log-variance suffers from a large bias if the realized measure is biased. To achieve satisfactory inference, we propose a two-stage estimation method. In the first stage we estimate a model without realized measures. In the second stage, we estimate other parameters by imposing identification conditions. We show by simulation that the two-stage method greatly reduces the estimation bias.

Unfortunately, estimating our MSV model with a moderate number of assets is not feasible. This is not surprising as our model is very flexible, allowing time-varying volatilities,

correlations, covariances, and in the meantime, always ensuring the positivity of variance-covariance matrix. However, estimation of our low-dimensional MSV model can help us to understand important features in the data and hence provide guidance to choose more restrictive MSV models. For example, after we estimate our three-dimensional MSV model, we find that it is critical to allow the pairwise correlation coefficient sequence to have different levels of persistence. A reasonable restrictive MSV model must 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 method based on the combination of particle filter and MCMC. We consider the incorporation of realized measures into our model in Section 4. Simulation results are provided in Section 5 and two empirical applications are discussed in Section 6. Section 7 concludes. Additional technical details can be found in the Appendix.

Throughout the paper, we let $diag(A)$ denote the vector of diagonal elements of a square matrix A (or the diagonal matrix whose diagonal elements are elements in A if A is a vector); $vech(A)$ 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 vector containing all lower off-diagonal elements of A (excluding the diagonal elements).

2 New Multivariate Stochastic Volatility Model

In this section, we first review existing MSV models with a focus on the parametrization of the covariance matrix. The generalized Fisher transformation proposed in Archakov and Hansen (2018) is then introduced, based on which a new MSV model is proposed.

2.1 Review of existing MSV models

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

[1]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.

The basic structure of the MSV model is

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

where r_t is a vector of asset returns. We aim at characterizing the dynamics of its variance-covariance matrix C_t. Clearly, C_t must be symmetric and positive-definite for all t. Different models rely on different techniques to ensure this positivity. 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, three methods have been considered. The first method is based on the matrix exponential. For example, Ishihara et al. (2016) assume that

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

and propose to model $vech(H_t)$ as a 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, Shirota et al. (2017) also use this decomposition to set up their MSV model. 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. 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 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 useing 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. A similar model specification is presented in Jin and Maheu (2013). Clearly, the dynamics of the volatilities and the correlations are not separated.

Models in the second group treat the volatilities and the correlation matrix separately. 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. The diagonal elements of R_t are ones. 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 positivity, 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

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

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 inference simple, it is too restrictive for modeling most financial data.

To allow for time-varying correlations, Asai and McAleer (2009) consider two models, both motivated by the dynamic conditional correlation (DCC) model of Engle (2002). 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 (2009) argue that the second one is preferred.

Following Yu and Meyer (2006), Yamauchi and Omori (2019) propose to model the pairwise correlations by the Fisher z-transformation. That is, let $R_t = \{\rho_{ij,t}\}$ and

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

they assume that $g_{ij,t}$ follows a random walk for any $i \neq j$. Although $|\rho_{ij,t}| < 1$ by construction, this element-wise operation does not guarantee the positivity of R_t. They further derive algebraic bounds for $\rho_{ij,t}$ that ensure the positivity of R_t. Note that 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 method.

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 Kurose and Omori (2016), R_t is positive-definite only if ρ_t is large than some lower bound, depending on the number of assets. This lower bound approaches zero as the number of assets goes to infinity.

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 complexity brought by this constraint in modeling, one can instead model Fisher's z-transformation of ρ, defined by $F(\rho)$ in (1a). 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 bivariate SV literature in Yu and Meyer (2006). 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 pair-wise transformation applied to each entry in a high-dimensional correlation matrix, though seems to be natural, is not a valid choice as it fails to 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 contribution made in Archakov and Hansen (2018). 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 (2018) 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 $q > 2$, R can easily be numerically obtained from q using an iterative algorithm as shown in Archakov and Hansen (2018).

When $p = 2$, Archakov and Hansen (2018) show that the above-defined transformation reduces to Fisher's z-transformation. The new transformation retains the advantages of Fisher's 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. Hence, it is reasonable to model elements of q via a Gaussian process. Third, this transformation is invariant to order 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 the inverse mapping 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 $q_t = (q_{1t}, ..., q_{dt})'$ denote the vector of latent variables at time t that underlie all the correlation coefficients, where $d = \frac{p(p-1)}{2}$. In particular, q_t is connected to the correlation

matrix R_t through the transformation detailed in Section 2.2. Our basic MSV model, which we refer to as MSV-GFT, is given by

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

$$V_t = diag\left(\exp(h_{1t}), ..., \exp(h_{pt})\right), \tag{3}$$

$$q_t = F(R_t), \tag{4}$$

$$h_{t+1} = \mu_h + \Phi_h(h_t - \mu_h) + \eta_{ht}, \text{ where } \eta_{ht} \sim N(0, \Sigma_h), \tag{5}$$

$$q_{t+1} = \mu_q + \Phi_q(q_t - \mu_q) + \eta_{qt}, \text{ where } \eta_{qt} \sim N(0, \Sigma_q), \tag{6}$$

$$h_0 \sim N\left(\mu_h, (I_p - \Phi_h^2)^{-1}\Sigma_h\right), \text{ and } q_0 \sim N\left(\mu_q, (I_d - \Phi_q^2)^{-1}\Sigma_q\right), \tag{7}$$

where $t = 1, ..., T$, $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{pt})'$, $\eta_{ht} = (\eta_{h1t}, ..., \eta_{hpt})'$, $\eta_{qt} = (\eta_{q1t}, ..., \eta_{qdt})'$, $\mu_h = (\mu_{h1}, ..., \mu_{hp})'$ and $\mu_q = (\mu_{q1}, ..., \mu_{qd})'$. 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_{ht}).

To reduce the number of parameters, we assume that both Φ_h and Φ_q are diagonal with diagonal elements $(\phi_{h1}, ..., \phi_{hp})' := \phi_h$ and $(\phi_{q1}, ..., \phi_{qd})' := \phi_q$, respectively. We further assume that Σ_h and Σ_q are diagonal with diagonal elements $(\sigma_{h1}^2, ..., \sigma_{hp}^2) := \sigma_h^2$ and $(\sigma_{q1}^2, ..., \sigma_{qd}^2)' := \sigma_q^2$. Here h_t is a p-dimensional latent variable that determines the volatilities via the expoential 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. 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 also consider a restricted model where we assume $\phi_{q1} = \cdots = \phi_{qd}$ and $\sigma_{q1}^2 = \cdots = \sigma_{qd}^2$.

(**Jun: would be more reasonable if we only assume $\phi_{q1} = \cdots = \phi_{qd}$ in the restricted model?**)

To fix some notations, let $r = (r_1', ..., r_T')'$, $h = (h_1', ..., h_T')'$, $q = (q_1', ..., q_T')'$, $x = (h', q')' := (x_1', ..., x_T')'$ 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. Sometimes we use $x_{1:T}$ to represent $(x_1', ..., x_T')'$. Let θ be the

set of parameters in the model and $p(r|\theta)$ be the likelihood function of the model.

3 Inference of MSV-GFT Model

Due to the difficulty of evaluating the likelihood function of MSV models, the literature on MSV models often relies on Bayesian methods to carry out statistical inference. In this paper, we take advantage of a recently proposed technique known as PMCMC, which builds an efficient, high-dimensional MCMC kernel. Specifically, we use an improved PG sampler that enjoys the good mixing property even with a small number of particles.

3.1 Review of estimation methods for MSV model

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 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, almost all MSV studies rely on the Bayesian MCMC method, combined with the data augmentation technique of Tanner and Wong (1987). The idea is to conduct the Bayesian posterior analysis based on $p(r|\theta, x)$ which is more tractable than $p(r|\theta)$. Hence, x is treated in the same way as θ and the posterior is given by $p(\theta, x|r)$. This parameter expansion technique is known as data augmentation.

There are a number of MCMC algorithms in the literature to estimate MSV models. Yu and Meyer (2006) use the single-move algorithm to estimate several bivariate MSV models. The algorithm is single-move in the sense that each latent variable is drawn one at a time, given all the other ones. This approach 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. Yamauchi and Omori (2019) use the single-move algorithm to estimate the pairwise-Fisher-transformation-based MSV model.

In light of the inefficiency in mixing for the single-move sampler, more recent papers, such as Ishihara and Omori (2012), Ishihara et al. (2016), Kurose and Omori (2016), have resorted multi-move algorithms to estimate different MSV models.[2] In contrast to the single-move sampler, multi-move algorithms samples a block of latent vectors simultaneously. 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 (2009) 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 all the above-reviewed methods are effective in drawing random samples from $p(\theta, x|r)$, additional efforts are needed to compute the marginal likelihood of the model, $p(r|\theta)$. While several MCMC-based methods, such as Chib (1995) and Chib and Jeliazkov (2001), are available to compute $p(r|\theta)$, they are numerically difficult to implement for MSV models.

3.2 Gibbs sampler based on particle filter

In this paper, instead of using above-mentioned conventional techniques, we apply a more efficient sampler, known as particle Gibbs (PG)[3]. This method is recently proposed in Andrieu et al. (2010) and it combines particle filter with Gibbs sampler. The intuition is to construct a high-dimensional efficient Markov kernel for latent processes using particle filter.

3.2.1 Introduction to Particle Gibbs approach

Consider a general state-space model given by

$$(r_t|x_t = x, \theta) \sim f(\cdot|x, \theta), \tag{8}$$

$$(x_{t+1}|x_t = x, \theta) \sim g(\cdot|x, \theta), \text{ and } x_1 \sim \mu_\theta(\cdot), \tag{9}$$

[2]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.

[3]Another particle-filter-based MCMC method potentially applicable is particle Metropolis-Hasting. It is not chosen, however, as it requires an accurate estimation of likelihood and hence a very large number of particles.

where r_t is the observable variable, x_t is the latent variable, and θ contains all parameters. To sample from the joint density $p(\theta, x_{1:T}|y_{1:T})$, typically we proceed by running a Gibbs sampler, which draws alternately from the two conditional densities, namely $p(\theta|x_{1:T}, r_{1:T})$ and $p(x_{1:T}|r_{1:T}, \theta)$. Usually, the former can be drawn either by imposing conjugate priors or through Metropolis-Hasting algorithm. The latter, on the other hand, can be handled by particle filter[4], 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 methodology of particle filter combines importance sampling and Monte Carlo simulations to approximate the target distribution. The key idea is to represent the distribution by a set of random samples with the corresponding weights and calculate the quantity of interest based on these samples and weights. To fix the idea, let $\{x_{1:t}^{(i)}, w_t^{(i)}\}_{i=1}^N$ be a random measure, where $\{x_{0:t}^{(i)}, i = 1, ..., N\}$ is a set of support points and $\{w_t^{(i)}, i = 1, ..., N\}$ are the associated weights. Here, we use $x_{1:t} = \{x_j, j = 1, ..., t\}$ to denote the set of all states up to time t. Each point is called a particle, and N is the number of particles used. The approximate distribution can then be written as

$$\hat{p}_\theta(dx_{1:t}|r_{1:t}) = \sum_{i=1}^{N} w_t^{(i)} \delta_{x_{1:t}^{(i)}}(dx_{1:t}), \tag{10}$$

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

To obtain the weights, we resort to importance sampling. That is, we sample 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 fol-

[4]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. The intuition is that this 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.

[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).

lows 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 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 joint likelihood[6] can then be easily obtained as $\hat{p}_\theta(r_{1:T}) = \prod_{t=2}^T \hat{p}_\theta(r_{1:t}|r_{1:t-1}) \cdot \hat{p}_\theta(r_1)$.

3.2.2 Advantages of PG sampler over traditional methods

A particle-filter-based MCMC procedure has a few desirable properties, which make it preferred compared with traditional methods.

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

Second, one can gain a lot from adopting sequential Monte Carlo when performing model comparison. As is discussed in Appendix C, 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). For both single-move and multi-move sampler, these quantities are hard to obtain. The reason is that for those methods, the latent variables are treated as extra parameters, c.f. Tanner and Wong (1987). As a result, the likelihood function at period t is defined as $p(r_t|x_t, \theta)$, whereas, in PG sampler, it

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

is defined as $p(r_t|\theta)$. The computation of both the Bayes factor and DIC requires $p(r_t|\theta)$, not $p(r_t|x_t, \theta)$. While the single-move sampler and the multi-move sampler can generate MCMC draws of $p(r_t|x_t, \theta)$ as a byproduct, x_t has to be integrated out from $p(r_t|x_t, \theta)$ to obtain MCMC draws of $p(r_t|\theta)$. This step is computationally expensive. On the other hand, the PG sampler produces an estimated $p(r_t|\theta)$ as a by-product, facilitating the calculation of Bayes factor and DIC.

3.2.3 Particle Gibbs with ancestor sampling

Though theoretically correct, PG has been shown to perform quite poorly when the underlying SMC sampler suffers from path degeneracy. As observed in Lindsten et al. (2014) and Chopin and Singh (2015), the mixing of the Markov kernel induced by PG is rather slow under those circumstances. What makes things worse is that for the high-dimensional problem, such as the one we consider in this paper, path degeneracy is inevitable. To overcome this severe drawback, Lindsten et al. (2014) propose a new method, which includes an additional step called ancestor sampling. While this is a small modification, the new PG with ancestor sampling (PGAS) 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. Lindsten et al. (2014) also show that for a state-space model, PGAS is probabilistically equivalent to the particle Gibbs sampler with a backward smoothing step under certain conditions.

For our purpose, a fast mixing under a small number of particles is highly desirable, as our likelihood function contains a component that has no closed-form solution and thus must be computed numerically. Although the cost for one-time computation is relatively low, it soon becomes infeasible when a vast number of particles are included in the system. Indeed, for MCMC with S iterations, if the sample size is T and N particles are used, $F^{-1}(\cdot)$ must be evaluated $S \times T \times N$ times. As S and T are usually quite large in the empirical application, we can gain a lot in terms of computational efficiency by choosing the PGAS approach. In summary, we believe that PGAS is a suitable estimation tool given our model setup. Its performance will be further examined through extensive simulation studies reported in Section 5.

3.3 Bayesian analysis 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, we choose the following prior distributions:

- $\mu_{hi} \sim N(m_{\mu 0}, s_{\mu 0}^2)$ and $\mu_{qj} \sim N(m_{\mu 0}, s_{\mu 0}^2)$;
- $\frac{\phi_{hi}+1}{2} \sim Beta(a, b)$ and $\frac{\phi_{qj}+1}{2} \sim Beta(a, b)$;
- $\sigma_{hi}^2 \sim IG(\frac{n_{m0}}{2}, \frac{d_{m0}}{2})$ and $\sigma_{hi}^2 \sim IG(\frac{n_{m0}}{2}, \frac{d_{m0}}{2})$,

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

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

Iteration over steps 2-5 consists of a complete sweep of MCMC sampler. 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.

4 Extension to Model with Realized Measures

In the past, modeling asset volatility was usually based on daily return data only. Recently, many papers take advantage of intra-daily high-frequency data to investigate return volatility. A combination of both has also emerged. Within the GARCH framework, the effort was

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

In the SV literature, fewer studies exist on joint modeling of daily returns and realized measures. Existing studies include Takahashi et al. (2009) and Koopman and Scharth (2012), which both consider the univariate model. The former was extended in several directions by Venter and de Jongh (2014), while Zheng and Song (2014) extend the latter paper by considering the Box-Cox transformation. Recent contributions to multivariate modeling include Shirota et al. (2017), Kurose and Omori (2020), and Yamauchi and Omori (2019). Models proposed in these papers incorporate a lot of stylized features such as the leverage effect, endogeneity, and block structure. To ensure the positive-definiteness of the covariance matrix, these studies use techniques such as the Cholesky decomposition, the matrix exponential, and the pairwise Fisher transformation. See also Jin and Maheu (2013) and Jin and Maheu (2016), where the Wishart autoregression is used to model the stochastic covariance matrix. Due to the high-dimensional latent variables involved, estimation and inference for such models are usually computationally intensive and rely heavily on simulation-based approaches, especially MCMC. Most papers mentioned above use single-move approach to construct their MCMC sampler.

4.1 RMSV-GFT model

Given the huge parameter space of MSV-GFT model, daily returns alone may not be enough to produce stable inference. When realized measures for the latent process h_t and q_t are available as well, we can incorporate them into our model and greatly improve parameter estimation efficiency as well as the fit of the model; see Hansen et al. (2012) and Hurn et al. (2020).

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

obtained through

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

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

$$\begin{aligned} h_t^r &= (h_{1t}^r, ..., h_{pt}^r)' = (\log \tilde{v}_{1t}, ..., \log \tilde{v}_{pt})', \\ q_t^r &= (q_{1t}^r, ..., q_{dt}^r)' = F(Q_t), \end{aligned} \tag{11}$$

where superscripts denote realized measures.

The relationship between the latent variables and their realized counterparts is modeled as

$$h_t^r = \psi_h + h_t + \xi_{ht}, \text{ where } \xi_{ht} \sim N(0, \Sigma_h^r),\ t = 1, ..., T-1, \tag{12}$$

$$q_t^r = \psi_q + q_t + \xi_{qt}, \text{ where } \xi_{qt} \sim N(0, \Sigma_q^r),\ t = 1, ..., T-1, \tag{13}$$

where $\psi_h = (\psi_{h1}, ..., \psi_{hp})'$ and $\psi_q = (\psi_{q1}, ..., \psi_{qd})'$ capture potential approximation biases in the realized measures, $\xi_h = (\xi_{h1}, ..., \xi_{hp})'$ and $\xi_q = (\xi_{q1}, ..., \xi_{qd})'$ are innovations which are normally distributed. These innovations are assumed to be independent of each other so that Σ_h^r and Σ_q^r are both diagonal matrices, with the diagonal entries being $\eta_h^2 = (\eta_{h1}^2, ...\eta_{hp}^2)$ and $\eta_q^2 = (\eta_{q1}^2, ..., \eta_{qd}^2)$. Combining equations (2)-(7) with equations (12)-(13), we have an extended model which will be referred to as realized MSV-GFT (RMSV-GFT) model in the following.

It can be seen that extra measurement equations are added to the basic MSV model. These additional equations are based on the transformation of realized measure and the same transformation applied to the latent covariance matrix. In the literature, it has been shown that the realized volatility converges to the integrated volatility and same applies to logarithm versions. However, Barndorff-Nielsen and Shephard (2002) argued that the approximation of log integrated volatility by the log realized volatility usually performs better in practice. This property has been used in Hansen and Huang (2016) to introduce a realized EGARCH model and in Phillips and Yu (2009) to construct a two-stage method to estimate continuous time models.

4.2 Bayesian analysis of RMSV-GFT

The inference of RMSV-GFT model is straightforward using our particle-filter-based Bayesian analysis. When realized measures are observed, the parameters of model are $\theta = (\mu_h, \mu_q, \phi_h, \phi_q, \sigma_h^2, \sigma_q^2, \psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)'$. For the first six, we assume the same priors as in Section 3.3. For the additional parameters, we impose following normal-inverse-gamma priors

- $\psi_{hi} \sim N(m_{\psi 0}, s_{\psi 0}^2)$ and $\psi_{qj} \sim N(m_{\psi 0}, s_{\psi 0}^2)$;
- $\eta_{hi}^2 \sim IG(\frac{n_{m0}}{2}, \frac{\eta_{m0}}{2})$ and $\eta_{qj}^2 \sim IG(\frac{n_{m0}}{2}, \frac{\eta_{m0}}{2})$.

4.2.1 Joint Estimation of RMSV-GFT

Let $h^r = (h_1^{r'}, ..., h_T^{r'})'$ and $q^r = (q_1^{r'}, ..., q_T^{r'})'$. Similar to the Gibbs sampler in Section 3.3, a natural way to make inference for the extended model with realized measures is as follows:

1. Initialize h, q and θ.
2. Draw $h, q|r, h^r, q^r, \theta$.
3. Draw $\mu_h, \mu_q|r, h, q, \theta_{/(\mu_h, \mu_q)}$.
4. Draw $\phi_h, \phi_q|r, h, q, \theta_{/(\phi_h, \phi_q)}$.
5. Draw $\sigma_h^2, \sigma_q^2|r, h, q, \theta_{/(\sigma_h^2, \sigma_q^2)}$.
6. Draw $\Sigma_h^r, \Sigma_q^r|h, q, h^r, q^r, \theta_{/(\Sigma_h^r, \Sigma_q^r)}$.
7. Draw $\psi_h, \psi_q|h, q, h^r, q^r, \theta_{/(\psi_h, \psi_q)}$.

Iterating over steps 2-7 consists of a complete sweep of MCMC sampler. We call this procedure joint estimation to distinguish from the two-stage method introduced in the next subsection. PGAS algorithm introduced in Section 3.2 is again used to sample latent variables h and q, conditional on not only daily returns but also realized measures. The sampling of model parameters is discussed in Appendix B.

4.2.2 Two-stage estimation of RMSV-GFT

Note that if we consider only the measurement equations 12 and 13 and the state equations for h_t and q_t, the system is in a linear Gaussian state space form and the Kalman filter can provide the straightforward likelihood-based inference. Unfortunately, in the same spirit as in Koopman and Scharth (2012), the setting precludes us to identify all parameters in the

RMSV-GFT model. In particular, (μ_h, μ_q) and (ψ_h, ψ_q) are not separately identified. To see this, let x_t denote either h_{it} or q_{jt} and x_t^r denote the realized measures of x_t. Combining the realized measurement equations with the dynamics of the latent variables, we can easily show that x_t^r admits following ARMA(1,1) representation

$$x_t^r = (\mu + \psi) + \phi(x_{t-1}^r - \mu - \psi) + \eta_t - \phi\eta_{t-1} + \xi_t. \tag{14}$$

From this representation, it is clear that μ and ψ cannot be separately identified under the case that only realized measures are observed. To deal with this lack of identification, the return equation is indispensable, as in Koopman and Scharth (2012) and Takahashi et al. (2009). Theoretically, therefore, the joint estimation in Section 4.2.1 should deliver correct inference. In the literature of SV models with realized measures, this issue has been treated in various ways. Takahashi et al. (2009), for instance, ignore this problem and estimate all the parameters jointly. Yamauchi and Omori (2019) also fails to take into account the restrictions related to log-variances, while circumventing the constraints on correlations by imposing driftless random walks. Koopman and Scharth (2012) propose a two-step estimation method, which uses realized measurements only in the first step and relies on the Kalman filter. They impose a zero bias in the first step as an identification condition and then estimate it in the second step.

With a small or moderate number of MCMC iterations, however, the joint estimation still fails to produce satisfactory results. To obtain reasonable estimates, we need a very large number of MCMC iterations. The reason is that the necessary amount of MCMC moves depends on the likelihood ratio between daily returns and realized measures, which is given by

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

The smaller is the likelihood ratio, the larger number of MCMC iteration is required. Note that the dimension of the numerator and the denominator is p and $\frac{p(p+1)}{2}$, respectively. As p goes up, the difference between two dimensions increases. It is hence expected that the likelihood ratio would decrease and as a result the mixing of Gibbs sampler would deteriorate.

To overcome this problem of slow convergence, we propose to use a two-stage estimation procedure, which estimates the bias parameters in the first stage and then keep them fixed in the second stage. The critical fact we rely on is that μ can be accurately estimated even without realized measures, which allows us to obtain μ first and then use it to estimate ψ. Specifically, in the first stage, we estimate the model without using realized measures and obtain the posterior mean of μ, denote as $\hat{\mu}$. The procedure for this step is same as in Section

3.3. Then ψ_h and ψ_q are estimated by

$$\begin{aligned} \hat{\psi}_{h_i} &= \bar{h}_i^r - \hat{\mu}_{h_i}, i = 1, ..., p, \\ \hat{\psi}_{q_j} &= \bar{q}_j^r - \hat{\mu}_{q_j}, j = 1, ..., qd, \end{aligned} \tag{15}$$

where $\bar{h}^r$ and $\bar{q}^r$ is the sample mean of h^r and q^r, respectively. In the second stage, we re-estimate the model using the joint likelihood by plugging in $\hat{\psi}$. That is, we sample from

$$p(\theta_{/(\psi_h,\psi_q)}, h, q | r, h^r, q^r, \hat{\psi}_h, \hat{\psi}_q) \tag{16}$$

by using the same procedure as in Section 3.3[7]. After sampling from (16), we estimate θ, h, q by their posterior means. In the end, ψ is updated by subtracting the new μ from average realized measures. By doing this two-stage estimation, we circumvent the slow convergence problem and are able to obtain a much less biased estimator of ψ and μ, as shown in the simulation studies reported below.

5 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 for 100 times. 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

[7]Note that the only impact of including realized measures into the likelihood function is to change the weights computed during the particle filter.

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

the Markov chain and thereby a higher sampling efficiency.

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

1. $\mu_{h1} = \mu_{h2} = \mu_{h3} = 0.3$ and $\mu_{q1} = \mu_{q2} = \mu_{q3} = 0.7$,
2. $\phi_{h1} = \phi_{h2} = \phi_{h3} = 0.9$ and $\phi_{q1} = \phi_{q2} = \phi_{q3} = 0.8$,
3. $\sigma^2_{h1} = \sigma^2_{h2} = \sigma^2_{h3} = 0.05$ and $\sigma^2_{q1} = \sigma^2_{q2} = \sigma^2_{q3} = 0.05$.

All the simulation results reported in this section is based on 5000 MCMC iterations, among which the first 1000 samples are discarded as burn-in period.[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.

Without realized measures The estimation results for the mean parameters μ_{h1} and μ_{q1} are reported in the first and fourth column of 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 though there is an upward bias for μ_h and a downward bias for μ_{q1}. Given a particular number of particles used, as sample size increases, bias in the estimator for μ_{q1} increases, but this is not the case for μ_{h1}. Nevertheless, as expected, the standard deviations for both parameters decrease as T increases. On the other hand, an increasing number of particles can reduce bias substantially. For example, when the sample size is 2000, bias in the posterior mean of μ_{h1} reduces from 0.016 to 0.008 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.

The second and fifth column of Table 1 show the simulation results for the persistence parameters ϕ_{h1} and ϕ_{q1}. These parameters can be estimated accurately, even with 500 observations and 50 particles. The estimates have very small bias and low standard deviation. When 200 particles are used, bias almost completely vanishes.

The third and last column of Table 1 present the simulation results related to parameters σ^2_{h1} and σ^2_{q1}. A substantial downward bias is observed for the variance estimator of both h_1s

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

and q_1s when 50 particles are used. This bias is insensitive to the number of observations. Fortunately, it can be improved by using more particles. Indeed, when N=200, bias becomes much smaller, although it seems that a larger number of particles are necessary to fully cancel this bias.

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

At last, we report the filtered h_1 and q_1, together with the 95% credible interval and their true values, in Figure 1 and 2, respectively. These figures show that the sampling of latent variables based on the proposed particle filter is reliable.

With realized measures We now investigate the performance of the proposed method for estimating the RMSV-GFT model that incorporates the realized measures. The true values of parameters are the same as above, with the following additional ones,

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

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

The sample size considered is 500. We use the joint as well as two-stage estimation to fit the RMSV-GFT model based on simulated data. The autocorrelation functions (ACFs) for μ_h and μ_q are displayed in Figure 5. It can be seen that the MCMC draws for μ_h and μ_q mix much better when the two-stage method is used compared with joint estimation.

The detailed results are reported in Table 4. It can be seen that when the joint estimation method is used, the posterior means of both μ and ϕ are biased. Specifically, the posterior means of μ_h and μ_q are biased downward. The posterior means of ψ_h and ψ_q are upward biased with a similar magnitude. Similarly, the posterior means of σ^2 are 0.01 and 0.015 lower than the true values for the log-variance and transformed correlations, respectively. The posterior means of η^2 are upward biased.

On the other hand, our two-stage estimation method works very well, with a very small bias for all the parameters. Unbiased parameter estimation is critical to the filtering of the latent processes, which influences the out-of-sample prediction performance. To this end, we can examine the in-sample performance of the filtered $\{h_t\}$ and $\{q_t\}$ as the true latent

processes are available for comparison in the simulation studies. Indeed, the average mean squared error of all six unobservables is 0.339 when the model is estimated jointly and is 0.327 when the two-stage method is used. In summary, our inference procedure produces accurate estimators for both the parameters and latent volatility/correlation.

6 Empirical Analysis

In this section, we consider two empirical applications, which serve to illustrate different aspects of our model. The first one shows that the flexibility brought by the new parameterization leads to a better in-sample fit of the data, while the second one presents out-of-sample improvement due to the incorporation of realized measures.

6.1 Foreign exchange rates

In the first empirical application, the data contains 1406 weekly mean-corrected log-returns of Euro, Pound sterling, and Swiss franc exchange rates, all against the US dollar, from January 13, 1993 to December 25, 2019. These series are expected to be correlated, as the underlying economies are closely connected. Three time series are plotted in Figure 3.

Following four models are considered:

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

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

$$\begin{aligned} r_t &= V_t^{1/2}\epsilon_t,\ \epsilon_t \sim N(0, R_t), \\ V_t &= \exp(H_t),\ h_t = diag(H_t),\ q_t = F(R_t), \\ h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht},\ \eta_{ht} \sim N(0, \sigma_h^2), \\ q_{t+1} &= \mu_q + \phi_q(q_t - \mu_q) + \eta_{qt},\ \eta_{qt} \sim N(0, \sigma_q^2), \end{aligned} \tag{17}$$

   where ϕ_h, ϕ_q, σ_h^2 and σ_q^2 are all scalars.

3. MSV-CC: Model with constant correlation matrix over time, similar to the one con-

sidered in Harvey et al. (1994).

$$\begin{aligned} r_t &= V_t^{1/2}\epsilon_t,\ \epsilon_t \sim N(0, R_t) \\ V_t &= \exp(H_t),\ h_t = diag(H_t),\ q_t = F(R_t), \\ h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht},,\ \eta_{ht} \sim N(0, \sigma_h^2), \\ q_{t+1} &= \mu_q. \end{aligned} \tag{18}$$

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

$$\begin{aligned} r_t &= V_t^{1/2}\epsilon_t,\ \epsilon_t \sim N(0, R_t), \\ V_t &= \exp(H_t),\ h_t = diag(H_t), \\ h_{t+1} &= \mu_h + \phi_h(h_t - \mu_h) + \eta_{ht},\ \eta_h \sim N(0, \sigma_h^2), \\ R_t &= tilde Q_t^{-1} Q_t \tilde{Q}_t^{-1}, \\ Q_{t+1}^{-1}|k, Q_t^{-1} &\sim Wishart\left(k, \frac{1}{k} Q_t^{-\phi/2} \Lambda Q_t^{-\phi/2}\right), \\ \Lambda &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}, \end{aligned}$$

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

For all models, the parametrization of the volatility movement is the same. Hence, we focus on alternative specifications of correlation dynamics. To estimate and compare these models in a unified framework, we treat all four models as nonlinear state-space models and use the proposed PGAS algorithm to estimate them. The competing models are compared using DIC criterion, see Appendix C for a review. For the MSV-GFT model, the filtered q sequences for all three pairs are plotted in Figure 4.

Table 2 reports the posterior statistics of parameters related to volatility for all four competing models, including the posterior mean, the posterior standard deviation, and the 95% credible interval. As expected, all h sequences have a very high level of persistence, with the autoregressive root close to 1.

Table 3 reports the posterior statistics of parameters related to correlation for all four competing models. In the second model, since the persistence levels and the standard deviations are restricted to be the same among all q sequences, we have only one ϕ_q and one σ_q^2 to

estimate. In the third model, no ϕ_q and one σ_q^2 is in presence as all correlations are assumed to be constant over time. An important finding from Table 3 is that, when three correlation sequences are allowed to have separate dynamics, q_1 has a very low level of persistence and hence, is highly stationary. Moreover, the posterior mean of ϕ_q is $(0.212, 0.911, 0.812)$ in this model. This finding is in sharp contrast to that in the MSV-GFT model with equi-persistence where one parameter is assumed to govern all the latent variables. In the latter case, the posterior mean of ϕ_q is 0.957. To see whether this additional flexibility leads to any statistical improvement, we compare the log marginal likelihoods and DIC values of all four models. The results are reported in the last row of Table 3. The fully flexible model has a much higher value of log marginal likelihood and much lower DIC than all the competing models. For example, the difference between the log marginal likelihood value for fully flexible model and that for each competing models are 190.9, 123.3 and 64.6, respectively. This indicates decisive evidence favoring the fully flexible model. We can then conclude that adopting the generalized Fisher transformation is indeed beneficial for modeling the multivariate exchange rate returns, at least in terms of in-sample fit.

6.2 Stock returns with realized measures

6.2.1 Data

In the second empirical application, we focus on daily stock returns and corresponding realized measures. We consider the daily log-returns, defined as demeaned close-to-close returns, of following three stocks ($p = 3$): Bank of America (BAC), JP Morgan (JPM) and American Express (AXP). The full sample period is from February 1, 2001 to December 31, 2009. The sample size is $T = 2232$. The realized covariance is constructed by using 5-minute returns with sub-sampling and the Parzen weight function. The detailed construction of is discussed in Noureldin et al. (2012). The daily return data and the corresponding realized covariance are downloaded from the data library of the Oxford Man Institute website.[10] The three time series are plotted in Figure 6. We transfer the realized variance and correlation by using equation (11).

Our model assumes that the counterparts of these realized measures follow independent Gaussian autoregressions specified in (12) and (13). The distribution of the log realized volatility had been carefully and extensively discussed in the literature; see Andersen et al. (2001a) and Andersen et al. (2001b). One way to check the validity of the model specification is to examine the distribution properties of the realized measures. In this paper, we report the Q-Q plot of q^r in Figure 7 and the scatter plot of each two q^rs in Figure 8. These plots

[10] https://realized.oxford-man.ox.ac.uk/data

suggest that the assumptions we impose are plausible empirically.

6.2.2 Model estimation and in-sample fit

We estimate MSV-GFT and RMSV-GFT based on the full sample data and compare the empirical result. The empirical results are reported in Tables 5. The posterior quantities of (μ_h, μ_q) are similar in two cases. For example, the posterior means of ϕ_h are close to 1. This finding is well-known in the literature. However, the posterior means of q are very different. In MSV-GFT, the posterior means of ϕ_q are $(0.586, 0.567, 0.585)$ which are close to the persistence levels of q^r, which are $(0.33, 0.358, 0.405)$. By adding the realized measures, the posterior means of ϕ_q in RMSV-GFT is higher than those in MSV-GFT. Moreover, the posterior means of ψ_h and ψ_q are negative which are consistent with that has been reported in the literature; see Koopman and Scharth (2012) and Yamauchi and Omori (2019). Furthermore, the posterior means of η_h^2 and η_q^2 are larger than those of σ_h^2 and σ_q^2. Furthermore, the chains mix better in RMSV-GFT than in MSV-GFT as the inefficiency factors are in general smaller.

To validate the estimation results, we compare the in-sample fit of the realized measures with the realized measures. Using the measurement equations in (12) and (13), we conduct the in-sample fit of the realized variance $\hat{v}_t$ and correlation $\hat{Q}_t$ as following

$$\begin{aligned} \hat{v}_t &= \int I_p * (\mathbb{1}_p \otimes \exp(h_t + \psi_h))\, f(h_t, \theta | D_{1:t}) dh_t d\theta, \\ \hat{Q}_t &= \int F(q_t + \psi_q)^{-1} f(q_t, \theta | D_{1:t}) dq_t d\theta, \end{aligned} \tag{19}$$

where I_p is a p-dimensional identity matrix, $\mathbb{1}_p$ is a p-dimensional column vector of ones and $D_{1:t}$ is data of the model from 1 to t. When the model is MSV-GFT, D is return r; when the model is RMSV-GFT, D is return and transfered realized measure (r, h^r, q^r).

We compare $\hat{v}_t$ and $\hat{Q}_t$ with v_t and Q_t. Note that v_t is a diagonal matrix and Q_t is a symmetric matrix with diagonal elements equal to 1. Hence, we only compare the diagonal elements in v_t and the off-diagonal ones in Q_t. Figure 9 and Figure 10 show the in-sample comparisons. For the realized variance, the two models provide a reasonably similar fit except during the great financial crisis. Nevertheless, MSV-GFT provides a less accurate fit than RMSV-GFT. However, for the realized correlations, RSMV-GFT provides a more reasonable fit than MSV-GFT.

6.2.3 Out-of-sample forecasting performance

We now turn to the comparison of out-of-sample performance of MSV models with and without realized measures. Two criteria are considered. The first one is their ability to forecast future realized measures. The second is the performance of optimal portfolios based on these two models. The out-of-sample period, which includes 480 trading days, is from February 7, 2008 to December 31, 2009. These two dates will be denoted as t_0 and t_1, respectively, in the following.

Forecasting Procedure To estimate two models, we always use the most recent 1,000 observations. As the new observation becomes available, we roll the window forward but keep the window size fixed at 1,000 when estimating the two models. The general procedure is summarized as follows.

1. Given data $D_{t-1000:t-1}$, we draw N samples from $p(\theta, h, q \mid D_{t-1000:t-1})$ by MSV-GFT and RMSV-GFT in Section 3.3 ans Section 4.2. We draw 5,000 MCMC iterations from the posterior distributions and discard first 1000 MCMC iterations.

2. Given the data from 1 to t-1, forecast the variance $\hat{V}_{t|t-1}$ and the correlation $\hat{R}_{t|-1}$ by using

$$\begin{aligned} \hat{V}_{t|t-1} &= \int I_p * (\mathbb{1}_p \otimes \exp(h_t)\, f(h_t|h_{t-1}, \theta) f(\theta, h_{t-1}|D_{1:t-1}) d\theta dh_{t-1}, \qquad (20) \\ \hat{R}_{t|t-1} &= \int F^{-1}(q_t) f(q_t|q_{t-1}, \theta) f(q_t, \theta|D_{1:t-1}) d\theta dq_{t-1}. \end{aligned}$$

   The out-of-sample forecast of the covariance is obtained as

$$\hat{\Sigma}_{t|t-1} = \hat{V}_{t|t-1}^{\frac{1}{2}} \hat{R}_{t|t-1} \hat{V}_{t|t-1}^{\frac{1}{2}}. \qquad (21)$$

3. Iterate steps 1 and 2 for t from t_0 to t_1. We end up with 480 one-step-ahead out-of-sample forecasts of h and q.

Forecasting Realized Measure In the literature, it is common to compare various volatility models based on their ability to predict future realized measures. Corsi et al. (2012), for instance, consider the HAR models, while Koopman and Scharth (2012) investigate the performance of SV models. Here, we examine the forecasting ability of MSV-GFT

and RMSV-GFT model. The predicted quantities are obtained by

$$\begin{aligned}\hat{v}_{t|t-1} &= \int I_p * (\mathbb{1}_p \otimes \exp(h_t + \psi_h))\, f(h_t|h_{t-1}, \theta) f(h_t, \theta|D_{1:t-1}) dh_{t-1} d\theta, \\ \hat{Q}_{t|t-1} &= \int F(q_t + \psi_q)^{-1} f(q_t|q_{t-1}, \theta) f(q_{t-1}, \theta|D_{1:t-1}) dq_{t-1} d\theta,\end{aligned} \tag{22}$$

where $\hat{v}_{t|t-1}$ is the forecast of the realized variance and $\hat{Q}_{t|t-1}$ is the forecast of the realized correlation. To evaluate the out-of-sample performance, we consider both the univariate and the multivariate setting. For the former case, Patton (2011) establish the robustness of MSE and Qlike. Unfortunately, Qlike is not available for realized correlations as they may take a negative value, in which case Qlike is not well-defined. So we report MSE, Qlike for the forecasts of realized variances and only MSE forrealized correlations. For the sake of robustness check, we also report the Mean Absolute Error (MAE). These metrics are computed as follows.

$$\begin{aligned}MSE_v &= \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} (v_t - \hat{v}_{t|t-1})^2 \\ MSE_Q &= \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} (Q_t - \hat{Q}_{t|t-1})^2, \\ Qlike_v &= \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} \left(\frac{\hat{v}_{t|t-1}}{v_t} - log(\frac{\hat{v}_{t|t-1}}{v_t}) - 1 \right), \\ MAE_v &= \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} |v_t - \hat{v}_{t|t-1}|, \\ MAE_Q &= \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} |Q_t - \hat{Q}_{t|t-1}|.\end{aligned}$$

To evaluate out-of-sample performance of alternative models on forecasting the realized covariance, we calculate the multivariate Qlike of Patton and Sheppard (2009) which is defined as

$$Qlike = \frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} \left(tr(\widehat{RCOV}_{t|t-1}^{-1} RCOV_t) - \log |\widehat{RCOV}_{t|t-1}^{-1} RCOV_t| - p \right), \tag{23}$$

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

Under univariate setting, Table 6 reports MSE, Qlike and MAE for both MSV-GFT and RMSV-GFT. According to all three metrics, RMSV-GFT provides more accurate forecasts

than MSV-GFT. This is by no means surprising as RMSV-GFT incorporates additional information from realized measures. As for multivariate case, Table 6 reports the Qlike in the last row. According to multivariate Qlike, again, RMSV-GFT performs better than MSV-GFT in terms of forecasting realized covariance.

Figure 11 plots the out-of-sample forecasts of the realized measures. For the realized variance, it shows MSV-GFT is less precise from Jan 2009 to Dec 2010. And MSV-GFT provides more downward outliers for the forecasts of realized correlation than RMSV-GFT.

Global Minimum Variance Portfolio Analysis Now, we examine the economic significance of alternative models. In particular, we use the predicted covariance matrix 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. In our application, at period $t-1$, we construct the GMV portfolio with optimal weights $w_t = (w_{1t}, ..., w_{pt})$, which minimize the variance of the portfolio return over next period. The optimal weights are computed as unique solution to

$$\min_{w_t \in \mathbb{R}^p} w_t' \hat{\Sigma}_{t|t-1} w_t, \quad \text{s.t. } w_t' \mathbb{1}_p = 1, \tag{24}$$

where $\hat{\Sigma}_{t|t-1}$ is the forecast of the covariance in (21), and $\mathbb{1}_p$ is a p-by-1 vector of ones. In general, negative weights are allowed so that short-sells are possible. The well-known solution of w_t is

$$w_t^\star = \frac{\hat{\Sigma}_{t|t-1}^{-1} \mathbb{1}_p}{\mathbb{1}_p' \hat{\Sigma}_{t|t-1}^{-1} \mathbb{1}_p}, \tag{25}$$

and the optimal portfolio return at time t is

$$R_t^p = w_t^{\star\prime} r_t. \tag{26}$$

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

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

To facilitate comparison, we assume all stocks have equal expected returns and focus only on the variance. The variance of the portfolio is measured by average squared return $\frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} (R_t^p)^2$. Since the average squared return is sensitive to outliers, we also compute average absolute return, $\frac{1}{t_1 - t_0} \sum_{t=t_0}^{t_1} |R_t^p|$ as a robustness check. The results are presented in Table 7. The equal-weighted portfolio shows a significantly larger variance. During the

out-of-sample period, the RMSV-GFT model generates the smallest portfolio variance. We also report the portfolio variance for both years in the out-of-sample period. It can be seen that the portfolio based on RMSV-GFT model in fact dominates in both subperiods.

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

7 Conclusion

We propose a novel multivariate stochastic volatility model in this paper, using a generalized version of Fisher's z-transformation to characterize the dynamics of correlation structure in a highly flexible manner. The leading features are that our model will automatically generate positive-definite correlation matrix and the driving forces underlying volatilities and correlations are fully separated. When realized measures are available, we also discuss how to incorporate these additional information. Different from many existing literature, when making inference, we apply the state-of-the-art particle-filter-based MCMC technique. We show through simulation that the approach we choose works well under our specific context.

Though our model enjoys good properties, there are still many potential improvements. First of all, we don't address the leverage effect in this paper. To accommodate this prominent empirical fact, one can extend the current model by adding correlations between innovations to returns and volatilities. Another fundamental task is to find an parsimonious yet plausible way to tackle the huge parameter space when a large pool of assets are considered simultaneously in our framework. Given the heavy computational burden, it's unrealistic to implement the model with many assets without modification. To this regard, special assumptions like (block) equicorrelation or latent factor structures must be imposed. These issues are left for our future work.

Appendix A. Details of PGAS algorithm

Consider a state-space model in the form of 8 and 9. 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 trejctory of $x_{1:T}$, which is a sample from last MCMC iteration. Let's denote that reference trejctory 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:
  - (a). Generate $\{\tilde{x}_{1:t-1}^{(i)}\}_{i=1}^{N-1}$ by sampling with replacement $N-1$ times from $\{x_{1:t-1}^{(i)}\}_{i=1}^{N}$ with probabilities propotional to the importance weights $\{w_{t-1}^{(i)}\}_{i=1}^{N}$.
  - (b). Draw J from $\{1, 2, ..., N\}$ with probabilities propotional to $w_{t-1}^{(i)} g(x'_t|x_{t-1}^{(i)})$ and then set $\tilde{x}_{1:t-1}^{(N)} = x_{1:t-1}^{(J)}$.
  - (c). Simulate $x_t^{(i)}$ from $q_t(x_t|\tilde{x}_{t-1}^{(i)}, y_t)$, for $i = 1, 2, ..., N-1$.
  - (d). Set $x_t^{(N)} = x'_t$.
  - (e). Set $x_{i:t}^{(i)} = (\tilde{x}_{1:t-1}^{(i)}, x_t^{(i)})$
  - (f). Set weight to be $w_t^{(i)} = f(y_t|x_t^{(i)}) g(x_t^{(i)}|\tilde{x}_{t-1}^{(i)})/q_t(x_t^{(i)}|\tilde{x}_{t-1}^{(i)}, y_t)$, for $i = 1, 2, ..., N$.
- Draw k from $\{1, 2, ..., N\}$ with probabilities propotional to $w_T^{(i)}$ and return $x_{1:T}^* = x_{1:T}^{(k)}$.

Note that this procedure is very similar to the original particle Gibbs sampler. The major modification is in the Step 4(b), where a new index is drawn and thus the N^{th} trajectory may not be the reference one from last iteration. In conditional PG, on the contrary, we fix the last particle to follow the input 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.

Appendix B. Details of Sampling Model Parameters

MSV-GFT Model The joint posterior distribution can be written as

$$
\begin{aligned}
p(\theta,h,q|r) &\propto p(r|\theta,h,q)p(\theta,h,q) \\
&= f(r|h,q)g_\theta(h)g_\theta(q)\pi(\theta) \\
&= f(r_1|h_1,q_1)g_\theta(h_1)g_\theta(q_1)\prod_{t=2}^{T}\left[f(r_t|h_t,q_t)g_\theta(h_t|h_{t-1})g_\theta(q_t|q_{t-1})\right]\pi(\theta) \\
&= \prod_{t=1}^{T}\left[\left(\sum_{i=1}^{p}h_{it}\right)|R_t|^{-1/2}\exp\left[-\frac{1}{2}r_t'(V_t^{1/2}R_tV_t^{1/2})^{-1}r_t\right]\right] \qquad (27) \\
&\quad\times\prod_{t=2}^{T}\prod_{i=1}^{p}\left[(\sigma_{hi}^2)^{-1/2}\exp\left(-\frac{1}{2\sigma_{hi}^2}\left(h_{it+1}-\mu_{hi}-\phi_{hi}(h_{it}-\mu_{hi})\right)^2\right)\right] \\
&\quad\times\prod_{t=2}^{T}\prod_{j=1}^{d}\left[(\sigma_{qj}^2)^{-1/2}\exp\left(-\frac{1}{2\sigma_{qj}^2}\left(q_{jt+1}-\mu_{qj}-\phi_{qj}(q_{jt}-\mu_{qj})\right)^2\right)\right] \\
&\quad\times\prod_{i=1}^{p}\left(\frac{\sigma_{hi}^2}{1-\phi_{hi}^2}\right)^{-1/2}\exp\left(-\frac{(h_{i1}-\mu_{h1})^2}{2\sigma_{hi}^2/(1-\phi_{hi}^2)}\right) \\
&\quad\times\prod_{j=1}^{d}\left(\frac{\sigma_{qj}^2}{1-\phi_{qj}^2}\right)^{-1/2}\exp\left(-\frac{(q_{j1}-\mu_{q1})^2}{2\sigma_{qj}^2/(1-\phi_{qj}^2)}\right)\pi(\theta).
\end{aligned}
$$

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 $j=1,...,d$,

$$
\mu_{hi}|r,h,q,\theta_{/\mu_{hi}} \sim N\left(\tilde{m}_{h\mu},\tilde{s}_{h\mu}^2\right) \text{ and } \mu_{qj}|r,h,q,\theta_{/\mu_{qj}} \sim N\left(\tilde{m}_{q\mu},\tilde{s}_{q\mu}^2\right) \qquad (28)
$$

   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

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

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

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

$$\begin{aligned}\log p(\phi_{hi}|y,h,q,\theta_{/\phi_{hi}}) &\propto \log p(h_i|\phi_{hi},\theta_{/\phi_{hi}}) + \log \pi(\phi_{hi}) \\ &= \log \pi(\phi_{hi}) - \frac{(h_{i1}-\mu_{hi})^2(1-\phi_{hi}^2)}{2\sigma_{hi}^2} + \frac{1}{2}\log(1+\phi_{hi}^2) \\ &\quad - \frac{\sum_{t=1}^{T-1}\left[(h_{it+1}-\mu_{hi}) - \phi_{hi}(h_{it}-\mu_{hi})\right]^2}{2\sigma_{hi}^2},\end{aligned} \tag{29}$$

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

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

Then, the draw is accepted with probability $\min\left[1, \exp\left\{g(\phi_{hi}^*)/g(\phi_{hi}^{(i-1)})\right\}\right]$, where $\phi_{hi}^{(i-1)}$ is the sample from last MCMC iteration and

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

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

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

$$\sigma_{hi}^2|y,h,q,\theta_{/\sigma_{hi}^2} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{d}_{hm}}{2}\right) \text{ and } \sigma_{qj}^2|y,h,q,\theta_{/\sigma_{qj}^2} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{d}_{qm}}{2}\right) \tag{30}$$

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

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

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

RMSV-GFT Model As (12) and (13) are two extra measurement equations independent of the original ones, the conditional likelihood can be written as

$$p\left(r, h^r, q^r | h, q, \theta\right) = p\left(r | h, q, \theta_{/(\psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)}\right) \cdot p\left(h^r, q^r | h, q, \psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r\right). \tag{31}$$

and θ, h, q can then be drawn from the full posterior distribution

$$p\left(\theta, h, q | r, h^r, q^r\right) \propto \left[p\left(r | h, q, \theta_{/(\psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r)}\right) \cdot p\left(h^r, q^r | h, q, \psi_h, \psi_q, \Sigma_h^r, \Sigma_q^r\right)\right] p(h, q, \theta). \tag{32}$$

All the parameters except for (ψ_{hj}, η_{hj}^2) can be sampled exactly the same as in (28), (29) and (30). As for new parameters, one can easily derive following conditional posterior distributions:

$$\psi_{hi} | h, h^r, \theta_{/\psi_{hi}} \sim N\left(\tilde{m}_{hi\psi}, \tilde{s}_{hi\psi}^2\right) \text{ and } \eta_{hi}^2 | h, h^r, \theta_{/\eta_{hi}^2} \sim IG\left(\frac{\tilde{n}_m}{2}, \frac{\tilde{\eta}_{hi}}{2}\right), \tag{33}$$

where

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

and

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

(ψ_{qj}, η_{qj}^2) can be dealt with similarly.

Appendix C. Model Comparison Based on DIC

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

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

$$\mathrm{D}IC = E_{\theta|r}[D(\theta)] + P_D \tag{34}$$

where $D(\theta) = -2 \log p(r|\theta)$ and $P_D = E_{\theta|r}[D(\theta)] - D(E_{\theta|r}[\theta])$. The first term measures the goodness of fit of the model while the second term measures the complexity of the model. Due to the use of the second term which penalizes the increasing size of a model, DIC can deal with the problem of over-fitting. The smaller the DIC value, the better the model.

To compute DIC based on MCMC output, the log-likelihood component in $D(\theta)$ can be

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

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

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

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

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

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

model M_i,

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

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

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

$$\log p(r|M_i) = \log p(r|\bar{\theta}M_i) + \log p(\bar{\theta}|M_i) - \log p(\bar{\theta}|r, M_i), \tag{35}$$

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

In the right-hand side of Equation (35), the first component $\log p(r|\bar{\theta}, M_i)$ is approximated by our proposed method. The second component is directly computed by the prior distribution. The posterior density is only known up to a normality constant. Follows as Kim et al. (1998), we approximate it by using the multivariate kernel density estimate. If a single-move or multi-move method is used, one would obtain $\log p(r|\bar{\theta}, h, q, M_i)$ by a by-product. Therefore, additional marginalization is required to integrated out h and q from $\log p(r|\bar{\theta}, h, q, M_i)$.

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

After DIC and the log marginal likelihood are obtained for competing models (either nested or non-nested), their values can be directly compared across models.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001a): "The distribution of realized stock return volatility," Journal of Financial Economics, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001b): "The distribution of realized exchange rate volatility," Journal of the American statistical association, 96, 42–55.

Andersen, T. G., T. Bollerslev, P. Frederiksen, and M. Ørregaard Nielsen (2010): "Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns," Journal of Applied Econometrics, 25, 233–261.

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

Archakov, I. and P. R. Hansen (2018): "A new parametrization of correlation matrices," Working paper, 1–32.

Archakov, I., P. R. Hansen, and A. Lunde (2020): "A Multivariate Realized GARCH Model," Working paper, 1–42.

Arulampalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002): "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Transactions on Signal Processing, 50, 174–188.

Asai, M. and M. McAleer (2006): "Asymmetric multivariate stochastic volatility," Econometric Reviews, 25, 453–473.

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

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

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

Bauwens, L., S. Laurent, and J. V. Rombouts (2006): "Multivariate GARCH models: a survey," Journal of Applied Econometrics, 21, 79–109.

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

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

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

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

Chopin, N. and S. S. Singh (2015): "On particle Gibbs sampling," Bernoulli, 21, 1855–1883.

Corsi, F., F. Audrino, and R. Renò (2012): HAR Modeling for Realized Volatility Forecasting, 363 – 382.

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

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

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

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

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

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

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

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

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

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

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

Hansen, P. R., A. Lunde, and V. Voev (2014): "Realized beta GARCH: A multivariate GARCH model with realized measures of volatility," Journal of Applied Econometrics, 29, 774–799.

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

Hurn, S., V. Martin, P. Phillips, and J. Yu (2020): Financial Econometric Modeling, Oxford University Press.

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

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

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

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

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

Kass, R. E. and A. E. Raftery (1995): "Bayes factors," Journal of the American statistical Association, 90, 773–795.

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

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

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

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

Li, Y., N. Wang, and Y. Jun (2020): "Improved marginal likelihood estimation via power posteriors and importance sampling," SMU Economics and Statistics Working Paper Series, Paper No. 16-2019.

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

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

Lopes, H. F., R. McCulloch, and R. Tsay (2010): "Cholesky stochastic volatility," Unpublished Technical Report, University of Chicago, Booth Business School, 2.

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

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

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

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

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

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

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

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

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

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002): "Bayesian measures of model complexity and fit," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64, 583–639.

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

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

Venter, J. H. and P. J. de Jongh (2014): "Extended stochastic volatility models incorporating realised measures," Computational Statistics and Data Analysis, 76, 687–707.

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

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

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

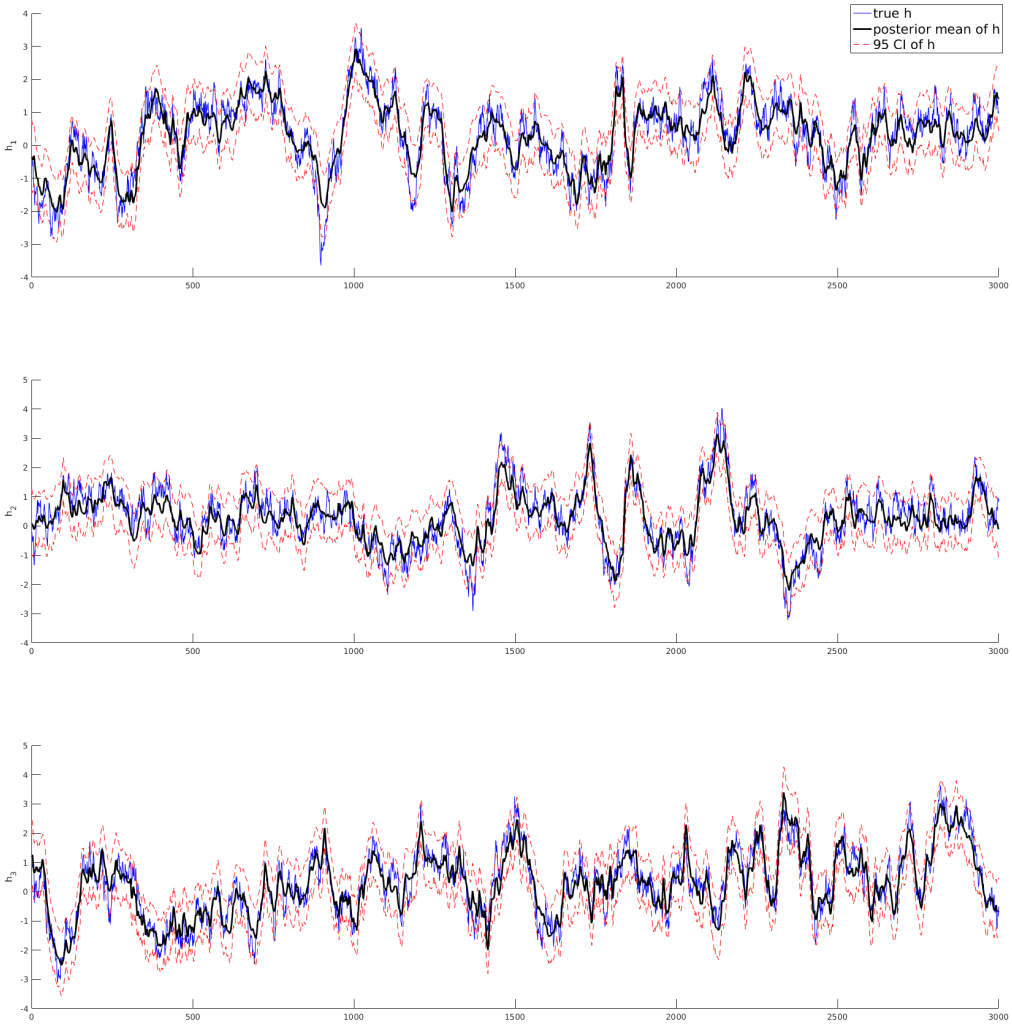
Zheng, T. and T. Song (2014): "A realized stochastic volatility model with Box–Cox transformation," Journal of Business & Economic Statistics, 32, 593–605.

Table 1: Simulation results without realized measures

			μ_{h1}	ϕ_{h1}	σ^2_{h1}	μ_{q1}	ϕ_{q1}	σ^2_{q1}
	N	True Value	0.3	0.9	0.05	0.7	0.8	0.05
T=500	50	Mean	0.311	0.894	0.026	0.632	0.784	0.027
		Std	0.094	0.033	0.013	0.063	0.051	0.012
		IF	11.5	56.4	173.6	17.3	56.9	124.2
	100	Mean	0.305	0.898	0.032	0.646	0.794	0.032
		Std	0.093	0.034	0.017	0.064	0.055	0.015
		IF	8.5	69.1	170.7	16.4	61.3	124.3
	200	Mean	0.304	0.902	0.036	0.657	0.798	0.037
		Std	0.092	0.033	0.019	0.064	0.059	0.018
		IF	6.4	61.3	138.0	12.2	60.7	121.8
T=1000	50	Mean	0.316	0.899	0.030	0.646	0.785	0.026
		Std	0.055	0.032	0.015	0.037	0.063	0.009
		IF	11.5	91.5	198.0	19.6	94.5	185.6
	100	Mean	0.310	0.900	0.036	0.661	0.793	0.032
		Std	0.053	0.032	0.016	0.036	0.060	0.011
		IF	9.0	103.86	182.3	17.2	101.0	180.43
	200	Mean	0.307	0.900	0.041	0.671	0.799	0.037
		Std	0.053	0.033	0.018	0.037	0.059	0.013
		IF	6.4	102.9	170.1	14.4	103.2	169.5
T=2000	50	Mean	0.316	0.900	0.030	0.651	0.793	0.025
		Std	0.036	0.022	0.010	0.028	0.047	0.008
		IF	10.1	112.4	21.2	198.9	128.2	240.4
	100	Mean	0.312	0.901	0.036	0.666	0.798	0.032
		Std	0.036	0.023	0.011	0.028	0.051	0.010
		IF	6.8	114.0	183.2	16.9	135.7	216.5
	200	Mean	0.308	0.900	0.041	0.675	0.801	0.037
		Std	0.037	0.023	0.012	0.028	0.049	0.012
		IF	6.3	103.6	161.8	15.7	134.9	205.7

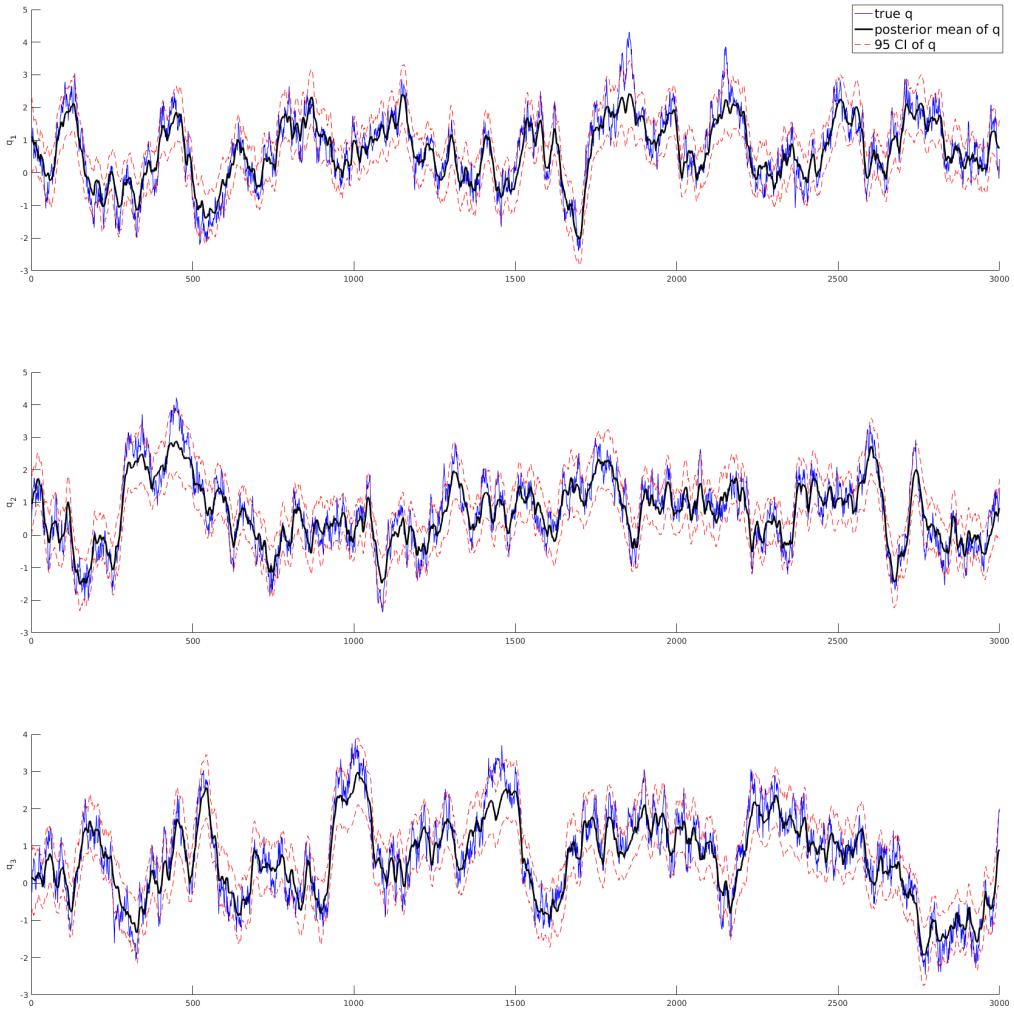
1. T is the number of observations for each asset and N is the number of particles used in PGAS.
2. Mean is the average posterior mean across replications.
3. Std is the standard error of the posterior mean across replications.
4. IF is the average inefficiency factor across replications calculated as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

Figure 1: True and filtered log-variances



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

Figure 2: True and filtered transformed correlations



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

Figure 3: Time series of exchange rate returns

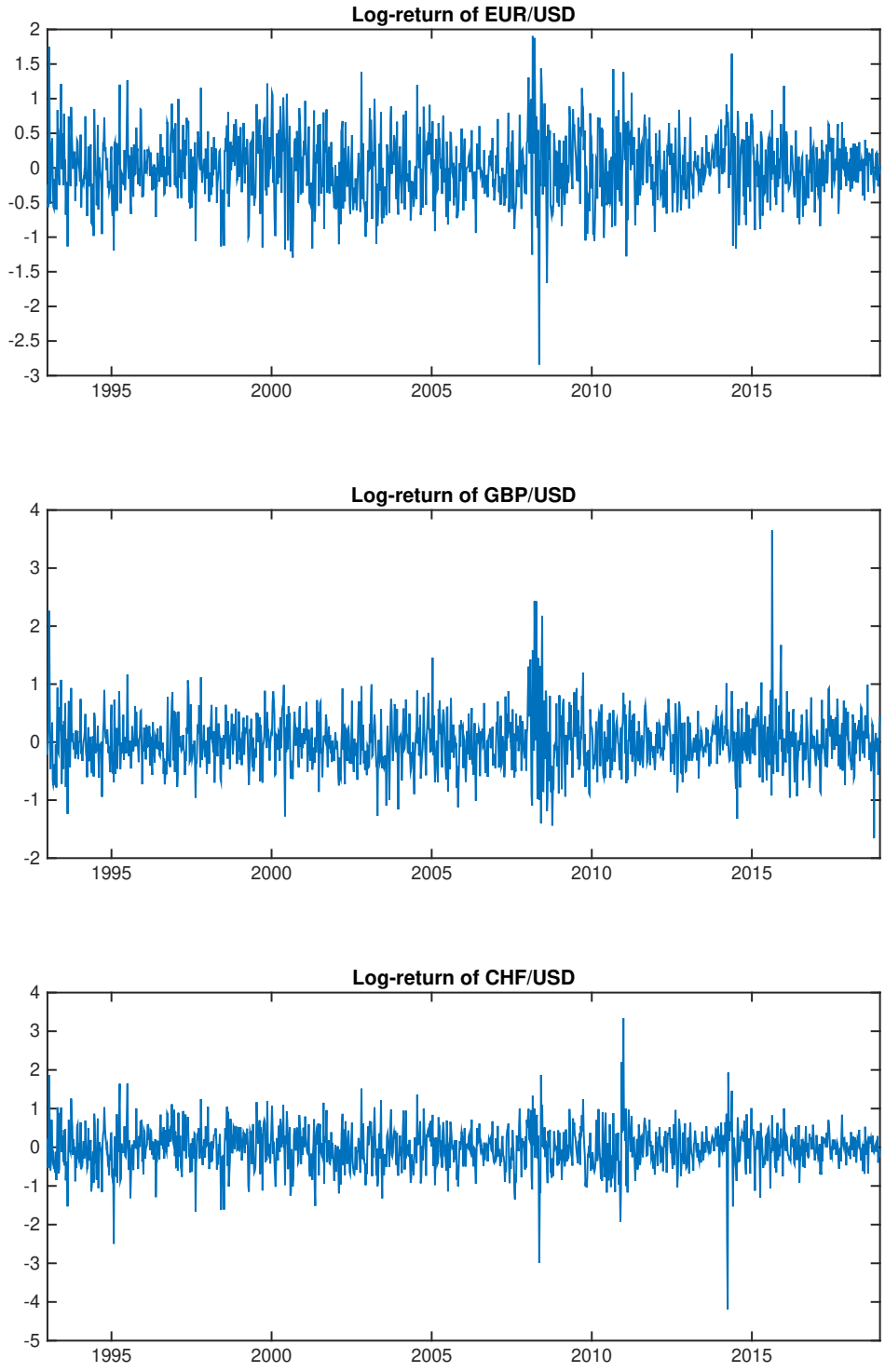


Table 2: Posterior statistics of parameters in the h sequences for competing models based on the exchange rate data

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

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

Table 3: Posterior statistics of parameters in the q sequences for competing models based on the exchange rate data

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

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

Figure 4: Filtered q sequences in MSV-GFT for the exchange rate returns

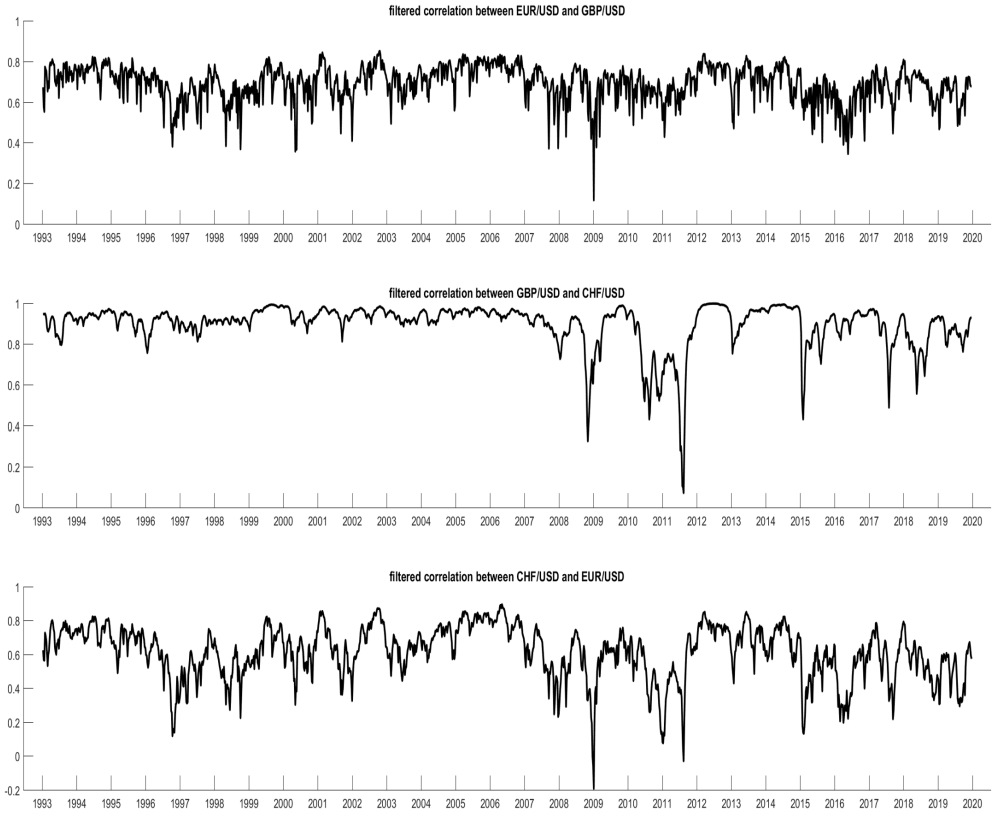
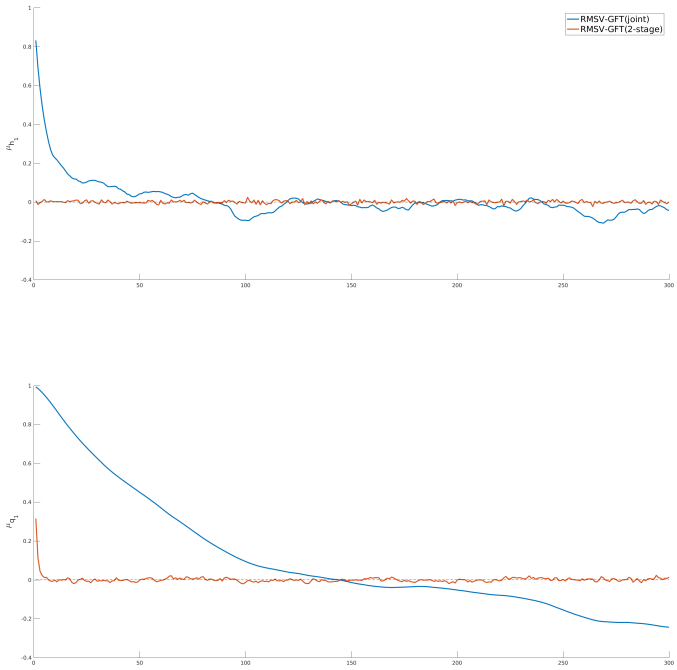


Figure 5: Comparison of ACFs of MCMC Draws for μ_h and μ_q



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

Table 4: Posterior statistics when the RMSV-GFT model is fitted to the simulated data

	μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}	ψ_{h1}	ψ_{h2}	ψ_{h3}	ψ_{q1}	ψ_{q2}	ψ_{q3}
Joint Estimation												
Mean	0.272	0.281	0.268	0.656	0.662	0.651	-0.474	-0.478	-0.475	-0.265	-0.269	-0.265
Std	0.088	0.084	0.081	0.063	0.049	0.057	0.073	0.068	0.069	0.05	0.037	0.052
IF	9.867	10.172	10.002	10.111	9.832	10.08	26.383	27.638	27.716	15.618	15.625	15.642
2-Stage Estimation												
Mean	0.299	0.307	0.295	0.65	0.657	0.648	-0.502	-0.505	-0.504	-0.255	-0.261	-0.259
Std	0.093	0.087	0.085	0.066	0.054	0.061	0.081	0.074	0.074	0.054	0.044	0.059
IF	1.272	1.307	1.021	2.106	1.878	1.936						
True Value	0.3	0.3	0.3	0.7	0.7	0.7	-0.5	-0.5	-0.5	-0.3	-0.3	-0.3
	ϕ_{h1}	ϕ_{h2}	ϕ_{h3}	ϕ_{q1}	ϕ_{q2}	ϕ_{q3}						
Joint estimation												
Mean	0.91	0.907	0.908	0.831	0.83	0.825						
Std	0.021	0.023	0.023	0.037	0.039	0.034						
IF	12.715	13.312	15.732	24.868	24.926	24.419						
2-Stage Estimation												
Mean	0.89	0.883	0.886	0.817	0.816	0.809						
Std	0.021	0.025	0.024	0.038	0.042	0.036						
IF	17.244	18.724	20.989	25.839	25.104	27.277						
True Value	0.9	0.9	0.9	0.8	0.8	0.8						
	σ^2_{h1}	σ^2_{h2}	σ^2_{h3}	σ^2_{q1}	σ^2_{q2}	σ^2_{q3}	η^2_{h1}	η^2_{h2}	η^2_{h3}	η^2_{q1}	η^2_{q2}	η^2_{q3}
Joint Estimation												
Mean	0.039	0.04	0.04	0.034	0.035	0.036	0.22	0.215	0.22	0.223	0.222	0.222
Std	0.009	0.008	0.01	0.009	0.009	0.01	0.018	0.015	0.018	0.017	0.02	0.017
IF	26.6	28.104	30.259	41.99	43.879	42.951	7.822	8.08	9.008	11.967	13.02	12.949
2-Stage Estimation												
Mean	0.048	0.051	0.05	0.034	0.036	0.037	0.211	0.198	0.207	0.228	0.226	0.224
Std	0.012	0.013	0.014	0.01	0.011	0.012	0.029	0.029	0.031	0.021	0.027	0.025
IF	40.446	44.711	43.358	52.335	52.938	53.244	14.36	16.557	13.835	27.365	28.418	29.568
True Value	0.05	0.05	0.05	0.05	0.05	0.05	0.2	0.2	0.2	0.2	0.2	0.2

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

Figure 6: Time series plot of stock returns

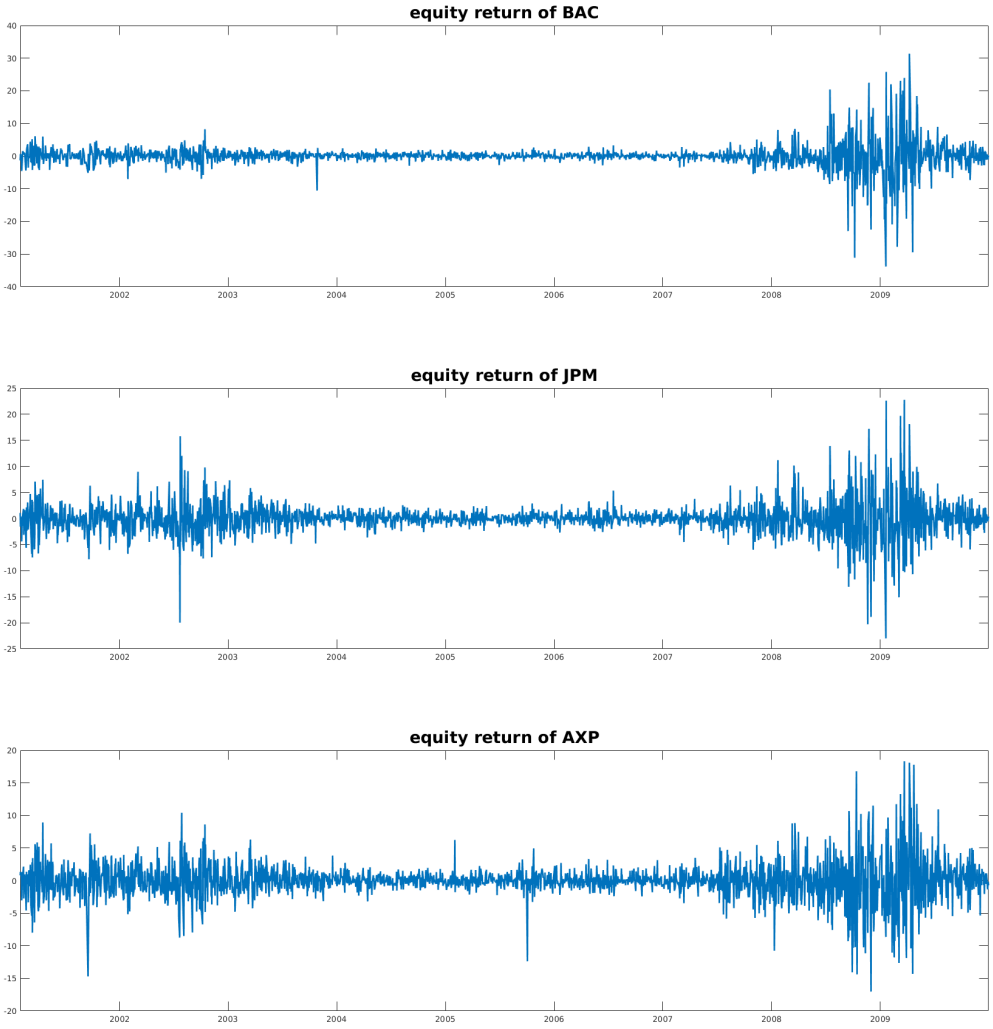
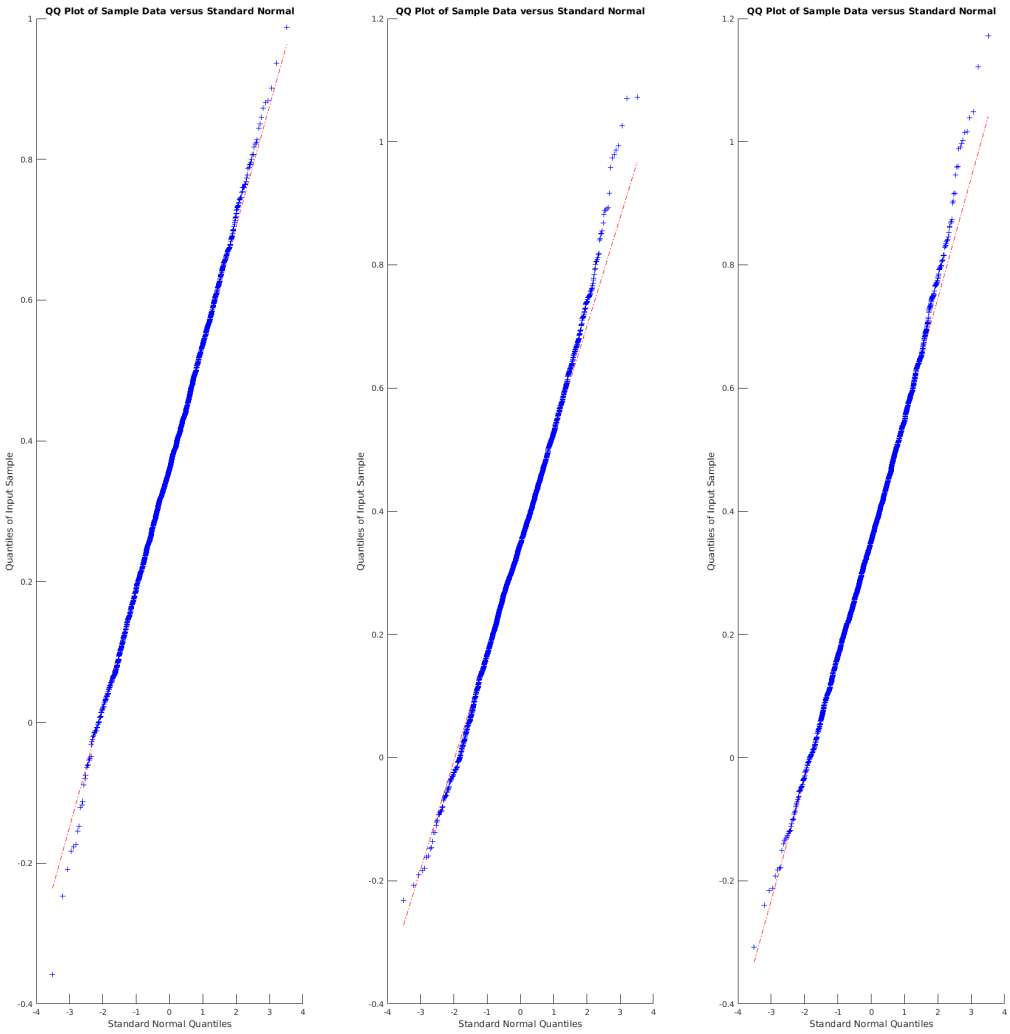
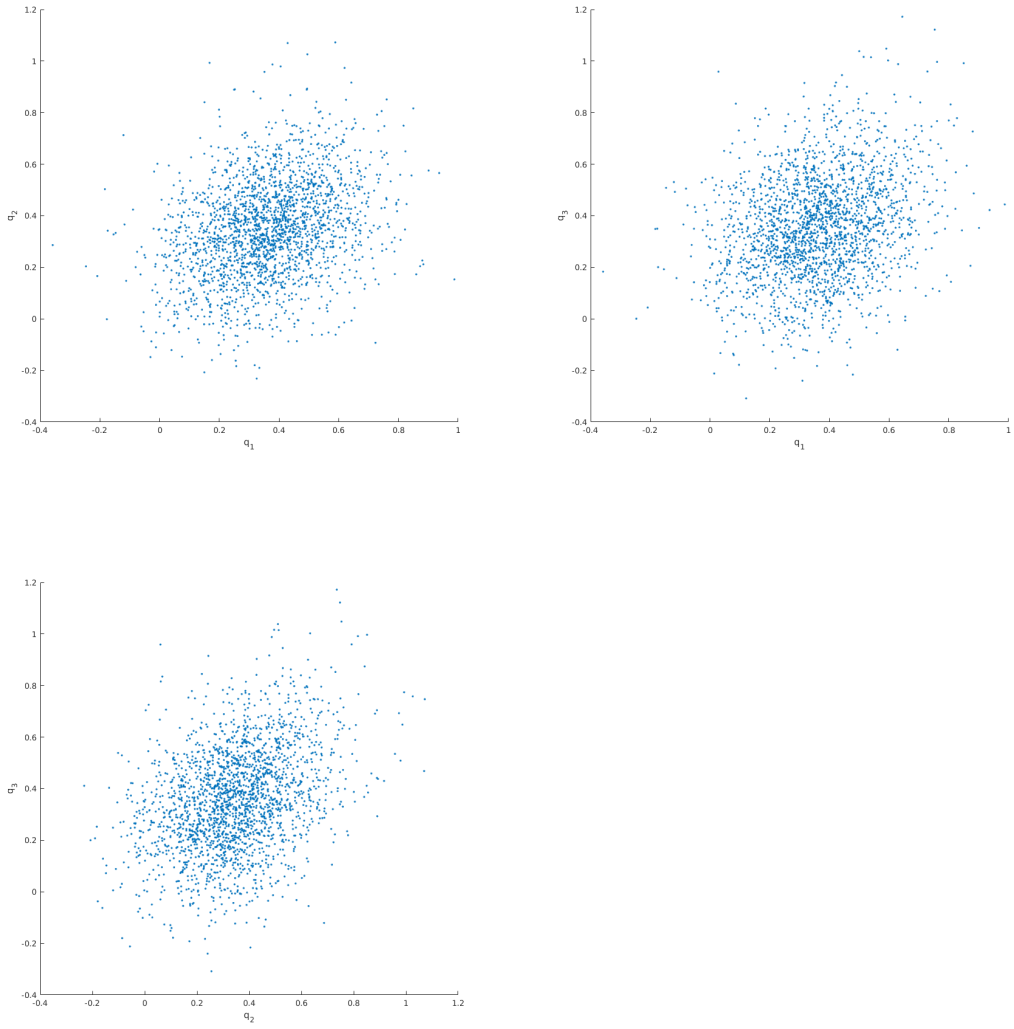


Figure 7: Q-Q plots of realized transformed correlations



Note: This is Q-Q plots for realized transformed correlitons q^r constructed using daily realized correlation matrix of BAC, JPM and AXP. The sample period is from 2 Jan 2001 to 31 Dec 2009.

Figure 8: Scatter plots of realized transformed correlations



Note: These are pairwise scatter plots of realized correlitons q^r constructed using daily realized correlation matrix of BAC, JPM and AXP. The sample period is from 2 Jan 2001 to 31 Dec 2009.

Table 5: Empirical Posterior parameter estimation without realized measures

a. without realized measures						
	μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}
mean	0.885	1.21	1.163	0.863	0.633	0.721
std	0.414	0.339	0.354	0.027	0.028	0.028
IF	8.476	9.397	8.998	26.738	34.805	26.225
	ϕ_{h1}	ϕ_{h2}	ϕ_{h3}	ϕ_{q1}	ϕ_{q2}	ϕ_{q3}
mean	0.993	0.993	0.992	0.586	0.567	0.585
std	0.002	0.002	0.002	0.132	0.089	0.099
IF	26.584	55.228	31.551	257.92	158.21	195.12
	σ^2_{h1}	σ^2_{h2}	σ^2_{h3}	σ^2_{q1}	σ^2_{q2}	σ^2_{q3}
mean	0.018	0.012	0.016	0.048	0.066	0.07
std	0.006	0.006	0.006	0.016	0.017	0.02
IF	80.872	109.16	83.834	257.76	250.05	213.49
b. with realized measures						
	μ_{h1}	μ_{h2}	μ_{h3}	μ_{q1}	μ_{q2}	μ_{q3}
mean	0.868	1.18	1.139	0.861	0.632	0.723
std	0.293	0.18	0.174	0.007	0.009	0.014
IF	8.714	8.915	8.59	10.549	9.537	9.762
	ϕ_{h1}	ϕ_{h2}	ϕ_{h3}	ϕ_{q1}	ϕ_{q2}	ϕ_{q3}
mean	0.988	0.979	0.981	0.773	0.842	0.924
std	0.004	0.007	0.007	0.047	0.038	0.031
IF	14.177	18.475	21.2	83.014	58.096	41.363
	σ^2_{h1}	σ^2_{h2}	σ^2_{h3}	σ^2_{q1}	σ^2_{q2}	σ^2_{q3}
mean	0.028	0.028	0.022	0.004	0.003	0.002
std	0.004	0.005	0.004	0.001	0.001	0.001
IF	55.265	69.463	84.922	89.295	54.492	39.511
	ψ_{h1}	ψ_{h2}	ψ_{h3}	ψ_{q1}	ψ_{q2}	ψ_{q3}
mean	-0.165	-1.103	-0.704	-0.501	-0.284	-0.364
	η^2_{h1}	η^2_{h2}	η^2_{h3}	η^2_{q1}	η^2_{q2}	η^2_{q3}
mean	0.4	0.207	0.211	0.022	0.024	0.028

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

Table 6: Out-of-Sample Prediction of Realized Measures

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

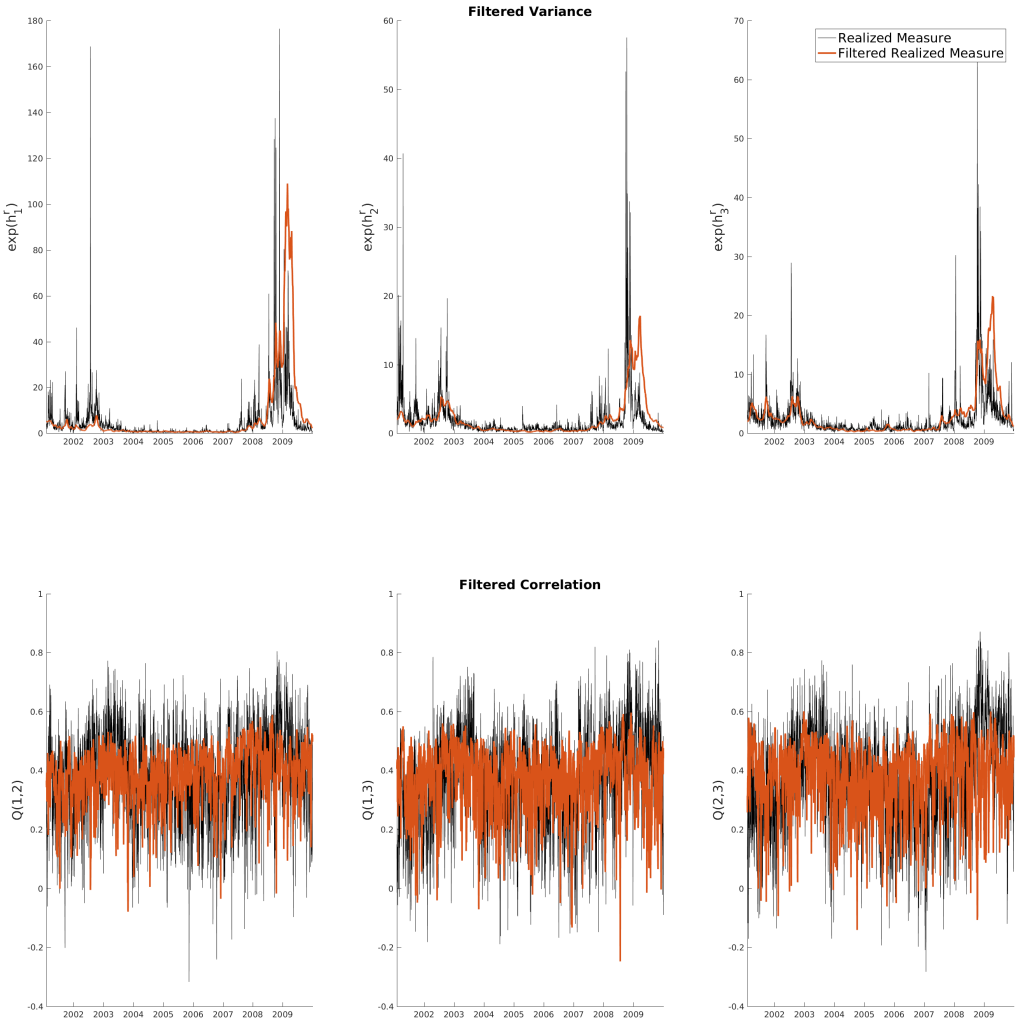
1.$\exp(h_i^r)$ is the i^{th} diagonal element in the realized variance.
2.Q(i,j) denotes the (i,j) components in the realized correlation matrix, while the diagonal elements are all 1.

Table 7: Out-of-Sample Performance of Portfolio Construction

	period	Equal Weight	MSV-GFT	RMSV-GFT
Squared Return	2008-2010	26.0852	19.4177	**19.0173**
	2008-2009	25.8988	20.5548	**20.3477**
	2009-2010	25.914	18.0109	**17.5466**
Absolute Return	2008-2010	3.4972	3.2044	**3.1696**
	2008-2009	3.7123	3.4434	**3.3787**
	2009-2010	3.2822	2.9654	**2.9605**

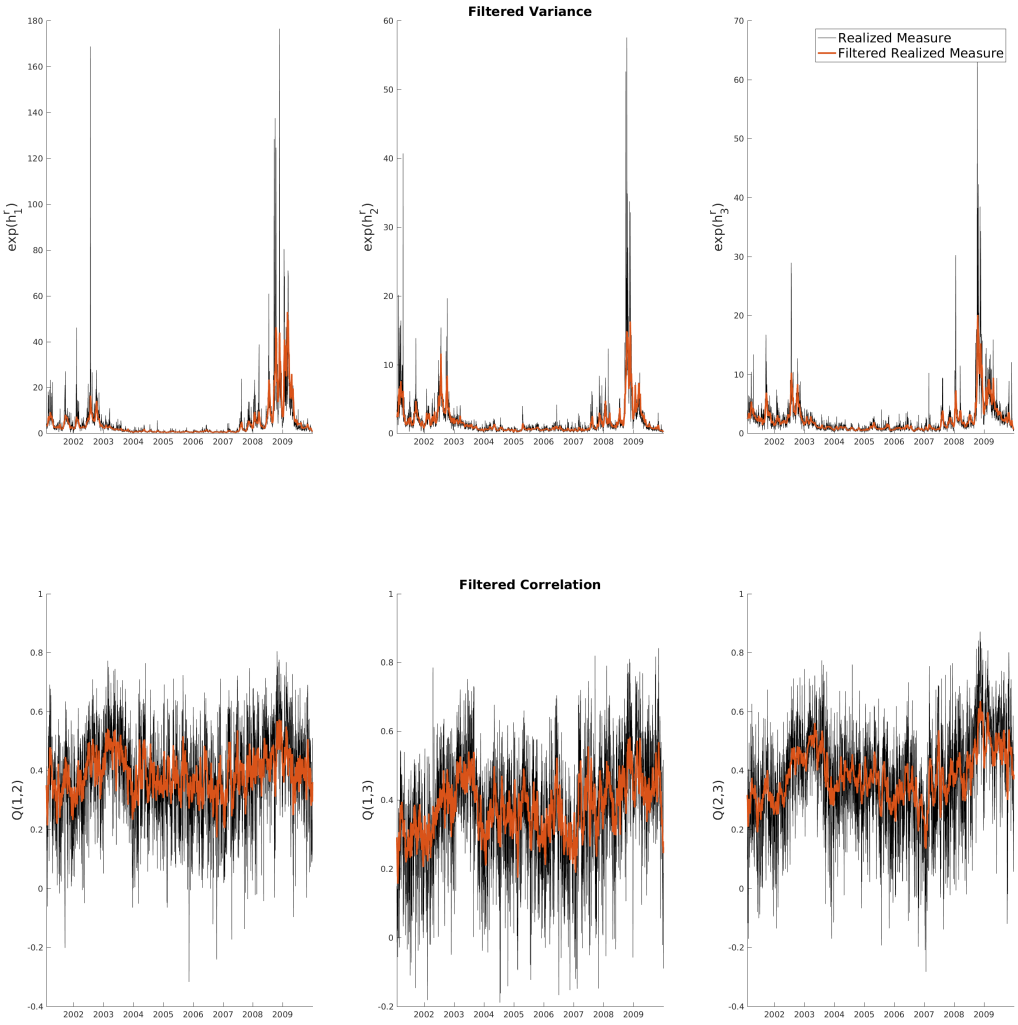
1.Bold figures indicates the optimal values
2.We report the result for $100 \times R_t^p$ return for precision.
3.Squared return is $\frac{1}{480}\sum_{t=1743}^{2242}(R_t^p)^2$ and Absolute Return is defined as $\frac{1}{480}\sum_{t=1743}^{2242}|R_t^p|$.

Figure 9: MSV-GFT: In-sample comparison with realized measure



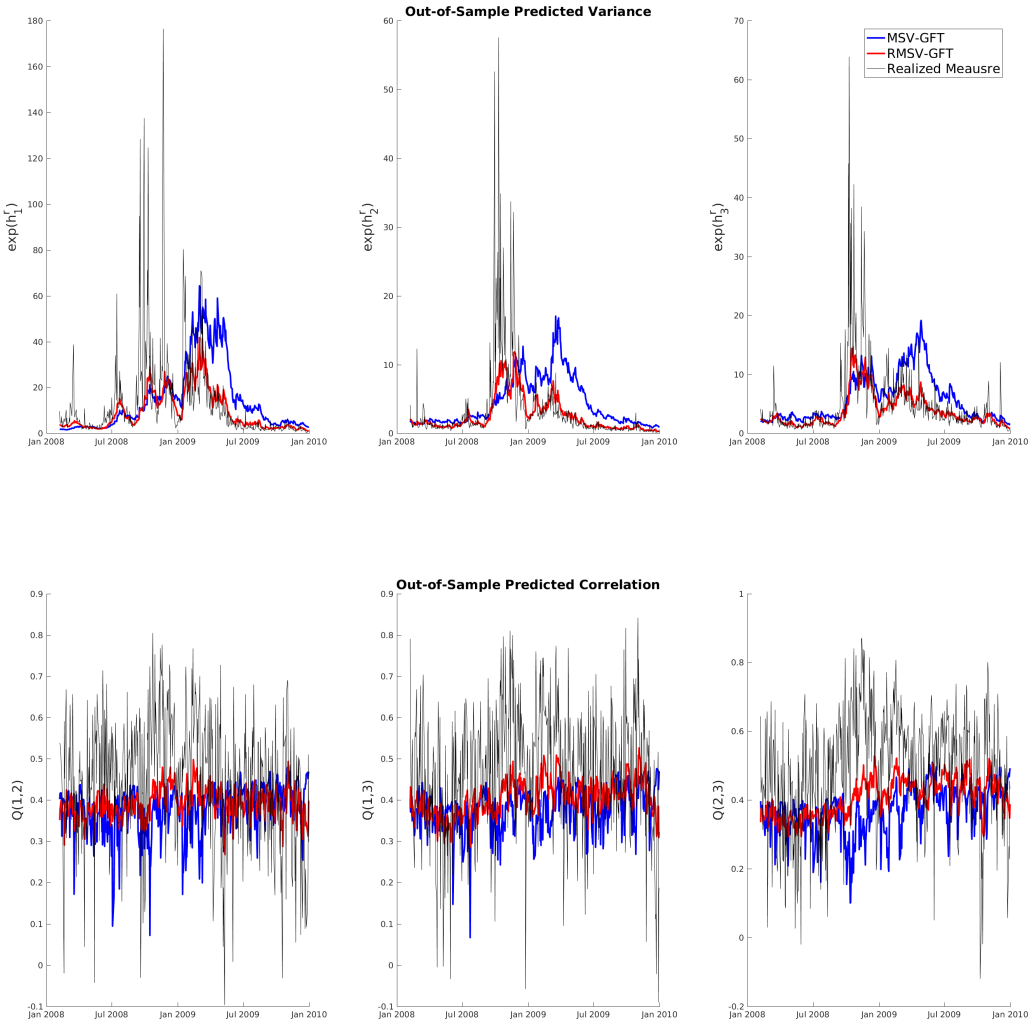
Note: The top three panels report in-sample fit of realized variance for BAC, JPM and AXP respectively; The bottom three panels report in-sample fit of the realized correlation for BAC and JPM, BAC and AXP, JPM and AXP respectively. The time period is from January 2, 2001 to December 31, 2009.

Figure 10: RMSV-GFT: In-sample comparison with realized measure



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

Figure 11: Out-of-sample prediction of realized measures



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