Practical estimation methods for high-dimensional multivariate stochastic volatility models [*]

Md. Nazmul Ahsan[†]
CIRANO

Jean-Marie Dufour [‡]
McGill University

August 03, 2022

Abstract

We propose computationally inexpensive and efficient estimators for multivariate stochastic volatility (MSV) models with cross dependence, Granger causality, and higher-order persistence in latent volatilities. The proposed class of estimators is based on a few moment equations derived from the VARMA representations of MSV models. Except for cross dependence parameters, closed-form expressions for the other parameters are derived where no numerical optimization procedure or choice of initial parameter values is required. To increase the stability and efficiency of volatility persistence parameter estimates, we suggest shrinkage type VARMA estimators where averaging or matrix-variate regression (MVR) is employed. We derive the asymptotic distribution of these estimators. Due to their computational simplicity, the VARMA estimators allow one to make reliable – even exact – simulation-based inference by applying Monte Carlo test techniques. In empirically realistic setups, simulation results show that the proposed shrinkage estimator based on MVR is superior to Bayesian and QML estimators in terms of bias and root mean square error. We examine the precision of the shrinkage estimator using large-scale simulated data where models up to 1,500 dimensions and 4,503,000 parameters are fitted and studied. The proposed estimators are applied to stock return data, and the effectiveness of the proposed estimators is assessed in two ways. First, we show the usefulness of the proposed models and methods in estimating high-frequency returns with many assets and observations. Second, in the context of dynamic minimum variance portfolio strategy, we find that unrestricted higher-order MSV models outperform existing alternatives, including multivariate GARCH-type models.

JEL Classification: C15, C32, C53, C55, C58, G12.

Keywords: Simple methods, matrix variate regression, Markov Chain Monte Carlo, Monte Carlo tests, multivariate stochastic volatility, asymptotic distribution.

[*] The authors thank Torben Andersen, Bertille Antoine, Russell Davidson, Prosper Dovonon, Hashem Pesaran, René Garcia, Eric Renault, Frank Kleibergen, Jean-Pierre Florens, Lynda Khalaf, Joann Jasiak, Richard Luger, Richard Baillie, John Galbraith, Victoria Zinde-Walsh, Denis Pelletier, Firmin Doko Tchatoka, Gabriel Rodriguez, Masaya Takano, Raida Alam, Endong Wang for their valuable comments and constructive discussions. Earlier versions of this paper were presented at CIREQ Montreal Econometrics Conference in Honor of Eric Renault, 37th Canadian Econometrics Study Group Meeting, CIREQ-Concordia Seminar, 11th RCEA Money-Macro-Finance Conference (Pandemic Crisis, Macro-Financial Distress, Risks and Opportunities), 7th RCEA Time Series Workshop (part of the keynote lecture), Bernoulli-IMS Symposium 2020. This paper was previously circulating under the title "Simple efficient estimators for large scale multivariate stochastic volatility models".

This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of excellence), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

[†] Centre interuniversitaire de recherche en analyse des organisations (CIRANO). Mailing address: 1130 Sherbrooke West, Suite 1400, Montreal, Quebec, H3A 2M8, Canada; e-mail: ahsann@cirano.qc.ca.

[‡] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com

1 Introduction

The dynamic conditional covariance matrices play a pivotal role in many areas of financial decision making, such as asset pricing, portfolio selection, option pricing, and risk management. For example, asset allocation depends on the covariance of the assets in a portfolio. Furthermore, multivariate volatility models have also been used to investigate volatility and correlation transmission and spillover effects across financial market contagion. The financial literature on modelling and forecasting the covariation among financial returns with parametric models are divided into two classes: (1) multivariate GARCH (MGARCH) class of models where volatility is modelled as a determinstic process given past observations [Engle (1982), Bollerslev (1986)] and (2) multivariate stochastic volatility (MSV) models where volatility is modelled as a latent stochastic process [Taylor (1982)].

The advantages of stochastic volatility models compared to GARCH-type models include: (1) SV models are directly connected to diffusion processes used in option pricing [Shephard and Andersen (2009)]; (2) SV models do not appear to require various ad hoc adjustments, like the addition of a random jump component or non-Gaussian innovations. These modifications improve the performance of the standard GARCH, but these are evidently unnecessary for SV models; (3) SV models provide more accurate forecasts than those offered by GARCH models [Kim, Shephard and Chib (1998); Koopman, Jungbacker and Hol (2005); Poon and Granger (2003); Yu (2002)]. However, in MSV models, the return variation dynamics is modeled as a latent autocorrelated stochastic process and the marginal likelihood of the model is given by a high dimensional integral which makes the estimation by conventional maximum likelihood infeasible.[1] Thus, estimating a multivariate stochastic volatility model is a challenging task and several alternative models and methods have developed.

Three broad classes of multivariate stochastic volatility models have emerged:

1. Multivariate extension of the univariate class of SV model; see Harvey et al. (1994), Danielsson (1998), Smith and Pitts (2006), Chan et al. (2006), and many others.

2. Multivariate factor stochastic volatility (MFSV); see Harvey et al. (1994), Jungbacker and Koopman

[1] This is a general feature of most nonlinear latent variable models because the latent variables must be integrated out of the joint density for the observed and latent processes, leading to an integral of high dimensionality. As a result, a variety of alternative methods have been proposed to estimate univariate SV models. Major references include: the Quasi-Maximum Likelihood (QML) [Harvey et al. (1994); Ruiz (1994)], the Generalized Method of Moments (GMM) [Melino and Turnbull (1990); Andersen and Sørensen (1996)], the Efficient Method of Moments (EMM) [Gallant and Tauchen (1996); Andersen et al. (1999)], the Maximum Likelihood Monte Carlo (MLMC) [Sandmann and Koopman (1998)], the Simulated Maximum Likelihood (SML) [Danielsson and Richard (1993); Danielsson (1994); Durham (2006); Richard and Zhang (2007)], method base on linear-representation (LiR) [Francq and Zakoïan (2006)], the closed-form moment-based estimator (DV) [Dufour and Valéry (2006)], the ARMA-based winsorized estimator (W-ARMA-SV) [Ahsan and Dufour (2019, 2021)] and Bayesian methods based on Markov Chain Monte Carlo (MCMC) methods [Jacquier et al. (1994), Kim et al. (1998), Chib et al. (2002), Flury and Shephard (2011)]. For a review of the SV literature, see Ghysels et al. (1996), Broto and Ruiz (2004), Ahsan and Dufour (2019, 2021).

(2006), Pitt and Shephard (1999), Aguilar and West (2000), So and Choi (2009), Chib et al. (2006), Han (2006), and many others.

3. Direct modeling of dynamic correlation matrices via matrix exponential transformations, Wishart processes, Cholesky decompositions and other means; see Uhlig (1997), Asai and McAleer (2009), Philipov and Glickman (2006*a*, 2006*b*), Gourieroux (2006), Gourieroux et al. (2009), and many others.

Despite recent advances, the estimation and forecasting of high-dimensional stochastic volatility models remain challenging problems. Major limitations of existing methods and models are:

- Curse of dimensionality: Portfolio management strategies often involve a large number of assets requiring the use of multivariate specifications. Most methods [QML, SML, MCL, SMM, and Bayesian MCMC], if not all, suffer from a common problem, the well-known *curse of dimensionality*, whereby methods become empirically infeasible if fitted to a number of series of moderate size (in some cases, the methods may become computationally intractable for even 5 or 6 assets). This also holds true when the sample size increased.

- Simplified restricted models: In order to match the need of introducing time-varying variances with practical computational problems, several restricted models are generally used. The introduction of strong restrictions can affect the interpretation and flexibility of the models, with a possibly impact on the purportedly improved performance they may provide and/or the appropriateness of the analysis based on their results. For example, many studies use multivariate AR(1) generalization for latent log volatilities, or multivariate log volatilities is a multivariate random walk. These models failed to capture the lead-lag (causal) relationship between time series and hence a non-diagonal VAR coefficient $\boldsymbol{\phi}$ should be useful in studying these relationships in variances. So and Choi (2009) observed that off-diagonal elements of $\boldsymbol{\phi}$ are statistically significant. Notable examples of restricted and inflexible models are the common stochastic volatility models [Carriero et al. (2016, 2018), Chan (2020)], and the Cholesky stochastic volatility models [Cogley and Sargent (2005), Carriero et al. (2019)].

- Inflexible estimation methods: Several estimation methods are inflexible across models, *i.e.,* not easy to generalize. For example, QML requires extremely tedious algebra to be involved in order to derive the analytical expression for the asymptotic variance-covariance matrix. Further, in the case of MCL method, when the cross-section size is large, it is not that obvious to write the analytical derivatives, and if instead one considers numerical derivatives in this case, then the derivatives calculated with respect to large state vectors could be very time consuming and numerically unstable;

see Jungbacker and Koopman (2006).

- Robustness of Bayes methods: Bayesian methods based on MCMC techniques are inherently less robust than analytic econometric methods and heavily depend on prior distributions. Further, when we consider a data set with many assets and observations, Bayes methods become extremely time consuming and empirically infeasible.

Practitioners may encounter several issues with the former classes of models, including modelling inflexibility, speed of estimation, and the lack of causal inferences (e.g., spillover effects). In this paper, we study and propose computationally simple and statistically efficient estimators for large-scale multivariate stochastic volatility models with cross dependence, multi-period Granger causality, and higher-order persistence in latent volatilities by exploiting the VARMA representation. Proposed estimators (denoted as VARMA estimators) are analytically tractable, computationally simple, and remarkably accurate. These estimators can readily be implemented without using numerical optimization methods (except for cross dependence parameters) and do not require choosing arbitrary initial values for the parameters or auxiliary models.

Our proposed high-dimensional modelling framework is also important in macroeconomic research, especially it can accommodate more sectors, parameters, endogenous variables, exogenous shocks, and observables than other small to medium scale models; recent studies have shown that stochastic covariances are more realistic in macroeconomic modelling [e.g., density forecasting in dynamic stochastic general equilibrium (DSGE) models (Diebold et al. (2017)) and measurements of uncertainty (Jurado et al. (2015))], can provide better statistical fits [see Primiceri (2005), Koop and Korobilis (2013), and many others], and vitally important for large macroeconomic VAR systems [see Clark (2011), D'Agostino et al. (2013), Koop and Korobilis (2013), Clark and Ravazzolo (2015), Cross et al. (2020) and Chan and Eisenstat (2018)].

The proposed simple moment-based VARMA estimator may violate stationarity conditions and fails to produce a positive semi-definite covariance matrix of the volatility innovations in the presence of outliers or in small samples. To circumvent this problem, we suggest restricted estimation where the estimates are restrained on the space of acceptable parameter solutions of $\boldsymbol{\phi}$ by adjusting the eigenvalues that lie on or outside the unit circle. Further, one can apply the modified Cholesky algorithm of Cheng and Higham (1998) to obtain the desired positive definite matrix.

We also suggest winsorized versions (shrinkage type) of the VARMA estimator (*W-VARMA* estimators), which substantially increases the probability of getting acceptable values and also improves efficiency. In proposing winsorized methods, autoregressive parameters of the latent volatility process [these parameters capture the volatility clustering of a financial time series] are estimated using a combination of several ratios of sample autocovariance matrices, including weighted averages, or a *matrix variate regression* (MVR) based weighting. This computationally simple adjustment improves the stability and accuracy of the estimators. Indeed, we show in simulations that W-VARMA estimators improve the precision and positive definiteness. Especially, the MVR-based W-VARMA estimator outperforms the Bayesian estimator in terms of bias and RMSE.

The proposed computationally inexpensive W-VARMA-MVR estimator can be useful in several contexts:

- Since MSV models are parametric models involving only a finite number of unknown parameters, using these estimators, one can construct simulation-based tests [even exact tests as opposed to procedures based on establishing asymptotic distributions] and/or prediction intervals based on the Monte Carlo test (MCT) technique [see Dufour (2006)]. In particular, exact tests obtained in this way do not depend on stationarity assumptions, and consequently are useful when the latent volatility process has a unit root (or is close to this structure).

- Due to computational simplicity, these estimators are helpful for estimation schemes which require repeated estimation based on a rolling window method, for example, out-of-sample forecasting or backtesting of risk measures (such as Value-at-Risk or Expected Shortfall) in the context of risk management.

We derive the asymptotic properties of the proposed simple estimator under standard regularity assumptions, showing consistency and asymptotic normality when the fourth moment of the latent volatility process exists. Due to the $\sqrt{T}$-consistency, our simple estimators can be effortlessly applied to very large samples, which are not rare in empirical finance. In these situations, estimators based on simulation technique and/or numerical optimization often require substantial computational effort to achieve convergence. So instead of using computationally costly or intractable estimators, one may prefer to use the W-VARMA-MVR estimator that are available in analytical form.

We study the statistical properties of our estimators and compare them with Bayesian and QML estimators. The simulation results confirm that the W-VARMA-MVR estimator has excellent statistical properties in terms of bias and RMSE. It outperforms all other estimators, including the Bayesian estimator, regarding bias and RMSE. This result holds across different simulation designs. Furthermore, the simple estimators are highly efficient in terms of computation time, compared to other estimators. We also simulate models upto 1,500 dimensions with 4,503,500 parameters and find that the W-VARMA-MVR estimator is extremely reliable in both small and large samples.

The proposed estimators are applied to stock return data, and the effectiveness of the proposed estimators is assessed in two ways. First, we show the usefulness of the proposed models and methods in estimating high-frequency returns with many assets and observations. Second, in the context of dynamic minimum variance portfolio strategy, we find that unrestricted higher-order MSV models outperform existing alternatives, including multivariate GARCH-type models.

The paper proceeds as follows. Section 2 specifies the model and its assumptions. Section 3 proposes simple estimators. Section 4 develops asymptotic theories for simple estimators. Section 5 discusses the simulation-based inference procedures, and Section 6 presents the simulation study, and Section 8 presents the empirical applications. We conclude in Section 9. The mathematical proofs, other discussions, figures and tables are given in the Technical Appendix.

2 Framework

We consider a discrete-time multivariate SV process, which is described below following Harvey, Ruiz and Shephard (1994). $\mathbb{N}_0$ refers to the non-negative integers.

Assumption 2.1. MULTIVARIATE STOCHASTIC VOLATILITY MODEL OF ORDER p. *Let $\boldsymbol{y}_t := (y_{1t},\ldots,y_{mt})'$ be a set of observations at time t on m financial variables and $\boldsymbol{h}_t := (h_{1t},\ldots,h_{mt})'$ be the corresponding vector of log volatilities. Then the process $\{\boldsymbol{y}_t : t \in \mathbb{N}_0\}$ follows a MSV(p) model of the type:*

$$\boldsymbol{y}_t = \boldsymbol{V}_t^{1/2}\boldsymbol{u}_t\,, \tag{2.1}$$

$$\boldsymbol{h}_t = \mu + \sum_{j=1}^{p}\boldsymbol{\phi}_j(\boldsymbol{h}_{t-j}-\mu) + \boldsymbol{v}_t\,, \tag{2.2}$$

where

$$\boldsymbol{h}_0,\ldots,\boldsymbol{h}_{-(p-1)} \sim \boldsymbol{N}_m(\mu\,,\boldsymbol{\Sigma}_0), \quad \begin{pmatrix}\boldsymbol{u}_t\\ \boldsymbol{v}_t\end{pmatrix}|\boldsymbol{h}_t \sim \boldsymbol{N}_{2m}(\boldsymbol{0},\boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_u & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\Sigma}_v\end{pmatrix}$$

and $\boldsymbol{V}_t^{1/2} = \mathrm{diag}[\exp(h_{1t}/2),\ldots,\exp(h_{mt}/2)]$, $\boldsymbol{h}_t = (h_{1t},\ldots,h_{mt})'$, $\boldsymbol{u}_t = (u_{1t},\ldots,u_{mt})'$, $\boldsymbol{v}_t = (v_{1t},\ldots,v_{mt})'$, $\mu = (\mu_1,\ldots,\mu_m)'$.

Remark 1. *Features of the MSV(p) model.*

1. *In this model, the return shocks are allowed to be correlated; so are the volatility shocks. Consequently, both returns and volatilities are cross-dependent.*

2. *Composition of $\boldsymbol{\phi}_j$'s are as follows:*

   - *Non-zero off-diagonal elements of $\boldsymbol{\phi}_j$ capture multilateral Granger causalities in volatility between assets — this also generates cross-dependence in asset return volatilities (cross leverage in asset volatilities). Which specification is more appropriate is an interesting empirical question.*

- *For $j > 1$, non-zero off-diagonal elements of $\boldsymbol{\phi}_j$ capture multi-horizon causalities in volatility between assets.*
- *Non-zero diagonal elements of $\boldsymbol{\phi}_j$ capture persistence in log volatilities of returns.*
- *For $j > 1$, non-zero diagonal elements of $\boldsymbol{\phi}_j$ capture higher-order persistence in log volatilities of returns.*

3. *For identification purposes, the diagonal elements of $\boldsymbol{\Sigma}_u$ must be 1, which means that the matrix $\boldsymbol{\Sigma}_u = (\rho)_{ij}$ is a correlation matrix.*
4. *For the VAR(p) model in (2.2), if $\boldsymbol{\Sigma}_u$ is not a diagonal matrix, then log volatilities $\boldsymbol{h}_t$'s are instantaneously correlated (or contemporaneously correlated). In this case, $\boldsymbol{h}_t$'s have instantaneous causality, which goes in both ways.*

To reduce the computational load, especially when m is large, the log volatilities can be assumed to be conditionally independent. In that case, $\boldsymbol{\phi} = \text{diag}[\phi_1, \ldots, \phi_m]$ and $\boldsymbol{\Sigma}_{\boldsymbol{v}} = \text{diag}[\sigma^2_{1,v}, \ldots, \sigma^2_{m,v}]$ are both diagonal matrices. Analyses of these restricted models are given by Harvey et al. (1994), Danielsson (1998), Smith and Pitts (2006) and Chan et al. (2006).

Assumption 2.2. STATIONARITY. *All eigenvalues of the companion matrix A of (2.2) have modulus smaller than one. The matrix A is given below:*

$$A := \begin{pmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \ldots & \boldsymbol{\phi}_{p-1} & \boldsymbol{\phi}_p \\ \boldsymbol{I}_m & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_m & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{I}_m & \mathbf{0} \end{pmatrix}.$$

Remark 2. *Number of parameters. For first-order : $2m^2 + m$ For higher-order : $(p+1)m^2 + m$*

Now transforming $\boldsymbol{y}_t$ by taking logarithms of the squares and using $\mathbb{E}(\log\circ(\boldsymbol{u}_t^2)) = (-1.27)\iota$, we can write the measurement equation of the model as

$$\begin{aligned} \log\circ(\boldsymbol{y}_t^2) &= \boldsymbol{h}_t + \log\circ(\boldsymbol{u}_t^2) && (2.3) \\ &= \mathbb{E}(\log\circ(\boldsymbol{u}_t^2)) + \boldsymbol{h}_t + [\log\circ(\boldsymbol{u}_t^2) - \mathbb{E}(\log\circ(\boldsymbol{u}_t^2))] && (2.4) \\ &= (-1.27)\iota + \boldsymbol{h}_t + \epsilon_t, && (2.5) \end{aligned}$$

where $\circ$ is the element-wise product and on setting

$$\boldsymbol{X}_t = \log\circ(\boldsymbol{y}_t^2) + (1.27)\iota, \quad \epsilon_t = [\log\circ(\boldsymbol{u}_t^2) - \mathbb{E}(\log\circ(\boldsymbol{u}_t^2))], \tag{2.6}$$

we have:

$$\boldsymbol{X}_t = \boldsymbol{h}_t + \epsilon_t. \tag{2.7}$$

Under the standard normality assumption of $\boldsymbol{u}_t$, the transformed error, ϵ_t, is a m-variate mean zero random variable with the covariance matrix Σ_ϵ. Harvey et al. (1994) showed that the (i, j)th element of the covariance matrix of $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})'$ is given by $(\pi^2/2)\rho^*_{ij}$, where $\rho^*_{ii} = 1$ and

$$\rho^*_{ij} = \frac{2}{\pi^2}\sum_{n=1}^{\infty}\frac{(n-1)!}{\left[\prod_{k=1}^{n}(1/2+k-1)\right]n}\rho_{ij}^{2n}. \tag{2.8}$$

Note that the absolute values of the unknown parameters in $\boldsymbol{\Sigma}_u$ namely the ρ_{ij}'s, the cross-correlations between different u_{it}'s, may be estimated, but their signs may not, because the relevant information is lost when the observations are squared. However, as suggested by Harvey et al. (1994), the signs of these estimates may be obtained by returning to the untransformed observations and noting that the sign of each of the pairs $u_i u_j, i, j = 1, \ldots, T$, will be the same as the corresponding pair of observed values $y_i y_j$. So the sign of ρ_{ij} is estimated as positive if more than one-half of the pairs $y_i y_j$ are positive.

Combining (2.2) and (2.7), the MSV(p) model can be written as a linear state space model of the type:

$$\boldsymbol{h}_t = \mu + \sum_{j=1}^{p}\boldsymbol{\phi}_j(\boldsymbol{h}_{t-j} - \mu) + \boldsymbol{v}_t, \quad \text{(State Transition Equation)} \tag{2.9}$$

$$\boldsymbol{X}_t = \boldsymbol{h}_t + \epsilon_t, \quad \text{(Measurement Equation)} \tag{2.10}$$

where $\boldsymbol{h}_t$ is the logarithm of latent daily volatilities, $\boldsymbol{X}_t$ is the logarithm of the daily squared return corrected by its mean, $(-1.27)\iota$. The $\boldsymbol{v}_t$ is the structural error given in Assumption 2.1 and ϵ_t is the transformed error that we explain above.[2]

3 Simple estimation methods

In this section, we propose simple estimators for MSV(p) models, including the corresponding recursive procedures. Besides, we also suggest alternative methods to improve the performance of these simple estimators.

[2] Note that several methods have been proposed in the literature that exploits the state space form of SV models; see Harvey, Ruiz and Shephard (1994), Ruiz (1994), Shephard (1994), Breidt and Carriquiry (1996), Harvey and Shephard (1996), Kim, Shephard and Chib (1998), Sandmann and Koopman (1998), Steel (1998), Chib, Nardari and Shephard (2002), Knight, Satchell and Yu (2002), Francq and Zakoïan (2006), Omori, Chib, Shephard and Nakajima (2007).

3.1 VARMA-based estimation

In this section, we propose a simple closed-form estimator for MSV(p) model, where we exploited the $VARMA$ representation of the observed process $\boldsymbol{X}_t$. The $VARMA$ representation and the autocovariance structure of $\boldsymbol{X}_t$ are given in the following Lemmas.

Proposition 3.1. VARMA REPRESENTATION OF MSV(P) MODELS. *Under the Assumptions 2.1–2.2, the process $\boldsymbol{X}_t$ defined in (2.6) has the following VARMA(p,p) representation:*

$$\boldsymbol{X}_t = \mu + \sum_{j=1}^{p} \boldsymbol{\phi}_j (\boldsymbol{X}_{t-j} - \mu) + \eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j}, \tag{3.1}$$

where $\eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-1} = \boldsymbol{v}_t + \epsilon_t - \sum_{j=1}^{p} \boldsymbol{\phi}_j \epsilon_{t-1}$ admits a VMA(p) process, and the error processes $\boldsymbol{v}_t$ and ϵ_t are independent random vectors with covariance matrices $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_\epsilon$, respectively.

From the above proposition, we have simple expressions for the autocovariances and parameters of the MSV(p) model, and these are given in following corollaries.

Corollary 3.2. AUTOCOVARIANCES OF THE OBSERVED PROCESS. *Under the assumptions of Proposition 3.1, the autocovariances of the observed process $\boldsymbol{X}_t$ defined in (2.6) satisfy the following equations:*

$$\operatorname{Cov}\big((\boldsymbol{X}_t - \mu)(\boldsymbol{X}_{t-k} - \mu)'\big) := \boldsymbol{\Gamma}_k = \begin{cases} \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon; & \text{if } k = 0, \\ \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} - \boldsymbol{\phi}_k \boldsymbol{\Sigma}_\epsilon; & \text{if } 1 \le k \le p, \\ \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p}; & \text{if } k > p. \end{cases} \tag{3.2}$$

Corollary 3.3. CLOSED-FORM EXPRESSIONS FOR MSV(P) PARAMETERS. *Under the assumptions of Proposition 3.1, we have:*

$$\begin{aligned} \boldsymbol{\phi}_{(p)} &= \boldsymbol{\Gamma}_{(p+j)} \boldsymbol{\Gamma}_{[p+j-1]}^{-1} \quad j \ge 1, \\ \boldsymbol{\Sigma}_\epsilon &= \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}_j' - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1, \\ \boldsymbol{\Sigma}_v &= \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1 - \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}_j' - \sum_{j=1}^{p} \boldsymbol{\phi}_j \boldsymbol{\Gamma}_j', \\ \mu &= \mathbb{E}(\boldsymbol{X}_t) \end{aligned} \tag{3.3}$$

where

$$\underset{m\times pm}{\boldsymbol{\phi}_{(p)}} := \left[\underset{m\times m}{\boldsymbol{\phi}_1} \quad \underset{m\times m}{\boldsymbol{\phi}_2} \quad \cdots \quad \underset{m\times m}{\boldsymbol{\phi}_p} \right], \quad \underset{m\times pm}{\boldsymbol{\Gamma}_{(p+j)}} := \left[\underset{m\times m}{\boldsymbol{\Gamma}_{p+j}} \quad \underset{m\times m}{\boldsymbol{\Gamma}_{p+j+1}} \quad \cdots \quad \underset{m\times m}{\boldsymbol{\Gamma}_{2p+j-1}} \right],$$

$$\underset{pm\times pm}{\boldsymbol{\Gamma}_{[p+j-1]}} := \begin{bmatrix} \underset{m\times pm}{\boldsymbol{\Gamma}_{(p+j-1)}} \\ \underset{m\times pm}{\boldsymbol{\Gamma}_{(p+j-2)}} \\ \vdots \\ \underset{m\times pm}{\boldsymbol{\Gamma}_{(j)}} \end{bmatrix} = \begin{bmatrix} \underset{m\times m}{\boldsymbol{\Gamma}_{p+j-1}} & \underset{m\times m}{\boldsymbol{\Gamma}_{p+j}} & \cdots & \underset{m\times m}{\boldsymbol{\Gamma}_{2p+j-2}} \\ \underset{m\times m}{\boldsymbol{\Gamma}_{p+j-2}} & \underset{m\times m}{\boldsymbol{\Gamma}_{p+j-1}} & \cdots & \underset{m\times m}{\boldsymbol{\Gamma}_{2p+j-3}} \\ \vdots & \vdots & & \vdots \\ \underset{m\times m}{\boldsymbol{\Gamma}_{j}} & \underset{m\times m}{\boldsymbol{\Gamma}_{j+1}} & \cdots & \underset{m\times m}{\boldsymbol{\Gamma}_{p+j-1}} \end{bmatrix}.$$

and p is the MSV order with $\boldsymbol{\Gamma}_k = \mathrm{Cov}\big((\boldsymbol{X}_t - \mu)(\boldsymbol{X}_{t-k} - \mu)'\big)$.

Given the data, it is natural to estimate $\boldsymbol{\Gamma}_k$, and μ by the corresponding empirical moments:

$$\hat{\boldsymbol{\Gamma}}_k = \frac{1}{T-k}\sum_{t=1}^{T-k}[(\boldsymbol{X}_{t+k} - \hat{\mu})(\boldsymbol{X}_t - \hat{\mu})'], \quad \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t \tag{3.4}$$

Setting $j = 1$ in (3.3) and replacing theoretical moments by the corresponding empirical moments yields the following closed-form VARMA estimator of MSV(p) coefficients:

$$\hat{\boldsymbol{\phi}}_{(p)} = \hat{\boldsymbol{\Gamma}}_{(p+1)}\hat{\boldsymbol{\Gamma}}_{[p]}^{-1}, \tag{3.5}$$

$$\hat{\boldsymbol{\Sigma}}_\epsilon = \hat{\boldsymbol{\Gamma}}_0 + \hat{\boldsymbol{\phi}}_1^{-1}\sum_{j=1}^{p-1}\hat{\boldsymbol{\phi}}_{j+1}\hat{\boldsymbol{\Gamma}}_j' - \hat{\boldsymbol{\phi}}_1^{-1}\hat{\boldsymbol{\Gamma}}_1, \quad \hat{\boldsymbol{\Sigma}}_v = \hat{\boldsymbol{\phi}}_1^{-1}\hat{\boldsymbol{\Gamma}}_1 - \hat{\boldsymbol{\phi}}_1^{-1}\sum_{j=1}^{p-1}\hat{\boldsymbol{\phi}}_{j+1}\hat{\boldsymbol{\Gamma}}_j' - \sum_{j=1}^{p}\hat{\boldsymbol{\phi}}_j\hat{\boldsymbol{\Gamma}}_j{}', \quad \hat{\mu} = \mathbb{E}(\boldsymbol{X}_t). \tag{3.6}$$

Note that, given $\hat{\boldsymbol{\Sigma}}_\epsilon$, we compute cross dependence parameters using the method proposed by Harvey et al. (1994).

In small samples, the estimate of $\hat{\boldsymbol{\phi}}_{(p)}$ may yield a solution outside the admissible area, *i.e.*, some of the eigenvalues of the latent volatility process [it is a VAR(p) process] may lie outside the unit circle or equal to unity. This issue can arise especially in small samples or in the presence of outliers. Further, the values $\hat{\boldsymbol{\phi}}_{(p)}$ computed by choosing different sets of Yule-Walker equations will in general be different. When this happens, a simple fix is projecting the estimate on the space of acceptable parameter solutions by altering the eigenvalues that lie on or outside the unit circle. Further, the estimate $\hat{\boldsymbol{\Sigma}}_v$ may yield a non-positive definite matrix. In this case, we can apply modified Cholesky algorithm of Cheng and Higham (1998) to $\hat{\boldsymbol{\Sigma}}_v$ and obtain a positive definite matrix.

3.2 VARMA-based winsorized estimation

We can achieve better stability and efficiency of VARMA estimator by using "winsorization" which exploits (3.3). Winsorization (censoring) substantially increases the probability of getting admissible values. From (3.3), it is easy to see that:

$$\boldsymbol{\phi}_{(p)} = \sum_{j=1}^{\infty}\omega_j\boldsymbol{\Gamma}_{(p+j)}\boldsymbol{\Gamma}_{[p+j-1]}^{-1} \tag{3.7}$$

for any ω_j sequence of scalers with $\sum_{j=1}^{\infty}\omega_j = 1$ (convex combination). Using (3.7), we can define a more general class of estimators for $\boldsymbol{\phi}_{(p)}$ by taking a weighted average of several sample analogs of the ratio $\boldsymbol{\Gamma}_{(p+j)}\boldsymbol{\Gamma}^{-1}_{[p+j-1]}$:

$$\tilde{\phi}_{(p)} = \sum_{j=1}^{J}\omega_j\hat{\boldsymbol{\Gamma}}_{(p+j)}\hat{\boldsymbol{\Gamma}}^{-1}_{[p+j-1]}, \tag{3.8}$$

where $1 \le J \le T-p$ with $\sum_{j=1}^{J}\omega_j = 1$ and T is the length of time series. We can expect that a sufficiently general class of weights may improve the efficiency of the VARMA estimators.

Using (3.8), we can propose alternative estimators of $\boldsymbol{\phi}_{(p)}$ and we call these estimators *winsorized VARMA* estimators (or *W-VARMA* estimators). Other (possibly nonlinear) averaging methods, such as the median, may also be used.

In the simulation section, we consider two types of winsorized estimators based on the expression given by (3.8):

- An arithmetic mean of sample ratios (equal weights), denoted by $\hat{\phi}^{AM}_{(p)}$, where we set

$$\omega_j = 1/J, \quad j = 1, \ldots, J, \tag{3.9}$$

- A *matrix-variate regression* (MVR) using the following system of equations with equal weights:

$$\hat{\boldsymbol{\Gamma}}_{(p+j)} = \hat{\boldsymbol{\phi}}_{(p)} \cdot \hat{\boldsymbol{\Gamma}}_{[p+j-1]}, \quad j = 1, \cdots, J. \tag{3.10}$$

  The solution of the above system is given by

$$\hat{\phi}^{MVR}_{(p,J)} = \hat{\boldsymbol{\Gamma}}_{(p+1,\mathbb{J})}\hat{\boldsymbol{\Gamma}}'_{[p,\mathbb{J}]}\left(\hat{\boldsymbol{\Gamma}}_{[p,\mathbb{J}]}\hat{\boldsymbol{\Gamma}}'_{[p,\mathbb{J}]}\right)^{-1}, \tag{3.11}$$

  where

$$\underset{m\times pmJ}{\hat{\boldsymbol{\Gamma}}_{(p+1,\mathbb{J})}} := \left[\begin{array}{cccc} \underset{m\times pm}{\hat{\boldsymbol{\Gamma}}_{(p+1)}} & \underset{m\times pm}{\hat{\boldsymbol{\Gamma}}_{(p+2)}} & \cdots & \underset{m\times pm}{\hat{\boldsymbol{\Gamma}}_{(p+J)}} \end{array}\right], \quad \underset{pm\times pmJ}{\hat{\boldsymbol{\Gamma}}_{[p,\mathbb{J}]}} := \left[\begin{array}{cccc} \underset{pm\times pm}{\hat{\boldsymbol{\Gamma}}_{[p]}} & \underset{pm\times pm}{\hat{\boldsymbol{\Gamma}}_{[p+1]}} & \cdots & \underset{pm\times pm}{\hat{\boldsymbol{\Gamma}}_{[p+J-1]}} \end{array}\right],$$

  and J is the winsorize truncation parameter.

These type of estimators are also considered by Ahsan and Dufour (2021) in the context of univariate estimation of SV models. All these estimators depend on J and for $J = 1$, they are equivalent to the simple VARMA estimator which is given in (3.5).

4 Asymptotic distributional theory

We will now study the asymptotic distribution of the simple VARMA estimator under the following set of assumptions.

Assumption 4.1. Distribution of the error processes. *The error processes u_t and v_t are mutually independent and $\{u_t\}$ is a sequence of i.i.d. real-valued random variables, independent of h_0. The probability distribution of u_t has a continuous density with respect to Lebesgue measure on the real line, and its density is positive on $(-\infty, +\infty)$. The transformed error ϵ_t satisfies $\mathbb{E}(|\epsilon_t|^s) < \infty$, where s is an integer such that $s \geq 1$.*

Assumption 4.2. Stationarity and mixing. *The process $\boldsymbol{y}_t$ is ergodic, β-mixing and strictly stationary.*

Assumption 4.3. Existence of moments and positive definiteness.

A. $\boldsymbol{V}_t$ is positive definite almost surely for each t.

B. The s-th moments of the observed process $\boldsymbol{y}_t$ exist and are finite, i.e., $\mathbb{E}|\boldsymbol{y}_t|^s < \infty$, where s is a positive integer.

Under Assumptions 4.1 – 4.3 with $s = 4$, the process $\{\boldsymbol{X}_t\}$ is strictly stationarity and geometrically ergodic with exponential β-mixing with finite second moments, *i.e.*, $\mathbb{E}\left[(\boldsymbol{X}_t)^2\right] < \infty$. In the following lemma, using ergodicity, we prove the consistency of the empirical moments in (3.4).

Lemma 4.1. Consistency of empirical moments. *Under the Assumptions 4.1 – 4.3 with $s = 4$ and for any $p \geq 0$, the estimators of*

$$\Lambda(p) := \left[\mu,\ vec\boldsymbol{\Gamma}_0,\ vec\ \boldsymbol{\Gamma}_1,\ \ldots,\ vec\ \boldsymbol{\Gamma}_{p+1}\right]',$$

defined by (3.4) satisfy:

$$\hat{\Lambda}(p) \xrightarrow{p} \Lambda(p). \tag{4.1}$$

Assumptions 4.1 – 4.3 with $s = 8$ are sufficient for the MSV model to have a strictly stationary solution with a finite fourth moment of $\boldsymbol{X}_t$, *i.e.*, $\mathbb{E}\left[\boldsymbol{X}_t^4\right] < \infty$. Note that the fourth moment of $\boldsymbol{X}_t$ translates into the eighth moment of $\boldsymbol{y}_t$. This solution will be β-mixing with geometrically decreasing mixing coefficients. In the following lemma, using a Central Limit Theorem for stationary ergodic processes (Lindeberg-Levy theorem for dependent processes), we give the asymptotic distribution of the empirical moments in (3.4).

Lemma 4.2. Asymptotic distribution of empirical moments. *Under Assumptions 4.1 – 4.3 with $s = 8$, the estimators $\hat{\Lambda}(p)$ defined by (3.4) satisfy:*

$$\sqrt{T}\left(\hat{\Lambda}(p) - \Lambda(p)\right) \xrightarrow{d} N(0, \mathbb{V}), \tag{4.2}$$

where $\mathbb{V} = Cov[\Lambda(p)]$.

This in turn yields the asymptotic distribution of the simple VARMA-type estimator.

Theorem 4.3. Asymptotic distribution of simple VARMA estimator. *Under Assumptions 4.1 – 4.3 with* $s = 4$, *the estimator* $\hat{\theta} := (\hat{\mu}, vec(\hat{\boldsymbol{\phi}}), vec(\hat{\boldsymbol{\Sigma}}_v), vec(\hat{\boldsymbol{\Sigma}}_\epsilon)) = \mathbf{F}(\Lambda)$ *is consistent, i.e.* $\hat{\theta} \xrightarrow{p} \theta$.

In addition, if Assumptions 4.1 – 4.3 with $s = 8$ *hold, we have*

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathbb{V}_\theta),$$

with $\mathbb{V}_\theta = \mathbf{F}_1(\Lambda)\mathbb{V}\mathbf{F}_1(\Lambda)'$, *where* $\mathbf{F}_1(\Lambda) = \partial\mathbf{F}(\Lambda)/\partial\Lambda$.

Theorem 4.3 covers the simplest VARMA estimator. The asymptotic distribution of more general winsorized estimators can be derived in the same way upon using Lemmas 4.1 – 4.2.

5 Simulation-based Inference

In this section, we discuss simulation-based inference procedures for MSV models. The simulation-based methods are more attainable in the context of this study for two reasons: (1) the MSV model is a parametric model, and we can easily simulate this model; (2) we can simulate the test statistic of MSV parameters that based on a computationally inexpensive estimator. However, if the estimator is computationally expensive, then we cannot simulate the test statistic easily, and the simulation will run forever. Using our proposed computationally simple estimator, one can construct more reliable finite sample inference.

5.1 Monte Carlo tests

We now examine the usefulness of our simple estimator in the context of simulation-based inference, i.e., Monte Carlo tests. The technique of Monte Carlo tests was originally been proposed by Dwass (1957) for implementing permutation tests and did not involve nuisance parameters. This technique was also independently proposed by Barnard (1963); for a review, see Dufour and Khalaf (2001) and for a general discussion and proofs, see Dufour (2006). It has the great attraction of providing exact (randomized) tests based on any statistic whose finite-sample distribution may be intractable but can be simulated. One can replace the unknown or intractable theoretical distribution $F(S|\theta)$ by its sample analogue based on the statistics $S_1(\theta), \ldots, S_N(\theta)$ simulated under the null hypothesis.

Let us first consider the pivotal statistics case, i.e. the case where the distribution of the test statistic under the null hypothesis does not depend on nuisance parameters. We can then proceed as follows to obtain an exact critical region.

1. Let S_0 be the observed test statistic (based on data).

2. By Monte Carlo methods, draw N $i.i.d.$ replications of S, denoted by $S(N) = (S_1, \ldots, S_N)$ under H_0, independently of S_0, i.e., $S_0, S_1, \ldots, S_N$ be exchangeable.

3. From the simulated samples compute the MC p-value $\hat{p}_N[S] \equiv p_N[S_0; S(N)]$, where

$$p_N[x, S(N)] \equiv \frac{NG_N[x; S(N)] + 1}{N+1} \tag{5.1}$$

$$G_N[x; S(N)] \equiv \frac{1}{N}\sum_{i=1}^{N} I_{[0,\infty)}(S_i - x), \qquad I_{[0,\infty)}(x) = \begin{cases} 1 & \text{if } x \in [0,\infty), \\ 0 & \text{if } x \notin [0,\infty). \end{cases} \tag{5.2}$$

In other words, $p_N[S_0; S(N)] = (NG_N[S_0; S(N)] + 1)/(N+1)$ where $NG_N[S_0; S(N)]$ is the number of simulated values which are greater than or equal to S_0 . When $S_0, S_1, \ldots, S_N$ are all distinct [an event with probability one when the vector $(S_0, S_1, \ldots, S_N)'$ has an absolutely continuous distribution], $\hat{R}_N(S_0) = N + 1 - NG_N[S_0; S(N)]$ is the rank of S_0 in the series $S_0, S_1, \ldots, S_N$.

4. The MC critical region is: $\hat{p}_N[S] \leq \alpha$, $0 < \alpha < 1$. If $\alpha(N+1)$ is an integer and the distribution of S is continuous under the null hypothesis, then under null,

$$P[\hat{p}_N[S] \leq \alpha] = \alpha; \tag{5.3}$$

see Dufour (2006).

We will now study the case where the distribution of the test statistic depends on nuisance parameters. In other words, we consider a model $\{(\Xi, \mathbb{A}_{\Xi}, P_\theta) : \theta \in \Omega\}$ where we assume that the distribution of S is determined by $P_{\bar{\theta}}$, where $\bar{\theta}$ represents the true parameter vector. To deal with this complication, the MC test procedure can be modified as follows.

1. To test the null hypothesis

$$H_0 : \bar{\theta} \in \Omega_0,$$

where $\Omega_0 \subset \Omega$, we calculate the relevant test statisticS_0 based on data.

2. For each $\theta \in \theta_0$, by Monte Carlo methods, we generate N i.i.d. replications of $S : S(N, \theta) = [(S_1(\theta), \ldots, S_N(\theta)]$.

3. Using these simulated test statistics, we compute the MC p-value $\hat{p}_N[S|\theta] \equiv p_N[S_0; S(N, \theta)]$, where

$$p_N[x; S(N,\theta)] \equiv \frac{NG_N[x; S(N,\theta)] + 1}{N+1}. \quad (5.4)$$

4. The p-value function $\hat{p}_N[S|\theta]$ as a function of θ is maximized over the parameter values compatible with the Ω_0, i.e., the null hypothesis, and H_0 is rejected if

$$\sup_{\theta \in \Omega_0} \hat{p}_N[S|\theta] \leq \alpha. \quad (5.5)$$

   If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:

$$P[\sup_{\theta \in \Omega_0} \{\hat{p}_N[S|\theta]\} \leq \alpha] \leq \alpha, \quad (5.6)$$

   The test defined by $\hat{p}_N[S|\theta] \leq \alpha$ has size α for known θ . Treating θ as a nuisance parameter and Ω_0 is a nuisance parameter set consistent with null, the test is *exact at level* α; for a proof, see Dufour (2006).

Because of the maximization in the critical region of (5.5), the test is called a *maximized Monte Carlo* (MMC) test. MMC tests provide valid inference under general regularity conditions such as almost-unidentified models or time series processes involving unit roots. In particular, even though the moment conditions defining the estimator are derived under the stationarity assumption, this does not question in any way the validity of maximized MC tests, unlike the parametric bootstrap whose distributional theory is based on strong regularity conditions. Only the power of MMC tests may be affected. However, the simulated p-value function is not continuous, so standard gradient based methods cannot be used to maximize it. But search methods applicable to non-differentiable functions are applicable, e.g. simulated annealing [see Goffe, Ferrier and Rogers (1994)].

A simplified approximate version of the MMC procedure can alleviate its computational load whenever a consistent point or set estimate of θ is available. To do this, we reformulate the setup in order to allow for an increasing sample size, i.e., now the test statistic depends on a sample of size T, $S = S_T$.

1. Let S_{T0} be the observed test statistic (based on data) and the distribution of S involves nuisance parameters under the null and $\bar{\theta} \in \Omega_0$ with $\Omega_0 \subset \Omega$ and $\Omega_0 \neq \emptyset$.

2. we have a consistent set estimator C_T of θ (under H_0) such that

$$\lim_{T\to\infty} P[\bar{\theta} \in C_T] = 1 \text{ under } H_0. \quad (5.7)$$

3. For each $\theta \in C_T$, by Monte Carlo methods, we generate N i.i.d. replications of $S : S_T(N,\theta) = [(S_{T1}(\theta),\ldots,S_{TN}(\theta)]$.

4. Using these simulations we compute the MC p-value $\hat{p}_{TN}[S_T|\theta] \equiv p_{TN}[S_{T0}; S_T(N,\theta)]$, where

$$p_{TN}[x; S_T(N,\theta)] \equiv \frac{NG_{TN}[x; S_T(N,\theta)] + 1}{N+1}. \tag{5.8}$$

5. The p-value function $\hat{p}_{TN}[S_T|\theta]$ as a function of θ is maximized with respect to θ in C_T, and H_0 is rejected if

$$\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha. \tag{5.9}$$

If the number of simulated statistics N is chosen so that $\alpha(N+1)$ is an integer, then we have under H_0:

$$\lim_{T\to\infty} P[\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha] \leq \alpha, \tag{5.10}$$

$i.e.$, we control for the level asymptotically.

In practice, it is easy to find a consistent set estimate of $\bar{\theta}$, whenever a *consistent* point estimate $\hat{\theta}_T$ of $\bar{\theta}$ available (e.g. a GMM estimator).

For instance, any set of the form

$$C_T = \{\theta \in : \left\|\hat{\theta}_T - \theta\right\| < d\} \tag{5.11}$$

with d a fixed positive constant independent of T, satisfies (5.7). The consistent set estimate MMC (CSEMMC) method is especially useful when the distribution of the test statistic is highly sensitive to nuisance parameters. Here, possible discontinuities in the asymptotic distribution are automatically overcome through a numerical maximization over a set that contains the true value of the nuisance parameter with probability one asymptotically (while there is no guarantee for the point estimate to converge sufficiently fast to overcome the discontinuity). It is worth noting that there is no need to maximize the p-value function with respect to unidentified parameters under the null hypothesis. Thus, parameters which are unidentified under the null hypothesis can be set to any fixed value and the maximization be performed only over the remaining identified nuisance parameters. When there are several nuisance parameters, one can use simulated annealing, an optimization algorithm which does not require differentiability. Indeed the simulated p-value function is not continuous, so standard gradient based methods cannot be used to maximize it. For an example where this is done on a VAR model involving a large number of nuisance parameters, see Dufour and Jouini (2006).

In Dufour and Khalaf (2002) call the test based on simulations using a point nuisance parameter estimate a *local* MC (LMC) test. The term local reflects the fact that the underlying MC p-value is based on a specific choice for the nuisance parameter. Here if the set C_T in (5.9) is reduced to a single point estimate $\hat{\theta}_T$, $i.e.$ $C_T = \{\hat{\theta}_T\}$, we get a LMC test

$$\hat{p}_{TN}[S_T|\hat{\theta}_T] \leq \alpha, \tag{5.12}$$

which can be interpreted as a parametric bootstrap test. Note that no asymptotic argument on the number N of MC replications is required to obtain this result; this is the fundamental difference between the latter procedure and the parametric bootstrap method.

Even if $\hat{\theta}_T$ is a consistent estimate of θ (under the null hypothesis), the condition (5.7) is not usually satisfied in this case, so additional assumptions are needed to show that the parametric bootstrap procedure yields an asymptotically valid test. It is computationally less costly but clearly less robust to violations of regularity conditions than the MMC procedure; for further discussion, see Dufour (2006). Furthermore, the LMC non-rejections are *exactly* conclusive in the following sense: if $\hat{p}_N[S|\hat{\theta}_0] > \alpha$, then the exact *Maximized Monte Carlo* (MMC) test is clearly not significant at level α.

5.2 Implicit standard error

In this section, we propose the *implicit standard error* (ISE) of θ that can be derived from simulation-based confidence interval. The asymptotic standard error can be markedly different and may be quite unreliable in finite samples. To construct more reliable standard error, we derive the ISE in the following way:

1. Calculate $\hat{\theta}_0$ based on observed data ($\mathbb{Y}_0$).

2. By Monte Carlo methods, draw N $i.i.d.$ replications of $\mathbb{Y}$, denoted by $\mathbb{Y}(N) = (\mathbb{Y}_1, \ldots, \mathbb{Y}_N)$ under H_0, independently of $\mathbb{Y}_0$, i.e., $\mathbb{Y}_0, \mathbb{Y}_1, \ldots, \mathbb{Y}_N$ be exchangeable.

3. Calculate $\hat{\theta}(N) = \hat{\theta}_1, \ldots, \hat{\theta}_N$ from $\mathbb{Y}(N) = (\mathbb{Y}_1, \ldots, \mathbb{Y}_N)$.

4. The confidence interval $[C_{\alpha_L}, C_{\alpha_H}]$, with coverage $1-\alpha$ [$= \alpha_H - \alpha_L$] is constructed using the empirical α_L quantile and the empirical α_H quantile of $\hat{\theta}(0;N) = \hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_N$. Furthermore, if $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_N$ are ordered from smallest to largest such that $\hat{\theta}^*_1, \ldots, \hat{\theta}^*_{N+1}$, and $\alpha_L(N+1)$ and $\alpha_H(N+1)$ are integers, then the confidence interval is

$$\left[\hat{\theta}^*_{\alpha_L(N+1)},\ \hat{\theta}^*_{\alpha_L(N+1)}\right].$$

5. Since our simple estimator is normally distributed, one can easily compute the ISE in the following

way:

$$\left[ISE_L,\ ISE_H\right] = \left[\frac{\hat{\theta}_0 - C_{\alpha_L}}{Z_{\alpha/2}}, \frac{C_{\alpha_H} - \hat{\theta}_0}{Z_{\alpha/2}}\right].$$

6. Finally, a conservative ISE is

$$ISE = \min\{ISE_L,\ ISE_H\}.$$

6 Simulation study

This section examines the properties of the proposed estimators in terms of bias and root mean square error (RMSE) through simulation.

6.1 Comparison of different VARMA estimators

We first study the finite-sample properties of the winsorized VARMA estimators, which are discussed in Section 3.2. For this comparison, the true DGP is given by (2.1)-(2.2) with $m = 2$, thus we have 10 parameters in the system $(\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \mu_1, \mu_2, \sigma^2_{v,1}, \sigma^2_{v,2}, \sigma_{v,12}, \rho) = (0.95, -0.2, 0.1, 0.95, -2, 2, 1, 1, 0.9, 0.9)$. We generate 1000 replications and consider different sample sizes $T = (500, 1000, 2000)$. Both shrinkage estimators depend on the truncation parameter J, so we set $J = 10$. Table A1 reports simulation results and illustrates a number of findings:

First, the MVR estimator is highly robust and outperforms the other estimators (mean, CF, where CF stands for the uncensored VARMA estimator) in terms of bias and RMSE, across different sample sizes particularly in small samples.

Second, The CF estimator performs very poorly and produces a large number of inadmissible values (NIV) for the parameter estimates. Its inferior performance may be due to the high variability of estimated ACF. However, it should be emphasized that the CF estimator did not produce any unbound solutions in this parameter setting with larger samples. This fact implies that the variability of estimated ACF goes down as the sample size increases.

6.2 Comparison with the QML Method

We compare the proposed estimators to QML [Harvey et al. (1994)] where the true DGP is given by (2.1)-(2.2) with $m = 2$, $p = 1$, diagonal $\boldsymbol{\phi}_1$, and no constant term in the volatility equation. Thus we have 6 parameters in the system $(\phi_{11}, , \phi_{22}, \sigma^2_{v,1}, \sigma^2_{v,2}, \sigma_{v,12}, \rho) = (0.98, 0.98, 0.5, 0.5, 0.5, -0.8)$. We generate 100 replications and consider different sample sizes $T = (500, 2000, 5000)$. These restrictions are also considered by Harvey et al. (1994). Table A2 reports the results. From the results, we find that W-VARMA-MVR estimators uniformly outperform the QML method. The QML estimates of ρ is severely biased. In most cases , MVR with $J = 10$ has the superior performance.

6.3 Comparison with the Bayes Method

Globally, there is no uniform ranking between the different estimators, but the Bayesian estimator's performance remains superior among the competing methods in the context of MSV models. As a result, we compare our proposed VARMA estimators (CF, MVR) to the Bayes estimator [Kim et al. (1998), Omori et al. (2007), Kastner and Frühwirth-Schnatter (2014), Kastner (2019)].

6.3.1 Simulated DGPs

For this comparison, the true DGP is given by (2.1)-(2.2) with $m = (2, 5, 10)$, $p = 1$. We set $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = 0.4\boldsymbol{I}_m$, $\mu = -0.5\mathbb{1}$, and stationary $\boldsymbol{\phi}_1$ has the following formations:

1. For $m = 2$, $\boldsymbol{\phi}_1$ is non-diagonal with $(\phi_{11},\ \phi_{12},\ \phi_{21},\ \phi_{22}) = (0.98,\ -0.1,\ 0.12,\ 0.99)$.

2. For $m = 5$, $\boldsymbol{\phi}_1$ is non-diagonal where diagonal elements are $(0.98, 0.985, 0.8, 0.985, 0.98)'$ and upper off-diagonal elements are set to -0.15 and lower off-diagonal elements are set to 0.1.

3. For $m = 10$, $\boldsymbol{\phi}_1$ is non-diagonal where diagonal elements are (0.95, 0.98, 0.98, 0.99, 0.8, 0.8, 0.99, 0.98, $0.98, 0.95)'$ and upper off-diagonal elements are set to -0.11 and lower off-diagonal elements are set to 0.055.

The number of parameters (NP) in these models are $m^2 + 2m$. We simulate 100 samples from each simulated model. The parameters in the system above are assumed to be unknown and need to be estimated.

Following a Bayesian perspective, we assume that the parameters are not completely unknown, but they follow some prior distributions. Then, using the prior distributions and the information provided by the data, we can make inferences about the parameters from their posterior distributions.

6.3.2 Priors and MCMC algorithm

We specify prior distributions for the set of parameters $\theta = (\mu,\ \boldsymbol{\phi},\ \boldsymbol{\Sigma}_v)$ of the MSV model. The mean vector μ follows a flat multivariate normal distribution $\mu \sim \mathcal{MVN}(\boldsymbol{0},\ 1000\boldsymbol{I}_m)$. Each entry of the persistence matrix $\phi_{ij} \in (-1, 1)$ is assumed to follow a beta distribution, $(\phi_{ij} + 1)/2 \sim Beta(20, 1.5)$ [see Kim et al. (1998)]. The beta prior distribution ensures that the entries of the persistent matrix are between -1 and 1, which guarantees the stationarity of the volatility process. For the scale parameter of volatility innovation, we utilize a flat gamma prior, $\sigma_{i,v} \sim \Gamma(1/2, 1/2 \times 10)$ [see Kastner and Frühwirth-Schnatter (2014)]. We choose hyperparameters which provide relatively flat prior distributions. In addition, the choice of the hyperparameters has little effect on the posterior estimates of the parameters in large samples.

We estimated the latent volatility processes and the parameters of the MSV model using a Metropolis-within-Gibbs sampler [Kim et al. (1998), Omori et al. (2007), Kastner and Frühwirth-Schnatter (2014)]. We derive an MCMC algorithm, which consists of Metropolis-Hastings steps within a Gibbs sampler, for estimating the latent volatility process and its parameters. As before, let $\boldsymbol{y}_t$ denote the latent multivariate volatility time-series and $\theta = (\mu,\ \boldsymbol{\phi},\ \boldsymbol{\Sigma}_v)$ the parameters of the MSV model. Following Kim et al. (1998) and Omori et al. (2007), the log-squared transformation of (2.1) is given by:

$$\log(y_{i,t}^2) = h_{i,t} + \log(u_{i,t}^2), \tag{6.1}$$

Note that (6.1) is linear but non-Gaussian. To ameliorate the non-Gaussianity problem, we approximate the log-transformed error term $\log(u_{i,t}^2) \sim \log(\chi_1^2)$ by a mixture of 10 normal distributions as in Omori et al. (2007):

$$\log(\chi_1^2) \sim \sum_{k=1}^{10} p_k \mathcal{N}(m_k,\ v_k^2)\ .$$

The values of p_k, m_k and v_k are tabulated in Omori et al. (2007). As a result, we introduce a latent mixture component indicator variable, $d_{i,t}$, for $y_{i,t}$ at time t such that $\log(u_{i,t}^2)|(d_{i,t} = k) \sim N(m_k,\ v_k^2)$. The indicator variable is also estimated in the MCMC sampler. Given the mixture indicator $\boldsymbol{d}_t$ and the vector parameter θ, the latent volatility series $\boldsymbol{h}_t$ can be sampled using a forward-filtering and backward-sampling (FFBS) procedure, Carter and Kohn (1994), Frühwirth-Schnatter (1994). The mixture indicator $\boldsymbol{d}_t$ can be sampled from a multinomial distribution

$$p(d_{i,t} = k|\boldsymbol{h}_t,\ \theta) \propto p(d_{i,t} = k)\frac{1}{v_k}\exp\left\{-\frac{(\tilde{y}_{i,t} - h_{i,t} - m_k)^2}{2v_k^2}\right\}, \tag{6.2}$$

where $\tilde{y}_{i,t}$ is defined to be $\log(y_{i,t}^2 + c)$ with a fixed offset constant $c = 10^{-4}$ to avoid values equal to 0.

Finally, the vector parameter θ can be sampled using an ancillarity-sufficiency interweaving strategy (ASIS) Yu and Meng (2011); Kastner and Frühwirth-Schnatter (2014) which involves sampling the vector parameter θ given the unstandardized volatility series $h_{i,t}$ via a Metropolis-Hasting step (non-centered step) and then sampling θ again given the standardized volatility series $\tilde{h}_{j,t} = \frac{h_{i,t}-\mu_i}{\sigma_{i,v}}$ (centered step). Yu and Meng (2011) argued that by alternating between the non-centered and centered steps, we obtain a more efficient MCMC sampler that has a better mixing rate and converges faster. In addition, Kastner and Frühwirth-Schnatter (2014) showed that the ASIS can accurately sample latent volatility time-series that have low persistences.

We apply the above Bayesian method to the simulated datasets and draw 40,000 observations from the posterior distributions. The Bayes estimator is approximated by the average of the last 20,000 draws. For maximal computational effectiveness, all sampling algorithms are implemented in the compiled language C++ with the help of the R package Rcpp; see Eddelbuettel and François (2011). Matrix computations make use of the efficient C++ template library Armadillo [Sanderson and Curtin (2016)] through the R package RcppArmadillo [Eddelbuettel and Sanderson (2014)].

6.3.3 Simulation results: Bayes vs W-VARMA-MVR

Table A3 reports bias and RMSE for each parameter separately when $m = 2$ and Table A4 reports the average bias and RMSE for ϕ, diag[ϕ], μ, $\boldsymbol{\Sigma}_v$ when $m = (2, 5, 10)$. The number of inadmissible values of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100. Several conclusions emerge from these tables:

First, MVR estimators outperform the Bayes estimator in terms of bias and RMSE in all setups. From Table A3, for each parameter, MVR estimators produce either the smallest or second smallest bias and RMSE. However, the CF estimator gives several inadmissible values of ϕ's in small samples and NPDs, whereas MVR estimators yield acceptable parameter solutions in almost every estimation. When $m = 2$, the MVR estimator produces 0 NIV and 2 NPDs out of 1,500 estimations across samples. These results show that MVR estimator not only improves stability but also increases efficiency. We also find that Bayes estimates are heavily biased in all setups.

Second, for the larger samples ($T = 2000, 5000$) and higher dimensions ($m = 5, 10$), we have almost identical results for all the parameter estimates as with $T = 1000$ and $m = 2$. Again MVR estimators outperform the Bayes in terms of bias and RMSE. RMSEs of MVR estimates decrease as the sample size increases that show the consistency of these estimators.

Third, the simulation results show that the Bayes method is extremely unreliable when the volatility process is close to the unit root. Whereas, on the contrary, the MVR method produces accurate parameter estimates. We find that the Bayes estimator leads to non-convergence and yields a substantial bias for μ and $\boldsymbol{\Sigma}_v$ parameter estimates, indicating this sampling scheme's vulnerability. Note that we use the best available prior and sampling scheme for the Bayes method with maximal computational efficiency. Another prior and/or sampler may not provide better performance in terms of bias and RMSE. Further, the choice of prior for a higher-order model ($p > 1$) gets complicated as the order of model changes, which is a challenging task and requires an extensive Monte Carlo study. Note that the computational cost increases as well with p.

Finally, the simple estimators are highly time-efficient, and the margin of time efficiency is enormous compared to the Bayes method.

6.4 High-dimensional simulation study

Now we study the finite-sample properties of the W-VARMA-MVR ($J = 10, 100$) estimators in high-dimensional (HD) setups. For this comparison, the true DGP is given by (2.1)-(2.2) with $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = \boldsymbol{I}_m$, $\mu = -\mathbb{1}$, and stationary $\boldsymbol{\phi}$'s are both diagonal and non-diagonal.

Four DGPs are considered based on volatility persistences:

1. HD_1: We consider first-order persistence in latent volatilities and stationary $\boldsymbol{\phi}_1$ is diagonal where the diagonal elements are equal to 0.95. NP $= 3m$ is the number of parameters in the simulated model.

2. HD_2: We consider first-order persistence in latent volatilities and stationary $\boldsymbol{\phi}_1$ is non-diagonal where for $\boldsymbol{\phi}_1$, diagonal elements are equal to 0.9 and off-diagonal elements are set between $(-0.01, 0.01)$. NP $= m^2 + 2m$ is the the number of parameters in the simulated model.

3. HD_3: We consider second-order persistence in latent volatilities and stationary $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are non-diagonal where for $\boldsymbol{\phi}_1$, diagonal elements are equal to 0.9 and off-diagonal elements are set between $(-0.01, 0.01)$, and for $\boldsymbol{\phi}_2$, diagonal elements are equal to -0.85 and off-diagonal elements are set between $(-0.01, 0.01)$. NP $= 2m^2 + 2m$ is the the number of parameters in the simulated model.

4. Ultra high-dimensional setup (UHD): We consider $P = 2$, and stationary $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are non-diagonal where for $\boldsymbol{\phi}_1$, diagonal elements are equal to 0.9 and off-diagonal elements are set between $(-0.001, 0.001)$, and for $\boldsymbol{\phi}_2$, diagonal elements are equal to -0.8 and off-diagonal elements are set between $(-0.001, 0.001)$. NP $= 2m^2 + 2m$ is the the number of parameters in the simulated model.

We set $m \in (2:250)$ for HD_1-HD_3 models and $m \in (500:1500)$ for the UHD model. We simulate 100 samples from each simulated model with different sample sizes. Tables A5-A8 report the multivariate bias and RMSE of the corresponding matrices $(\boldsymbol{\phi}, \mu, \boldsymbol{\Sigma}_v)$. Several conclusions emerge from these tables:

First, MVR estimators perform extremely well in large scale setups and produce accurate estimates. From Table A8, we can see that the MVR method is extremely reliable even with a model that has 1,500 dimensions and 4,503,000 parameters. This holds even with a moderate sample size $T = 10,000$.

Second, we find that the winsorize parameter J plays several key roles for high-dimensional models:

- A higher level of J (*e.g.*, 100) reduces the number of non-positive definiteness of $\boldsymbol{\Sigma}_v$ matrix.

- A higher level of J reduces the RMSE, especially in small samples; compare column of $J = 10$ to $J = 100$ in Tables A5-A7.

- For high-dimensional setups, when $m = 100, 250$, the RMSEs (biases) go down significantly, implies that the MVR method improves the precession of all parameters.

Third, in small samples, RMSEs of high-dimensional models ($m = 25, 50, 100, 250$) are slightly larger compared to low-dimensional models ($m = 2, 5, 10$). This dispersion goes down as the sample size increases. We find that $T = 1000$ is not an adequate sample size for high-dimensional models. The same conclusion also holds for the UHD model.

Fourth, we find that the W-VARMA-MVR estimator is not only statistically efficient but also extremely time-efficient. For a large sample $T = 10{,}000$, MVR ($J = 10$) can compute a model with 250 dimensions and 125,500 parameters in just 2.717 seconds on a 2.8 GHz Intel Core i7 processor (with 16 GB 2133 MHz DDR3 memory); see Table A7.

7 Forecasting with MSV(p) models

As discussed earlier, MSV(p) models can be written as a linear state-space model without losing any information. The state-space representation of MSV(p) models is given by

$$\begin{aligned} \boldsymbol{X}_t - \mu &= \boldsymbol{h}_t - \mu + \epsilon_t, \\ \boldsymbol{h}_t - \mu &= \sum_{j=1}^{p} \boldsymbol{\phi}_j (\boldsymbol{h}_{t-j} - \mu) + \boldsymbol{v}_t, \end{aligned} \tag{7.1}$$

where the distribution ϵ_t, is approximated by a m-variate normal distribution with mean zero and covariance matrix Σ_ϵ. The model defined in (7.1) can be rewritten as following:

$$\begin{aligned} \tilde{\boldsymbol{X}}_t &= C\tilde{\boldsymbol{h}}_t + \epsilon_t, \\ \tilde{\boldsymbol{h}}_{t+1} &= A\tilde{\boldsymbol{h}}_t + \tilde{\boldsymbol{v}}_{t+1}, \end{aligned} \tag{7.2}$$

where

$$\tilde{\boldsymbol{X}}_t := \boldsymbol{X}_t - \mu$$

$$\underset{pm\times 1}{\tilde{\boldsymbol{h}}_{t+1}} := \begin{pmatrix} \underset{m\times 1}{\boldsymbol{h}_{t+1} - \mu} \\ \underset{m\times 1}{\boldsymbol{h}_t - \mu} \\ \vdots \\ \underset{m\times 1}{\boldsymbol{h}_{t-p+2} - \mu} \end{pmatrix}, \quad \underset{pm\times m}{\tilde{\boldsymbol{v}}_{t+1}} := \begin{pmatrix} \underset{m\times m}{\boldsymbol{v}_{t+1}} \\ \underset{m\times m}{\boldsymbol{0}} \\ \underset{m\times m}{\boldsymbol{0}} \\ \vdots \\ \underset{m\times m}{\boldsymbol{0}} \end{pmatrix}, \quad \underset{pm\times pm}{A} = \begin{pmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \dots & \boldsymbol{\phi}_{p-1} & \boldsymbol{\phi}_p \\ \boldsymbol{I}_m & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m & \dots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{I}_m & \boldsymbol{0} \end{pmatrix},$$

$$\underset{pm\times pm}{B} := \mathbb{E}(\tilde{\boldsymbol{v}}_{t+1}\tilde{\boldsymbol{v}}'_{t+1}) = \begin{pmatrix} \Sigma_v & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \underset{m\times pm}{C} := \begin{pmatrix} \boldsymbol{I}_m & \mathbf{0}_m & \dots & \mathbf{0}_m \end{pmatrix}, \quad \underset{m\times m}{D} := \mathbb{E}(\epsilon_t\epsilon'_t) = \Sigma_\epsilon$$

Now using (7.2), the Kalman filter can be applied as follows:

- Initialization:

$$\begin{aligned} \hat{\tilde{\boldsymbol{h}}}_{1|0} &= \mathbb{E}(\tilde{\boldsymbol{h}}_1) = \mathbf{0}_{(pm\times 1)}, \\ \mathbf{P}_{1|0} &= \mathbb{E}([(\tilde{\boldsymbol{h}}_1 - \mathbb{E}(\tilde{\boldsymbol{h}}_1)][(\tilde{\boldsymbol{h}}_1 - \mathbb{E}(\tilde{\boldsymbol{h}}_1)]') = \mathrm{diag}[\Sigma_v, \dots, \Sigma_v]_{(pm\times pm)}, \end{aligned} \tag{7.3}$$

  where $\mathbf{P}_{1|0}$ is the MSE associated with $\hat{\tilde{\boldsymbol{h}}}_{1|0}$.

- Sequential updating:

$$\begin{aligned} \hat{\tilde{\boldsymbol{h}}}_{t|t} &= \hat{\tilde{\boldsymbol{h}}}_{t|t-1} + \mathbf{P}_{t|t-1}C'(C\mathbf{P}_{t|t-1}C' + D)^{-1} \times (\tilde{\boldsymbol{X}}_t - C\hat{\tilde{\boldsymbol{h}}}_{t|t-1}), \\ \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}C'(C\mathbf{P}_{t|t-1}C' + D)^{-1}C\mathbf{P}_{t|t-1}. \end{aligned} \tag{7.4}$$

- In-sample prediction:

$$\begin{aligned} \hat{\tilde{\boldsymbol{h}}}_{t+1|t} &= A\hat{\tilde{\boldsymbol{h}}}_{t|t-1} + A\mathbf{P}_{t|t-1}C'(C\mathbf{P}_{t|t-1}C' + D)^{-1} \times (\tilde{\boldsymbol{X}}_t - C\hat{\tilde{\boldsymbol{h}}}_{t|t-1}), \\ \mathbf{P}_{t+1|t} &= A\mathbf{P}_{t|t}A' + B. \end{aligned} \tag{7.5}$$

  Given (7.5), the forecast of $\tilde{\boldsymbol{X}}_{t+1}$ and the MSE of forecast error are given by

$$\begin{aligned} \hat{\tilde{\boldsymbol{X}}}_{t+1|t} &= C\hat{\tilde{\boldsymbol{h}}}_{t+1|t}, \\ \mathbb{E}([\tilde{\boldsymbol{X}}_{t+1} - \hat{\tilde{\boldsymbol{X}}}_{t+1|t}][\tilde{\boldsymbol{X}}_{t+1} - \hat{\tilde{\boldsymbol{X}}}_{t+1|t}]') &= C\mathbf{P}_{t+1|t}C' + D. \end{aligned} \tag{7.6}$$

- Out-of-sample s-step-ahead forecasting:

$$\begin{aligned} \hat{\tilde{\boldsymbol{h}}}_{t+s|t} &= A^s\hat{\tilde{\boldsymbol{h}}}_{t|t}, \\ \hat{\tilde{\boldsymbol{X}}}_{t+s|t} &= C\hat{\tilde{\boldsymbol{h}}}_{t+s|t} = CA^s\hat{\tilde{\boldsymbol{h}}}_{t|t} = A^s\hat{\tilde{\boldsymbol{h}}}_{t|t}. \end{aligned} \tag{7.7}$$

The h-step-ahead forecast is computed by (7.7) with the parameter estimates plugged in.

8 Application to stock returns

In this section, we apply our models and methods to daily and high-frequency stock returns. First, we investigate the estimation of our proposed specification with high-frequency data. Second, using daily data, we investigate the effect of shrinkage control parameter. Third, we compare the forecasting performance of our specifications with alternative stochastic covariance models, including multivariate GARCH models.

8.1 Estimation with high-frequency data

We use the estimated five-minute high-frequency stock returns constructed using the price data from the TAQ millisecond database for the period 2004 to 2016. We construct the log prices for five-minute sampling, which gives us on average 250 days per year with 77 daily increments. The first observation is the volume-weighted trading price in the exact second of 9:30:00. For the remaining 78 observations, the volume-weighted trading prices are calculated for each second, and the last observations in each 300-second interval are taken. For a significant portion of the stocks, there is no trade in the first seconds of the day, and thus we start sampling at 9:35 am. We use the price of the trade at or immediately preceding each five-minute mark. For each year, we take the intersection of the stocks traded each day and the stocks in the S&P 500 index over the period 1993 to 2012 based on the Bloomberg terminal. This gives us a cross-section of around 500-600 firms for each year. Based on stocks still present in S&P 500 index in the period 2020, we have an intersection of 228 stocks for all 13 years. The standard data-cleaning procedures are applied where we delete all entries with a time stamp outside 9:30 am to 4 pm and entries with a transaction price equal to zero, retain entries originating from a single exchange and delete entries with corrected trades and abnormal sale condition.

In our balanced panel, we have 228 stocks and 252,021 observations. Using this dataset, we estimate unrestricted MSV(1) and MSV(2) models. The W-VARMA-MVR($J = 10$) method can estimate an MSV(1) [MSV(2)] model with this dataset in just 19.22 [48.05] seconds on a 2.8 GHz Intel Core i7 processor with 16 GB 2133 MHz DDR3 memory. The MSV(1) model has 52,440 parameters, while MSV(2) model has 104,424 parameters.

Figures A1-A4 show the high-dimensional heatmap plots of the estimated volatility persistent matrix with different values of shrinkage control parameter. Several conclusions emerge from these heatmaps. *First*, when $J = 1$, we find that both $\hat{\boldsymbol{\phi}}_1$ and $\hat{\boldsymbol{\phi}}_2$ matrices are unstable, i.e., some of the eigenvalues of the companion matrix have a modulus greater or equal to one. However, with higher values of the shrinkage control parameter ($J = 10, 50, 100$), the W-VARMA-MVR produces stable solutions. *Second*, we find that many of the off-diagonal values of $\hat{\boldsymbol{\phi}}_1$ and $\hat{\boldsymbol{\phi}}_2$ are significantly different from zero. The non-zero off-diagonal elements of $\hat{\boldsymbol{\phi}}_1$ imply that there are volatility spillover effects (or multilateral Granger causalities in volatility) between assets at a higher frequency. In addition to that, the non-zero off-diagonal elements of $\hat{\boldsymbol{\phi}}_2$ show that we have multi-horizon volatility spillover effects (or multi-horizon causalities in volatility) between assets at a higher frequency. *Third*, the non-zero diagonal elements of $\hat{\boldsymbol{\phi}}_1$ and $\hat{\boldsymbol{\phi}}_2$ capture first-order and second-order persistence in log volatilities of returns. Finally, from Figures A2 and A4, it is easy to see that the above patterns hold true even in small samples when $T = 126,011$.

8.2 Effect of shrinkage control parameter

The choice of shrinkage control parameter J plays a crucial role in W-VARMA-MVR estimation; therefore, we scrutinize the possible choices of J with a small sample with many assets. The data set in this estimation contains daily observations from 2008/01/02 to 2020/12/31 of all stocks listed in NYSE. After filtering and balancing, the NYSE dataset includes 1,184 stocks with 3,247 time series observations.

For this sample, we consider MSV(p) models with $p = (1,2)$ and plot the estimated (by W-VARMA-MVR) modulus of the max eigenvalue of the companion matrix of the volatility persistence parameters estimated with different number of assets [$m = (2,10,50,100,250,500,1000,1184)$] and for $J = (1,\ldots,100)$.

We present plots in Figures A5-A6. From the evolution of the modulus of the maximum eigenvalue of the companion matrix as a function of J and m, we find that both MSV(p) models with $p = (1,2)$ provide stable solutions when $J > 5$. Furthermore, we get stable solutions with many assets, even with $m = 1184$. It should be noted that the truncation parameter J plays a smoothing role, and our results (simulation and empirical) suggest that there is a bias-variance trade-off for ϕ estimators as J increases. Therefore, a moderate level of winsorizing is sufficient.

8.3 Forecasting performance and comparison to other models

In this section, we evaluate the out-of-sample performance of MSV(p) models in the context of asset allocation strategies. The competitors are: the Risk Metrics 2006 methodology [Zumbach (2007)] denoted as RM, the dynamic conditional correlation with composite likelihood [Engle (2002), Pakel et al. (2021)] denoted as cDCC, the conditional covariance estimation using the dynamic factor model approach with finite-dimensional space [Alessi et al. (2009), Aramonte et al. (2013)] denoted as fDCC, the generalized orthogonal GARCH model of van der Weide (2002) denoted as GO-GARCH with three different specification [standard GARCH with multivariate normal distribution (GO-GARCH-S) and multivariate affine normal inverse Gaussian distribution (GO-GARCH-S-MANIG), and GJR with multivariate normal distribution (GO-GARCH-GJR)], unrestricted MSV(p) models (U-MSV(p)) and restricted MSV(p) models (R-MSV(p))

where we consider uncorrelated SV(p) structure. While this choice is not exhaustive, it includes many of the most widely used approaches in practice.

For comparison, we consider the global minimum variance (GMV) portfolio where the one step ahead conditional forecast of the covariance matrix $\Sigma_{t+1|t}$ uniquely defines the optimal portfolio weights:

$$\hat{\mathbf{w}}_{t+1} = \frac{\hat{\Sigma}^{-1}_{t+1|t}\iota_m}{\iota'_m \hat{\Sigma}^{-1}_{t+1|t}\iota_m},$$

where ι_m denotes an m-variate vector of ones. The key issue is to forecast $\Sigma_{t+1|t}$, and it is estimated at time t using each of the conditional volatility models described above.

We computed out-of-sample forecasts of $\Sigma_{t+1|t}$ using rolling (moving) window method and computed for a forecast horizon equal to 1-day. In this rolling forecasts setup, an initial sample using data from $t = 1,\ldots,T$ is used to determine a window width T, to estimate the models, and to form 1−step ahead out-of-sample forecasts starting at time T. Then the window is moved ahead one time period, the models are re-estimated using data from $t = 2,\ldots,T+1$, and 1−step ahead out-of-sample forecasts are produced starting at time $T+1$. This process is repeated until no more 1−step ahead forecasts can be computed.

Using the forecast of $\Sigma_{t+1|t}$, we compute GMV weights and using these weights, we compute the corresponding realized portfolio returns of a given portfolio p, denoted by $r_{p,t+1}$, is derived as $r_{p,t+1} = \hat{\mathbf{w}}'_{t+1}\mathbf{r}_{t+1}$, where $\mathbf{r}_{t+1} = (r_{1,t+1},\ r_{2,t+1},\ldots,\ r_{m,t+1})$ and $r_{i,t+1}$ gives the return of asset i between day t and day $t+1$.

The evaluation of the portfolio's performance based on the one-step-ahead excess portfolio returns over the equal weighted portfolio is denoted by $\hat{r}_{p,t+1}$.[3] Following Gasbarro et al. (2007), DeMiguel and Nogales (2009), DeMiguel et al. (2009), Zakamouline and Koekebakker (2009), Behr et al. (2013) and Hautsch and Voigt (2019), to measure the performance of the GMV asset-allocation strategy by adjusting it for its risk, we use the well-known Sharpe ratio ($\mathcal{SR}$) index defined as the ratio between the sample mean of the out-of-sample excess returns over those obtained from the equal weight portfolio and their sample standard deviation:

$$\mathcal{SR} = \frac{\frac{1}{T-1}\sum_{t=1}^{T-1}\hat{r}_{p,t+1}}{\sqrt{\frac{1}{T-2}\sum_{t=1}^{T-1}\left(\hat{r}_{p,t+1} - \frac{1}{T-1}\sum_{t=1}^{T-1}\hat{r}_{p,t+1}\right)^2}}. \tag{8.1}$$

A superior covariance forecasting model should provide portfolios with higher $\mathcal{SR}$.

The data set in the estimation contains daily observations from 2009/01/02 to 2020/12/31 of all stocks formed the S&P 500 index. Two portfolios are considered: one with 20 stocks and the other with 50 stocks; see the composition of these portfolios in Section A.2. We select stocks that have larger weights in the S&P index. The data are in a log difference form, adjusted for dividends, and obtained from CRSP through

[3] We compute the equal weighted portfolio with rebalancing to maintain the naive 1/m or strategic weight for each asset over time.

Wharton Research Data Services. In the out-of-sample forecast experiment, we consider an initial in-sample period from 2009/01/02 to 2018/12/31, and the out-of-sample is from 2019/01/02 to 2020/12/31. Thus, the out-of-sample includes a highly volatile period, *i.e.*, the Covid-19 pandemic. The forecasts of MSV(p) models are based on the W-VARMA-MVR estimator with $J = 50$.

Table A9 reports the annualized $\mathcal{SR}$ of excess minimum-variance portfolio returns, which have been computed over the out-of-sample period after daily rebalancing. Our findings can be summarized as follows.

First, the results reported in A9 show that the $\mathcal{SR}$ for the portfolios constructed using the U-MSV(p) estimates of the covariances consistently provides the highest $\mathcal{SR}$s, regardless of the number of assets and different sub-samples. These findings indicate that higher-order persistence in volatilities and higher-order volatility spillover effects are crucial for asset allocation; compare $\mathcal{SR}$s of U-MSV(p) and R-MSV(p) models. Most of the high-dimensional (or moderate-dimensional) conditional covariance matrix estimators proposed recently do not consider these features. Our findings suggest that the consequences of neglecting these features in the covariance estimation process may be detrimental.

Second, we find that the U-MSV(p) models provide larger $\mathcal{SR}$ than those provided by multivariate GARCH models. This suggests that modelling volatility as a latent stochastic VAR(p) process is vital for forecasting covariance matrix. In addition to that, we also find numerical instability with GO-GARCH models [GO-GARCH-S, GO-GARCH-GJR] when $m = 50$.

Third, in both of our out-of-sample experiments, higher-order U-MSV models also outperform the first-order U-MSV model. This finding suggests that additional lag terms in the latent volatility equation are essential for forecasting volatility. This result is consistent with findings of Ahsan and Dufour (2020).

Finally, when we forecast a highly volatile period, i.e., the Covid-19 pandemic, the performance of U-MSV(p) models are better than other competing models (this holds across a different number of assets). The U-MSV(3) model produces the superior forecast in terms of $\mathcal{SR}$ criteria in both experiments. On the other hand, performances of RM, GARCH and R-MSV models are poor, and these models failed to outperform the equal-weighted portfolio; $\mathcal{SR}$s of these models are negative.

9 Conclusion

We have examined the problem of estimating unrestricted higher-order MSV models and proposed several computationally simple estimators. We also study the asymptotic distribution of these simple estimators. We show that simple estimators are especially convenient for use in simulation-based inference techniques, *i.e.*, Bootstrap or Monte Carlo tests. The W-VARMA-MVR method that we have proposed outperforms all other estimators, including an extremely optimized Bayesian estimator, in terms of bias and statistical efficiency. This conclusion holds across different simulation designs. Furthermore, proposed simple estimators are highly time-efficient compared to other estimators. In the context of dynamic minimum variance portfolio strategy, we find that unrestricted higher-order MSV models outperform existing alternatives, including multivariate GARCH-type models.

One can exploit the W-VARMA-MVR in the context of measuring multivariate macroeconomic uncertainty. Other potential extensions of interest include: MSV models with non-Gaussian innovations, allowance for leverage effects, and incorporating factor structure in the mean equation. In cases of leverage and heavy-tailed distributions, we can develop estimation methods similar to Ahsan (2021). These are the objects of ongoing research.

References

Aguilar, O. and West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business and Economic Statistics* **18**(3), 338–357, doi:10.1080/07350015.2000.10524875.

Ahsan, M. N. (2021), Almost closed-form efficient estimators for stochastic volatility models with leverage and heavy-tailed distributions, Technical report, Concordia University.

Ahsan, M. N. and Dufour, J.-M. (2019), A simple efficient moment-based estimator for the stochastic volatility model, *in* I. Jeliazkov and J. Tobias, eds, 'Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling', Vol. 40 of *Advances in Econometrics*, Emerald, Bingley, U.K., pp. 157–201.

Ahsan, M. N. and Dufour, J.-M. (2020), Volatility forecasting and option pricing with higher-order stochastic volatility models, Technical report, McGill University.

Ahsan, M. N. and Dufour, J.-M. (2021), 'Simple estimators and inference for higher-order stochastic volatility models', *Journal of Econometrics* **224**(1), 181–197, doi:10.1016/j.jeconom.2021.03.008.

Alessi, L., Barigozzi, M. and Capasso, M. (2009), Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors, SSRN Scholarly Paper ID 1502684, Social Science Research Network, Rochester, NY.

Andersen, T. G. and Sørensen, B. E. (1996), 'GMM estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Business and Economic Statistics* **14**(3), 328–352, doi:10.2307/1392446.

Andersen, T. G., Chung, H.-J. and Sørensen, B. E. (1999), 'Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Econometrics* **91**, 61–87, doi:10.1016/S0304-4076(98)00049-9.

Aramonte, S., del Giudice Rodriguez, M. and Wu, J. (2013), 'Dynamic factor Value-at-Risk for large heteroskedastic portfolios', *Journal of Banking & Finance* **37**(11), 4299–4309, doi:10.1016/j.jbankfin.2013.07.038.

Asai, M. and McAleer, M. (2009), 'The structure of dynamic correlations in multivariate stochastic volatility models', *Journal of Econometrics* **150**(2), 182–192, doi:10.1016/j.jeconom.2008.12.012.

Barnard, G. A. (1963), 'Comment on 'The spectral analysis of point processes' by M. S. Bartlett', *Journal of the Royal Statistical Society, Series B* **25**, 294.

Behr, P., Guettler, A. and Miebs, F. (2013), 'On portfolio optimization: Imposing the right constraints', *Journal of Banking & Finance* **37**(4), 1232–1242, doi:10.1016/j.jbankfin.2012.11.020.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327, doi:10.1016/0304-4076(86)90063-1.

Breidt, F. J. and Carriquiry, A. L. (1996), Improved quasi-maximum likelihood estimation for stochastic volatility models, *in* W. O. Johnson, J. C. Lee and A. Zellner, eds, 'Modelling and Prediction: Honouring Seymour Geisel', Springer-Verlag, New York, pp. 228–247.

Broto, C. and Ruiz, E. (2004), 'Estimation methods for stochastic volatility models: a survey', *Journal of Economic Surveys* **18**(5), 613–649, doi:10.1111/j.1467-6419.2004.00232.x.

Carriero, A., Clark, T. E. and Marcellino, M. (2016), 'Common drifting volatility in large Bayesian VARs', *Journal of Business & Economic Statistics* **34**(3), 375–390.

Carriero, A., Clark, T. E. and Marcellino, M. (2018), 'Measuring uncertainty and its impact on the economy', *Review of Economics and Statistics* **100**(5), 799–815.

Carriero, A., Clark, T. E. and Marcellino, M. (2019), 'Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors', *Journal of Econometrics* **212**(1), 137–154.

Carter, C. K. and Kohn, R. (1994), 'On Gibbs sampling for state space models', *Biometrika* **81**(3), 541–553, doi:10.1093/biomet/81.3.541.

Chan, D., Kohn, R. and Kirby, C. (2006), 'Multivariate stochastic volatility models with correlated errors', *Econometric Reviews* **25**(2-3), 245–274, doi:10.1080/07474930600713309.

Chan, J. C. (2020), 'Large bayesian vars: A flexible kronecker error covariance structure', *Journal of Business & Economic Statistics* **38**(1), 68–79.

Chan, J. C. and Eisenstat, E. (2018), 'Bayesian model comparison for time-varying parameter VARs with stochastic volatility', *Journal of Applied Econometrics* **33**(4), 509–532.

Cheng, S. H. and Higham, N. J. (1998), 'A modified cholesky algorithm based on a symmetric indefinite factorization', *SIAM Journal on Matrix Analysis and Applications* **19**(4), 1097–1110, doi:10.1137/S0895479896302898.

Chib, S., Nardari, F. and Shephard, N. (2002), 'Markov chain Monte Carlo methods for stochastic volatility models', *Journal of Econometrics* **108**, 281–316, doi:10.1016/S0304-4076(01)00137-3.

Chib, S., Nardari, F. and Shephard, N. (2006), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **134**(2), 341–371, doi:10.1016/j.jeconom.2005.06.026.

Clark, T. E. (2011), 'Real-time density forecasts from bayesian vector autoregressions with stochastic volatility', *Journal of Business & Economic Statistics* **29**(3), 327–341.

Clark, T. E. and Ravazzolo, F. (2015), 'Macroeconomic forecasting performance under alternative specifications of time-varying volatility', *Journal of Applied Econometrics* **30**(4), 551–575.

Cogley, T. and Sargent, T. J. (2005), 'Drifts and volatilities: monetary policies and outcomes in the post WWII US', *Review of Economic Dynamics* **8**(2), 262–302.

Cross, J. L., Hou, C. and Poon, A. (2020), 'Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity', *International Journal of Forecasting* **36**(3), 899–915.

D'Agostino, A., Gambetti, L. and Giannone, D. (2013), 'Macroeconomic forecasting and structural change', *Journal of applied econometrics* **28**(1), 82–101.

Danielsson, J. (1994), 'Stochastic volatility in asset prices estimation with simulated maximum likelihood', *Journal of Econometrics* **64**(1-2), 375–400, doi:10.1016/0304-4076(94)90070-1.

Danielsson, J. (1998), 'Multivariate stochastic volatility models: Estimation and a comparison with VGARCH models', *Journal of Empirical Finance* **5**(2), 155–173, doi:10.1016/S0927-5398(97)00016-9.

Danielsson, J. and Richard, J.-F. (1993), 'Accelerated Gaussian importance sampler with application to dynamic latent variable models', *Journal of Applied Econometrics* **8**, S153–S173, doi:10.1002/jae.3950080510.

Davidson, J. (1994), *Stochastic limit theory*, Oxford University Press.

DeMiguel, V. and Nogales, F. J. (2009), 'Portfolio selection with robust estimation', *Operations Research* **57**(3), 560–577, doi:10.1287/opre.1080.0566.

DeMiguel, V., Garlappi, L. and Uppal, R. (2009), 'Optimal versus naive diversification: how inefficient is the 1/ n portfolio strategy?', *Review of Financial Studies* **22**(5), 1915–1953, doi:10.1093/rfs/hhm075.

Diebold, F. X., Schorfheide, F. and Shin, M. (2017), 'Real-time forecast evaluation of DSGE models with stochastic volatility', *Journal of Econometrics* **201**(2), 322–332.

Dufour, J.-M. (2006), 'Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics', *Journal of Econometrics* **133**(2), 443–477, doi:10.1016/j.jeconom.2005.06.007.

Dufour, J.-M. and Jouini, T. (2006), 'Finite-sample simulation-based tests in VAR models with applications to Granger causality testing', *Journal of Econometrics* **135**(1-2), 229–254, doi:10.1016/j.jeconom.2005.07.025.

Dufour, J.-M. and Khalaf, L. (2001), Monte Carlo test methods in econometrics, *in* B. Baltagi, ed., 'A Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 23, pp. 494–519.

Dufour, J.-M. and Khalaf, L. (2002), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **111**(2), 303–322, doi:10.1016/S0304-4076(02)00108-2.

Dufour, J.-M. and Valéry, P. (2006), 'On a simple two-stage closed-form estimator for a stochastic volatility in a general linear regression', *Advances in Econometrics* **20**, 259–288, doi:10.1016/S0731-9053(05)20010-5.

Durham, G. B. (2006), 'Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models', *Journal of Econometrics* **133**, 273–305, doi:10.1016/j.jeconom.2005.03.016.

Dwass, M. (1957), 'Modified randomization tests for nonparametric hypotheses', *Annals of Mathematical Statistics* **28**, 181–187, doi:10.1214/aoms/1177707045.

Eddelbuettel, D. and François, R. (2011), 'Rcpp: Seamless r and c++ integration', *Journal of Statistical Software* **40**(8), 1–18, doi:10.18637/jss.v040.i08.

Eddelbuettel, D. and Sanderson, C. (2014), 'Rcpparmadillo: Accelerating r with high-performance c++ linear algebra', *Computational Statistics and Data Analysis* **71**, 1054–1063, doi:10.1016/j.csda.2013.02.005.

Engle, R. (2002), 'Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business and Economic Statistics* **20**(3), 339–350, doi:10.1198/073500102288618487.

Engle, R. F. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**(4), 987–1008, doi:10.2307/1912773.

Flury, T. and Shephard, N. (2011), 'Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models', *Econometric Theory* **27**(05), 933–956, doi:10.1017/S0266466610000599.

Francq, C. and Zakoïan, J.-M. (2006), 'Linear-representation based estimation of stochastic volatility models', *Scandinavian Journal of Statistics* **33**(4), 785–806, doi:10.1111/j.1467-9469.2006.00495.x.

Frühwirth-Schnatter, S. (1994), 'Data augmentation and dynamic linear models', *Journal of Time Series Analysis* **15**(2), 183–202, doi:10.1111/j.1467-9892.1994.tb00184.x.

Gallant, A. R. and Tauchen, G. (1996), 'Which moments to match?', *Econometric Theory* **12**, 657 – 681.

Gasbarro, D., Wong, W.-K. and Kenton Zumwalt, J. (2007), 'Stochastic dominance analysis of ishares', *The European Journal of Finance* **13**(1), 89–101, doi:10.1080/13518470601025243.

Ghysels, E., Harvey, A. and Renault, E. (1996), Stochastic volatility, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics: Statistical Methods in Finance', Vol. 14, North-Holland, Amsterdam, pp. 119–191.

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), 'Global optimization of statistical functions with simulated annealing', *Journal of Econometrics* **60**, 65–99, doi:10.1016/0304-4076(94)90038-8.

Gourieroux, C. (2006), 'Continuous time wishart process for stochastic risk', *Econometric Reviews* **25**(2-3), 177–217, doi:10.1080/07474930600713234.

Gourieroux, C., Jasiak, J. and Sufana, R. (2009), 'The Wishart Autoregressive process of multivariate stochastic volatility', *Journal of Econometrics* **150**(2), 167–181, doi:10.1016/j.jeconom.2008.12.016.

Granger, C. W. J. and Morris, M. J. (1976), 'Time series modelling and interpretation', *Journal of the Royal Statistical Society. Series A (General)* **139**(2), 246–257, doi:10.2307/2345178.

Han, Y. (2006), 'Asset allocation with a high dimensional latent factor stochastic volatility model', *The Review of Financial Studies* **19**(1), 237–271, doi:10.1093/rfs/hhj002.

Harvey, A. C. and Shephard, N. (1996), 'Estimation of an asymmetric stochastic volatility model for asset returns', *Journal of Business and Economic Statistics* **14**(4), 429–434, doi:10.2307/1392251.

Harvey, A., Ruiz, E. and Shephard, N. (1994), 'Multivariate stochastic variance models', *The Review of Economic Studies* **61**(2), 247–264, doi:10.2307/2297980.

Hautsch, N. and Voigt, S. (2019), 'Large-scale portfolio allocation under transaction costs and model uncertainty', *Journal of Econometrics* **212**(1), 221–240, doi:10.1016/j.jeconom.2019.04.028.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business and Economic Statistics* **12**(4), 371–389, doi:10.1080/07350015.1994.10524553.

Jungbacker, B. and Koopman, S. J. (2006), 'Monte carlo likelihood estimation for three multivariate stochastic volatility models', *Econometric Reviews* **25**(2-3), 385–408, doi:10.1080/07474930600712848.

Jurado, K., Ludvigson, S. C. and Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.

Kastner, G. (2019), 'Sparse Bayesian time-varying covariance estimation in many dimensions', *Journal of Econometrics* **210**(1), 98–115, doi:10.1016/j.jeconom.2018.11.007.

Kastner, G. and Frühwirth-Schnatter, S. (2014), 'Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models', *Computational Statistics & Data Analysis* **76**, 408–423, doi:10.1016/j.csda.2013.01.002.

Kim, S., Shephard, N. and Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *The Review of Economic Studies* **65**, 361–393, doi:10.1111/1467-937X.00050.

Knight, J. L., Satchell, S. E. and Yu, J. (2002), 'Estimation of the stochastic volatility model by the empirical characteristic function method', *Australian and New Zealand Journal of Statistics* **44**(3), 319–335, doi:10.1111/1467-842X.00234.

Koop, G. and Korobilis, D. (2013), 'Large time-varying parameter VARs', *Journal of Econometrics* **177**(2), 185–198.

Koopman, S. J., Jungbacker, B. and Hol, E. (2005), 'Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements', *Journal of Empirical Finance* **12**(3), 445–475, doi:10.1016/j.jempfin.2004.04.009.

Melino, A. and Turnbull, S. M. (1990), 'Pricing foreign currency options with stochastic volatility', *Journal of Econometrics* **45**(1), 239–265, doi:10.1016/0304-4076(90)90100-8.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007), 'Stochastic volatility with leverage: Fast and efficient likelihood inference', *Journal of Econometrics* **140**(2), 425–449, doi:10.1016/j.jeconom.2006.07.008.

Pakel, C., Shephard, N., Sheppard, K. and Engle, R. F. (2021), 'Fitting vast dimensional time-varying covariance models', *Journal of Business & Economic Statistics* **39**(3), 652–668, doi:10.1080/07350015.2020.1713795.

Philipov, A. and Glickman, M. E. (2006*a*), 'Factor multivariate stochastic volatility via Wishart processes', *Econometric Reviews* **25**(2-3), 311–334, doi:10.1080/07474930600713366.

Philipov, A. and Glickman, M. E. (2006*b*), 'Multivariate stochastic volatility via Wishart processes', *Journal of Business and Economic Statistics* **24**(3), 313–328, doi:10.1198/073500105000000306.

Pitt, M. K. and Shephard, N. (1999), Time-varying covariances: A factor stochastic volatility approach, *in* J. O. Berger, A. P. Dawid and A. Smith, eds, 'JM Bernardo', Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting, Oxford University Press, pp. 547–570.

Poon, S.-H. and Granger, C. W. (2003), 'Forecasting volatility in financial markets: A review', *Journal of Economic Literature* **41**(2), 478–539, doi:10.1257/jel.41.2.478.

Primiceri, G. E. (2005), 'Time varying structural vector autoregressions and monetary policy', *The Review of Economic Studies* **72**(3), 821–852.

Richard, J.-F. and Zhang, W. (2007), 'Efficient high-dimensional importance sampling', *Journal of Econometrics* **141**(2), 1385–1411, doi:10.1016/j.jeconom.2007.02.007.

Ruiz, E. (1994), 'Quasi-maximum likelihood estimation of stochastic variance models', *Journal of Econometrics* **63**, 284–306, doi:10.1016/0304-4076(93)01569-8.

Sanderson, C. and Curtin, R. (2016), 'Armadillo: A template-based c++ library for linear algebra', *Journal of Open Source Software* **1**, 26–18, doi:10.21105/joss.00026.

Sandmann, G. and Koopman, S. J. (1998), 'Estimation of stochastic volatility models via Monte Carlo maximum likelihood', *Journal of Econometrics* **87**(2), 271–301, doi:10.1016/S0304-4076(98)00016-5.

Shephard, N. (1994), 'Partial non-Gaussian state space', *Biometrika* **81**(1), 115–131, doi:10.1093/biomet/81.1.115.

Shephard, N. and Andersen, T. G. (2009), Stochastic volatility: Origins and overview, *in* T. Mikosch, J.-P. Kreiß, A. R. Davis and G. T. Andersen, eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin, Heidelberg, pp. 233–254.

Smith, M. and Pitts, A. (2006), 'Foreign exchange intervention by the Bank of Japan: Bayesian analysis using a bivariate stochastic volatility model', *Econometric Reviews* **25**(2-3), 425–451, doi:10.1080/07474930600712897.

So, M. K. P. and Choi, C. Y. (2009), 'A threshold factor multivariate stochastic volatility model', *Journal of Forecasting* **28**(8), 712–735, doi:10.1002/for.1123.

Steel, M. F. (1998), 'Bayesian analysis of stochastic volatility models with flexible tails', *Econometric Reviews* **17**(2), 109–143, doi:10.1080/07474939808800408.

Taylor, S. J. (1982), Financial returns modelled by the product of two stochastic processes – a study of daily sugar prices, *in* O. D. Anderson, ed., 'Time Series Analysis: Theory and Practice', Vol. 1, North-Holland, Amsterdam, pp. 203–226.

Uhlig, H. (1997), 'Bayesian vector autoregressions with stochastic volatility', *Econometrica* **65**(1), 59–73, doi:10.2307/2171813.

van der Weide, R. (2002), 'GO-GARCH: a multivariate generalized orthogonal GARCH model', *Journal of Applied Econometrics* **17**(5), 549–564, doi:10.1002/jae.688.

Yu, J. (2002), 'Forecasting volatility in the New Zealand stock market', *Applied Financial Economics* **12**(3), 193–202, doi:10.1080/09603100110090118.

Yu, Y. and Meng, X.-L. (2011), 'To center or not to center: that is not the questionan ancillaritysufficiency interweaving strategy (Asis) for boosting mcmc efficiency', *Journal of Computational and Graphical Statistics* **20**(3), 531–570, doi:10.1198/jcgs.2011.203main.

Zakamouline, V. and Koekebakker, S. (2009), 'Portfolio performance evaluation with generalized Sharpe ratios: Beyond the mean and variance', *Journal of Banking & Finance* **33**(7), 1242–1254, doi:10.1016/j.jbankfin.2009.01.005.

Zumbach, G. O. (2007), A gentle introduction to the rm2006 methodology, SSRN Scholarly Paper ID 1420183, Social Science Research Network, Rochester, NY.

Practical estimation methods for high-dimensional multivariate stochastic volatility models

Md. Nazmul Ahsan[†] Jean-Marie Dufour[‡]

August 03, 2022

Technical Appendix

This Technical Appendix contains: (i) Mathematical proofs. (ii) Composition of portfolios. (iii) Figures. (iv) Tables.

A.1 Mathematical proofs

PROOF OF PROPOSITION 3.1 From (2.9-2.10), we have

$$\boldsymbol{\phi}(B)\boldsymbol{h}_t = \boldsymbol{\phi}(B)\mu + \boldsymbol{v}_t, \quad \text{and} \quad \boldsymbol{X}_t = \boldsymbol{h}_t + \epsilon_t,$$

where $\boldsymbol{\phi}(B) = (\boldsymbol{I}_m - \sum_{j=1}^{p}\boldsymbol{\phi}_j B^j)$. Furthermore, Assumption 2.1 implies that $\boldsymbol{v}_t$'s and ϵ_t's are independent. Now, applying $\boldsymbol{\phi}(B)$ to both sides of (2.10) yields

$$\boldsymbol{\phi}(B)\boldsymbol{X}_t = \boldsymbol{\phi}(B)\boldsymbol{h}_t + \boldsymbol{\phi}(B)\epsilon_t = \boldsymbol{\phi}(B)\mu + \boldsymbol{v}_t + \boldsymbol{\phi}(B)\epsilon_t.$$

$$\boldsymbol{\phi}(B)(\boldsymbol{X}_t - \mu) = \boldsymbol{v}_t + \boldsymbol{\phi}(B)\epsilon_t. \tag{A.1}$$

Consider the right hand side of (A.1). This is clearly a covariance stationary process. By the Wold decomposition theorem it must have a vector moving average representation. Since the autocovariance function cuts off for lags $k > p$ it must be a VMA(p) process, say $\theta(B)\eta_t = (\boldsymbol{I}_m - \sum_{j=1}^{p}\theta_j B^j)\eta_t$. Hence, $\boldsymbol{X}_t$ must be an VARMA(p, p) process (for univariate case, see Granger and Morris (1976)). The moving average parameters $(\theta_1, \theta_2, \ldots, \theta_p)$ and the noise variance $\boldsymbol{\Sigma}_\eta$ of this VARMA(p, p) process can be found by equating the autocovariance function of the right hand side of (A.1) with that of $\theta(B)\eta_t$ for lags $k = 0, 1, \ldots, p$ and solving the $p+1$ following non-linear equations

$$\begin{aligned}
\boldsymbol{\Sigma}_\eta + \theta_1\boldsymbol{\Sigma}_\eta\theta_1' + \theta_2\boldsymbol{\Sigma}_\eta\theta_2' + \cdots + \theta_p\boldsymbol{\Sigma}_\eta\theta_p' &= \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon + \boldsymbol{\phi}_1\boldsymbol{\Sigma}_\epsilon\boldsymbol{\phi}_1' + \boldsymbol{\phi}_2\boldsymbol{\Sigma}_\epsilon\boldsymbol{\phi}_2' + \cdots + \boldsymbol{\phi}_p\boldsymbol{\Sigma}_\epsilon\boldsymbol{\phi}_p' \\
-\theta_1\boldsymbol{\Sigma}_\eta + \theta_1\boldsymbol{\Sigma}_\eta\theta_2' + \cdots + \theta_{p-1}\boldsymbol{\Sigma}_\eta\theta_p' &= -\boldsymbol{\phi}_1\boldsymbol{\Sigma}_\epsilon + \boldsymbol{\phi}_1\boldsymbol{\Sigma}_\epsilon\boldsymbol{\phi}_2' + \cdots + \boldsymbol{\phi}_{p-1}\boldsymbol{\Sigma}_\epsilon\boldsymbol{\phi}_p' \\
&\vdots \\
-\theta_{p-1}\boldsymbol{\Sigma}_\eta + \theta_1\boldsymbol{\Sigma}_\eta\theta_p' &= -\boldsymbol{\phi}_{p-1}\boldsymbol{\Sigma}_\epsilon + \boldsymbol{\phi}_1\boldsymbol{\Sigma}_\epsilon\boldsymbol{\phi}_p' \\
-\theta_p\boldsymbol{\Sigma}_\eta &= -\boldsymbol{\phi}_p\boldsymbol{\Sigma}_\epsilon.
\end{aligned}$$

[†]Centre interuniversitaire de recherche en analyse des organisations (CIRANO). Mailing address: 1130 Sherbrooke West, Suite 1400, Montreal, Quebec, H3A 2M8, Canada; e-mail: ahsann@cirano.qc.ca.

[‡]William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com

Note that there may be multiple solutions, only some of which result in an invertible process. □

Proof of Corollary 3.2 From Proposition 3.1, the observed process $\{\boldsymbol{X}_t\}$ satisfies the following equation:

$$\boldsymbol{X}_t - \mu = \sum_{j=1}^{p} \boldsymbol{\phi}_j (\boldsymbol{X}_{t-j} - \mu) + \eta_t - \sum_{j=1}^{p} \theta_j \eta_{t-j}, \tag{A.2}$$

or

$$\boldsymbol{X}_t - \mu = \sum_{j=1}^{p} \boldsymbol{\phi}_j (\boldsymbol{X}_{t-j} - \mu) + \boldsymbol{v}_t + \epsilon_t - \sum_{j=1}^{p} \boldsymbol{\phi}_j \epsilon_{t-j}. \tag{A.3}$$

Multiply both sides of (A.3) by $(\boldsymbol{X}_{t-k} - \mu)'$ and taking expectation we get:

$$\boldsymbol{\Gamma}_k = \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \mathbb{E}[\boldsymbol{v}_t (\boldsymbol{X}_{t-k} - \mu)'] + \mathbb{E}[\epsilon_t (\boldsymbol{X}_{t-k} - \mu)'] - \sum_{j=1}^{p} \boldsymbol{\phi}_j \mathbb{E}[\epsilon_{t-j} (\boldsymbol{X}_{t-k} - \mu)'].$$

Setting $k = 0$, we get

$$\begin{aligned}
\boldsymbol{\Gamma}_0 &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_1 + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_p + \mathbb{E}[\boldsymbol{v}_t (\boldsymbol{X}_t - \mu)'] + \mathbb{E}[\epsilon_t (\boldsymbol{X}_t - \mu)'] - \sum_{j=1}^{p} \boldsymbol{\phi}_j \mathbb{E}[\epsilon_{t-j} (\boldsymbol{X}_t - \mu)'] \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_1 + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_p + \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon - \sum_{j=1}^{p} \boldsymbol{\phi}_j \mathbb{E}[\epsilon_{t-j} ((\boldsymbol{X}_{t-j} - \mu)' - \epsilon'_{t-j})] \boldsymbol{\phi}'_j \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_1 + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_p + \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon + \sum_{j=1}^{p} \boldsymbol{\phi}_j [\boldsymbol{\Sigma}_\epsilon - \boldsymbol{\Sigma}_\epsilon] \boldsymbol{\phi}'_j \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_1 + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_p + \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon.
\end{aligned} \tag{A.4}$$

Setting $1 \le k \le p$, we get

$$\begin{aligned}
\boldsymbol{\Gamma}_k &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \mathbb{E}[\boldsymbol{v}_t (\boldsymbol{X}_{t-k} - \mu)'] + \mathbb{E}[\epsilon_t (\boldsymbol{X}_{t-k} - \mu)'] - \sum_{j=1}^{p} \boldsymbol{\phi}_j \mathbb{E}[\epsilon_{t-j} (\boldsymbol{X}_{t-k} - \mu)'] \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + 0 + 0 - \boldsymbol{\phi}_k \boldsymbol{\Sigma}_\epsilon. \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} - \boldsymbol{\phi}_k \boldsymbol{\Sigma}_\epsilon.
\end{aligned} \tag{A.5}$$

Setting $k > p$, we get

$$\begin{aligned}
\boldsymbol{\Gamma}_k &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \mathbb{E}[\boldsymbol{v}_t (\boldsymbol{X}_{t-k} - \mu)'] + \mathbb{E}[\epsilon_t (\boldsymbol{X}_{t-k} - \mu)'] - \sum_{j=1}^{p} \boldsymbol{\phi}_j \mathbb{E}[\epsilon_{t-j} (\boldsymbol{X}_{t-k} - \mu)'] \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + 0 + 0 - 0 \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p}.
\end{aligned} \tag{A.6}$$

Combining (A.4-A.6), we get the autocovariance structure of the observed process stated in Corollary 3.2. □

Proof of Corollary 3.3 The estimator of $\phi_{(\boldsymbol{p})}$ is based on the autocovariance structure of the process

$\boldsymbol{X}_t$. This is the solution of p-system of equations from (3.2) with $k = p+1, \ldots, 2p$. So

$$\begin{bmatrix} \boldsymbol{\Gamma}_{p+1} & \boldsymbol{\Gamma}_{p+2} & \cdots & \boldsymbol{\Gamma}_{2p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_p \end{bmatrix} . \begin{bmatrix} \boldsymbol{\Gamma}_p & \boldsymbol{\Gamma}_{p+1} & \cdots & \boldsymbol{\Gamma}_{2p-1} \\ \boldsymbol{\Gamma}_{p-1} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_{2p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \end{bmatrix},$$

or

$$\begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_{p+1} & \boldsymbol{\Gamma}_{p+2} & \cdots & \boldsymbol{\Gamma}_{2p} \end{bmatrix} . \begin{bmatrix} \boldsymbol{\Gamma}_p & \boldsymbol{\Gamma}_{p+1} & \cdots & \boldsymbol{\Gamma}_{2p-1} \\ \boldsymbol{\Gamma}_{p-1} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_{2p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \end{bmatrix}^{-1},$$

or

$$\boldsymbol{\phi}_{(p)} = \boldsymbol{\Gamma}_{(p+1)} \boldsymbol{\Gamma}_{[p]}^{-1}, \tag{A.7}$$

where

$$\boldsymbol{\phi}_{(p)} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_p \end{bmatrix}, \quad \boldsymbol{\Gamma}_{(p+1)} = \begin{bmatrix} \boldsymbol{\Gamma}_{p+1} & \boldsymbol{\Gamma}_{p+2} & \cdots & \boldsymbol{\Gamma}_{2p} \end{bmatrix},$$

$$\boldsymbol{\Gamma}_{[p]} = \begin{bmatrix} \boldsymbol{\Gamma}_p & \boldsymbol{\Gamma}_{p+1} & \cdots & \boldsymbol{\Gamma}_{2p-1} \\ \boldsymbol{\Gamma}_{p-1} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_{2p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \end{bmatrix}.$$

Note that (A.7) is also valid for any $j \geq 1$ such that

$$\boldsymbol{\phi}_{(p)} = \boldsymbol{\Gamma}_{(p+j)} \boldsymbol{\Gamma}_{[p+j-1]}^{-1}, \tag{A.8}$$

where

$$\boldsymbol{\Gamma}_{(p+j)} = \begin{bmatrix} \boldsymbol{\Gamma}_{p+j} & \boldsymbol{\Gamma}_{p+j+1} & \cdots & \boldsymbol{\Gamma}_{2p+j-1} \end{bmatrix}, \quad \boldsymbol{\Gamma}_{[p+j-1]} = \begin{bmatrix} \boldsymbol{\Gamma}_{p+j-1} & \boldsymbol{\Gamma}_{p+j} & \cdots & \boldsymbol{\Gamma}_{2p+j-2} \\ \boldsymbol{\Gamma}_{p+j-2} & \boldsymbol{\Gamma}_{p+j-1} & \cdots & \boldsymbol{\Gamma}_{2p+j-3} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_j & \boldsymbol{\Gamma}_{j+1} & \cdots & \boldsymbol{\Gamma}_{p+j-1} \end{bmatrix}.$$

Now from (3.2) with $k = 1$ we have:

$$\boldsymbol{\Gamma}_1 = \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_2 \boldsymbol{\Gamma}_{-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{1-p} - \boldsymbol{\phi}_1 \boldsymbol{\Sigma}_\epsilon = \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_2 \boldsymbol{\Gamma}_1' + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{p-1}' - \boldsymbol{\phi}_1 \boldsymbol{\Sigma}_\epsilon.$$

Hence,

$$\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}_j' - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1, \tag{A.9}$$

and substitute the above expression of $\boldsymbol{\Sigma}_\epsilon$ into (3.2) with $k = 0$, we have

$$\boldsymbol{\Gamma}_0 \quad = \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{-p} + \boldsymbol{\Sigma}_v + \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}_j' - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1$$

$$= \sum_{j=1}^{p} \boldsymbol{\phi}_j \boldsymbol{\Gamma}_{-j} + \boldsymbol{\Sigma}_v + \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1$$
$$= \sum_{j=1}^{p} \boldsymbol{\phi}_j \boldsymbol{\Gamma}'_j + \boldsymbol{\Sigma}_v + \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1 . \tag{A.10}$$

Hence

$$\boldsymbol{\Sigma}_v = \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1 - \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \sum_{j=1}^{p} \boldsymbol{\phi}_j \boldsymbol{\Gamma}'_j . \tag{A.11}$$

Finally,

$$\mu = \mathbb{E}(\boldsymbol{X}_t). \tag{A.12}$$

□

PROOF OF LEMMA 4.1 Under the Assumptions 4.1 – 4.3 with $s = 4$, the observed process $\{\boldsymbol{X}_t\}$ is strictly stationarity and geometrically ergodic with $\mathbb{E}[\boldsymbol{X}_t] < \infty$ and $\mathbb{E}[\boldsymbol{X}_t \boldsymbol{X}_{t+k}] < \infty$. So the consistency is a simple application of the Law of Large Numbers for stationary and ergodic processes, *i.e.*, the Ergodic theorem; see Theorem 13.12 and Corollary 13.14 of Davidson (1994). □

PROOF OF LEMMA 4.2 Under the Assumptions 4.1 – 4.3 with $s = 8$, $\Lambda(p)$ is finite and consistent given the law of large numbers for stationary and ergodic processes. Finally, the joint distribution converges to a normal distribution by the central limit theory for strongly mixing sequences (see Davidson (1994), Theorem 24.5, p. 385). □

PROOF OF THEOREM 4.3 The first part of the theorem follows from the continuous mapping theorem and the later part follows from the standard result for differentiable transformations of asymptotically normally distributed variables together with the application of the multivariate delta method. Further, it is easy to see that $\mathbf{F}(\Lambda)$ is a continuously differentiable mapping of Λ; see Ahsan and Dufour (2021) for a similar proof. □

A.2 Composition of portfolios

- **Ticker symbols for 20 assets portfolio:** AAPL, ABT, ADBE, AMZN, CMCSA, CSCO, CVX, HD, JNJ, JPM, MSFT, NFLX, NVDA, PEP, PFE, PG, TMO, UNH, VZ, XOM.
- **Ticker symbols for 50 assets portfolio:** AAPL, ABT, ADBE, AMD, AMT, AMZN, BLK, BMY, C, CMCSA, CMS, COST, CSCO, CVX, DHR, GE, GS, HD, HON, INTC, INTU, ISRG, JNJ, JPM, KO, LLY, LOW, MCD, MDT, MSFT, NFLX, NKE, NVDA, ORCL, PEP, PFE, PG, QCOM, SBUX, T, TGT, TMO, UNH, UNP, UPS, VRTX, VZ, WFC, WMT, XOM.

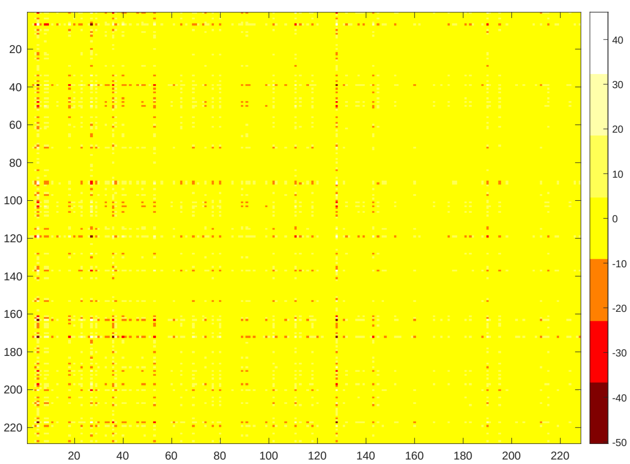
A.3 Figures

Figure A1

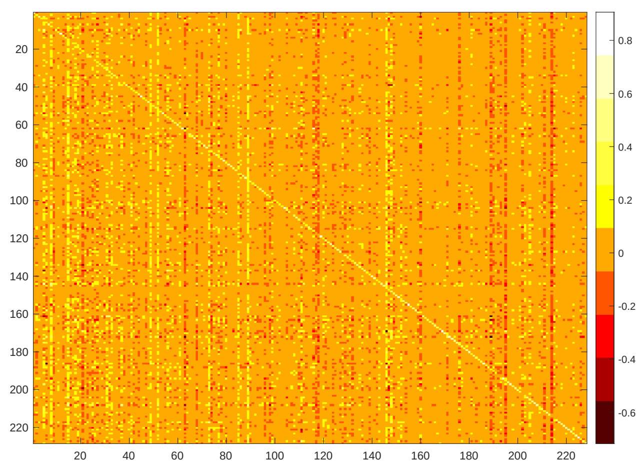
High-dimensional heatmap plots of the estimated first-order volatility persistent matrix with different values of shrinkage control parameter

Number of asset $m = 228$ and number of high-frequency observations $T = 252,021$

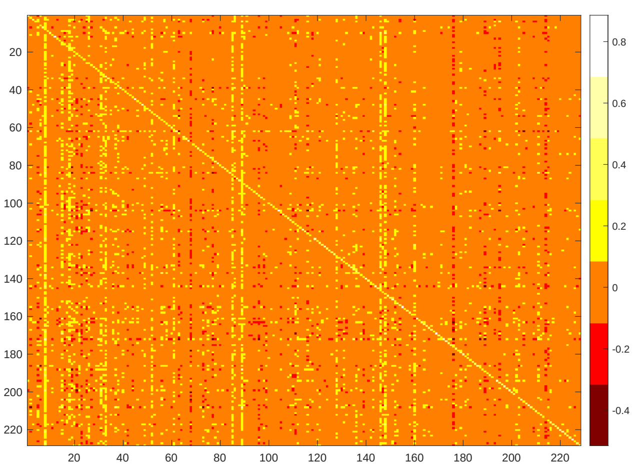
Figure A1 shows the heatmap plots of $\hat{\boldsymbol{\phi}}_{(1)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$). The sample period is from January 01, 2004 to December 31, 2016 and the number of five minute high-frequency observations is $T = 252,021$. The balanced panel includes 228 S&P 500 ticker components which are selected based on their market capitalization. We consider first-order persistence in latent volatilities with non-diagonal $\boldsymbol{\phi}_1$ and diagonal $\boldsymbol{\Sigma}_v$. The model has $m = 228$ number of time series and the number of parameters (NP) are $(= m^2 + 2m = 52,440)$ and $\boldsymbol{\phi}_{(1)}$ is a $m \times m$ matrix, which has $51,984$ parameters.



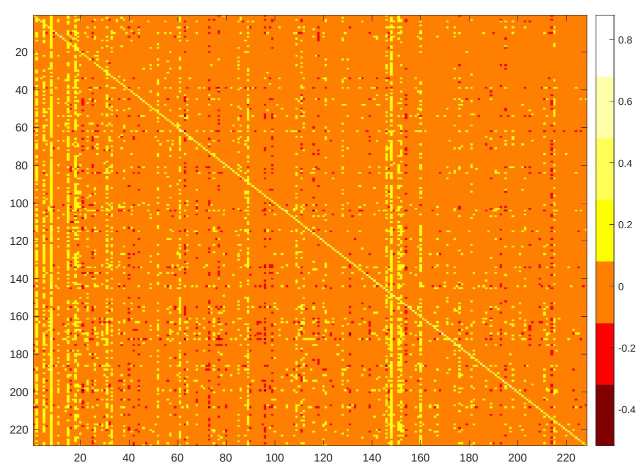
(a) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 1$ (unstable)



(b) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 10$ (stable)



(c) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 50$ (stable)



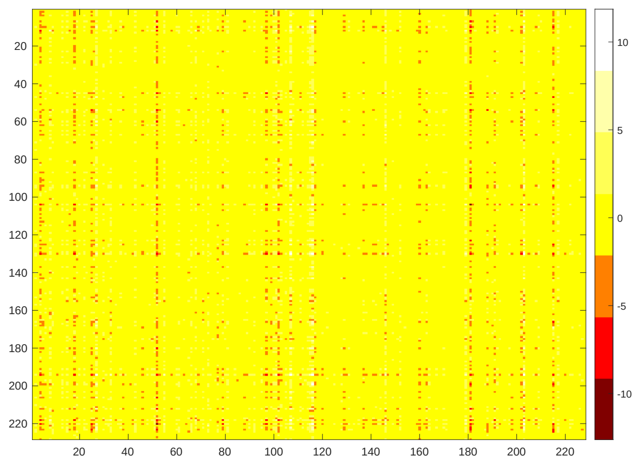
(d) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 100$ (stable)

Figure A2

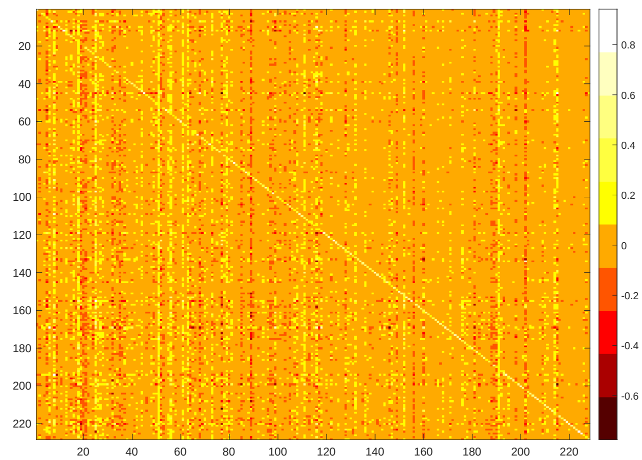
High-dimensional heatmap plots of the estimated first-order volatility persistent matrix with different values of shrinkage control parameter

Number of asset $m = 228$ and number of high-frequency observations $T = 126{,}011$

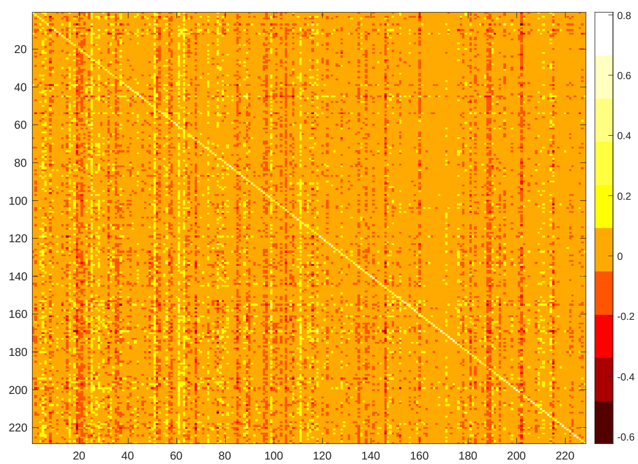
Figure A2 shows the heatmap plots of $\hat{\boldsymbol{\phi}}_{(1)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$). The sample period is from July 01, 2009 to December 31, 2016 and the number of five minute high-frequency observations is $T = 126{,}011$. The balanced panel includes 228 S&P 500 ticker components which are selected based on their market capitalization. We consider first-order persistence in latent volatilities with non-diagonal $\boldsymbol{\phi}_1$ and diagonal $\boldsymbol{\Sigma}_v$. The model has $m = 228$ number of time series and the number of parameters (NP) are ($= m^2 + 2m = 52{,}440$) and $\boldsymbol{\phi}_{(1)}$ is a $m \times m$ matrix, which has 51,984 parameters.



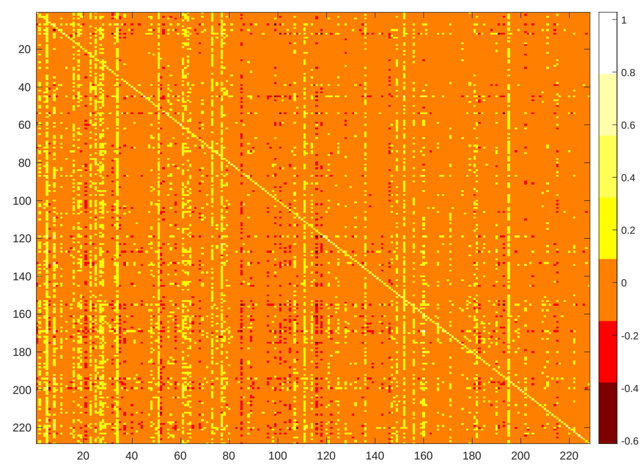
(a) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 1$ (unstable)



(b) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 10$ (stable)



(c) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 50$ (stable)



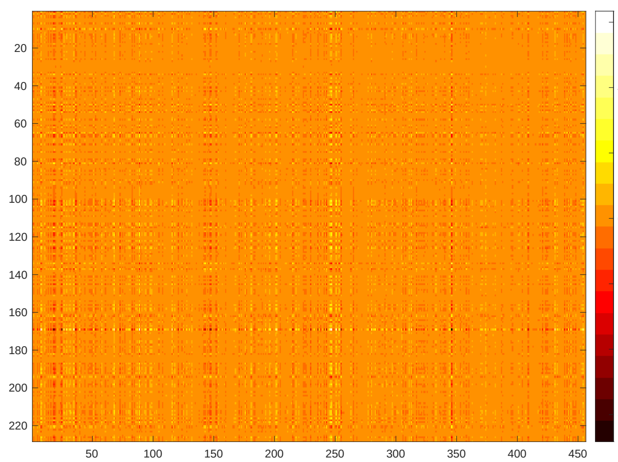
(d) $\hat{\boldsymbol{\phi}}_{(1)}$ with $J = 100$ (stable)

Figure A3

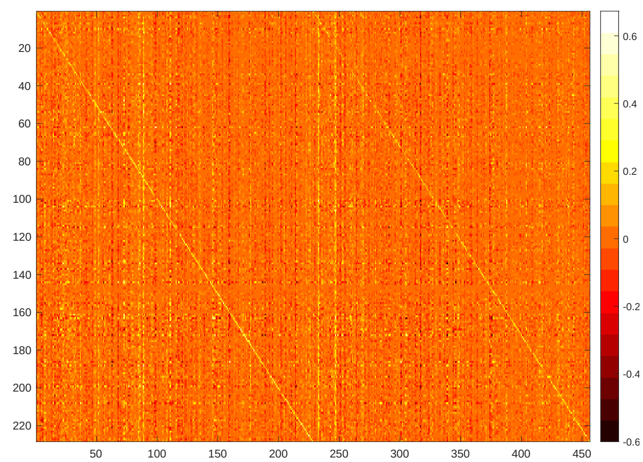
High-dimensional heatmap plots of the estimated second-order volatility persistent matrix with different values of shrinkage control parameter

Number of asset $m = 228$ and number of high-frequency observations $T = 252,021$

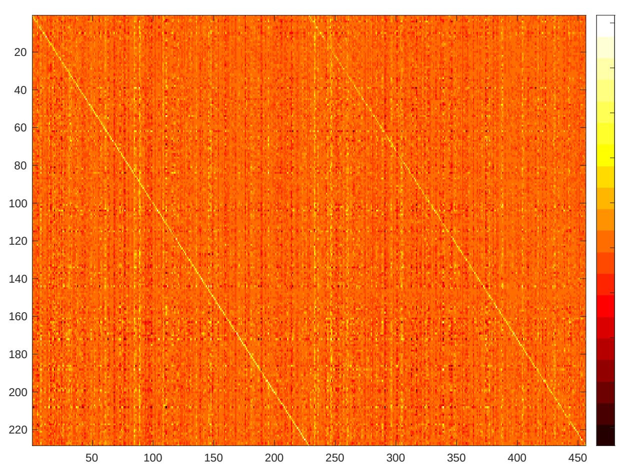
Figure A3 shows the heatmap plots of $\hat{\boldsymbol{\phi}}_{(2)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$).The sample period is from July 01, 2009 to December 31, 2016 and the number of five minute high-frequency observations is $T = 252,021$. The balanced panel includes 228 S&P 500 ticker components which are selected based on their market capitalization. We consider second-order persistence in latent volatilities with non-diagonal $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, and diagonal $\boldsymbol{\Sigma}_v$. The model has $m = 228$ number of time series and the number of parameters (NP) are ($= 2m^2 + 2m = 104,424$) and $\boldsymbol{\phi}_{(2)}$ is a $m \times 2m$ matrix, which has 103,968 parameters.



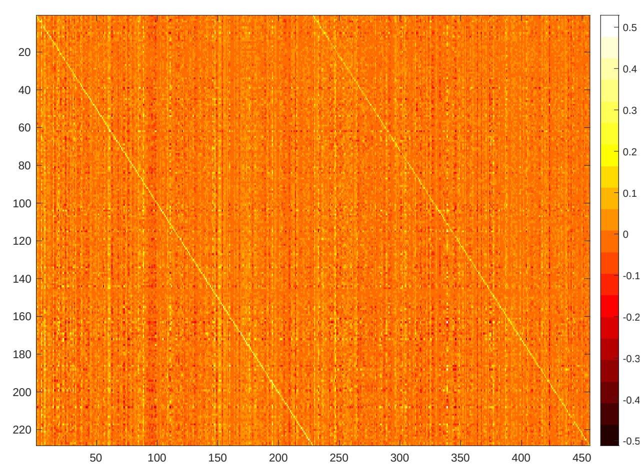
(a) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 1$ (unstable)



(b) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 10$ (stable)



(c) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 50$ (stable)



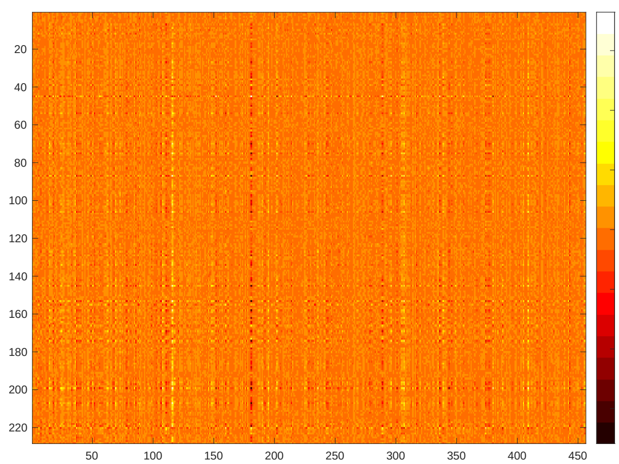
(d) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 100$ (stable)

Figure A4

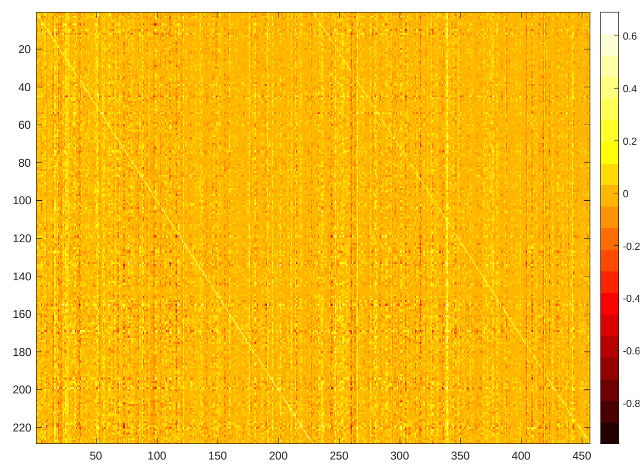
High-dimensional heatmap plots of the estimated second-order volatility persistent matrix with different values of shrinkage control parameter

Number of asset $m = 228$ and number of high-frequency observations $T = 126,011$

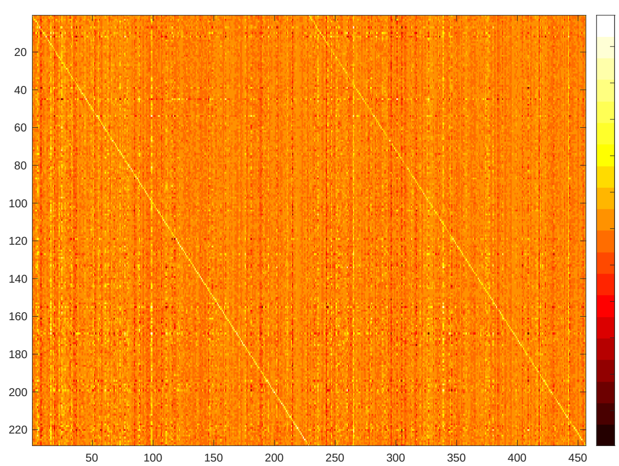
Figure A4 shows the heatmap plots of $\hat{\boldsymbol{\phi}}_{(2)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$). The sample period is from July 01, 2009 to December 31, 2016 and the number of five minute high-frequency observations is $T = 126,011$. The balanced panel includes 228 S&P 500 ticker components which are selected based on their market capitalization. We consider second-order persistence in latent volatilities with non-diagonal $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$, and diagonal $\boldsymbol{\Sigma}_v$. The model has $m = 228$ number of time series and the number of parameters (NP) are $(= 2m^2 + 2m = 104,424)$ and $\boldsymbol{\phi}_{(2)}$ is a $m \times 2m$ matrix, which has 103,968 parameters.



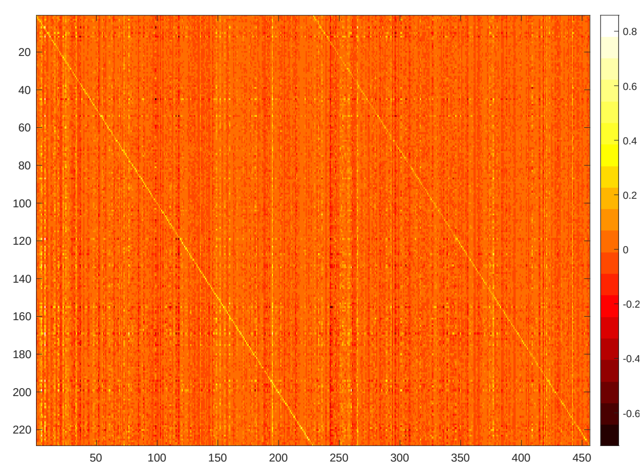
(a) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 1$ (unstable)



(b) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 10$ (stable)



(c) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 50$ (stable)



(d) $\hat{\boldsymbol{\phi}}_{(2)}$ with $J = 100$ (stable)

Figure A5

Modulus of the maximum eigenvalue of the companion matrix associated with a first-order unrestricted MSV model

The dataset includes 1,184 NYSE stocks with 3,247 daily time series observations

Figure A5 shows the evolution of the modulus of the maximum eigenvalue of the companion matrix associated with a first-order unrestricted MSV model with different number of assets $m = (2, 10, 50, 100, 250, 500, 1000, 1184)$ and different values of the shrinkage control parameter $J = (1, \ldots, 100)$.

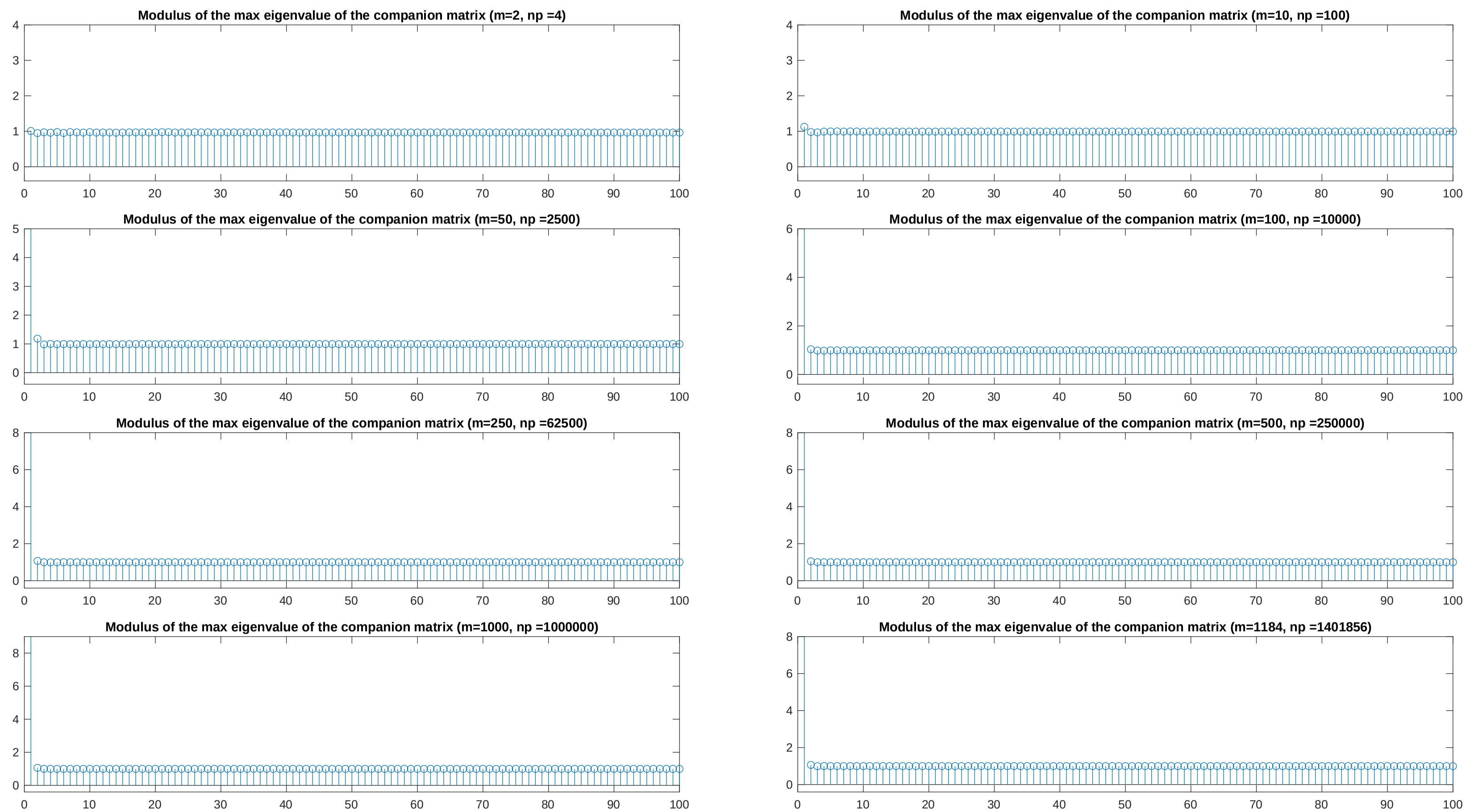
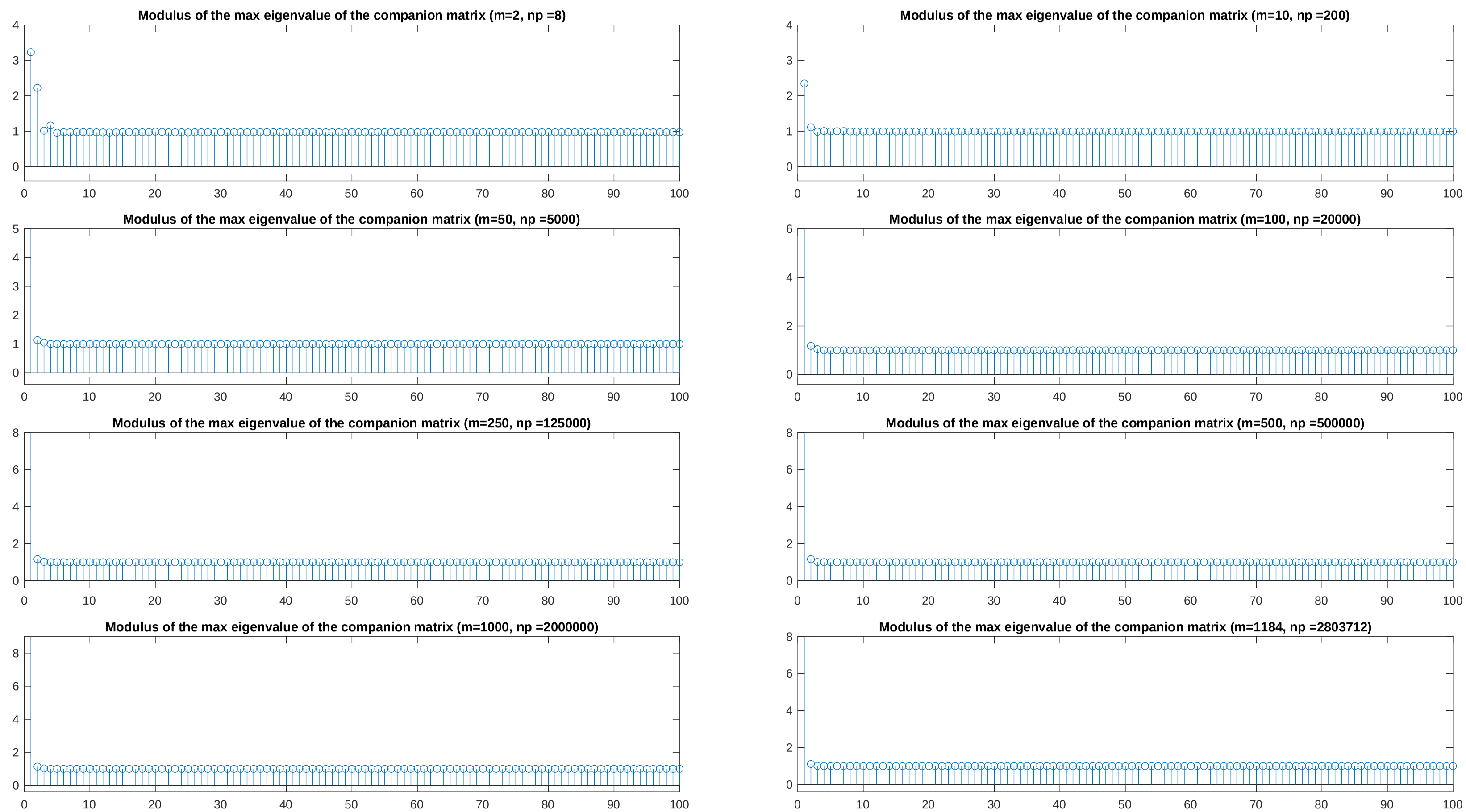


Figure A6

Modulus of the maximum eigenvalue of the companion matrix associated with a second-order unrestricted MSV model

The dataset includes 1,184 NYSE stocks with 3,247 daily time series observations

Figure A6 shows the evolution of the modulus of the maximum eigenvalue of the companion matrix associated with a second-order unrestricted MSV model with different number of assets $m = (2, 10, 50, 100, 250, 500, 1000, 1184)$ and different values of the shrinkage control parameter $J = (1, \ldots, 100)$.



A.4 Tables

Table A1.
Comparison of different winsorized VARMA estimators (W-VARMA)

Table A1 reports the performance of different winsorized estimators. The estimators compared are the uncensored (CF) and winsorized estimators with ($J = 10$) defined in Section 3. For $J = 1$, both the winsorized estimators reduce to the CF estimator. RMSE is the estimated root mean square error based on the simulation. NIV stands for the number of inadmissible parameter values produced by the estimators (over 1000).

	Parameter		ϕ_1	ϕ_2	ϕ_{12}	ϕ_{21}	μ_1	μ_2	$\sigma_{v,11}$	$\sigma_{v,22}$	$\sigma_{v,12}$	ρ
	True value		0.95	0.95	-0.2	0.1	-2	2	1	1	0.9	0.9
Estimators	T	NIV					Bias					
CF-VARMA	500	39	-0.0007	-0.0052	-0.0052	0.0002	0.0175	-0.0219	0.0192	0.0410	0.0237	-0.0104
	2000	0	-0.0009	-0.0012	-0.0018	0.0000	0.0000	0.0038	0.0211	0.0082	0.0123	-0.0035
	5000	0	0.0002	0.0002	0.0000	0.0004	-0.0003	-0.0012	-0.0044	-0.0089	-0.0074	-0.0006
VARMA-MEAN	500	0	0.0007	-0.0003	-0.0024	0.0012	0.0175	-0.0219	-0.0325	-0.0740	-0.0292	-0.0022
	2000	0	-0.0001	-0.0002	-0.0006	0.0002	0.0000	0.0038	-0.0044	-0.0178	-0.0094	-0.0010
	5000	0	0.0002	-0.0001	-0.0003	0.0003	-0.0003	-0.0012	-0.0046	-0.0034	-0.0019	-0.0004
W-VARMA-MVR	500	0	-0.0004	-0.0037	-0.0034	0.0003	0.0175	-0.0219	-0.0047	-0.0046	-0.0081	-0.0035
	2000	0	-0.0004	-0.0008	-0.0009	0.0000	0.0000	0.0038	0.0027	-0.0035	-0.0034	-0.0014
	5000	0	0.0001	-0.0003	-0.0003	0.0002	-0.0003	-0.0012	-0.0022	0.0005	-0.0009	-0.0005
Estimators	T	NIV					RMSE					
CF-VARMA	500	39	0.0294	0.0380	0.0378	0.0299	0.3277	0.3133	0.7037	0.7385	0.5804	0.0535
	2000	0	0.0127	0.0187	0.0173	0.0136	0.1667	0.1588	0.3394	0.3911	0.2886	0.0260
	5000	0	0.0082	0.0111	0.0106	0.0086	0.1037	0.0959	0.2169	0.2400	0.1834	0.0159
VARMA-MEAN	500	0	0.0171	0.0209	0.0234	0.0147	0.3282	0.3141	0.2780	0.3454	0.2776	0.0350
	2000	0	0.0075	0.0104	0.0109	0.0070	0.1667	0.1588	0.1365	0.1614	0.1331	0.0177
	5000	0	0.0048	0.0060	0.0065	0.0045	0.1037	0.0959	0.0897	0.0989	0.0836	0.0111
W-VARMA-MVR	500	0	0.0174	0.0202	0.0213	0.0150	0.3282	0.3141	0.2737	0.2839	0.2482	0.0348
	2000	0	0.0075	0.0097	0.0102	0.0070	0.1667	0.1588	0.1342	0.1381	0.1236	0.0178
	5000	0	0.0048	0.0057	0.0062	0.0045	0.1037	0.0959	0.0881	0.0870	0.0790	0.0111

Table A2

Simulation results for estimated MSV processes: QML vs W-VARMA-MVR estimator

True value: $(\phi_{11}, , \phi_{22}, \sigma^2_{v,1}, \sigma^2_{v,2}, \sigma_{v,12}, \rho) = (0.98, 0.98, 0.5, 0.5, 0.5, -0.8)$

Table A2 reports the performance of QML and W-VARMA-MVR estimators in bi-variate setups. VARMA-type estimator are proposed in Section 3. QML estimator is proposed by (Harvey et al., 1994). We consider first-order persistence in latent volatilities. The true DGP is given by (2.1)-(2.2) with $(\phi_{11}, , \phi_{22}, \sigma^2_{v,1}, \sigma^2_{v,2}, \sigma_{v,12}, \rho) = (0.98, 0.98, 0.5, 0.5, 0.5, -0.8)$. We simulate 100 samples from each simulated model. RCT stands for the relative computational time w.r.t. the CF-VARMA estimator. The simulated model has $m = 2$ number of time series and the number of parameters (NP) are 6. The number of inadmissible values (NIV) of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100. Boldface font highlights the smallest bias and RMSE with no NIV/NC. Boldface font also highlights the estimator, which has the best overall performance.

					Bias						RMSE					
T	Estimators	RCT	NIV	NPD	ϕ_{11}	ϕ_{22}	$\sigma^2_{v,1}$	$\sigma^2_{v,2}$	$\sigma_{v,12}$	ρ	ϕ_{11}	ϕ_{22}	$\sigma^2_{v,1}$	$\sigma^2_{v,2}$	$\sigma_{v,12}$	ρ
500	QML	7880.30	0	3	-0.0157	-0.0163	-0.1257	-0.1394	-0.0909	0.1677	0.0233	0.0316	0.1911	0.2041	0.1234	0.2188
	CF-VARMA	1.00	39	46	**-0.0025**	**-0.0049**	-0.0299	0.0348	-0.0094	**0.0297**	0.0419	0.0406	0.4635	0.4944	0.2526	0.0696
	W-VARMA-MVR($J = 5$)	1.02	1	1	-0.0086	-0.0091	-0.0392	-0.0216	**0.0088**	0.0301	**0.0184**	0.0187	0.1731	0.1780	0.1257	0.0609
	W-VARMA-MVR($J = 10$)	1.08	0	0	-0.0106	-0.0092	**-0.0128**	**-0.0160**	0.0163	0.0305	0.0196	**0.0166**	**0.1465**	**0.1409**	**0.1162**	**0.0600**
	W-VARMA-MVR($J = 25$)	1.28	0	0	-0.0131	-0.0121	0.0267	0.0308	0.0410	0.0328	0.0224	0.0196	0.1531	0.1583	0.1362	0.0615
	W-VARMA-MVR($J = 50$)	1.60	0	0	-0.0147	-0.0137	0.0567	0.0621	0.0582	0.0344	0.0234	0.0207	0.1694	0.1715	0.1516	0.0624
	W-VARMA-MVR($J = 100$)	2.21	0	0	-0.0133	-0.0131	0.0289	0.0395	0.0497	0.0336	0.0233	0.0221	0.1656	0.1515	0.1523	0.0616
2000	QML	9683.44	0	0	-0.0050	-0.0050	-0.1374	-0.1451	-0.0951	0.1842	0.0071	0.0071	0.1798	0.1709	0.1076	0.2500
	CF-VARMA	1.00	12	16	-0.0041	**0.0010**	0.0125	-0.0729	-0.0173	0.0312	0.0145	0.0131	0.2648	0.2493	0.1267	0.0435
	W-VARMA-MVR($J = 5$)	1.04	0	0	**-0.0028**	-0.0018	-0.0229	-0.0407	-0.0128	**0.0310**	0.0065	**0.0063**	0.0872	0.0998	0.0700	0.0405
	W-VARMA-MVR($J = 10$)	1.10	0	0	-0.0030	-0.0022	-0.0194	-0.0349	-0.0103	0.0312	**0.0061**	0.0064	**0.0750**	**0.0813**	**0.0688**	**0.0402**
	W-VARMA-MVR($J = 25$)	1.26	0	0	-0.0032	-0.0029	-0.0140	-0.0204	-0.0055	0.0316	0.0063	0.0069	0.0791	0.0754	0.0685	0.0403
	W-VARMA-MVR($J = 50$)	1.52	0	0	-0.0035	-0.0033	**-0.0065**	-0.0116	**-0.0001**	0.0321	0.0067	0.0073	0.0929	0.0866	0.0758	0.0408
	W-VARMA-MVR($J = 100$)	2.06	0	0	-0.0033	-0.0032	-0.0090	-0.0132	-0.0009	0.0320	0.0065	0.0071	0.0936	0.0879	0.0771	0.0408
5000	QML	14883.96	0	0	-0.0032	-0.0029	-0.1106	-0.1205	-0.0771	0.1286	0.0048	0.0045	0.1392	0.1489	0.0878	0.1600
	CF-VARMA	1.00	2	3	-0.0014	-0.0013	**-0.0269**	-0.0276	**-0.0167**	0.0326	0.0091	0.0090	0.1849	0.1926	0.0828	0.0380
	W-VARMA-MVR($J = 5$)	1.03	0	0	-0.0008	**-0.0009**	-0.0398	-0.0405	-0.0223	**0.0319**	0.0043	0.0037	0.0681	0.0661	0.0450	0.0361
	W-VARMA-MVR($J = 10$)	1.07	0	0	**-0.0007**	-0.0012	-0.0415	-0.0337	-0.0213	0.0320	**0.0038**	**0.0034**	**0.0596**	0.0525	**0.0417**	**0.0360**
	W-VARMA-MVR($J = 25$)	1.24	0	0	-0.0009	-0.0014	-0.0375	-0.0297	-0.0194	0.0322	0.0039	0.0036	0.0620	**0.0519**	0.0419	0.0361
	W-VARMA-MVR($J = 50$)	1.50	0	0	-0.0010	-0.0014	-0.0356	-0.0280	-0.0187	0.0322	0.0042	0.0040	0.0681	0.0623	0.0427	0.0362
	W-VARMA-MVR($J = 100$)	2.00	0	0	-0.0010	-0.0015	-0.0346	**-0.0260**	-0.0179	0.0323	0.0045	0.0042	0.0748	0.0677	0.0452	0.0363

Table A3

Simulation results for estimated MSV processes: Bayes vs W-VARMA-MVR estimator ($m = 2, p = 1$)

True value: $(\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \mu_1, \mu_2, \sigma^2_{v,1}, \sigma^2_{v,2}) = (0.98, -0.1, 0.12, 0.99, -0.5, -0.5, 0.4, 0.4)$

Table A3 reports the performance of Bayes and W-VARMA-MVR estimators in bi-variate setups. VARMA-type estimator are proposed in Section 3. Bayes estimator is based on a Metropolis-within-Gibbs sampler [Kim et al. (1998), Omori et al. (2007), Kastner and Frühwirth-Schnatter (2014)]. Prior distributions, MCMC algorithm and effective computations are discussed in Section 6.3.2. The posteriors are based on 20000 draws of the sampler, after discarding 20000 draws. We consider first-order persistence in latent volatilities. The true DGP is given by (2.1)-(2.2) with $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = 0.4\boldsymbol{I}_m$, $\mu = -0.5\mathbb{1}$, and stationary $\boldsymbol{\phi}_1$ is non-diagonal with $(\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}) = (0.98, -0.1, 0.12, 0.99)$. We simulate 100 samples from each simulated model. RCT stands for the relative computational time w.r.t. the CF-VARMA estimator. The simulated model has $m = 2$ number of time series and the number of parameters (NP) are ($= m^2 + 2m = 8$). The number of inadmissible values (NIV) of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100. Boldface font highlights the smallest bias and RMSE with no NIV/NC. Boldface font also highlights the estimator, which has the best overall performance.

					Bias								RMSE							
T	Estimators	RCT	NIV	NPD	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	μ_1	μ_2	$\sigma^2_{v,1}$	$\sigma^2_{v,2}$	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	μ_1	μ_2	$\sigma^2_{v,1}$	$\sigma^2_{v,2}$
	Bayes	576674.51	0	0	0.0043	0.0032	-0.0186	-0.0036	0.7691	0.7081	1.0506	1.0143	0.0056	0.0065	0.0205	0.0049	0.8952	0.8319	1.2595	1.1956
	CF-VARMA	1.00	9	35	-0.0008	-0.0001	-0.0005	-0.0007	-0.0086	0.0090	-0.0075	0.0194	0.0129	0.0105	0.0131	0.0103	0.2011	0.2040	0.3421	0.3283
	W-VARMA-MVR($J = 5$)	1.02	0	0	-0.0026	-0.0002	-0.0008	-0.0013	**-0.0086**	**0.0090**	-0.0045	-0.0055	0.0055	**0.0048**	0.0061	0.0049	**0.2011**	**0.2040**	0.1137	0.1241
1000	**W-VARMA-MVR($J = 10$)**	1.08	0	0	-0.0019	-0.0009	**0.0000**	-0.0010	**-0.0086**	**0.0090**	**-0.0311**	-0.0240	**0.0050**	0.0049	0.0058	**0.0046**	**0.2011**	**0.2040**	**0.0944**	**0.0985**
	W-VARMA-MVR($J = 25$)	1.28	0	0	-0.0014	-0.0008	-0.0009	-0.0005	**-0.0086**	**0.0090**	-0.0517	-0.0421	0.0054	**0.0048**	**0.0056**	**0.0046**	**0.2011**	**0.2040**	0.1032	0.1124
	W-VARMA-MVR($J = 50$)	1.60	0	0	-0.0012	-0.0012	-0.0016	**-0.0001**	**-0.0086**	**0.0090**	-0.0609	-0.0636	0.0059	0.0050	0.0058	0.0049	**0.2011**	**0.2040**	0.1220	0.1417
	W-VARMA-MVR($J = 100$)	2.21	0	2	-0.0010	-0.0011	-0.0017	0.0002	-0.0086	0.0090	-0.0697	-0.0728	0.0063	0.0056	0.0064	0.0054	0.2011	0.2040	0.1515	0.1759
	Bayes	618345.38	0	0	0.0032	0.0020	-0.0143	-0.0040	0.3421	0.2818	0.5231	0.5289	0.0049	0.0051	0.0160	0.0054	0.4184	0.3506	0.6409	0.6386
	CF-VARMA	1.00	2	13	-0.0025	-0.0007	-0.0023	-0.0011	0.0348	-0.0193	0.0273	-0.0143	0.0094	0.0077	0.0087	0.0083	0.1298	0.1628	0.2607	0.2659
	W-VARMA-MVR($J = 5$)	1.04	0	0	-0.0021	**-0.0003**	-0.0017	-0.0015	**0.0348**	**-0.0193**	**-0.0012**	**-0.0081**	0.0046	**0.0038**	0.0041	0.0044	**0.1298**	**0.1628**	0.0758	0.0813
2000	**W-VARMA-MVR($J = 10$)**	1.10	0	0	-0.0018	-0.0010	**-0.0008**	-0.0011	**0.0348**	**-0.0193**	-0.0157	-0.0266	0.0044	**0.0038**	**0.0037**	**0.0041**	**0.1298**	**0.1628**	**0.0616**	**0.0654**
	W-VARMA-MVR($J = 25$)	1.26	0	0	-0.0011	-0.0008	-0.0010	-0.0010	**0.0348**	**-0.0193**	-0.0381	-0.0320	**0.0042**	**0.0038**	0.0038	**0.0041**	**0.1298**	**0.1628**	0.0706	0.0732
	W-VARMA-MVR($J = 50$)	1.52	0	0	-0.0008	-0.0015	-0.0016	-0.0006	**0.0348**	**-0.0193**	-0.0510	-0.0515	0.0045	0.0040	0.0042	0.0044	**0.1298**	**0.1628**	0.0871	0.1011
	W-VARMA-MVR($J = 100$)	2.06	0	0	**-0.0006**	-0.0013	-0.0015	**-0.0003**	**0.0348**	**-0.0193**	-0.0557	-0.0616	0.0048	0.0044	0.0046	0.0046	**0.1298**	**0.1628**	0.1087	0.1185
	Bayes	632532.49	0	0	0.0028	0.0020	-0.0098	-0.0039	0.1578	0.1287	0.2494	0.2607	0.0040	0.0038	0.0114	0.0052	0.1848	0.1707	0.3032	0.3118
	CF-VARMA	1.00	0	2	-0.0006	0.0011	-0.0037	-0.0016	**0.0089**	**-0.0203**	-0.0137	0.0072	0.0051	0.0038	0.0063	0.0045	**0.0823**	**0.0918**	0.1805	0.1745
	W-VARMA-MVR($J = 5$)	1.03	0	0	-0.0016	**-0.0002**	-0.0019	-0.0017	**0.0089**	**-0.0203**	0.0095	**0.0061**	0.0027	0.0024	0.0035	0.0025	**0.0823**	**0.0918**	0.0554	0.0515
5000	**W-VARMA-MVR($J = 10$)**	1.07	0	0	-0.0012	-0.0006	**-0.0008**	-0.0012	**0.0089**	**-0.0203**	**-0.0092**	-0.0173	0.0026	0.0022	**0.0029**	**0.0021**	**0.0823**	**0.0918**	**0.0453**	**0.0395**
	W-VARMA-MVR($J = 25$)	1.24	0	0	-0.0006	-0.0007	-0.0010	-0.0009	**0.0089**	**-0.0203**	-0.0317	-0.0282	**0.0025**	**0.0021**	**0.0029**	0.0022	**0.0823**	**0.0918**	0.0552	0.0481
	W-VARMA-MVR($J = 50$)	1.50	0	0	**-0.0001**	-0.0013	-0.0016	**-0.0005**	**0.0089**	**-0.0203**	-0.0513	-0.0482	0.0027	0.0026	0.0032	0.0024	**0.0823**	**0.0918**	0.0745	0.0738
	W-VARMA-MVR($J = 100$)	2.00	0	0	**0.0001**	-0.0013	-0.0018	**-0.0005**	**0.0089**	**-0.0203**	-0.0572	-0.0477	0.0029	0.0029	0.0035	0.0025	**0.0823**	**0.0918**	0.0872	0.0849

Table A4

Simulation results for estimated MSV processes: Bayes vs W-VARMA-MVR estimator

True value: $(\phi_{11},\ \phi_{12},\ \phi_{21},\ \phi_{22},\ \mu_1,\ \mu_2,\ \sigma^2_{v,1},\ \sigma^2_{v,2}) = (0.98,\ -0.1,\ 0.12,\ 0.98,\ -0.5,\ -0.5,\ 0.4,\ 0.4)$

Table A4 reports the performance of Bayes and W-VARMA-MVR estimators in bi-variate setups. VARMA-type estimator are proposed in Section 3. Bayes estimator is based on a Metropolis-within-Gibbs sampler [Kim et al. (1998), Omori et al. (2007), Kastner and Frühwirth-Schnatter (2014)]. Prior distributions, MCMC algorithm and effective computations are discussed in Section 6.3.2. The posteriors are based on 20000 draws of the sampler, after discarding 20000 draws. We consider first-order persistence in latent volatilities. The true DGP is given by (2.1)-(2.2) with $m = (2, 5, 10)$, $p = 1$. We set $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = 0.4\boldsymbol{I}_m$, $\mu = -0.5\mathbb{1}$, and stationary $\boldsymbol{\phi}_1$ has the following formations: (1) for $m = 2$, $\boldsymbol{\phi}_1$ is non-diagonal with $(\phi_{11},\ \phi_{12},\ \phi_{21},\ \phi_{22}) = (0.98,\ -0.1,\ 0.12,\ 0.99)$, (2) for $m = 5$, $\boldsymbol{\phi}_1$ is non-diagonal where diagonal elements are $(0.98, 0.985, 0.8, 0.985, 0.98)'$ and upper off-diagonal elements are set to -0.15 and lower off-diagonal elements are set to 0.1, and (3) for $m = 10$, $\boldsymbol{\phi}_1$ is non-diagonal where diagonal elements are $(0.95, 0.98, 0.98, 0.99, 0.8, 0.8, 0.99, 0.98, 0.98, 0.95)'$ and upper off-diagonal elements are set to -0.11 and lower off-diagonal elements are set to 0.055. We simulate 100 samples from each simulated model. RCT stands for the relative computational time w.r.t. the CF-VARMA estimator. The simulated model has $m = 2$ number of time series and the number of parameters (NP) are ($= m^2 + 2m = 8$). The number of inadmissible values (NIV) of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100. Boldface font highlights the smallest bias and RMSE with no NIV/NC. Boldface font also highlights the estimator, which has the best overall performance.

$m=2,\ p=1,\ NP=8$					**Bias**				**RMSE**			
T	Estimators	RCT	NIV	NPD	ϕ	diag[ϕ]	μ	$\boldsymbol{\Sigma}_v$	ϕ	diag[ϕ]	μ	$\boldsymbol{\Sigma}_v$
1000	Bayes	461089.9	0	0	-0.0037	**0.0004**	0.7559	1.0624	0.0092	0.0052	0.9059	1.2668
	CF-VARMA	1.0	9	35	-0.0005	-0.0008	0.0002	0.0059	0.0117	0.0116	0.2025	0.3352
	W-VARMA-MVR($J=10$)	1.1	0	0	**-0.0009**	-0.0014	**0.0002**	**-0.0275**	**0.0050**	**0.0048**	**0.2025**	**0.0965**
	W-VARMA-MVR($J=100$)	1.9	0	2	-0.0009	-0.0004	0.0002	-0.0713	0.0059	0.0058	0.2025	0.1637
2000	Bayes	528485.7	0	0	-0.0032	-0.0001	0.3641	0.5964	0.0084	0.0055	0.4384	0.7244
	CF-VARMA	1.0	2	13	-0.0017	-0.0018	0.0078	0.0065	0.0085	0.0089	0.1463	0.2633
	W-VARMA-MVR($J=10$)	1.1	0	0	-0.0011	-0.0014	**0.0078**	**-0.0212**	**0.0040**	**0.0042**	**0.1463**	**0.0635**
	W-VARMA-MVR($J=100$)	1.8	0	0	**-0.0009**	**-0.0005**	**0.0078**	-0.0587	0.0046	0.0047	**0.1463**	0.1136
5000	Bayes	556495.7	0	0	-0.0022	**-0.0002**	0.1403	0.2681	0.0062	0.0047	0.1857	0.3326
	CF-VARMA	1.0	0	2	-0.0012	-0.0011	-0.0057	-0.0033	0.0049	0.0048	0.0871	0.1775
	W-VARMA-MVR($J=10$)	1.2	0	0	**-0.0009**	-0.0012	**-0.0057**	**-0.0133**	**0.0024**	**0.0023**	**0.0871**	**0.0424**
	W-VARMA-MVR($J=100$)	1.9	0	0	**-0.0009**	**-0.0002**	**-0.0057**	-0.0524	0.0029	0.0027	**0.0871**	0.0861
$m=5,\ p=1,\ NP=35$					**Bias**				**RMSE**			
T	Estimators	RCT	NIV	NPD	ϕ	diag[ϕ]	μ	$\boldsymbol{\Sigma}_v$	ϕ	diag[ϕ]	μ	$\boldsymbol{\Sigma}_v$
1000	Bayes	824292.8	0	0	**0.0002**	-0.0237	0.2157	0.9016	0.0466	0.0477	0.4360	1.0496
	CF-VARMA	1.0	36	48	0.0004	-0.0025	0.0063	0.0593	0.0373	0.0380	0.3185	0.3604
	W-VARMA-MVR($J=10$)	1.1	0	0	**0.0002**	**-0.0086**	**0.0063**	**0.0083**	**0.0167**	**0.0185**	**0.3185**	**0.1241**
	W-VARMA-MVR($J=100$)	1.9	0	20	0.0003	-0.0081	0.0063	-0.0047	0.0201	0.0233	0.3185	0.2124
2000	Bayes	1421792.2	0	0	0.0004	-0.0169	0.0992	0.5548	0.0312	0.0330	0.2694	0.6438
	CF-VARMA	1.0	16	23	0.0000	-0.0044	0.0085	0.0573	0.0270	0.0265	0.2426	0.2809
	W-VARMA-MVR($J=10$)	1.1	0	0	**0.0002**	**-0.0054**	**0.0085**	**-0.0001**	**0.0114**	**0.0123**	**0.2426**	**0.0955**
	W-VARMA-MVR($J=100$)	2.0	0	5	0.0003	-0.0056	0.0085	-0.0118	0.0146	0.0167	0.2426	0.1680
5000	Bayes	1472299.8	0	0	**0.0003**	-0.0094	0.0232	0.2818	0.0167	0.0191	0.1505	0.3261
	CF-VARMA	1.0	3	2	-0.0001	-0.0037	-0.0026	0.0535	0.0156	0.0158	0.1476	0.1793
	W-VARMA-MVR($J=10$)	1.2	0	0	**0.0003**	**-0.0022**	**-0.0026**	**-0.0117**	**0.0071**	**0.0066**	**0.1476**	**0.0525**
	W-VARMA-MVR($J=100$)	1.9	0	0	**0.0003**	-0.0025	**-0.0026**	-0.0256	0.0093	0.0094	**0.1476**	0.1026
$m=10,\ p=1,\ NP=120$					**Bias**				**RMSE**			
T	Estimators	RCT	NIV	NPD	ϕ	diag[ϕ]	μ	$\boldsymbol{\Sigma}_v$	ϕ	diag[ϕ]	μ	$\boldsymbol{\Sigma}_v$
1000	Bayes	1211101.1	0	0	0.0015	**-0.0226**	0.2297	0.5322	0.0838	0.0648	0.8050	0.7140
	CF-VARMA	1.0	71	74	0.0012	-0.0042	-0.0066	0.1232	0.0556	0.0557	0.6828	0.3932
	W-VARMA-MVR($J=10$)	1.1	0	0	**0.0011**	-0.0355	**-0.0066**	**0.1105**	**0.0285**	**0.0455**	**0.6828**	**0.2227**
	W-VARMA-MVR($J=100$)	2.0	0	30	0.0016	-0.0097	-0.0066	0.0065	0.0238	0.0264	0.6828	0.2181
2000	Bayes	1458465.7	0	0	0.0014	**-0.0093**	0.1394	0.2849	0.0618	0.0420	0.5597	0.3976
	CF-VARMA	1.0	42	36	0.0013	-0.0076	-0.0046	0.0905	0.0353	0.0362	0.4862	0.2971
	W-VARMA-MVR($J=10$)	1.1	0	0	**0.0012**	-0.0187	**-0.0046**	**0.0508**	**0.0183**	**0.0260**	**0.4862**	**0.1290**
	W-VARMA-MVR($J=100$)	2.1	0	14	0.0016	-0.0068	-0.0046	-0.0040	0.0198	0.0218	0.4862	0.1759
5000	Bayes	1607087.0	0	0	0.0007	-0.0041	0.0708	0.1734	0.0388	0.0249	0.3291	0.2350
	CF-VARMA	1.0	14	4	0.0013	-0.0026	-0.0036	0.0790	0.0215	0.0216	0.3137	0.2021
	W-VARMA-MVR($J=10$)	1.1	0	0	**0.0011**	-0.0087	**-0.0036**	0.0228	**0.0115**	**0.0132**	**0.3137**	**0.0731**
	W-VARMA-MVR($J=100$)	1.9	0	0	0.0013	**-0.0039**	**-0.0036**	**-0.0127**	0.0151	0.0161	**0.3137**	0.1192

Table A5

Large scale simulation with W-VARMA-MVR estimator: Bias and RMSE

Model: HD_1

Table A5 reports the performance of W-VARMA-MVR($J = 10, 100$) estimator in high-dimensional setups. This method is proposed in Section 3. We consider upto 100 dimensions and first-order persistence in latent volatilities. The true DGP is given by (2.1)-(2.2) with $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = \boldsymbol{I}_m$, $\mu = -\mathbb{1}$, and stationary $\boldsymbol{\phi}_1$ is diagonal where the diagonal elements are equal to 0.95. We simulate 100 samples from each simulated model. Computational time per replication is given in seconds. NP (= $3m$) is the number of parameters in the simulated model and m denotes the number of time series. The number of inadmissible values (NIV) of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100.

			$J = 10$									$J = 100$								
						Bias			RMSE						Bias			RMSE		
T	m	NP	Time	NIV	NPD	ϕ	μ	Σ_v	ϕ	μ	Σ_v	Time	NIV	NPD	ϕ	μ	Σ_v	ϕ	μ	Σ_v
1000	2	6	0.007	0	0	-0.0069	0.0023	-0.0326	0.0159	0.5952	0.0935	0.009	0	0	-0.0041	0.0023	-0.0588	0.0154	0.5952	0.1190
	5	15	0.007	0	0	-0.0066	-0.0624	-0.0402	0.0155	0.6241	0.0926	0.010	0	0	-0.0039	-0.0624	-0.0644	0.0149	0.6241	0.1172
	10	30	0.014	0	0	-0.0060	-0.0236	-0.0385	0.0151	0.6286	0.0900	0.022	0	0	-0.0030	-0.0236	-0.0661	0.0145	0.6286	0.1171
	25	75	0.017	0	0	-0.0061	-0.0151	-0.0383	0.0159	0.6340	0.0963	0.031	0	0	-0.0032	-0.0151	-0.0659	0.0156	0.6340	0.1226
	50	150	0.029	0	0	-0.0059	-0.0082	-0.0411	0.0155	0.6324	0.0970	0.061	0	0	-0.0029	-0.0082	-0.0694	0.0152	0.6324	0.1242
	100	300	0.052	0	0	-0.0061	-0.0155	-0.0397	0.0154	0.6359	0.0950	0.136	0	0	-0.0032	-0.0155	-0.0674	0.0153	0.6359	0.1218
	250	750	0.128	0	0	-0.0060	0.0034	-0.0405	0.0156	0.6324	0.0963	0.505	0	0	-0.0030	0.0034	-0.0686	0.0154	0.6324	0.1234
2500	2	6	0.018	0	0	-0.0038	-0.0580	-0.0469	0.0089	0.4021	0.0687	0.017	0	0	-0.0034	-0.0580	-0.0518	0.0097	0.4021	0.0795
	5	15	0.020	0	0	-0.0020	-0.0100	-0.0506	0.0083	0.3967	0.0732	0.024	0	0	-0.0010	-0.0100	-0.0612	0.0090	0.3967	0.0905
	10	30	0.035	0	0	-0.0027	-0.0104	-0.0463	0.0089	0.4044	0.0702	0.048	0	0	-0.0014	-0.0104	-0.0590	0.0094	0.4044	0.0887
	25	75	0.043	0	0	-0.0023	0.0011	-0.0478	0.0089	0.4040	0.0722	0.076	0	0	-0.0013	0.0011	-0.0587	0.0094	0.4040	0.0897
	50	150	0.072	0	0	-0.0025	-0.0058	-0.0478	0.0088	0.4061	0.0724	0.144	0	0	-0.0013	-0.0058	-0.0593	0.0094	0.4061	0.0901
	100	300	0.133	0	0	-0.0026	0.0065	-0.0472	0.0088	0.4017	0.0718	0.378	0	0	-0.0014	0.0065	-0.0591	0.0094	0.4017	0.0898
	250	750	0.326	0	0	-0.0023	-0.0009	-0.0481	0.0087	0.4024	0.0722	1.441	0	0	-0.0011	-0.0009	-0.0600	0.0093	0.4024	0.0902
5000	2	6	0.033	0	0	-0.0016	-0.0009	-0.0478	0.0063	0.3120	0.0601	0.038	0	0	-0.0011	-0.0009	-0.0532	0.0068	0.3120	0.0715
	5	15	0.030	0	0	-0.0015	0.0054	-0.0472	0.0060	0.3024	0.0595	0.041	0	0	-0.0009	0.0054	-0.0528	0.0067	0.3024	0.0714
	10	30	0.067	0	0	-0.0013	-0.0016	-0.0502	0.0060	0.2871	0.0627	0.098	0	0	-0.0009	-0.0016	-0.0550	0.0068	0.2871	0.0740
	25	75	0.074	0	0	-0.0014	-0.0046	-0.0497	0.0060	0.2926	0.0627	0.134	0	0	-0.0009	-0.0046	-0.0559	0.0068	0.2926	0.0751
	50	150	0.143	0	0	-0.0014	0.0035	-0.0501	0.0059	0.2839	0.0628	0.342	0	0	-0.0007	0.0035	-0.0567	0.0066	0.2839	0.0754
	100	300	0.277	0	0	-0.0013	-0.0004	-0.0509	0.0060	0.2844	0.0637	0.719	0	0	-0.0007	-0.0004	-0.0569	0.0067	0.2844	0.0754
	250	750	0.665	0	0	-0.0013	0.0019	-0.0504	0.0060	0.2851	0.0633	2.609	0	0	-0.0007	0.0019	-0.0565	0.0066	0.2851	0.0749
10000	2	6	0.061	0	0	-0.0010	0.0099	-0.0504	0.0044	0.2069	0.0575	0.072	0	0	-0.0007	0.0099	-0.0532	0.0047	0.2069	0.0633
	5	15	0.062	0	0	-0.0008	-0.0082	-0.0513	0.0041	0.2042	0.0573	0.082	0	0	-0.0004	-0.0082	-0.0549	0.0047	0.2042	0.0653
	10	30	0.126	0	0	-0.0010	-0.0011	-0.0505	0.0043	0.2073	0.0577	0.182	0	0	-0.0007	-0.0011	-0.0534	0.0048	0.2073	0.0643
	25	75	0.156	0	0	-0.0008	0.0062	-0.0515	0.0042	0.2042	0.0579	0.289	0	0	-0.0005	0.0062	-0.0550	0.0047	0.2042	0.0650
	50	150	0.305	0	0	-0.0008	0.0027	-0.0522	0.0042	0.2006	0.0588	0.705	0	0	-0.0004	0.0027	-0.0556	0.0047	0.2006	0.0659
	100	300	0.555	0	0	-0.0007	0.0055	-0.0517	0.0041	0.2029	0.0583	1.817	0	0	-0.0004	0.0055	-0.0550	0.0047	0.2029	0.0652
	250	750	1.377	0	0	-0.0008	0.0022	-0.0514	0.0041	0.2005	0.0580	5.650	0	0	-0.0005	0.0022	-0.0546	0.0047	0.2005	0.0648

Table A6

Large scale simulation with W-VARMA-MVR estimator: Bias and RMSE

Model: HD_2

Table A6 reports the performance of W-VARMA-MVR ($J = 10, 100$) estimators in high-dimensional setups. This method is proposed in Section 3. We consider upto 100 dimensions and first-order persistence in latent volatilities. The true DGP is given by (2.1)-(2.2) with $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = \boldsymbol{I}_m$, $\mu = -\mathbb{1}$, and stationary $\boldsymbol{\phi}_1$ is non-diagonal where for $\boldsymbol{\phi}_1$, diagonal elements are equal to 0.9 and off-diagonal elements are set between $(-0.01, 0.01)$. We simulate 100 samples from each simulated model. Computational time per replication is given in seconds. NP ($= m^2 + 2m$) is the the number of parameters in the simulated model and m denotes the number of time series. The number of inadmissible values (NIV) of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100.

			$J = 10$									$J = 100$								
						Bias			RMSE						Bias			RMSE		
T	m	NP	Time	NIV	NPD	ϕ	μ	$\boldsymbol{\Sigma}_v$	ϕ	μ	$\boldsymbol{\Sigma}_v$	Time	NIV	NPD	ϕ	μ	$\boldsymbol{\Sigma}_v$	ϕ	μ	$\boldsymbol{\Sigma}_v$
1000	2	8	0.006	0	0	-0.0074	-0.0026	-0.0023	0.0266	0.3028	0.1055	0.009	0	0	-0.0038	-0.0026	-0.0291	0.0260	0.3028	0.1035
	5	35	0.007	0	0	-0.0045	-0.0412	0.0216	0.0289	0.3255	0.1003	0.010	0	0	-0.0024	-0.0412	-0.0273	0.0254	0.3255	0.1029
	10	120	0.014	0	0	-0.0044	-0.0145	0.0719	0.0330	0.3221	0.1249	0.023	0	0	-0.0016	-0.0145	-0.0319	0.0238	0.3221	0.0934
	25	675	0.016	0	0	-0.0050	0.0077	0.2138	0.0451	0.3332	0.2522	0.033	0	0	-0.0020	0.0077	-0.0209	0.0259	0.3332	0.0804
	50	2600	0.028	0	25	-0.0039	-0.0022	0.3052	0.0489	0.3485	0.4197	0.062	0	0	-0.0020	-0.0022	-0.0159	0.0311	0.3485	0.0884
	100	10200	0.051	0	100	-0.0030	0.0037	0.3246	0.0475	0.3889	1.3338	0.147	0	99	-0.0023	0.0037	-0.0809	0.0382	0.3889	0.6806
	250	63000	0.128	0	100	-0.0021	0.0024	0.5394	0.0471	0.7174	2.0263	0.512	0	100	-0.0020	0.0024	0.5924	0.0437	0.7174	1.9162
2500	2	8	0.013	0	0	-0.0027	-0.0321	-0.0248	0.0154	0.2055	0.0652	0.019	0	0	-0.0032	-0.0321	-0.0240	0.0163	0.2055	0.0657
	5	35	0.015	0	0	-0.0011	-0.0058	-0.0152	0.0150	0.2031	0.0611	0.024	0	0	-0.0004	-0.0058	-0.0338	0.0153	0.2031	0.0735
	10	120	0.030	0	0	-0.0018	-0.0071	0.0063	0.0167	0.2098	0.0639	0.048	0	0	-0.0008	-0.0071	-0.0398	0.0151	0.2098	0.0752
	25	675	0.036	0	0	-0.0018	0.0004	0.0637	0.0201	0.2135	0.0932	0.070	0	0	-0.0006	0.0004	-0.0390	0.0144	0.2135	0.0677
	50	2600	0.063	0	0	-0.0016	-0.0070	0.1593	0.0256	0.2239	0.1763	0.141	0	0	-0.0005	-0.0070	-0.0421	0.0149	0.2239	0.0653
	100	10200	0.123	0	0	-0.0016	0.0022	0.2735	0.0293	0.2459	0.2920	0.351	0	0	-0.0007	0.0022	-0.0550	0.0171	0.2459	0.0743
	250	63000	0.339	0	100	-0.0012	-0.0006	0.3397	0.0288	0.4457	1.5323	1.432	0	97	-0.0008	-0.0006	-0.1864	0.0228	0.4457	0.3559
5000	2	8	0.029	0	0	-0.0009	-0.0021	-0.0282	0.0100	0.1573	0.0511	0.036	0	0	0.0001	-0.0021	-0.0367	0.0106	0.1573	0.0599
	5	35	0.030	0	0	-0.0010	0.0021	-0.0229	0.0103	0.1547	0.0493	0.044	0	0	-0.0004	0.0021	-0.0385	0.0105	0.1547	0.0619
	10	120	0.062	0	0	-0.0008	-0.0014	-0.0164	0.0109	0.1478	0.0467	0.091	0	0	-0.0003	-0.0014	-0.0428	0.0107	0.1478	0.0638
	25	675	0.073	0	0	-0.0008	-0.0044	0.0118	0.0120	0.1522	0.0481	0.134	0	0	-0.0002	-0.0044	-0.0468	0.0103	0.1522	0.0650
	50	2600	0.141	0	0	-0.0007	0.0035	0.0616	0.0142	0.1580	0.0775	0.351	0	0	-0.0002	0.0035	-0.0500	0.0101	0.1580	0.0646
	100	10200	0.260	0	0	-0.0007	0.0026	0.1499	0.0177	0.1761	0.1591	0.710	0	0	-0.0002	0.0026	-0.0679	0.0102	0.1761	0.0772
	250	63000	0.654	0	0	-0.0007	0.0025	0.2490	0.0201	0.3148	0.2647	2.621	0	0	-0.0003	0.0025	-0.1321	0.0121	0.3148	0.1382
10000	2	8	0.056	0	0	-0.0006	0.0042	-0.0315	0.0073	0.1051	0.0447	0.067	0	0	-0.0004	0.0042	-0.0340	0.0077	0.1051	0.0473
	5	35	0.057	0	0	-0.0004	-0.0048	-0.0311	0.0072	0.1046	0.0426	0.082	0	0	-0.0002	-0.0048	-0.0396	0.0076	0.1046	0.0511
	10	120	0.122	0	0	-0.0004	0.0000	-0.0257	0.0073	0.1064	0.0407	0.173	0	0	0.0000	0.0000	-0.0422	0.0075	0.1064	0.0553
	25	675	0.148	0	0	-0.0004	0.0021	-0.0129	0.0078	0.1077	0.0335	0.285	0	0	0.0000	0.0021	-0.0489	0.0075	0.1077	0.0594
	50	2600	0.276	0	0	-0.0003	0.0003	0.0096	0.0085	0.1132	0.0339	0.760	0	0	0.0000	0.0003	-0.0545	0.0073	0.1132	0.0634
	100	10200	0.549	0	0	-0.0004	0.0008	0.0577	0.0102	0.1230	0.0669	1.674	0	0	-0.0001	0.0008	-0.0707	0.0072	0.1230	0.0768
	250	63000	1.311	0	0	-0.0004	0.0017	0.1527	0.0130	0.2196	0.1580	5.558	0	0	0.0000	0.0017	-0.1374	0.0072	0.2196	0.1403

Table A7

Large scale simulation with W-VARMA-MVR estimator: Bias and RMSE

Model: HD_3

Table A7 reports the performance of W-VARMA-MVR ($J = 10, 100$) estimators in high-dimensional setups. This method is proposed in Section 3. We consider upto 100 dimensions and second-order persistence in latent volatilities. The true DGP is given by (2.1)-(2.2) with $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = \boldsymbol{I}_m$, $\mu = -\mathbb{1}$, and stationary $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are non-diagonal where for $\boldsymbol{\phi}_1$, diagonal elements are equal to 0.9 and off-diagonal elements are set between $(-0.01, 0.01)$, and for $\boldsymbol{\phi}_2$, diagonal elements are equal to -0.85 and off-diagonal elements are set between $(-0.01, 0.01)$. We simulate 100 samples from each simulated model. Computational time per replication is given in seconds. NP ($= 2m^2 + 2m$) is the the number of parameters in the simulated model and m denotes the number of time series. The number of inadmissible values (NIV) of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100.

			$J = 10$									$J = 100$								
						Bias			RMSE						Bias			RMSE		
T	m	NP	Time	NIV	NPD	ϕ	μ	$\boldsymbol{\Sigma}_v$	ϕ	μ	$\boldsymbol{\Sigma}_v$	Time	NIV	NPD	ϕ	μ	$\boldsymbol{\Sigma}_v$	ϕ	μ	$\boldsymbol{\Sigma}_v$
1000	2	12	0.010	0	0	0.0014	-0.0032	-0.0019	0.0388	0.0732	0.1771	0.016	0	0	-0.0029	-0.0032	0.1575	0.0603	0.0732	0.2179
	5	60	0.011	0	0	0.0005	-0.0100	0.0861	0.0506	0.0791	0.1827	0.017	0	0	-0.0024	-0.0100	0.2558	0.0580	0.0791	0.2889
	10	220	0.021	0	0	0.0000	-0.0047	0.2183	0.0687	0.0750	0.2769	0.043	0	0	-0.0011	-0.0047	0.3096	0.0524	0.0750	0.3337
	25	1300	0.026	0	4	0.0004	-0.0006	0.3346	0.0807	0.0772	0.4100	0.066	0	0	-0.0005	-0.0006	0.2954	0.0480	0.0772	0.3198
	50	5100	0.046	0	77	0.0000	0.0007	0.3129	0.0698	0.0783	0.7092	0.148	0	1	-0.0001	0.0007	0.1907	0.0489	0.0783	0.2473
	100	20200	0.097	0	100	0.0001	0.0008	0.4163	0.0586	0.0780	1.8869	0.360	0	98	0.0001	0.0008	0.0103	0.0495	0.0780	0.5152
	250	125500	0.290	0	100	0.0006	0.0010	0.3023	0.0570	0.0777	1.5174	1.576	0	100	0.0006	0.0010	0.4476	0.0536	0.0777	1.7774
2500	2	12	0.023	0	0	-0.0002	-0.0068	-0.0335	0.0216	0.0477	0.1099	0.032	0	0	-0.0022	-0.0068	0.0582	0.0330	0.0477	0.1125
	5	60	0.023	0	0	0.0002	-0.0041	0.0158	0.0247	0.0511	0.1112	0.043	0	0	-0.0016	-0.0041	0.1602	0.0349	0.0511	0.1855
	10	220	0.050	0	0	0.0001	-0.0016	0.0732	0.0289	0.0509	0.1253	0.094	0	0	-0.0009	-0.0016	0.2352	0.0343	0.0509	0.2512
	25	1300	0.058	0	0	0.0000	0.0008	0.2135	0.0422	0.0508	0.2389	0.149	0	0	-0.0006	0.0008	0.2979	0.0308	0.0508	0.3083
	50	5100	0.110	0	0	0.0000	0.0001	0.3136	0.0494	0.0501	0.3376	0.328	0	0	-0.0003	0.0001	0.2915	0.0286	0.0501	0.3015
	100	20200	0.238	0	2	0.0000	0.0012	0.3237	0.0449	0.0498	0.3691	0.876	0	0	-0.0001	0.0012	0.2034	0.0287	0.0498	0.2213
	250	125500	0.642	0	100	0.0008	-0.0001	0.2614	0.0374	0.0497	1.2035	3.447	0	99	0.0007	-0.0001	-0.0450	0.0316	0.0497	0.5568
5000	2	12	0.041	0	0	-0.0003	-0.0037	-0.0235	0.0144	0.0341	0.0782	0.057	0	0	-0.0012	-0.0037	0.0259	0.0191	0.0341	0.0766
	5	60	0.044	0	0	0.0002	0.0007	-0.0077	0.0152	0.0371	0.0747	0.084	0	0	-0.0009	0.0007	0.0891	0.0215	0.0371	0.1115
	10	220	0.093	0	0	0.0000	0.0005	0.0148	0.0170	0.0354	0.0774	0.188	0	0	-0.0007	0.0005	0.1568	0.0232	0.0354	0.1705
	25	1300	0.119	0	0	0.0000	0.0000	0.1006	0.0224	0.0354	0.1243	0.364	0	0	-0.0006	0.0000	0.2490	0.0226	0.0354	0.2563
	50	5100	0.231	0	0	0.0000	0.0008	0.2156	0.0301	0.0352	0.2291	1.061	0	0	-0.0003	0.0008	0.2854	0.0209	0.0352	0.2912
	100	20200	0.480	0	0	0.0000	0.0008	0.2967	0.0344	0.0351	0.3094	1.868	0	0	-0.0001	0.0008	0.2590	0.0195	0.0351	0.2650
	250	125500	1.377	0	64	0.0010	0.0006	0.2514	0.0315	0.0350	0.5182	6.637	0	0	0.0008	0.0006	0.0989	0.0222	0.0350	0.1260
10000	2	12	0.079	0	0	0.0001	-0.0003	-0.0223	0.0101	0.0258	0.0572	0.107	0	0	-0.0004	-0.0003	0.0009	0.0120	0.0258	0.0512
	5	60	0.088	0	0	0.0000	0.0001	-0.0261	0.0101	0.0247	0.0596	0.151	0	0	-0.0006	0.0001	0.0339	0.0134	0.0247	0.0612
	10	220	0.193	0	0	0.0001	0.0009	-0.0112	0.0109	0.0249	0.0528	0.329	0	0	-0.0004	0.0009	0.0868	0.0151	0.0249	0.0986
	25	1300	0.234	0	0	0.0000	0.0008	0.0340	0.0128	0.0246	0.0615	0.623	0	0	-0.0004	0.0008	0.1806	0.0160	0.0246	0.1863
	50	5100	0.474	0	0	0.0000	0.0005	0.1023	0.0162	0.0247	0.1148	1.882	0	0	-0.0003	0.0005	0.2378	0.0157	0.0247	0.2418
	100	20200	1.010	0	0	0.0000	0.0004	0.2072	0.0219	0.0246	0.2143	4.558	0	0	-0.0001	0.0004	0.2513	0.0144	0.0246	0.2548
	250	125500	2.717	1	2	0.0011	0.0004	0.2561	0.0270	0.0245	0.2742	14.464	0	0	0.0009	0.0004	0.1935	0.0162	0.0245	0.1999

Table A8

Large scale simulation with W-VARMA-MVR estimator: Bias and RMSE

Model: UHD

Table A8 reports the performance of W-VARMA-MVR ($J = 10, 100$) estimators in ultra high-dimensional setups. This method is proposed in Section 3. We consider upto 100 dimensions and second-order persistence in latent volatilities. The true DGP is given by (2.1)-(2.2) with $\boldsymbol{\Sigma}_u = \boldsymbol{I}_m$, $\boldsymbol{\Sigma}_v = \boldsymbol{I}_m$, $\mu = -\mathbb{1}$, and stationary $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are non-diagonal where for $\boldsymbol{\phi}_1$, diagonal elements are equal to 0.9 and off-diagonal elements are set between $(-0.001, 0.001)$, and for $\boldsymbol{\phi}_2$, diagonal elements are equal to -0.8 and off-diagonal elements are set between $(-0.001, 0.001)$. We simulate 100 samples from each simulated model. Computational time per replication is given in seconds. NP ($= 2m^2 + 2m$) is the the number of parameters in the simulated model and m denotes the number of time series. The number of inadmissible values (NIV) of $\boldsymbol{\phi}$ and non-positive definiteness (NPD) of $\boldsymbol{\Sigma}_v$ are also reported, these are out of 100.

$J = 10$

						Bias			**RMSE**		
T	m	NP	Time	NIV	NPD	ϕ	μ	$\boldsymbol{\Sigma}_v$	ϕ	μ	$\boldsymbol{\Sigma}_v$
5000	500	501000	4.69	0	14	0.0000	-0.0001	-0.0062	0.0235	0.0387	0.1689
	750	1126500	9.37	0	100	0.0000	0.0001	0.1230	0.0231	0.0386	1.8448
	1000	2002000	21.41	0	100	0.0000	0.0002	0.0061	0.0236	0.0387	1.6758
	1250	3127500	35.99	0	100	0.0000	0.0001	0.1610	0.0246	0.0386	2.6293
	1500	4503000	60.83	0	100	0.0000	0.0002	-0.1271	0.0264	0.0386	1.3389
10000	500	501000	42.11	0	0	0.0000	0.0003	0.1610	0.0180	0.0274	0.1717
	750	1126500	77.61	0	0	0.0000	0.0002	0.0851	0.0170	0.0274	0.1109
	1000	2002000	71.74	0	4	0.0000	0.0001	-0.0017	0.0164	0.0274	0.1136
	1250	3127500	124.14	0	100	0.0000	0.0001	-0.0659	0.0162	0.0273	1.0176
	1500	4503000	138.32	0	100	0.0000	0.0000	0.0167	0.0162	0.0273	1.4972
15000	500	501000	132.06	0	0	0.0000	0.0002	0.1903	0.0152	0.0223	0.1956
	750	1126500	162.46	0	0	0.0000	0.0001	0.1612	0.0146	0.0222	0.1683
	1000	2002000	172.26	0	0	0.0000	0.0000	0.1122	0.0140	0.0221	0.1244
	1250	3127500	224.94	0	0	0.0000	0.0000	0.0571	0.0136	0.0222	0.0853
	1500	4503000	261.58	0	0	0.0000	0.0000	-0.0006	0.0134	0.0222	0.0886

$J = 100$

						Bias			**RMSE**		
T	m	NP	Time	NIV	NPD	ϕ	μ	$\boldsymbol{\Sigma}_v$	ϕ	μ	$\boldsymbol{\Sigma}_v$
5000	500	501000	10.69	0	0	0.0000	-0.0001	-0.2145	0.0192	0.0387	0.2253
	750	1126500	20.69	0	100	0.0000	0.0001	-0.3619	0.0205	0.0386	0.6218
	1000	2002000	45.60	0	100	0.0000	0.0002	0.2923	0.0216	0.0387	3.3618
	1250	3127500	74.80	0	100	0.0000	0.0001	-0.0405	0.0230	0.0386	1.3195
	1500	4503000	124.82	0	100	0.0000	0.0002	-0.0930	0.0250	0.0386	1.3094
10000	500	501000	92.57	0	0	0.0000	0.0003	-0.0195	0.0115	0.0274	0.0401
	750	1126500	161.02	0	0	0.0000	0.0002	-0.1207	0.0127	0.0274	0.1271
	1000	2002000	153.56	0	0	0.0000	0.0001	-0.2160	0.0135	0.0274	0.2214
	1250	3127500	253.47	0	0	0.0000	0.0001	-0.2897	0.0140	0.0273	0.2970
	1500	4503000	287.77	0	93	0.0000	0.0000	-0.3654	0.0144	0.0273	0.5451
15000	500	501000	278.95	0	0	0.0000	0.0002	0.0372	0.0086	0.0223	0.0462
	750	1126500	339.46	0	0	0.0000	0.0001	-0.0201	0.0094	0.0222	0.0349
	1000	2002000	365.87	0	0	0.0000	0.0000	-0.0874	0.0100	0.0221	0.0927
	1250	3127500	471.17	0	0	0.0000	0.0000	-0.1554	0.0105	0.0222	0.1592
	1500	4503000	554.78	0	0	0.0000	0.0000	-0.2171	0.0109	0.0222	0.2207

Table A9
Portfolio performance under different conditional volatility models
Annualized Sharpe Ratio

Table A9 reports the performance of conditional volatility models, i.e., the annualized Sharpe Ratio, in the context of asset allocation. We consider the global minimum variance (GMV) portfolio where the one step ahead conditional forecast of the covariance matrix $\Sigma_{t+1|t}$ uniquely defines the optimal portfolio weights $\hat{\mathbf{w}}_{t+1} = \hat{\Sigma}^{-1}_{t+1|t}\iota_m / \iota'_m \hat{\Sigma}^{-1}_{t+1|t}\iota_m$, where ι_m denotes an m-variate vector of ones. The competing models are: the Risk Metrics 2006 methodology [Zumbach (2007)] denoted as RM, the dynamic conditional correlation with composite likelihood [Engle (2002), Pakel et al. (2021)] denoted as cDCC, the conditional covariance estimation using the dynamic factor model approach with finite-dimensional space [Alessi et al. (2009), Aramonte et al. (2013)] denoted as fDCC, the generalized orthogonal GARCH model of van der Weide (2002) denoted as GO-GARCH with three different specification [standard GARCH with multivariate normal distribution (GO-GARCH-S) and multivariate affine normal inverse Gaussian distribution (GO-GARCH-S-MANIG), and GJR with multivariate normal distribution (GO-GARCH-GJR)], unrestricted MSV(p) models (U-MSV(p)) and restricted MSV(p) models (R-MSV(p)) where we consider uncorrelated SV(p) structure. The initial in-sample period from 2009/01/02 to 2018/12/31 and the out-of-sample is from 2019/01/02 to 2020/12/31. The out-of-sample includes a highly volatile period, *i.e.*, the Covid-19 pandemic. The forecasts of MSV(p) models are based on the W-VARMA-MVR estimator with $J = 50$. Boldface font highlights the models which has higher annualized Sharpe Ratio over the equal weight portfolio and Boldface color font highlights the best model.

	20 Asset			50 Asset		
	2019-2020	2019	2020	2019-2020	2019	2020
RM	-1.00	-1.27	-1.04	-0.49	-2.04	-0.04
cDCC	-1.40	-0.60	-2.02	-1.10	-1.05	-1.23
fDCC	-1.10	-1.44	-1.04	-1.01	-1.54	-0.83
GOGARCH-S	-0.87	-1.19	-0.87	-	-	-
GOGARCH-GJR	-1.09	-1.12	-1.23	-	-	-
GOGARCH-S-MANIG	-0.97	-1.17	-1.02	-1.10	-0.93	-1.31
U-MSV(1)	-0.64	-0.91	**0.07**	**0.80**	**1.02**	**0.48**
U-MSV(2)	-0.18	-0.14	-1.10	**0.25**	-0.37	**0.68**
U-MSV(3)	**0.67**	-0.16	**1.04**	**0.48**	**1.36**	-0.49
U-MSV(4)	-0.71	-0.84	-0.86	**0.19**	**0.58**	-0.12
U-MSV(5)	**1.41**	**1.50**	**1.52**	**0.14**	-1.00	**1.23**
R-MSV(1)	-1.26	-1.08	-1.41	-1.17	-1.38	-1.09
R-MSV(2)	-1.29	-1.10	-1.46	-1.20	-1.36	-1.13
R-MSV(3)	-1.26	-1.11	-1.41	-1.22	-1.35	-1.16
R-MSV(4)	-1.32	-1.02	-1.57	-1.14	-1.31	-1.06
R-MSV(5)	-1.33	-1.06	-1.56	-0.95	-1.23	-0.81