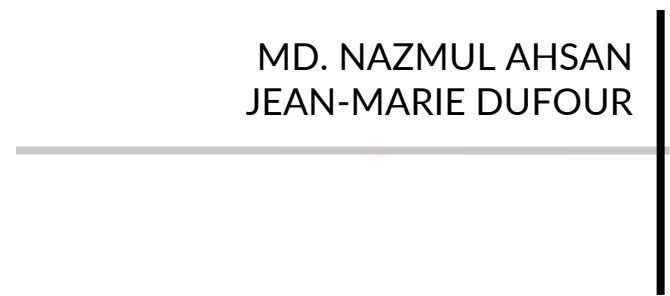


PRACTICAL ESTIMATION METHODS
FOR HIGH-DIMENSIONAL
MULTIVARIATE STOCHASTIC
VOLATILITY MODELS

MD. NAZMUL AHSAN
JEAN-MARIE DUFOUR



Les **cahiers de la série scientifique** visent à rendre accessibles les résultats des recherches effectuées par des chercheurs membres du CIRANO afin de susciter échanges et commentaires. Ces cahiers sont rédigés dans le style des publications scientifiques et n'engagent que leurs auteurs.

The purpose of the Working Papers is to disseminate the results of research conducted by CIRANO research members in order to solicit exchanges and comments. These reports are written in the style of scientific publications. The ideas and opinions expressed in these documents are solely those of the authors.

Le **CIRANO** est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO – CIRANO Partners

Partenaires Corporatifs - Corporate Partners

Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque Nationale du Canada
Bell Canada
BMO Groupe financier
La Caisse
Énergir
Hydro-Québec
Intact Corporation Financière
Mouvement Desjardins
Power Corporation du Canada
Pratt & Whitney Canada
VIA Rail Canada

Partenaires gouvernementaux - Governmental partners

Ministère des Finances du Québec
Ministère de l'Économie, de l'Innovation et de l'Énergie
Innovation, Sciences et Développement Économique
Canada
Ville de Montréal

Partenaires universitaires - University Partners

École de technologie supérieure
École nationale d'administration publique
HEC Montréal
Institut national de la recherche scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web. CIRANO collaborates with many centres and university research chairs; list available on its website.

© Octobre 2026. Md. Nazmul Ahsan, Jean-Marie Dufour. Tous droits réservés. *All rights reserved.* Reproduction partielle permise avec citation du document source, incluant la notice ©. *Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*

Les idées et les opinions émises dans cette publication sont sous l'unique responsabilité des auteurs et ne représentent pas les positions du CIRANO ou de ses partenaires. *The observations and viewpoints expressed in this publication are the sole responsibility of the authors; they do not represent the positions of CIRANO or its partners.*

ISSN 2292-0838 (version en ligne)

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

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

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 small set of

*The authors thank the Editor Viktor Todorov, two anonymous referees, Torben Andersen, Robert Engle, Stephen Taylor, Neil Shephard, Christian Gourin roux, Peter Hansen, Nour Meddahi, Alastair Hall, Massimiliano Marcellino, Dennis Kristensen, Bertille Antoine, Russell Davidson, Prosper Dovonon, Hashem Pesaran, Ren  Garcia, Eric Renault, Frank Kleibergen, Jean-Pierre Florens, Lynda Khalaf, Boris Kozyrev, Joann Jasiak, Richard Luger, Richard Baillie, John Galbraith, Victoria Zinde-Walsh, Denis Pelletier, Firmin Doko Tchatoka, Pipat Wongsart, Maria Grith, Gabriel Rodriguez-Rondon, Masaya Takano, Raida Alam for their valuable comments and constructive discussions. Earlier versions of this paper were presented at the Latin American Meeting of the Econometric Society - LAMES 2026 (Lima, Peru), IWH Macro Network Day 2026, Annual Conference of the International Association for Applied Econometrics - IAAE 2026 (Lisbon, Portugal), 7th Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance, University of Mannheim (Mannheim, Germany), Econometric Society World Congress - ESWC 2025 (Seoul, South Korea), Annual Conference of the International Association for Applied Econometrics - IAAE 2025 (Torino, Italy), Conference to Honor Frank Diebold 65 + 35 (University of Pennsylvania), Annual Conference of the International Association for Applied Econometrics - IAAE 2024 (Thessaloniki, Greece), 58th Annual Meetings of the Canadian Economics Association, CIREQ-CMP Econometrics Conference in Honor of Eric Ghysels, Financial Econometrics Conference (Toulouse School of Economics), CIREQ Econometrics Conference (Recent Developments in Econometrics), Econometrics for Finance (JEF 2023, Rabat, Morocco), Financial Econometrics Conference to mark Stephen Taylor's Retirement (Lancaster University), 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), and Bernoulli-IMS Symposium 2020.

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 Universidad 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 soci t  et la culture (Qu bec).

[†] Canada Mortgage and Housing Corporation (CMHC), 700 Montreal Road, Ottawa, Ontario K1A 0P7, Canada; e-mail: mahsan@cmhc-schl.gc.ca.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the position of the Canada

Mortgage and Housing Corporation.

[‡] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en  conomie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montr al, Qu bec 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>

moment equations derived from the weak VARMA representations of MSV models. Except for the cross-dependence parameters, closed-form expressions are obtained for the remaining parameters, so that no numerical optimization is required. To improve the stability and efficiency of the volatility- persistence parameter estimates, we develop shrinkage-type VARMA estimators based on averaging different simple estimates (winsorization) and matrix-variate regressions (MVR). We also derive the asymptotic distribution of these estimators. Due to their computational simplicity, VARMA estimators make reliable — possibly exact — simulation-based inference tractable through Monte Carlo test techniques. We formulate volatility commonality as a testable common-C-space restriction on the MSV transition matrices, and develop finite-sample Monte Carlo tests for this hypothesis. Simulation results show that the proposed shrinkage-type VARMA-MVR estimators compare favorably with Bayesian and QML benchmarks in terms of bias, RMSE, and computational cost. They also remain feasible in very high dimensions, and, using auxiliary particle filtering, deliver accurate latent-state recovery and covariance forecasts even under model misspecification. In empirical applications to stock returns, we find that unrestricted higher-order MSV models perform well for covariance forecasting and portfolio building, while the data provide little support for an exact low-dimensional common dynamic subspace.

Résumé

Nous proposons des estimateurs efficaces et peu coûteux en ressources de calcul pour les modèles de volatilité stochastique multivariée (MSV) intégrant des dépendances croisées, une causalité au sens de Granger et une persistance d'ordre supérieur des volatilités latentes. La classe d'estimateurs proposée repose sur un nombre restreint d'équations de moments dérivées des représentations VARMA faibles des modèles MSV. À l'exception des paramètres de dépendance croisée, des expressions analytiques explicites sont obtenues pour tous les autres paramètres, ce qui élimine la nécessité de recourir à une optimisation numérique. Afin d'améliorer la stabilité et l'efficacité des estimations des paramètres de persistance de la volatilité, nous développons des estimateurs VARMA de type shrinkage (rétrécissement), fondés sur la combinaison de différentes estimations simples (winsorisation) et sur des régressions matricielles (matrix-variate regressions, MVR). Nous établissons également la distribution asymptotique de ces estimateurs.

Grâce à leur simplicité de calcul, les estimateurs VARMA permettent de réaliser des inférences statistiques fiables, voire exactes, fondées sur la simulation au moyen de techniques de tests de Monte-Carlo. Nous formulons l'existence d'une composante commune aux volatilités sous la forme d'une restriction d'espace C commun (common-C-space) sur les matrices de transition des modèles MSV, et élaborons des tests de Monte-Carlo en échantillon fini pour vérifier cette hypothèse.

Les résultats des simulations montrent que les estimateurs VARMA-MVR de type shrinkage proposés se comparent favorablement aux méthodes de référence bayésiennes et à celles fondées sur le quasi-maximum de vraisemblance (QML), tant sur le plan du biais et de la racine de l'erreur quadratique moyenne (RMSE) que du coût de calcul. Ils demeurent également applicables à des modèles de très grande dimension et permettent, grâce au filtrage

particulaire auxiliaire, de reconstituer avec précision les états latents et de produire des prévisions fiables des covariances, même lorsque le modèle est mal spécifié.

Dans les applications empiriques aux rendements boursiers, nous constatons que les modèles MSV non restreints d'ordre supérieur offrent de bonnes performances pour la prévision des covariances et la construction de portefeuilles, tandis que les données fournissent peu d'éléments à l'appui de l'existence d'un sous-espace dynamique commun exact de faible dimension.

Mots-clés/Keywords : Méthodes simples, régression matricielle, méthodes de Monte-Carlo par chaînes de Markov, tests de Monte-Carlo, volatilité stochastique multivariée, distribution asymptotique / Simple methods, matrix-variate regression, Markov Chain Monte Carlo, Monte Carlo tests, multivariate stochastic volatility, asymptotic distribution.

Codes JEL/JEL Codes : C15, C32, C53, C55, C58, G12.

Pour citer ce document / To quote this document

Ahsan, M. N., & Dufour, J.-M. (2026). Practical estimation methods for high-dimensional multivariate stochastic volatility models (2026s-20, Cahiers scientifiques, CIRANO.)

<https://doi.org/10.54932/HZOV4270>

1 Introduction

Dynamic conditional covariance matrices play a pivotal role in various aspects of financial decision-making, encompassing asset pricing, portfolio selection, option pricing, and risk management. Asset allocation, for instance, heavily relies on asset covariances within a portfolio. Multivariate volatility models also allow one to investigate volatility and covariance transmission and spillover effects across financial market contagion. The literature on modelling and forecasting covariation among financial returns with parametric models can be categorized into two broad classes: (1) multivariate GARCH models (MGARCH), where volatility is modelled as a deterministic process based on past observations [Engle (1982), Bollerslev (1986)], and (2) multivariate stochastic volatility (MSV) models, wherein volatility is modeled as a latent stochastic process [Taylor (1982), Harvey et al. (1994)]. An important distinctive feature of MSV models (with respect to GARCH-type models) lies in the inclusion of additional error terms in the latent volatility process, which capture new market information. Consequently, given the past information set $\mathcal{F}_{t-1}$, volatilities are not known functions of past information, but remain unobserved random variables.

The advantages of stochastic volatility (SV) models compared to GARCH models include the following. (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. While these modifications improve standard GARCH models, they may not be needed for SV models [Carnero et al. (2004), Chan and Grant (2016)]. (3) SV models provide more accurate forecasts than those offered by GARCH models [Kim et al. (1998), Koopman et al. (2005), Poon and Granger (2003), Yu (2002)].

In MSV models, the return variation dynamics is modeled as a latent autocorrelated stochastic process, and the likelihood of the model is given by a high-dimensional integral which makes estimation by conventional maximum likelihood difficult. This is a general feature of nonlinear latent variable models, because the latent variables must be integrated out of the joint density for the observed and latent processes, leading to a high-dimensional integral; see the discussions and reviews in Ghysels et al. (1996), Broto and Ruiz (2004), Ahsan and Dufour (2019, 2021). Thus, estimating a multivariate stochastic volatility model is a challenging task, and several alternative models and methods have been developed.

Three broad classes of multivariate stochastic volatility models have emerged:

1. Multivariate extensions of the univariate class of SV models; see Harvey et al. (1994), Danielsson (1998), Smith and Pitts (2006), Chan et al. (2006), and others.
2. Multivariate factor stochastic volatility (MFSV); see Harvey et al. (1994), Jungbacker and Koopman (2006), Pitt and Shephard (1999), Aguilar and West (2000), So and Choi (2009), Chib et al. (2006), Han (2006), and 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 (2006a, 2006b), Gouriou (2006), Gouriou et al. (2009), and others.

Despite recent advances, the estimation and forecasting of high-dimensional stochastic volatility models remain challenging problems. Limitations of existing methods and models include the following.

- The 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], if not all, suffer from a common problem, the well-known *curse of dimensionality*, whereby methods become inaccurate or infeasible when applied to several variables. In some cases, the methods may become computationally intractable for even 5 or 6 assets.
- Simplified restricted models – In order to meet the need of introducing time-varying variances with practical computational problems, restricted models are generally used. The introduction of strong restrictions can affect the interpretation and flexibility of the models, with a possible impact on the purportedly improved performance they may provide. For example, many studies focus on a multivariate AR(1) specification for latent log volatilities, or a multivariate random walk. These models fail to capture the lead-lag relationships between time series, causality, spillovers, etc. It is also important to note that orthogonality restrictions can be misleading. So and Choi (2009) observed that off-diagonal elements of ϕ 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 derive the analytical expression for the asymptotic variance-covariance matrix. For the MCL method, with a large cross-section, analytical derivatives are difficult to compute, while numerical derivatives (with a large state vector) tend to be 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 depend heavily on prior distributions (which may lead to large biases). Further, for data sets with many assets and observations, Bayes methods become extremely time consuming and empirically impractical.

In this paper, we develop 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. The proposed estimators (based on the VARMA representation) 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 simple estimators can be interpreted as analytically tractable moment-based methods for the MSV(p) class studied in this paper, rather than as fully model-agnostic procedures for all MSV parameterizations. Their tractability relies on the log-squared transformation of returns and on the weak VARMA representation of the transformed observed process, as in multivariate extensions of the standard univariate SV model. The present paper extends this class to higher-order, high-dimensional, and cross-dependent MSV settings. Alternative parameterizations, such as triangular or Cholesky-type SV models, generally imply different transformed moment restrictions. Direct application of the proposed approach to such models would therefore require either a suitable mapping to an MSV-type innovation process or new model-specific moment equations.

Our modeling framework is particularly relevant for macroeconomic research because it can accommodate substantially richer systems than conventional small- or medium-scale models, including larger sets of sectors, parameters, endogenous variables, exogenous shocks, and observables. Recent work shows that stochastic covariance structures offer a more realistic description of macroeconomic dynamics, with important applications to density forecasting in DSGE models [Diebold et al. (2017)] and the measurement of uncertainty [Jurado et al. (2015)]. These covariances also yield improved statistical fit [see, among others, Primiceri (2005) and Koop and Korobilis (2013)], and they are central to the analysis of large-scale 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)]. In this setting, the proposed method can be interpreted as a scalable second-moment module: one may first estimate a conditional-mean block, for example in a VAR, FAVAR, or related macroeconomic system, construct pseudo-innovations, and then fit the MSV model to the resulting innovation sequence.

The proposed closed-form VARMA-type estimators may violate stationarity conditions or fail 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 ϕ by adjusting the eigenvalues which lie on or outside the unit circle. Further, one can apply the modified Cholesky algorithm of Cheng and Higham (1998) to obtain positive definite covariance matrices.

We also suggest winsorized versions of the VARMA estimator (*W-VARMA* estimators), which substantially increase the probability of getting acceptable values and improve 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, particularly the W-VARMA-MVR (abbreviated as MVR) method, enhances the stability of parameter solutions, improves estimation precision, and ensures the positive definite estimates of volatility innovation covariance matrices (confirmed in our simulations).

The closed-form VARMA estimator has computational complexity of order $\mathcal{O}((pm)^3)$, where m denotes

the dimension and p the VAR lag order, because it requires solving a block linear system of dimension pm . The MVR estimator retains the same cubic dependence on pm , since the matrix inversion [see (3.14) below] is still carried out on a $(pm) \times (pm)$ matrix. Its total computational burden nevertheless increases with the truncation parameter J , because the stacked block moment matrices entering the associated cross-product calculations grow with J . Accordingly, a convenient upper-bound description of its overall cost is $\mathcal{O}(J(pm)^3)$, although the factor J reflects the construction of the stacked moment matrices rather than an increase in the dimension of the matrix inversion itself. Despite this additional cost, the MVR method offers greater statistical robustness and can reduce bias, especially in settings with small samples or outliers, by incorporating $J - 1$ additional block moment conditions. It also remains computationally less expensive than Bayesian, QML, or GMM estimation methods for the MSV(p) model. This advantage arises because such alternative methods rely on iterative optimization or simulation-based procedures, with repeated updates of latent states and/or parameters over the sample, whereas the MVR estimator is obtained from a one-shot block linear-algebra calculation based on second-order autocovariances. The paper also shows how the estimated MSV(p) parameters can be used to recover latent log-volatility states and to forecast conditional covariance matrices using Gaussian and particle-based filtering methods.

We derive the asymptotic properties of the proposed estimators under standard regularity assumptions, showing consistency and asymptotic normality. The MVR estimator introduced in this paper can be regarded as a parsimonious moment-based estimator, based on a limited number of carefully chosen moments. In the context of moment-based (or GMM) inference, an excessive number of moment conditions can incur considerable costs in terms of estimation efficiency and forecasting.

The MVR estimator offers several practical benefits. Since MSV models are parametric with a finite number of parameters, these simple estimators enable the construction of simulation-based tests – including exact tests, in contrast to methods relying on establishing asymptotic distributions – and/or prediction intervals using the Monte Carlo test (MCT) technique [see [Dufour \(2006\)](#)]. Tests derived in this manner do not require stationarity assumptions, hence remain transparently applicable when the latent volatility process has a unit root (or is close to this structure). Additionally, the estimators are ideal for procedures requiring repeated estimation using rolling windows, such as backtesting risk measures like Value-at-Risk or Expected Shortfall in risk management. Our simple estimators can be effortlessly applied to very large samples with many assets, common in empirical finance (e.g., high-frequency financial returns). By contrast, estimators based on simulation techniques and/or numerical optimization often require substantial computational effort to achieve convergence.

We evaluate the statistical properties of the proposed estimators through Monte Carlo comparisons with Bayesian and QML benchmarks. Relative to these benchmarks, the MVR estimator delivers lower bias and RMSE in the simulations, while retaining a substantial computational advantage. Additional Monte Carlo results show that particle-filtering procedures, especially the auxiliary particle filter, improve latent-state recovery and one-step-ahead forecasting relative to the Gaussian Kalman filter. Comparisons

with a benchmark factor stochastic volatility (FSV) model also indicate stronger latent-state recovery and competitive or superior one-step-ahead volatility forecasting in the designs considered. Finally, high-dimensional simulations show that the MVR estimator remains reliable for models with up to 1,500 dimensions and 4,503,000 parameters.

We examine the empirical relevance of the proposed estimators using stock return data across several applications. *First*, we consider high-frequency panels with many assets and observations, and show that the proposed estimation methods can be used to estimate rich volatility dynamics in large financial panels. *Second*, in dynamic minimum-variance portfolio applications, the best covariance-forecasting performance is delivered by unrestricted higher-order MSV models, indicating more broadly that the main empirical gains come from a flexible higher-order MSV covariance structure. *Third*, we examine volatility commonality through a common dynamic C -space restriction using finite-sample LMC/MMC procedures. The empirical evidence indicates that any low-dimensional concentration in the unrestricted volatility dynamics is not strong enough to support an exact common dynamic subspace. *Finally*, we compare the fitted asset-level volatility paths implied by unrestricted MSV and FSV specifications and find that the MSV- and FSV-implied volatility paths are not close substitutes overall.

The paper is organized as follows. Section 2 specifies the model and its assumptions. Section 3 proposes simple estimators and discusses conditional covariance forecasting under MSV(p). Section 4 develops asymptotic theory for the simple estimators. Section 5 develops the common- C -space restriction, and the associated finite-sample Monte Carlo tests. Section 6 presents the simulation study, Section 7 reports the empirical applications, and Section 8 discusses scope, two-step implementations for models with nontrivial conditional means, modular dependence extensions based on MSV-standardized residuals, and other practical extensions. We conclude in Section 9. The proofs, other discussions, figures and tables are given in a Technical Appendix.

2 Framework

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

Assumption 2.1. MULTIVARIATE STOCHASTIC VOLATILITY MODEL OF ORDER p . *Let $\mathbf{y}_t := (y_{1t}, \dots, y_{mt})'$ denote the m -dimensional vector of observations at time t , and $\mathbf{h}_t := (h_{1t}, \dots, h_{mt})'$ the corresponding vector of log volatilities. The process $\{\mathbf{y}_t : t \in \mathbb{N}_0\}$ follows an MSV(p) model of the form*

$$\mathbf{y}_t = \mathbf{V}_t^{1/2} \mathbf{u}_t, \quad \mathbf{V}_t^{1/2} := \text{diag}[\exp(h_{1t}/2), \dots, \exp(h_{mt}/2)], \quad (2.1)$$

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

where $\mathbf{u}_t := (u_{1t}, \dots, u_{mt})'$, $\mathbf{v}_t := (v_{1t}, \dots, v_{mt})'$, and $\boldsymbol{\mu} := (\mu_1, \dots, \mu_m)'$. The innovation sequence $\{(\mathbf{u}_t', \mathbf{v}_t')' : t \in \mathbb{N}_0\}$

$\mathbb{Z}$ is independent and identically distributed over time, with

$$\begin{pmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{pmatrix} \sim \mathcal{N}_{2m}[\mathbf{0}, \Sigma], \quad \Sigma = \begin{pmatrix} \Sigma_u & \mathbf{0} \\ \mathbf{0} & \Sigma_v \end{pmatrix}, \quad (2.3)$$

where Σ_u and Σ_v are positive definite. For identification, the diagonal elements of Σ_u are normalized to one, so that $\Sigma_u = (\rho_{ij})$ is a positive-definite correlation matrix.

Remark 2.1. FEATURES OF THE MSV(p) MODEL.

1. The elements of $\mathbf{u}_t$ (typically return shocks) are allowed to be mutually correlated, and similarly for the elements of $\mathbf{v}_t$ (the volatility shocks).
2. The $\boldsymbol{\phi}_j$'s allow for the following features.
 - Non-zero off-diagonal elements of $\boldsymbol{\phi}_j$ capture Granger causalities in volatility between assets — this also generates cross-dependence in asset return volatilities.
 - For $j > 1$, non-zero off-diagonal elements of $\boldsymbol{\phi}_j$ capture causalities in volatilities.
 - Non-zero diagonal elements of $\boldsymbol{\phi}_j$ capture persistence in log volatilities.
 - For $j > 1$, non-zero diagonal elements of $\boldsymbol{\phi}_j$ capture higher-order persistence in log volatilities.
3. If Σ_v is not a diagonal matrix in the VAR(p) model (2.2), then log volatilities $\mathbf{h}_t$'s are instantaneously correlated (or contemporaneously correlated). In this case, $\mathbf{h}_t$'s have instantaneous causality, which goes in both ways.

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

Assumption 2.2. STATIONARITY. *The companion matrix $\mathcal{A}$ associated with (2.2), defined in (E7), is stable, that is, all its eigenvalues lie strictly inside the unit circle. Equivalently, the autoregressive matrix polynomial $\boldsymbol{\phi}(z) := \mathbf{I}_m - \sum_{j=1}^p \boldsymbol{\phi}_j z^j$ satisfies $\det[\boldsymbol{\phi}(z)] \neq 0$ for all $|z| \leq 1$.*

Under Assumptions 2.1-2.2, the state process $\{\mathbf{h}_t : t \in \mathbb{Z}\}$ admits a unique causal strictly stationary solution.

Remark 2.2. NUMBER OF PARAMETERS. The number of parameters for the first-order model is $2m^2 + m$, while for the higher-order model, it is $(p+1)m^2 + m$.

Transforming $\mathbf{y}_t$ by taking logarithms of the squares and using $\mathbb{E}(\log \circ (\mathbf{u}_t^2)) = (-1.2704)\iota_m$ [where ι_m is an $m \times 1$ vector of ones], we can write the measurement equation of the model:

$$\log \circ (\mathbf{y}_t^2) = \mathbf{h}_t + \log \circ (\mathbf{u}_t^2) = \mathbb{E}(\log \circ (\mathbf{u}_t^2)) + \mathbf{h}_t + [\log \circ (\mathbf{u}_t^2) - \mathbb{E}(\log \circ (\mathbf{u}_t^2))] = (-1.2704)\iota_m + \mathbf{h}_t + \boldsymbol{\epsilon}_t \quad (2.4)$$

where the logarithm and square are applied componentwise. Setting $\mathbf{X}_t = \log \circ (\mathbf{y}_t^2) + (1.2704)\iota_m$ and $\epsilon_t = [\log \circ (\mathbf{u}_t^2) - \mathbb{E}(\log \circ (\mathbf{u}_t^2))]$, and combining with (2.2), the MSV(p) model can be written as a linear state space model of the type:

$$\text{(State Transition Equation)} \quad \mathbf{h}_t = \mu + \sum_{j=1}^p \phi_j (\mathbf{h}_{t-j} - \mu) + \mathbf{v}_t, \quad (2.5)$$

$$\text{(Measurement Equation)} \quad \mathbf{X}_t = \mathbf{h}_t + \epsilon_t, \quad (2.6)$$

where $\mathbf{h}_t$ is the logarithm of latent daily volatilities, $\mathbf{X}_t$ is the logarithm of the daily squared return corrected by its mean, $(-1.2704)\iota_m$, and $\mathbf{v}_t$ is the error given in Assumption 2.1. Under the standard normality assumption on $\mathbf{u}_t$, the transformed error ϵ_t is an m -variate mean-zero $\log(\chi_1^2)$ random variable with covariance matrix Σ_ϵ . Harvey et al. (1994) showed that the (i, j) th element of the covariance matrix of $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{mt})'$ is given by $(\pi^2/2)\rho_{ij}^*$, where $\rho_{ii}^* = 1$ and

$$\rho_{ij}^* = \frac{2}{\pi^2} \sum_{n=1}^{\infty} \frac{(n-1)!}{[\prod_{k=1}^n (1/2 + k - 1)] n} \rho_{ij}^{2n}. \quad (2.7)$$

The absolute values of the unknown parameters in Σ_u , namely the ρ_{ij} 's (the cross-correlations between different u_{it} 's), can be estimated, but their signs cannot be directly determined because the relevant information is lost when the observations are squared. However, as suggested by Harvey et al. (1994), the signs of these estimates can be recovered from the untransformed observations. Specifically, for each pair of series $i, j = 1, \dots, m$, the sign of $u_{it}u_{jt}$ coincides with that of the corresponding observed product $y_{it}y_{jt}$, for $t = 1, \dots, T$. Accordingly, the sign of ρ_{ij} is estimated as positive if more than half of the observed products $y_{it}y_{jt}$ are positive, and negative otherwise.

3 Simple estimation methods

This section develops computationally simple estimation methods for MSV(p) models, introduces refinements to improve stability and finite-sample performance, and then extends the framework to covariance forecasting under the MSV(p) specification.

Our estimation strategy exploits the weak VARMA representation of the transformed observed process. We first use this VARMA representation to derive block Yule-Walker-type moment equations for the latent volatility-transition matrices. These moment equations yield closed-form estimators of the MSV parameters, avoiding numerical optimization except for the cross-dependence parameters. We then introduce winsorized estimators, including the MVR version, which combine additional higher-lag moment equations to improve stability and efficiency. The resulting parameter estimates provide the inputs for latent-state filtering and conditional covariance forecasting.

3.1 VARMA-based estimation

This section proposes a simple closed-form estimator for MSV(p) based on the weak VARMA representation of $\mathbf{X}_t$. This representation and its autocovariances are given in the following proposition and corollaries.

Proposition 3.1. VARMA REPRESENTATION OF MSV(p) MODELS. *Under Assumptions 2.1-2.2, the process $\mathbf{X}_t$ admits a causal and invertible weak VARMA(p, p) representation of the form*

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

where $\{\eta_t\}$ is an m -dimensional weak white-noise process with positive definite covariance matrix $\boldsymbol{\Sigma}_\eta$, and $\eta_t - \sum_{j=1}^p \boldsymbol{\theta}_j \eta_{t-j} = \mathbf{v}_t + \epsilon_t - \sum_{j=1}^p \boldsymbol{\phi}_j \epsilon_{t-j}$. Writing $\boldsymbol{\phi}(z) = \mathbf{I}_m - \sum_{j=1}^p \boldsymbol{\phi}_j z^j$ and $\boldsymbol{\theta}(z) = \mathbf{I}_m - \sum_{j=1}^p \boldsymbol{\theta}_j z^j$, we have $\det[\boldsymbol{\phi}(z)] \neq 0$ and $\det[\boldsymbol{\theta}(z)] \neq 0$ for $|z| \leq 1$. Moreover, if the latent VAR is of exact order p , so that $\boldsymbol{\phi}_p \neq \mathbf{0}$, then $\boldsymbol{\theta}_p \neq \mathbf{0}$, and the moving-average polynomial is also of exact order p . In addition, the polynomial pair $[\boldsymbol{\phi}(z), \boldsymbol{\theta}(z)]$ is left coprime, so no nontrivial common left polynomial factor can be cancelled.

We now get simple expressions for the autocovariances and the parameters of the MSV(p) model.

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

$$\text{Cov}(\mathbf{X}_t - \mu, \mathbf{X}_{t-k} - \mu) := \boldsymbol{\Gamma}_k = \begin{cases} \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon; & \text{if } k = 0, \\ \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} - \boldsymbol{\phi}_k \boldsymbol{\Sigma}_\epsilon; & \text{if } 1 \leq k \leq p, \\ \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p}; & \text{if } k > p. \end{cases} \quad (3.2)$$

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

$$\boldsymbol{\phi}_{(p)} = \boldsymbol{\Gamma}_{(p+j)} \boldsymbol{\Gamma}_{[p+j-1]}^{-1} \quad j \geq 1, \quad \boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1, \quad (3.3)$$

$$\boldsymbol{\Sigma}_v = \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1 - \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\Gamma}'_j, \quad \mu = \mathbb{E}(\mathbf{X}_t), \quad (3.4)$$

where $\boldsymbol{\Gamma}_k = \text{Cov}(\mathbf{X}_t - \mu, \mathbf{X}_{t-k} - \mu)$, and

$$\boldsymbol{\phi}_{(p)} := \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_p \\ m \times m & m \times m & & m \times m \end{bmatrix}, \quad \boldsymbol{\Gamma}_{(p+j)} := \begin{bmatrix} \boldsymbol{\Gamma}_{p+j} & \boldsymbol{\Gamma}_{p+j+1} & \cdots & \boldsymbol{\Gamma}_{2p+j-1} \\ m \times m & m \times m & & m \times m \end{bmatrix}, \quad (3.5)$$

$$\mathbf{\Gamma}_{[p+j-1]} := \begin{bmatrix} \mathbf{\Gamma}_{(p+j-1)} \\ \mathbf{\Gamma}_{(p+j-2)} \\ \vdots \\ \mathbf{\Gamma}_{(j)} \end{bmatrix}_{pm \times pm} = \begin{bmatrix} \mathbf{\Gamma}_{p+j-1} & \mathbf{\Gamma}_{p+j} & \cdots & \mathbf{\Gamma}_{2p+j-2} \\ \mathbf{\Gamma}_{p+j-2} & \mathbf{\Gamma}_{p+j-1} & \cdots & \mathbf{\Gamma}_{2p+j-3} \\ \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_j & \mathbf{\Gamma}_{j+1} & \cdots & \mathbf{\Gamma}_{p+j-1} \end{bmatrix}_{m \times m}. \quad (3.6)$$

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

$$\hat{\mathbf{\Gamma}}_k = \frac{1}{T-k} \sum_{t=1}^{T-k} (\mathbf{X}_{t+k} - \hat{\mu})(\mathbf{X}_t - \hat{\mu})', \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t. \quad (3.7)$$

On setting $j = 1$ in (3.3) and replacing theoretical moments by the corresponding empirical moments, we get the following closed-form estimators of the MSV(p) coefficients:

$$\hat{\boldsymbol{\phi}}_{(p)} = \hat{\mathbf{\Gamma}}_{(p+1)} \hat{\mathbf{\Gamma}}_{[p]}^{-1}, \quad \hat{\boldsymbol{\Sigma}}_\epsilon = \hat{\mathbf{\Gamma}}_0 + \hat{\boldsymbol{\phi}}_1^{-1} \sum_{j=1}^{p-1} \hat{\boldsymbol{\phi}}_{j+1} \hat{\mathbf{\Gamma}}_j' - \hat{\boldsymbol{\phi}}_1^{-1} \hat{\mathbf{\Gamma}}_1, \quad (3.8)$$

$$\hat{\boldsymbol{\Sigma}}_v = \hat{\boldsymbol{\phi}}_1^{-1} \hat{\mathbf{\Gamma}}_1 - \hat{\boldsymbol{\phi}}_1^{-1} \sum_{j=1}^{p-1} \hat{\boldsymbol{\phi}}_{j+1} \hat{\mathbf{\Gamma}}_j' - \sum_{j=1}^p \hat{\boldsymbol{\phi}}_j \hat{\mathbf{\Gamma}}_j', \quad \hat{\mu} = \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t. \quad (3.9)$$

Given $\hat{\boldsymbol{\Sigma}}_\epsilon$, we compute cross-dependence parameters using the method proposed by [Harvey et al. \(1994\)](#).

In small samples, the estimate $\hat{\boldsymbol{\phi}}_{(p)}$ may yield an unstable VAR(p) specification, *i.e.*, one or more eigenvalues of the estimated companion matrix may lie on or outside the unit circle. This issue can arise especially in small samples or in the presence of outliers. The values $\hat{\boldsymbol{\phi}}_{(p)}$ computed by choosing different values of $j \geq 1$ [in (3.3)] may also yield different estimates. When this happens, a simple fix consists in projecting the estimate onto the space of admissible parameter values by adjusting 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 the modified Cholesky algorithm of [Cheng and Higham \(1998\)](#) to $\hat{\boldsymbol{\Sigma}}_v$ and obtain a positive definite matrix. This problem can also be alleviated by combining different estimates of $\hat{\boldsymbol{\phi}}_{(p)}$ obtained by choosing different values of $j \geq 1$ [in (3.3)]. We now discuss this approach.

3.2 VARMA-based winsorized estimation

Better stability and efficiency of VARMA estimator can be achieved through “winsorization,” which substantially increases the probability of obtaining admissible values. From (3.3), it is easy to see that:

$$\boldsymbol{\phi}_{(p)} = \sum_{j=1}^J \omega_j \mathbf{\Gamma}_{(p+j)} \mathbf{\Gamma}_{[p+j-1]}^{-1} \quad (3.10)$$

for any sequence of scalars ω_j with $\sum_{j=1}^J \omega_j = 1$ (convex combination) and any positive integer J . Using (3.10), 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 $\mathbf{\Gamma}_{(p+j)}\mathbf{\Gamma}_{[p+j-1]}^{-1}$:

$$\tilde{\boldsymbol{\phi}}_{(p)} = \sum_{j=1}^J \omega_j \hat{\mathbf{\Gamma}}_{(p+j)} \hat{\mathbf{\Gamma}}_{[p+j-1]}^{-1} \quad (3.11)$$

where $1 \leq J \leq T - 2p$ with $\sum_{j=1}^J \omega_j = 1$, and T is the length of the time series. We can expect that a sufficiently general class of weights may improve the efficiency of the VARMA estimators.

Using (3.11), 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 (3.11):

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

$$\omega_j = 1/J, \quad j = 1, \dots, J; \quad (3.12)$$

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

$$\hat{\mathbf{\Gamma}}_{(p+j)} = \hat{\boldsymbol{\phi}}_{(p)} \cdot \hat{\mathbf{\Gamma}}_{[p+j-1]}, \quad j = 1, \dots, J. \quad (3.13)$$

The solution of the above system is given by

$$\hat{\boldsymbol{\phi}}_{(p,J)}^{MVR} = \hat{\mathbf{\Gamma}}_{(p+1,J)} \hat{\mathbf{\Gamma}}'_{[p,J]} (\hat{\mathbf{\Gamma}}_{[p,J]} \hat{\mathbf{\Gamma}}'_{[p,J]})^{-1}, \quad (3.14)$$

$$\hat{\mathbf{\Gamma}}_{(p+1,J)} := \begin{bmatrix} \hat{\mathbf{\Gamma}}_{(p+1)} & \hat{\mathbf{\Gamma}}_{(p+2)} & \cdots & \hat{\mathbf{\Gamma}}_{(p+J)} \end{bmatrix}_{m \times pmJ}, \quad \hat{\mathbf{\Gamma}}_{[p,J]} := \begin{bmatrix} \hat{\mathbf{\Gamma}}_{[p]} & \hat{\mathbf{\Gamma}}_{[p+1]} & \cdots & \hat{\mathbf{\Gamma}}_{[p+J-1]} \end{bmatrix}_{pm \times pmJ},$$

where J is the winsorization truncation parameter.

Similar estimators are considered by [Ahsan and Dufour \(2021\)](#) for univariate SV models. All these estimators depend on J . For $J = 1$, we get the simple VARMA estimator in (3.8).

Remark 3.1. ROBUSTNESS TO OUTLIERS. The VARMA-based winsorized estimators tend to be robust to outliers from contaminated/corrupted data, or heavy-tailed error distributions. This class of estimators is based on higher-order autocovariances, which are less affected by outliers than lower-order autocovariances; see [Politis \(2009\)](#). It should be noted that the weak VARMA representation of $\mathbf{X}_t$ (given in Proposition 3.1) is non-Gaussian. In this case, the moment equation for $\hat{\mathbf{\Gamma}}_{(p+1)}$ [from (3.13)] – which yields the simple estimator – is not efficient under non-Gaussianity. It seems natural that in the context of second-order stationarity, a general estimator of $\hat{\boldsymbol{\phi}}_{(p)}$ and $\hat{\boldsymbol{\Sigma}}_v$ would be a function of all the second-order information available, i.e., $[\hat{\mathbf{\Gamma}}_{(p+1)}, \hat{\mathbf{\Gamma}}_{(p+2)}, \dots, \hat{\mathbf{\Gamma}}_{(T-2p)}]$. Since $\hat{\mathbf{\Gamma}}_{(k)}$ is unreliable for large k , i.e., when $T - k$ is small, it makes sense to base our estimator on $[\hat{\mathbf{\Gamma}}_{(p+1)}, \hat{\mathbf{\Gamma}}_{(p+2)}, \dots, \hat{\mathbf{\Gamma}}_{(p+J)}]$ where J is larger than 1, but smaller than $T - 2p$; in particular, we require that $T - 2p$ is large, i.e., $1 \leq J \leq c \times (T - 2p)$ for some $c \in (0, 1)$.

The robustness of the winsorized estimators should be understood in a moment-stabilization sense:

the estimators remain moment based and continue to rely on higher-order autocovariance restrictions implied by the transformed MSV(p) model, and their theoretical justification requires relevant moment and regularity conditions to hold. Thus, while winsorization and MVR reduce sensitivity to unstable low-lag empirical moments, isolated outliers, and moderate heavy-tailed behavior, severe misspecification or infinite-variance behavior may still affect finite-sample performance.

Remark 3.2. ASYMPTOTIC EFFICIENCY GAINS. Since the weak VARMA representation of $\mathbf{X}_t$ (given in Proposition 3.1) is non-Gaussian, additional $J-1$ higher-order moment conditions can provide substantial asymptotic efficiency gains over the simple estimator based on the first lower-order moment condition. In this case, the higher-order moment conditions are non-redundant and the winsorized estimator, based on these moment conditions, does not require strong assumptions on the noise; see Broze et al. (2001) for a further discussion on the non-redundancy of higher-order moment conditions for efficient moment-based estimation.

Remark 3.3. MVR AS A SHRINKAGE/REGULARIZE ESTIMATOR. The MVR estimator is defined by stacking multiple moment conditions at higher lags and then solving a single large linear system. Because more moment conditions are included, the resulting estimate typically exhibits the following:

Stable-region pull-in – With additional lags/blocks, the system *pushes* the parameter solution $\phi_{(p,J)}$ to satisfy more constraints. The truncation parameter J can be viewed as a shrinkage control parameter of the companion matrix $\mathcal{A}$ defined in (E7). In simulations and empirical sections, we observe that the modulus of the companion matrix’s maximum eigenvalue shrinks toward the admissible region as J increases. It is easy to see that this result is the direct consequence of asymptotic efficiency gains due to using non-redundant higher-order moment conditions. In addition, this holds for different sample sizes with many assets. Our simulation and empirical results also suggest a bias-variance trade-off for higher levels of J , indicating that a moderate level of J is sufficient.

Regularization effect – In classical shrinkage (e.g., ridge regression), extra penalization or constraints lead to smaller-norm solutions. Here, by adding higher-lag moment equations, we effectively *regularize* the system: any single outlier in the lower-lag equation is balanced by higher-lag conditions, preventing estimates from *overfitting* the first lag. This yields a parameter vector (or matrix) which is often smaller in norm or better-conditioned.

3.3 Conditional covariance matrix forecasting with MSV(p) Models

In most empirical applications of MSV(p) models, the primary objects of interest are not the static parameters themselves, but the latent log-volatility vector $\mathbf{h}_t$ and the implied conditional covariance matrix of returns. Under the MSV(p) specification, conditional on the latent state $\mathbf{h}_t$, the covariance matrix at time t is

$$\Sigma_t = \mathbf{V}_t^{1/2} \Sigma_u \mathbf{V}_t^{1/2}, \quad \mathbf{V}_t^{1/2} = \text{diag}\{\exp(\mathbf{h}_t/2)\}, \quad (3.15)$$

where Σ_u denotes the contemporaneous correlation matrix of return shocks. Hence, inference on Σ_t depends directly on recovery of the latent state $\mathbf{h}_t$.

Similarly, the s -step-ahead conditional covariance forecast is obtained by first forecasting the latent log-volatility state and then mapping it into the covariance matrix:

$$\hat{\Sigma}_{t+s|t} = \hat{V}_{t+s|t}^{1/2} \hat{\Sigma}_u \hat{V}_{t+s|t}^{1/2}, \quad \hat{V}_{t+s|t}^{1/2} = \text{diag}\{\exp(\hat{h}_{t+s|t}/2)\}. \quad (3.16)$$

Accordingly, more accurate estimation of the time-invariant parameters and more accurate filtering of $\mathbf{h}_t$ translate directly into more reliable covariance estimation and forecasting.

To recover the latent states, we work with the mean-corrected log-squared transformation, which yields a measurement equation linear in $\mathbf{h}_t$, but non-Gaussian in the disturbance term. Accordingly, the Kalman filter (KF) is used as a Gaussian benchmark rather than as an exact filter. We therefore also consider particle-based filtering methods. The bootstrap particle filter (BPF) propagates particles through the state transition equation and reweights them using a second-moment-compatible pseudo-likelihood for the transformed measurement density, following the spirit of simulation-based likelihood methods for non-Gaussian state-space models; see [Shephard and Pitt \(1997\)](#). The auxiliary particle filter (APF) further incorporates a look-ahead resampling step which improves particle allocation toward states which are more likely to match the next observation, thereby helping to mitigate weight degeneracy relative to the BPF; see [Pitt and Shephard \(1999\)](#). In our Monte Carlo experiments, the APF tends to deliver the most accurate latent-state extraction and covariance forecasts. Full details on filtering and forecasting under MSV(p) are provided in [Appendix F](#).

4 Asymptotic distributional theory

In this section, we develop asymptotic distribution theory for the simple VARMA-based estimator proposed above for the MSV(p) model, as well as for its winsorized extensions that incorporate additional higher-order moment conditions. We also show how using additional non-redundant higher-order moment conditions can yield asymptotic efficiency gains.

The following remarks follow from Assumptions 2.1-2.2: the observed process $\{\mathbf{X}_t\}$ is stationary, ergodic, and mixing, and it has finite moments, while additional higher-lag moment conditions may provide non-redundant information.

Remark 4.1. STATIONARITY AND ERGODICITY. Assumption 2.2 ensures that all eigenvalues of the VAR(p) companion matrix associated with $\mathbf{h}_t$ lie strictly inside the unit circle. This implies that $\{\mathbf{h}_t\}$ is strictly stationary and ergodic [see [Brockwell and Davis \(1996, Theorem 11.1\)](#)]. Under Assumption 2.1, the joint process $(\mathbf{h}_t', \mathbf{u}_t')'$ is stationary and ergodic. Since $\mathbf{X}_t = \mathbf{h}_t + \epsilon_t$ is a measurable function of $(\mathbf{h}_t, \mathbf{u}_t)$, it inherits strict stationarity and ergodicity.

Remark 4.2. MIXING AND FINITE MOMENTS. Under Assumptions 2.1-2.2, the latent process $\{\mathbf{h}_t\}$ is the

causal stationary solution of a stable Gaussian VAR(p) and therefore has a linear representation with geometrically decaying coefficients. Standard results for stable linear processes then imply geometric strong mixing [see Davidson (1994, Chapter 14)]. Since $\{u_t\}$ is i.i.d. and independent of the volatility-innovation history, and ϵ_t is a measurable function of u_t , the joint process $\{(\mathbf{h}'_t, u'_t)'\}$, and hence $\mathbf{X}_t = \mathbf{h}_t + \epsilon_t$, is geometrically strongly mixing. Moreover, $\mathbf{h}_t$ is Gaussian and $\log(u_{it}^2)$ has finite moments of every finite order. Consequently, $\mathbf{X}_t$ has finite moments of all orders.

Remark 4.3. NON-REDUNDANT HIGHER-ORDER MOMENT CONDITIONS. Since the weak VARMA representation of $\mathbf{X}_t$ (given in Proposition 3.1) is non-Gaussian, additional higher-order moment conditions are non-redundant.

Beyond Assumptions 2.1-2.2, we require the following assumptions for the asymptotic analysis.

Assumption 4.1. IDENTIFIABILITY. *The parameter vector Θ is locally identified by the autocovariance sequence $(\Gamma_0, \Gamma_1, \dots)$, that is, the corresponding parameter-to-moment mapping is one-to-one in a neighborhood of the true parameter.*

Assumption 4.2. NON-DEGENERACY. *All covariance matrices involved, such as Σ_v and Σ_ϵ , are positive definite.*

Assumption 4.3. SMOOTHNESS. *For $\Lambda(m) := [\mu', \text{vec}(\Gamma_0)', \text{vec}(\Gamma_1)', \dots, \text{vec}(\Gamma_m)']'$, the mapping $g : \Lambda(m) \mapsto \Theta$ is continuously differentiable at the true parameter vector, and the Jacobian $Dg[\Lambda(m)]$ has full row rank.*

Assumption 4.1 entails that distinct parameter vectors yield distinct second-order structures, ensuring uniqueness of $\Theta := [\mu', \text{vec}(\phi_{(p)})', \text{vec}(\Sigma_v)', \text{vec}(\Sigma_\epsilon)']'$ given the autocovariances. It is a local identification condition for the MSV(p) parameter vector Θ through the autocovariance sequence $\{\Gamma_k\}_{k \geq 0}$ of the transformed process $\mathbf{X}_t$. It should not be interpreted as identification of a generic unrestricted VARMA model: the VARMA(p, p) representation of $\mathbf{X}_t$ is induced by the structural MSV(p) state-space system and therefore obeys additional moment restrictions. Under these restrictions, the coefficient matrices ϕ_j are determined by the Yule-Walker system (3.2) whenever the relevant block moment matrix is nonsingular, while Σ_ϵ , Σ_v , and μ are recovered from the identities in (3.3). Assumption 4.1 thus rules out local observational equivalence, namely distinct parameter values in a neighborhood of the truth which generate the same second-order structure. In particular, it does not require diagonal restrictions on the matrices ϕ_j . The left-coprimeness result in Proposition 3.1 already rules out nontrivial AR-MA polynomial cancellations under Assumptions 2.1-2.2; Assumption 4.1 instead excludes remaining local degeneracies of the restricted MSV moment mapping, such as singular moment systems.

Assumption 4.2 rules out degeneracies in the covariance structure, while Assumption 4.3 ensures that the relevant moment-to-parameter mapping is sufficiently smooth for application of the delta method. Under Assumptions 2.1-2.2, we then have the following asymptotic results on empirical moments.

Lemma 4.1. CONSISTENCY OF EMPIRICAL MOMENTS. *Under Assumptions 2.1-2.2, the sequence $\{\mathbf{X}_t\}$ is strictly stationary, ergodic and has finite moments of all orders. Further, for any fixed $m \geq 1$, the estimator $\hat{\Lambda}(m) := [\hat{\mu}', \text{vec}(\hat{\Gamma}_0)', \text{vec}(\hat{\Gamma}_1)', \dots, \text{vec}(\hat{\Gamma}_m)']'$ defined by (3.7) satisfies: $\hat{\Lambda}(m) \xrightarrow{a.s.} \Lambda(m)$.*

Lemma 4.2. ASYMPTOTIC DISTRIBUTION OF EMPIRICAL MOMENTS. *Under Assumptions 2.1-2.2, the sequence $\{\mathbf{X}_t\}$ is strongly mixing with exponentially decaying mixing coefficients and has finite moments of all orders. As a result, a multivariate Central Limit Theorem (CLT) applies, yielding: $\sqrt{T}[\hat{\Lambda}(m) - \Lambda(m)] \xrightarrow{d} \mathcal{N}[0, \Sigma(\Lambda, m)]$, for each fixed integer m , where $\Sigma(\Lambda, m)$ is a positive semidefinite covariance matrix.*

Lemmas 4.1-4.2 provide the asymptotic distribution of the simple VARMA-type estimator derived in (3.8) and (3.9). Note that strong consistency in Lemma 4.1 is entailed by ergodicity.

Theorem 4.3. ASYMPTOTIC DISTRIBUTION OF SIMPLE VARMA ESTIMATOR. *Under Assumptions 2.1, 2.2, 4.1, 4.2, and 4.3, the estimator $\hat{\Theta} := [\hat{\mu}', \text{vec}(\hat{\Phi}_{(p)})', \text{vec}(\hat{\Sigma}_v)', \text{vec}(\hat{\Sigma}_e)']' = g[\hat{\Lambda}(2p)]$ is consistent, i.e., $\hat{\Theta} \xrightarrow{a.s.} \Theta$. In addition, $\sqrt{T}(\hat{\Theta} - \Theta) \xrightarrow{d} \mathcal{N}[0, \Omega]$, where $\Omega = (Dg[\Lambda(2p)])\Sigma(\Lambda, 2p)(Dg[\Lambda(2p)])'$ and $Dg[\Lambda(2p)]$ is the Jacobian matrix of g evaluated at the true parameter values.*

We now consider the asymptotic distribution of the *winsorized VARMA estimators* given in (3.9) and (3.11), which is expressed in vectorized form as: $\tilde{\Theta}_J := [\hat{\mu}', \text{vec}(\tilde{\Phi}_{(p)})', \text{vec}(\tilde{\Sigma}_{v,J})', \text{vec}(\tilde{\Sigma}_{e,J})']'$ where the estimator uses additional $J - 1$ *non-redundant* moment conditions, which involve $2p + J - 1$ empirical autocovariances, i.e., $[\hat{\Gamma}_0, \hat{\Gamma}_1, \dots, \hat{\Gamma}_{2p+J-1}]$.

Theorem 4.4. ASYMPTOTIC DISTRIBUTION OF THE VARMA-BASED WINSORIZED ESTIMATOR. *Suppose that Assumptions 2.1, 2.2, 4.1, 4.2, and 4.3 hold. Consider the winsorized VARMA estimator $\tilde{\Phi}_{(p)}$ or the MVR estimator $\hat{\Phi}_{(p,J)}^{MVR}$ defined in (3.13) for $J \geq 1$. Then $\sqrt{T}(\tilde{\Theta}_J - \Theta) \xrightarrow{d} \mathcal{N}[0, \Omega_J]$ for some positive semidefinite Ω_J . In particular, $\sqrt{T}[\tilde{\Phi}_{(p)} - \Phi_{(p)}] \xrightarrow{d} \mathcal{N}[0, \Omega_J^\Phi]$, and similarly for the MVR estimator.*

If the additional $J - 1$ moment conditions are *non-redundant* (as discussed in Remark 4.3 and Section 3.2), the moment conditions introduce new information beyond that available in the $J = 1$ case. This is analogous to incorporating additional independent instruments in a GMM framework, which improves efficiency.

Theorem 4.5. ASYMPTOTIC EFFICIENCY GAINS FROM NON-REDUNDANT HIGHER-ORDER MOMENT CONDITIONS. *Suppose that Assumptions 2.1, 2.2, 4.1, 4.2, and 4.3 hold. Let Ω_1 denote the asymptotic covariance matrix of the simple VARMA-based estimator (for $J = 1$), as established in Theorem 4.3. Similarly, let Ω_J denote the asymptotic covariance matrix of the winsorized VARMA-based estimator utilizing $J > 1$ moment conditions, as established in Theorem 4.4. If the additional $J - 1$ higher-order moment conditions are non-redundant, then $\Omega_J \leq \Omega_1$, where $\leq$ denotes the Loewner (partial) order.*

Typically, $\Omega_J < \Omega_1$, which means that Ω_J is strictly smaller in at least one direction. In particular, Ω_J reduces to Ω_1 when $J = 1$, and for $J > 1$, the presence of additional non-redundant moment conditions yields asymptotic efficiency gains, reflected in a smaller asymptotic variance matrix.

5 Volatility commonality and a common C-space

The literature on volatility commonality has long recognized that high-dimensional conditional covariance dynamics may be driven by a relatively small number of common directions. In factor-ARCH models, covariance variation is governed by time-varying factor variances along fixed directions [Engle et al. (1990)]. Common-volatility models ask whether several return series share common volatility movements [Engle and Susmel (1993)]. Factor stochastic-volatility models extend this idea to a latent-state setting by representing high-dimensional return variation through a small number of latent factors with stochastic volatilities [Aguilar and West (2000), Chib et al. (2006)].

Our contribution differs in emphasis. Rather than imposing low-dimensional structure directly on returns or conditional covariances for parsimony or forecasting, we formulate volatility commonality in an MSV(p) model as a testable dynamic common-subspace restriction on the transition matrices $\{\phi_j\}_{j=1}^p$. We ask whether these lag matrices share a common right r -dimensional subspace, and we develop feasible finite-sample LMC/MMC inference for this restriction. The proposed framework thus connects the volatility-commonality literature to reduced-rank dynamic modeling [Ahn and Reinsel (1988)] and to the common-feature perspective in multivariate time-series analysis [Engle and Kozicki (1993)], while relying on finite-sample Monte Carlo testing methods for models with nuisance parameters [Dufour (2006)].

Remark 5.1. VOLATILITY SIMILARITY VIA A COMMON DYNAMIC SUBSPACE. A parsimonious way to encode the empirically observed similarity of log-volatility paths across related macroeconomic or financial series is to assume that the propagation of volatility shocks occurs through a low-dimensional common subspace. In the MSV(p) transition equation (2.5), this is captured by the factorization $\phi_j = \mathbf{B}_j \mathbf{C}'$ with $r \ll m$, which implies $\text{rank}(\phi_j) \leq r$ for each lag and reduces the dynamic parameter dimension from order pm^2 to order $pmr + mr$, up to the orthonormality normalization on $\mathbf{C}$. The restriction is testable and can be imposed when supported by the data.

Assumption 5.1. COMMON C-SPACE MSV(p) VOLATILITY DYNAMICS. *There exists an integer $r \in \{1, \dots, m-1\}$ and matrices $\mathbf{C} \in \mathbb{R}^{m \times r}$ and $\mathbf{B}_1, \dots, \mathbf{B}_p \in \mathbb{R}^{m \times r}$ such that $\mathbf{C}'\mathbf{C} = \mathbf{I}_r$ and, for all $j = 1, \dots, p$,*

$$\phi_j = \mathbf{B}_j \mathbf{C}'. \quad (5.1)$$

Equivalently, letting $\mathbf{P}_C := \mathbf{C}\mathbf{C}'$, the restriction implies $\phi_j = \phi_j \mathbf{P}_C$ for all j .

5.1 Estimation under a common C-space restriction

Let $\hat{\phi}_1, \dots, \hat{\phi}_p \in \mathbb{R}^{m \times m}$ denote unrestricted estimators of $\{\phi_j\}_{j=1}^p$ (e.g., the W-VARMA-MVR estimator in Section 3). Define

$$\hat{\mathbf{M}}_C := \sum_{j=1}^p \hat{\phi}_j' \hat{\phi}_j \in \mathbb{R}^{m \times m}. \quad (5.2)$$

Proposition 5.1. CLOSED-FORM ESTIMATOR OF THE COMMON $\mathbf{C}$ -SPACE. Let $\hat{\mathbf{C}}$ collect the r leading eigenvectors of $\hat{\mathbf{M}}_C$ in (5.2), normalized so that $\hat{\mathbf{C}}'\hat{\mathbf{C}} = \mathbf{I}_r$, and define $\hat{\mathbf{P}}_C := \hat{\mathbf{C}}\hat{\mathbf{C}}'$. Then the restricted estimators are

$$\hat{\boldsymbol{\phi}}_j^{(r, \text{com})} := \hat{\boldsymbol{\phi}}_j \hat{\mathbf{P}}_C, \quad j = 1, \dots, p. \quad (5.3)$$

Moreover, $\hat{\mathbf{C}}$ solves the constrained least-squares problem (G.1) stated in Appendix G.

5.2 Testing volatility similarity via a common $\mathbf{C}$ -space

For a fixed rank r , consider the null hypothesis

$$H_0^{\text{com}}(r) : \exists \mathbf{C} \in \mathbb{R}^{m \times r}, \mathbf{C}'\mathbf{C} = \mathbf{I}_r, \exists \{\mathbf{B}_j\}_{j=1}^p \text{ s.t. } \boldsymbol{\phi}_j = \mathbf{B}_j\mathbf{C}', j = 1, \dots, p, \quad (5.4)$$

against the unrestricted MSV(p) alternative. Under $H_0^{\text{com}}(r)$, the volatility transition matrices share a common r -dimensional right subspace across lags. Thus, lagged log-volatilities affect current volatility only through the same r dynamic directions for all $j = 1, \dots, p$, although the loading matrices $\mathbf{B}_j$ may differ across lags. This restriction is stronger than lag-wise rank reduction, because it requires the reduced dynamic subspace to be common across lags. It is related to the reduced-rank autoregressive framework of Ahn and Reinsel (1988) and, more broadly, to the common-feature perspective of Engle and Kozicki (1993). In contrast with factor-ARCH, common-volatility, or factor-SV models, which impose common low-dimensional structure as part of the model specification, $H_0^{\text{com}}(r)$ treats this structure as a falsifiable restriction on the transition law itself. Our testing procedure, however, is based on finite-sample Monte Carlo methods for models with nuisance parameters; see Appendix B for the general simulation-based inference framework.

Given $\{\hat{\boldsymbol{\phi}}_j\}_{j=1}^p$, let $\hat{\mathbf{P}}_C = \hat{\mathbf{C}}\hat{\mathbf{C}}'$ be formed from (5.2), and let $\Pi_r(\cdot)$ denote the best rank- r approximation in the Frobenius norm. Setting

$$D_{\text{com}}(\hat{\boldsymbol{\phi}}; r) := \sum_{j=1}^p \|\hat{\boldsymbol{\phi}}_j - \hat{\boldsymbol{\phi}}_j \hat{\mathbf{P}}_C\|_F^2, \quad D_{\text{rr}}(\hat{\boldsymbol{\phi}}; r) := \sum_{j=1}^p \|\hat{\boldsymbol{\phi}}_j - \Pi_r(\hat{\boldsymbol{\phi}}_j)\|_F^2, \quad (5.5)$$

we then use

$$S_T^{\text{com}}(r) := T \cdot \max\{0, D_{\text{com}}(\hat{\boldsymbol{\phi}}; r) - D_{\text{rr}}(\hat{\boldsymbol{\phi}}; r)\}. \quad (5.6)$$

The truncation is a numerical safeguard, since finite-sample rounding may occasionally produce a slightly negative value inside the braces.

The finite-sample distribution of $S_T^{\text{com}}(r)$ depends on nuisance parameters in the MSV(p) model. Accordingly, inference is based on simulation under the null-restricted plug-in model obtained by imposing the common-space restriction via (5.3). Under the local Monte Carlo (LMC) procedure, we generate N pseudo-samples from this restricted model, re-estimate the unrestricted parameters for each pseudo-sample, and recompute the statistic, producing simulated values $\{S_T^{(i)}\}_{i=1}^N$. The associated LMC p -value is

$\hat{p}_{T,N}^{\text{LMC}}(r) = [1 + \sum_{i=1}^N \mathbb{I}\{S_T^{(i)} \geq S_T^{\text{obs}}\}] / (N + 1)$. We also report a maximized Monte Carlo (MMC) version, obtained by maximizing the simulated p -value over a small nuisance-parameter set, implemented here as a grid around the plug-in estimate. Relative to LMC, the main advantage of MMC is its stronger robustness of finite-sample size control in the presence of nuisance parameters. Because LMC conditions on a single plug-in value, it may be sensitive to nuisance-parameter estimation error or other finite-sample irregularities. By contrast, MMC guards against such sensitivity by maximizing the simulated p -value over a set of plausible nuisance values, thereby basing inference on the least favorable null configuration. As discussed in Appendix B, this typically makes MMC more conservative, but also more reliable for size-robust inference.

6 Simulation study

This section investigates finite-sample performance of proposed methods through Monte Carlo experiments on parameter estimation, latent log-variance filtering and forecasting, comparison with QML, Bayes, and FSV benchmarks, high-dimensional settings, and common C-space testing.

6.1 Parameter estimation

6.1.1 Comparison of different VARMA estimators

We first study the finite-sample properties of the winsorized VARMA estimators, introduced 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 set to $(\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \mu_1, \mu_2, \sigma_{v,1}^2, \sigma_{v,2}^2, \sigma_{v,12}, \rho) = (0.95, -0.2, 0.1, 0.95, -2, 2, 1, 1, 0.9, 0.9)$. We generate 1000 replications and consider different sample sizes $T = (500, 1000, 2000)$. Both shrinkage estimators depend on the truncation parameter J , whose value is set to $J = 10$. Detailed results are reported in Appendix Table A1. The main findings can be summarized as follows.

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 unbounded solutions in this parameter setting with larger samples. This fact implies that the variability of estimated ACF goes down as the sample size increases.

6.1.2 Comparison with QML estimator

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 ϕ_1 , and no constant term in the volatility equation. Thus we have 6

parameters in the system $(\phi_{11}, \phi_{22}, \sigma_{v,1}^2, \sigma_{v,2}^2, \sigma_{v,12}, \rho) = (0.98, 0.98, 0.5, 0.5, 0.5, -0.8)$. We generate 100 replications and consider the sample sizes $T = 500, 2000, 5000$. These restrictions are also considered by [Harvey et al. \(1994\)](#). Appendix Table A2 reports the results. These show that W-VARMA-MVR estimators deliver more accurate estimates than the QML method in this design. The QML estimates of ρ are severely biased. In most cases, the MVR estimator with $J = 10$ performs best.

6.1.3 Comparison with Bayes estimator

In the literature, there is no uniform ranking among the different estimators, but the performance of the Bayesian estimator remains superior to that of competing methods in the context of univariate stochastic volatility models. Consequently, we compare our proposed VARMA estimators (CF, MVR) to the Bayesian estimator, as discussed in [Kim et al. \(1998\)](#), [Omori et al. \(2007\)](#), [Kastner and Frühwirth-Schnatter \(2014\)](#), [Kastner \(2019\)](#). Adopting a Bayesian perspective, we assume that the parameters are not completely unknown but follow specific prior distributions. Using these priors and the data, we make inferences about the parameters from their posterior distributions. We define prior distributions for the parameter set $\theta = (\mu, \phi_1, \Sigma_v)$ of the MSV model. The mean vector μ follows a flat multivariate normal distribution $\mu \sim \mathcal{N}_m(\mathbf{0}, 1000\mathbf{I}_m)$. Each entry of the persistence matrix $\phi_{ij} \in (-1, 1)$ is assumed to follow a beta distribution, $(\phi_{ij} + 1)/2 \sim \text{Beta}(20, 1.5)$ [see [Kim et al. \(1998\)](#)]. We also impose the VAR stability restriction by restricting the MCMC chain to include only stable ϕ_1 matrices, ensuring the stationarity of the VAR(1) process throughout the sampling procedure. For the scale parameter of volatility innovation, we employ a flat gamma prior, $\sigma_{i,v}^2 \sim \Gamma(1/2, 1/2 \times 10)$ [see [Kastner and Frühwirth-Schnatter \(2014\)](#)], which is equivalent to $\pm\sqrt{\sigma_{i,v}^2} \sim \mathcal{N}(0, 10)$. We select hyperparameters which provide relatively flat prior distributions. These are also used in [Phan et al. \(2019\)](#); the choice of hyperparameters minimally affects the posterior estimates of the parameters in large samples. The MCMC algorithm is discussed in Appendix D.

For this comparison, the true data generating process (DGP) is defined by (2.1)-(2.2) with $m = (2, 5, 10)$, $p = 1$. We set $\Sigma_u = \mathbf{I}_m$, $\Sigma_v = 0.4\mathbf{I}_m$, $\mu = -0.5\mathbf{1}$, and stationary ϕ_1 configurations are as follows: for $m = 2$, ϕ_1 is non-diagonal with $(\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}) = (0.98, -0.1, 0.12, 0.99)$; for $m = 5$, ϕ_1 is non-diagonal with diagonal elements $(0.98, 0.985, 0.8, 0.985, 0.98)'$ and upper off-diagonal elements at -0.15 and lower off-diagonal elements at 0.1; for $m = 10$, ϕ_1 is non-diagonal with diagonal elements $(0.95, 0.98, 0.98, 0.99, 0.8, 0.8, 0.99, 0.98, 0.98, 0.95)'$ and upper off-diagonal elements at -0.11 and lower off-diagonal elements at 0.055. The number of parameters (NP) in these models is $m^2 + 2m$. We simulate 100 samples from each simulated model.

We apply the above Bayesian method to the simulated datasets and draw 40,000 observations from the posterior distributions. The Bayesian 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 assistance of the R package Rcpp [see [Edelbuettel and François \(2011\)](#)]. Matrix computations utilize the efficient C++ template library Armadillo [[Sanderson and Curtin \(2016\)](#)] through the R

Table 1
Simulation results for estimated MSV processes: Bayes versus W-VARMA-MVR estimators

Table 1 reports average bias and RMSE for Bayes, CF-VARMA, and W-VARMA-MVR estimators under the MSV(1) designs described in Section 6.1.3, for $m \in \{2, 5, 10\}$ and $T \in \{1000, 5000\}$. RCT denotes relative computational time with respect to CF-VARMA. NIV and NPD denote, respectively, the numbers of inadmissible estimates of ϕ and non-positive-definite estimates of Σ_v across 100 Monte Carlo replications. For the Bayes estimator, posterior summaries are based on 20,000 retained draws after discarding 20,000 burn-in draws. Boldface identifies the lowest absolute average bias and the lowest RMSE among estimators with NIV = 0 and NPD = 0 within each (m, T) design; ties are also reported in bold. For additional results related to this table, see Appendix Table A4.

T	Estimator	RCT	NIV	NPD	Average bias				Average RMSE			
					ϕ	diag[ϕ]	μ	Σ_v	ϕ	diag[ϕ]	μ	Σ_v
Panel A: $m = 2, p = 1, NP = 8$												
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
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
Panel B: $m = 5, p = 1, NP = 35$												
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
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
Panel C: $m = 10, p = 1, NP = 120$												
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
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

package RcppArmadillo [see [Eddelbuettel and Sanderson \(2014\)](#)].

Table 1 reports the average bias and RMSE for ϕ , diag[ϕ], μ , Σ_v [where the average operator applies over element-wise bias (or RMSE) for vectors and matrices] when $m = (2, 5, 10)$ and Appendix Table A3 reports bias and RMSE for each parameter separately when $m = 2$. The number of inadmissible values of ϕ and non-positive definiteness (NPD) of Σ_v are also reported. Several conclusions emerge from these tables.

First, the MVR estimators have smaller biases and RMSEs than the Bayes estimator. 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 stable parameter values 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 the 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, the MVR estimators deliver lower bias and RMSE than the Bayes estimator in these designs. The RMSEs of the MVR estimates decrease as the sample size increases, which is consistent with the asymptotic properties 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 in the estimates of μ and Σ_v , indicating this sampling scheme's vulnerability. Note that the Bayesian comparison uses a standard prior specification and an efficient sampling scheme. Alternative priors or samplers may lead to different finite-sample results; a systematic assessment of Bayesian specifications is beyond the scope of the present comparison. Further, the choice of prior for a higher-order model ($p > 1$) gets more complicated as the order of the 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 their computational-time advantage is substantial compared to the Bayes method.

6.1.4 High-dimensional simulation study

We now examine the finite-sample properties of the W-VARMA-MVR estimators for $J = 10, 100$ in high-dimensional (HD) settings. For this analysis, the true DGP is described by (2.1)-(2.2) with $\Sigma_u = I_m$, $\Sigma_v = I_m$, $\mu = -\mathbf{1}$, and stationary ϕ matrices; both diagonal and non-diagonal specifications are considered.

Four DGPs are considered, depending on volatility persistence.

1. HD_1 – First-order persistence in latent volatilities involves a stationary ϕ_1 matrix, which is diagonal and has elements equal to 0.95. The model has $NP = 3m$ parameters.
2. HD_2 – First-order persistence in latent volatilities is assumed with a stationary ϕ_1 which is non-diagonal, where diagonal elements are 0.9 and off-diagonal elements range between $(-0.01, 0.01)$. The model has $NP = m^2 + 2m$ parameters.
3. HD_3 – Second-order persistence in latent volatilities is assumed with non-diagonal stationary ϕ_1 and ϕ_2 matrices. Here, for ϕ_1 , diagonal elements are 0.9, and off-diagonal elements range between $(-0.01, 0.01)$. For ϕ_2 , diagonal elements are -0.85 , and off-diagonal elements also range between $(-0.01, 0.01)$. The model has $NP = 2m^2 + 2m$ parameters.
4. Ultra high-dimensional setup (UHD) – For $p = 2$, non-diagonal stationary ϕ_1 and ϕ_2 matrices are considered, with diagonal elements of ϕ_1 at 0.9 and off-diagonal elements ranging from

$(-0.001, 0.001)$. Similarly, ϕ_2 has diagonal elements at -0.8 with the same range for off-diagonal elements. The model has $NP = 2m^2 + 2m$ parameters.

We set m from 2 to 250 for the $HD_1 - HD_3$ models, and from 500 to 1500 for the UHD model. We simulate 100 samples from each model with varying sample sizes. Tables A5-A7 in the Appendix and Table 2 report the average bias and RMSE for the corresponding matrices ϕ , μ , Σ_v . Several conclusions emerge from these tables:

First, MVR estimators perform exceptionally well in large-scale setups, providing highly accurate estimates. From Table 2, it is evident that the MVR method remains extremely reliable, even for a model with 1,500 dimensions and 4,503,000 parameters, maintaining accuracy with a moderate sample size of $T = 15000$.

Second, we observe several crucial roles played by the winsorized parameter J in high-dimensional models: (1) Increasing J reduces RMSE, particularly in small samples. This is evident when comparing the $J = 10$ and $J = 100$ columns in Appendix Tables A5-A7; (2) in high-dimensional setups, particularly when $m = 100$ or 250, a higher J significantly lowers RMSEs (and biases), indicating that the MVR method improves the precision of all parameters; (3) a higher J value (e.g., 100) reduces the incidence of non-positive definite estimates of the Σ_v matrix.

Third, in small samples, RMSEs for high-dimensional models ($m = 25, 50, 100, 250$) are slightly larger compared to those for low-dimensional models ($m = 2, 5, 10$). However, this disparity decreases as the sample size increases. We find that $T = 1000$ is inadequate for high-dimensional models, including the UHD design.

Fourth, we find that the W-VARMA-MVR estimator is not only statistically efficient but also extremely time-efficient. For a large sample size of $T = 10,000$, the 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 Appendix Table A7.

6.2 Latent log-variance estimation and forecasting

6.2.1 Comparison of different filtering methods

We compare three state-inference methods for latent log-variance extraction under the log-squared measurement system: KF, BPF, and APF. The corresponding filtering and forecasting procedures are described in Section 3.3 and Appendix F. The simulation design follows the same DGP settings as in Section 6.1.3 for $m \in \{2, 5, 10\}$. All results are based on VARMA-MVR ($J = 10$) parameter estimates recomputed within each Monte Carlo replication, with $B = 100$ replications overall. The particle-filter implementations use $N_{\text{pf}} = 10,000$ particles. Table 3 reports Monte Carlo average losses for in-sample filtering (Panel A) and one-step-ahead out-of-sample forecasting (Panel B), while the corresponding series-level results are reported in Appendix Tables A9-A10.

Table 2
Ultra-high-dimensional simulation results for W-VARMA-MVR estimators

Table 2 reports Monte Carlo average bias and RMSE for the W-VARMA-MVR estimators with $J = 10$ and $J = 100$ under the UHD design described in Section 6.1.4. Results are based on $B = 100$ replications for $m \in \{500, 750, 1000, 1250, 1500\}$ and $T \in \{5000, 15000\}$. Time denotes computation time per replication in seconds, $NP = 2m^2 + 2m$ is the number of parameters, and NIV and NPD denote, respectively, the numbers of inadmissible estimates of ϕ and non-positive-definite estimates of Σ_ν , out of 100 replications. Reported bias and RMSE values are element-wise averages over the corresponding parameter matrices or vectors. For additional results related to this table, see Appendix Table A8.

T	m	NP	Time	NIV	NPD	Average bias			Average RMSE		
						ϕ	μ	Σ_v	ϕ	μ	Σ_v
Panel A: $J = 10$											
5000	500	501000	4.69	0	14	0.0000	-0.0001	-0.0062	0.0235	0.0387	0.1689
	1000	2002000	21.41	0	100	0.0000	0.0002	0.0061	0.0236	0.0387	1.6758
	1500	4503000	60.83	0	100	0.0000	0.0002	-0.1271	0.0264	0.0386	1.3389
15000	500	501000	132.06	0	0	0.0000	0.0002	0.1903	0.0152	0.0223	0.1956
	1000	2002000	172.26	0	0	0.0000	0.0000	0.1122	0.0140	0.0221	0.1244
	1500	4503000	261.58	0	0	0.0000	0.0000	-0.0006	0.0134	0.0222	0.0886
Panel B: $J = 100$											
5000	500	501000	10.69	0	0	0.0000	-0.0001	-0.2145	0.0192	0.0387	0.2253
	1000	2002000	45.60	0	100	0.0000	0.0002	0.2923	0.0216	0.0387	3.3618
	1500	4503000	124.82	0	100	0.0000	0.0002	-0.0930	0.0250	0.0386	1.3094
15000	500	501000	278.95	0	0	0.0000	0.0002	0.0372	0.0086	0.0223	0.0462
	1000	2002000	365.87	0	0	0.0000	0.0000	-0.0874	0.0100	0.0221	0.0927
	1500	4503000	554.78	0	0	0.0000	0.0000	-0.2171	0.0109	0.0222	0.2207

The superior performance of the particle-based filters over the Kalman filter indicates that the non-Gaussian features of the transformed measurement equation matter for latent-state recovery and forecasting. Whereas the KF relies on a Gaussian approximation, the BPF and APF provide more flexible approximations to the filtering distribution. The additional gain of the APF over the BPF reflects the value of look-ahead resampling in reducing weight degeneracy and improving finite-sample accuracy. This interpretation is consistent with the Monte Carlo evidence reported in Table 3. Panel A of Table 3 shows that particle-based methods deliver substantial improvements in latent-state extraction accuracy relative to the KF across all (T, m) configurations. Averaging over the nine (T, m) designs, the APF reduces the average loss from 1.503 under the KF to 1.129, corresponding to an overall gain of 24.9%. Two additional features are noteworthy. *First*, the proportional gains are broadly stable across $T \in \{1000, 2000, 5000\}$: APF gains are approximately 29–30% for $m = 2$, about 25% for $m = 5$, and about 19–21% for $m = 10$. *Second*, the APF–BPF gap is consistently small, suggesting that the main improvement arises from relaxing the Gaussian approximation, whereas the auxiliary reweighting step provides an additional, though comparatively modest, refinement. Appendix Table A9 further shows that these gains are broadly distributed across series rather than driven by a small subset of components.

Panel B of Table 3 reports one-step-ahead out-of-sample forecasting performance. This exercise is more demanding than in-sample filtering because it combines state-propagation uncertainty, measurement non-Gaussianity, and parameter re-estimation error. Even so, the particle-based methods remain uniformly superior to the KF. In particular, the APF reduces the overall average loss from 1.853 under the KF

Table 3
Simulation results on in-sample estimation and out-of-sample forecasting accuracy for latent log-variance

Table 3 presents Monte Carlo average mean squared errors for latent log-volatility estimation and forecasting under the same DGP settings as in Section 6.1.3, for $m \in \{2, 5, 10\}$. Panel A reports the in-sample filtering mean squared error, $\text{MSE}_{\text{fil}} = \frac{1}{mT} \sum_{i=1}^m \sum_{t=1}^T (h_{t,i} - \hat{h}_{t,i|t})^2$, while Panel B reports the out-of-sample one-step-ahead forecasting mean squared error, $\text{MSE}_{\text{fc}} = \frac{1}{m} \sum_{i=1}^m (h_{T+1,i} - \hat{h}_{T+1|T,i})^2$. All reported values are averages over $B = 100$ Monte Carlo replications. KF denotes the Gaussian Kalman filter, whereas BPF and APF denote the bootstrap and auxiliary particle filters, respectively. Boldface identifies the lowest loss among KF, BPF, and APF within each row, with ties also reported in bold. The column labeled “Gain” reports the percentage improvement of APF relative to KF, computed as $100 \times (\text{KF} - \text{APF})/\text{KF}$.

T	m	Panel A: In-sample filtering loss				Panel B: One-step-ahead OOS forecasting loss			
		KF	BPF	APF	Gain	KF	BPF	APF	Gain
1000	2	1.632	1.160	1.156	29.1%	1.737	1.350	1.355	22.0%
	5	1.314	0.996	0.993	24.5%	1.763	1.462	1.465	16.9%
	10	1.627	1.334	1.323	18.7%	2.175	1.876	1.896	12.8%
2000	2	1.624	1.145	1.143	29.6%	2.110	1.555	1.553	26.4%
	5	1.302	0.982	0.979	24.8%	1.705	1.280	1.244	27.0%
	10	1.545	1.233	1.222	20.9%	1.946	1.566	1.545	20.6%
5000	2	1.645	1.158	1.158	29.6%	1.816	1.402	1.399	22.9%
	5	1.285	0.963	0.962	25.1%	1.587	1.252	1.249	21.3%
	10	1.552	1.233	1.221	21.3%	1.840	1.528	1.491	19.0%
Overall (all T, m)		1.503	1.134	1.129	24.9%	1.853	1.475	1.466	20.9%

to 1.466, implying an average improvement of 20.9%, based on the overall mean losses. For the shortest sample size, $T = 1000$, the BPF is marginally best, whereas for $T \in \{2000, 5000\}$ the APF attains the lowest loss in every configuration. This pattern is consistent with auxiliary resampling improving particle efficiency once the relative importance of parameter re-estimation noise declines. Appendix Table A10 confirms that these average gains are not driven by any single series.

Overall, the comparison shows that the gains from particle-based filtering are not merely numerical but also structural: they reflect the value of respecting the non-Gaussian nature of the transformed measurement equation. The KF remains a useful and computationally convenient benchmark, but the particle-based procedures – especially the APF – provide a more accurate treatment of the latent volatility dynamics under the transformed MSV(p) measurement system. This pattern is reflected both in the in-sample recovery of the latent log-variance path and in the one-step-ahead forecasting results.

6.2.2 Filtering and forecast accuracy: MSV versus FSV

We compare the finite-sample performance of the proposed moment-based MSV procedure with a benchmark factor stochastic volatility model estimated by Bayesian MCMC in the spirit of Aguilar and West (2000). The comparison is conducted for $m = 5$ under both an MSV and an FSV data-generating process, with $T \in \{1000, 2000, 3000\}$. For the MSV case, the simulation design follows the same specification as in Section 6.1.3; full details on the FSV design, parameter settings, estimation procedure, prior structure, and additional results are provided in Appendix J.

Table 4 reports Monte Carlo average pooled RMSE ratios for latent log-variance recovery and one-

Table 4
Pooled RMSE ratio measures and relative computational time under the MSV and FSV data-generating processes

Table 4 reports Monte Carlo average pooled RMSE ratios for latent log-variance recovery and one-step-ahead conditional-volatility forecasting under the MSV and FSV data-generating processes, for $m = 5$ and $T \in \{1000, 2000, 3000\}$. For the MSV case, the simulation design follows the same specification as in Section 6.1.3. Full details on the FSV design, including the parameter settings, estimation procedure, prior structure, and hyperparameters, are reported in Appendix J. The reported RMSE ratios are defined as FSV/MSV, so that values above one favor MSV, values below one favor FSV, and values equal to one indicate a tie. The RCT entries are computed using the same FSV/MSV normalization. Runtime is measured as the total time required to estimate the model, extract the filtered latent states, and produce the one-step-ahead forecast. All entries are Monte Carlo averages based on $B = 100$ replications. The MSV procedure is implemented with $N_{\text{pt}} = 10,000$ particles to recover the filtered log-variance paths and produce the one-step-ahead forecasts, while the benchmark FSV procedure uses a total of 20,000 draws, with the first 10,000 discarded as burn-in and every tenth draw retained thereafter, yielding 1,000 posterior draws for inference. Performance measures and series-specific results are provided in Appendix J.

True DGP	Reported ratio	$T = 1000$		$T = 2000$		$T = 3000$		Relative computational time (RCT)		
		RMSE _{h}	RMSE _{σ}	RMSE _{h}	RMSE _{σ}	RMSE _{h}	RMSE _{σ}	$T = 1000$	$T = 2000$	$T = 3000$
MSV DGP	FSV/MSV	1.7222	0.8386	1.7387	1.3693	1.7173	1.3294	145.60	168.50	151.15
FSV DGP	FSV/MSV	1.2479	0.8803	1.2476	1.0000	1.2550	1.0844	1957.55	168.25	168.16

step-ahead conditional-volatility forecasting, together with the corresponding relative computational-time ratios. For each replication, the pooled path-recovery criterion is $\text{RMSE}_h = [(mT)^{-1} \sum_{t=1}^T \sum_{j=1}^m (\hat{h}_{j,t} - h_{j,t})^2]^{1/2}$, and the pooled one-step-ahead forecast criterion is $\text{RMSE}_\sigma = [m^{-1} \sum_{j=1}^m (\hat{\sigma}_{j,T+1|T} - \sigma_{j,T+1})^2]^{1/2}$, where $\sigma_{j,T+1} = \exp(h_{j,T+1}/2)$. All reported RMSE ratios are constructed as FSV/MSV, so that values above one favor MSV, values below one favor FSV, and values equal to one indicate a tie. The same FSV/MSV normalization is used for relative computational time, so values above one imply that the MSV procedure is faster.

The latent-state recovery evidence is clear and stable across the two data-generating environments. Under the MSV data-generating process, the pooled RMSE_h ratio equals 1.7222, 1.7387, and 1.7173 for $T = 1000$, 2000, and 3000, respectively. Under the FSV data-generating process, the corresponding ratios are 1.2479, 1.2476, and 1.2550. Since all of these values exceed one, the proposed MSV procedure delivers uniformly more accurate recovery of the latent log-variance paths than the FSV benchmark across all sample sizes and under both data-generating processes.

The one-step-ahead volatility-forecast evidence is more mixed. Under the MSV data-generating process, the pooled RMSE_σ ratio is 0.8386 at $T = 1000$, indicating an advantage for FSV at the smaller sample size, but rises to 1.3693 and 1.3294 at $T = 2000$ and $T = 3000$, indicating an advantage for MSV at the two larger sample sizes. Under the FSV data-generating process, the pooled forecast ratio is 0.8803 at $T = 1000$, 1.0 at $T = 2000$, and 1.0844 at $T = 3000$, implying that FSV performs better at the smallest sample size, the two procedures are essentially tied at $T = 2000$, and MSV holds a modest advantage at $T = 3000$.

The computational evidence strongly favors the MSV procedure throughout. The FSV/MSV runtime ratio ranges from 145.60 to 168.50 under the MSV data-generating process and from 168.16 to 1957.55 under the FSV data-generating process. These magnitudes imply very large computational-time savings

in favor of MSV. Taken together, simulation evidence indicates that the proposed MSV procedure is robust to model misspecification, delivering uniformly stronger latent-state recovery across designs, remaining competitive or superior for one-step-ahead volatility forecasting depending on design and sample size, and entailing substantially lower computational cost.

6.3 Finite-sample rejection behavior of the common C-space test

Data are generated from the MSV(p) model in (2.1)-(2.2), under the same measurement equation and latent volatility transition structure as in the main framework. Throughout the simulation study, we fix $(m, p, r) = (5, 2, 2)$ and consider sample sizes $T \in \{1000, 1500, 2000\}$. The remaining parameters are chosen to regulate persistence and innovation scale. Specifically, we set $\mu = -0.5 \mathbf{1}_m$ and $\Sigma_u = \mathbf{I}_m$, $\Sigma_v = \mathbf{I}_m$. The transition matrices are first constructed using a common scale factor $c = 0.35$ and are then jointly rescaled so that the spectral radius of the companion-form matrix satisfies $\rho(\mathcal{A}) \approx 0.95$.

Under the null DGP for $H_0^{\text{com}}(r)$ given in (5.4), the coefficient matrices satisfy (5.1). Under the alternatives, we introduce a monotone deviation indexed by $\kappa \geq 0$ by adding an orthogonal component to one lag while keeping the remaining lag(s) in the common space:

$$\phi_1 = c \mathbf{B}_1 \mathbf{C}', \quad \phi_2(\kappa) = c (\mathbf{B}_2 \mathbf{C}' + \kappa \mathbf{B}_2 \mathbf{C}'_{\perp}), \quad (6.1)$$

where $\mathbf{C}_{\perp}$ spans an orthonormal complement of $\mathbf{C}$. The pair $(\phi_1, \phi_2(\kappa))$ is then rescaled to preserve $\rho(\mathcal{A}) \approx 0.95$. The null benchmark used for local LMC calibration corresponds to $\kappa = 0$, and the rejection-frequency curves are evaluated over $\kappa \in \{0, 0.5, 1, 1.5, 2\}$.

All rejection frequencies are based on 1,000 outer Monte Carlo replications and are reported at the nominal level $\alpha = 5\%$. For both the LMC and MMC procedures, each test p -value is computed by an inner Monte Carlo procedure with $N = 99$ simulated samples. For the LMC procedure, we report locally level-calibrated rejection frequencies. These are obtained by calibrating the LMC p -value cutoff at the benchmark null DGP, corresponding here to $\kappa = 0$: for each sample size T , we simulate the model under the null, compute the resulting LMC p -values, and set the calibrated cutoff equal to their empirical α -quantile. Rejection under the alternative is then based on the rule $p_{\text{LMC}} \leq \tau_T$, where τ_T denotes this calibrated cutoff. The calibration itself is based on 1,000 simulated LMC p -values for each T .¹ For MMC, we report raw rejection frequencies based on the nominal rule $p_{\text{MMC}} \leq \alpha$.

The MMC procedure is implemented through a finite local nuisance grid centered at the null-restricted plug-in estimator. Specifically, for a fixed rank r , we approximate the nuisance-parameter set by relative Frobenius neighborhoods around the restricted transition matrices and the corresponding plug-in covari-

¹The qualifier “locally level-calibrated” is deliberate. A genuine size correction would require uniform control of the null rejection probability over the admissible nuisance-parameter space, which would in turn require a numerical search not implemented here.

ance estimates,

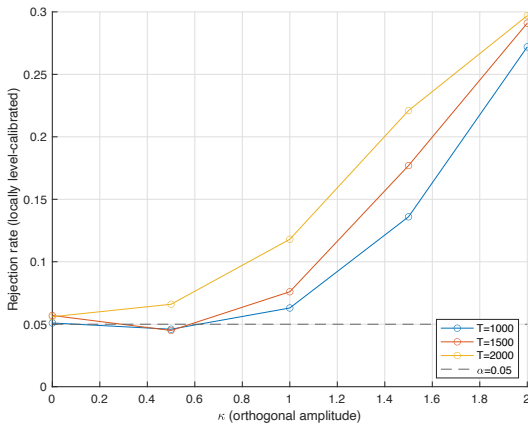
$$\begin{aligned} \mathcal{C}_T^{\text{MMC}}(r) := \{(\{\boldsymbol{\phi}_j\}_{j=1}^p, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_u) : \|\boldsymbol{\phi}_j - \hat{\boldsymbol{\phi}}_j^{(r, \text{com})}\|_F \leq c_\phi \|\hat{\boldsymbol{\phi}}_j^{(r, \text{com})}\|_F, j = 1, \dots, p, \\ \|\boldsymbol{\Sigma}_v - \hat{\boldsymbol{\Sigma}}_v\|_F \leq c_v \|\hat{\boldsymbol{\Sigma}}_v\|_F, \quad \|\boldsymbol{\Sigma}_u - \hat{\boldsymbol{\Sigma}}_u\|_F \leq c_u \|\hat{\boldsymbol{\Sigma}}_u\|_F\} \end{aligned} \quad (6.2)$$

with $c_\phi = 0.05$ and $c_v = c_u = 0.10$. In implementation, candidate points in $\mathcal{C}_T^{\text{MMC}}(r)$ are retained only when they satisfy the admissibility constraints of the MSV(p) model, in particular stability of the volatility dynamics and positive definiteness of the covariance matrices. The reported MMC p -value is the supremum of the corresponding Monte Carlo p -values over the associated finite local grid.

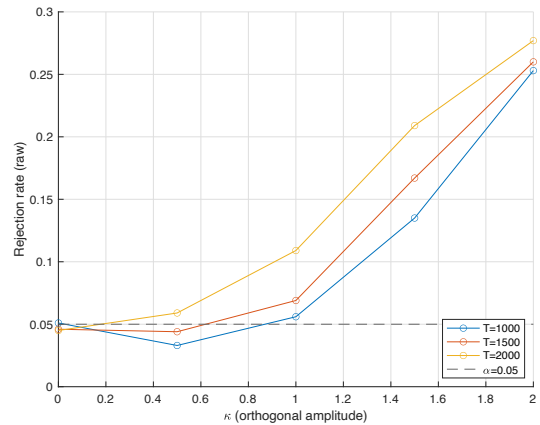
Figure 1 reports rejection-frequency curves under monotone deviations indexed by κ , with Panel (a) showing LMC locally level-calibrated rejection frequencies and Panel (b) showing MMC raw rejection frequencies. At the null benchmark $\kappa = 0$, both procedures are close to the nominal 5% level, with the LMC curve lying very near the nominal line by construction. In the near-null region, represented here by $\kappa = 0.5$, rejection frequencies remain close to α and are not yet ordered uniformly by T , indicating limited sensitivity to very small deviations from the common-space restriction. Once the departure becomes moderate, however, rejection frequencies rise steadily with κ and become increasingly ordered by sample size, so that larger T delivers stronger discrimination against the alternative. Across the entire grid, the MMC curves lie below the corresponding LMC curves by a modest margin, consistent with the more conservative behavior induced by nuisance-parameter maximization in MMC.

Figure 1
Rejection-frequency curves for the common C-space test under monotone deviations indexed by κ

Figure 1 reports rejection frequencies for the common C-space test under monotone deviations from $H_0^{\text{com}}(r)$ given in (5.4), indexed by κ , for $T \in \{1000, 1500, 2000\}$. Panel (a) shows LMC locally level-calibrated rejection frequencies, while Panel (b) shows MMC raw rejection frequencies. The horizontal dashed line marks the nominal 5% level. For both the LMC and MMC procedures, p -values are computed using $N = 99$ replications.



(a) LMC locally level-calibrated rejection frequencies



(b) MMC raw rejection frequencies

7 Empirical applications to stock returns

This section applies the proposed MSV methods to daily and high-frequency stock return data. We examine high-dimensional estimation, the effects of shrinkage choices, the forecasting performance relative to competing volatility models, and evidence on cross-sectional volatility commonality. Finally, we compare the fitted asset-level volatility paths implied by MSV and FSV models.

7.1 Estimation with high-frequency data

We use the estimated five-minute high-frequency stock returns constructed from price data in the TAQ millisecond database for the period 2004 to 2016. Log prices are constructed for five-minute sampling, yielding an average of 250 trading days per year with 77 daily increments. The first observation is the volume-weighted trading price at exactly 9:30:00 AM. For the remaining 78 observations, the volume-weighted trading prices are calculated for each second, and the last price within each 300-second interval is taken. For a significant portion of stocks, there is no trade during the first seconds of the day; thus, sampling begins 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 stocks traded daily and those included in the S&P 500 index over the period 1993 to 2012, based on Bloomberg terminal data. This results in a cross-section of approximately 500–600 firms each year. Using stocks still present in the S&P 500 index as of 2020, we obtain an intersection of 228 stocks over all 13 years. Standard data-cleaning procedures are applied: we delete all entries with timestamps outside the 9:30 AM to 4 PM window, entries with transaction prices equal to zero, and entries originating from a single exchange. Further, we remove entries with corrected trades and abnormal sale conditions.

Before turning to estimation, we remove the deterministic intraday (diurnal) volatility pattern in the 5-minute returns, following [Todorov and Tauchen \(2012\)](#). The adjustment is performed separately for each asset; to simplify notation, we suppress the asset index. For P_u the continuous-time log-price, the variable $R_{b,\ell} \equiv \Delta_{b,\ell}^{n^d} P := P_{\ell-1+b/n^d} - P_{\ell-1+(b-1)/n^d}$ represents the raw 5-minute return for intraday bin $b = 1, \dots, n^d$ on trading day $\ell = 1, \dots, N^d$. We estimate the deterministic intraday variance pattern nonparametrically, and denote the estimated variance factor for bin b by $\hat{d}_b$. The raw and standardized returns are related by

$$R_{b,\ell}^{\text{st}} \equiv \frac{R_{b,\ell}}{\sqrt{\hat{d}_b}} = \frac{\Delta_{b,\ell}^{n^d} P}{\sqrt{\hat{d}_b}}, \quad b = 1, \dots, n^d, \quad \ell = 1, \dots, N^d. \quad (7.1)$$

The construction of $\hat{d}_b$, including the continuous-time motivation and the jump-robust truncation procedure, is described in [Appendix K](#). For each asset, the standardized returns $R_{b,\ell}^{\text{st}}$, rather than the raw returns $R_{b,\ell}$, are stacked chronologically to form the return series used in the MSV estimation, with $T = n^d N^d$ observations.

Our balanced panel comprises 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 involves 52,440 parameters, while the MSV(2) model has 104,424 parameters.

Appendix Figures A1-A4 show high-dimensional heatmap plots of the estimated volatility persistence matrix with different values of shrinkage control parameter. Several conclusions emerge from these heatmaps. *First*, with $J = 1$, we find that both $\hat{\phi}_1$ and $\hat{\phi}_2$ matrices are unstable, i.e., some of the eigenvalues of the companion matrix have a modulus greater than 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 $\hat{\phi}_1$ and $\hat{\phi}_2$ are significantly different from zero. The non-zero off-diagonal elements of $\hat{\phi}_1$ imply that there are volatility spillover effects (or multilateral Granger causalities in volatility) between assets at a higher frequency. In addition to this, the non-zero off-diagonal elements of $\hat{\phi}_2$ 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 $\hat{\phi}_1$ and $\hat{\phi}_2$ 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 $T = 126,011$.

7.2 Effect of shrinkage parameter

The choice of shrinkage 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 dataset in this estimation contains daily observations from 2008/01/02 to 2020/12/31 for stocks listed on the 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 maximum eigenvalue of the companion matrix of the volatility persistence parameters estimated with different numbers of assets [$m = 2, 10, 50, 100, 250, 500, 1000, 1184$] and $J = 1, \dots, 100$.

We present these plots in Appendix 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 degree of winsorization is sufficient.

7.3 Forecasting performance and comparison to other models

We now study the out-of-sample performance of MSV(p) models in the context of asset allocation strategies. The competitors are: the Risk Metrics 2006 methodology [denoted RM, Zumbach (2007)], the dynamic conditional correlation model estimated by composite likelihood [cDCC, Engle (2002)],

Pakel et al. (2021)], the conditional covariance estimator based on a dynamic-factor structure with finite-dimensional correlation dynamics [fDCC, Alessi et al. (2009), Aramonte et al. (2013)], the generalized orthogonal GARCH model (GO-GARCH) of van der Weide (2002) with four different specifications [standard GARCH with multivariate normal distribution (GO-GARCH-S), standard GARCH with multivariate affine normal inverse Gaussian distribution (GO-GARCH-S-MANIG), GJR with multivariate normal distribution (GO-GARCH-GJR), and GJR with multivariate affine normal inverse Gaussian distribution (GO-GARCH-GJR-MANIG)], a benchmark factor stochastic volatility model [FSV, Aguilar and West (2000)], unrestricted MSV(p) models [U-MSV(p)], and restricted MSV(p) models [R-MSV(p)] with an 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:

$$\hat{\mathbf{w}}_{t+1} = \frac{\hat{\Sigma}_{t+1|t}^{-1} \iota_m}{\iota_m' \hat{\Sigma}_{t+1|t}^{-1} \iota_m} \quad (7.2)$$

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 compute one-step-ahead out-of-sample forecasts of $\Sigma_{t+1|t}$ using a rolling-window scheme. Specifically, an initial estimation sample covering $t = 1, \dots, T$ is used to estimate each model and generate the first one-step-ahead covariance forecast at time T . The estimation window is then rolled forward by one period, so that the models are re-estimated using observations $t = 2, \dots, T+1$, yielding the next forecast. This procedure is repeated sequentially until the end of the sample, producing a sequence of one-step-ahead out-of-sample covariance forecasts. Using the forecast of $\Sigma_{t+1|t}$, we compute the GMV weights, and then form the corresponding realized portfolio return, $r_{p,t+1} = \hat{\mathbf{w}}_{t+1}' \mathbf{r}_{t+1}$, where $\mathbf{r}_{t+1} = (r_{1,t+1}, r_{2,t+1}, \dots, r_{m,t+1})'$ collects the asset returns between day t and day $t+1$.

The one-step-ahead excess portfolio return relative to the equal-weighted portfolio is denoted by $\hat{r}_{p,t+1}$.² Following 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 of the sample mean of the out-of-sample excess returns over those obtained from the equal-weight portfolio and their sample standard deviation:

$$\mathcal{SR} = \frac{\frac{1}{T-1} \sum_{t=1}^{T-1} \hat{r}_{p,t+1}}{\sqrt{\frac{1}{T-2} \sum_{t=1}^{T-1} (\hat{r}_{p,t+1} - \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{r}_{p,t+1})^2}}. \quad (7.3)$$

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

The dataset in the estimation contains daily observations from 2008/01/02 to 2020/12/31 of all stocks included in the S&P 500 index. We consider portfolios of 20, 50, and 100 stocks respectively, with compo-

²We compute the equal-weighted portfolio with rebalancing to maintain the naive $1/m$ or strategic weight for each asset over time.

sitions detailed in Section E, prioritizing stocks with higher weights in the S&P index. The data are in a log difference form, adjusted for dividends, and obtained from CRSP through Wharton Research Data Services. In the out-of-sample forecast experiment, we consider an initial in-sample period from 2008/01/02 to 2018/12/31, and the out-of-sample runs 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 $MSV(p)$ forecasts are obtained using the W-VARMA-MVR estimator with $J = 100$ and the APF filter with $N_{pf} = 10,000$.

Table 5 reports the annualized $\mathcal{SR}$ of excess minimum-variance portfolio returns, computed over the out-of-sample period after daily rebalancing. Our findings can be summarized as follows.

First, the results reported in Table 5 show that the strongest portfolio-forecasting performance is delivered by the unrestricted higher-order U- $MSV(p)$ family. This indicates that rich latent volatility persistence and cross-series volatility spillovers are first-order ingredients for covariance forecasting. By contrast, the restricted R- $MSV(p)$ specifications [which impose an uncorrelated $SV(p)$ structure] are uniformly weaker.

Second, relative to the other multivariate volatility benchmarks, the unrestricted U- $MSV(p)$ specifications remain clearly competitive. They often dominate RM, cDCC, fDCC, GO-GARCH, and FSV in terms of annualized Sharpe ratio. In addition, we encountered numerical instability with the GO-GARCH specifications [GO-GARCH-S, GO-GARCH-GJR, GO-GARCH-GJR-MANIG] and the fDCC model when $m = 50, 100$. In these cases, we do not report the corresponding Sharpe ratios. This finding underscores the value of using an analytically tractable and computationally efficient estimator, such as the W-VARMA estimator, which can be implemented directly without numerical optimization.

Third, the results continue to support the inclusion of higher-order volatility dynamics. Across the unrestricted MSV family, specifications with $p > 1$ frequently outperform the first-order versions, which suggests that additional lag terms in latent volatility are empirically relevant for covariance forecasting.

Finally, during the highly volatile out-of-sample period that includes the Covid-19 episode, the best-performing forecasts still come from the unrestricted higher-order MSV class. Our broader conclusion is therefore maintained: the main empirical advantage comes from modeling latent volatilities through a flexible higher-order MSV structure.

7.4 Empirical assessment of volatility commonality

We investigate volatility commonality across three sets of return portfolios with increasing cross-sectional dimension, $m \in \mathcal{M} := \{20, 50, 100\}$, using the unrestricted $MSV(\hat{p}_m)$ specifications selected from the out-of-sample GMV portfolio exercise reported in Table 5. For each portfolio dimension, the lag order $\hat{p}_m$ is set equal to the lag order of the specification yielding the highest annualized Sharpe ratio for the corresponding GMV portfolio. The selected lag orders are $\hat{p}_{20} = 2$, $\hat{p}_{50} = 4$, and $\hat{p}_{100} = 2$. Common-C inference is therefore conducted conditionally on these unrestricted benchmark models.

For each m , we construct $\hat{\mathbf{M}}_{C,m} := \sum_{j=1}^{\hat{p}_m} \hat{\boldsymbol{\phi}}'_{j,m} \hat{\boldsymbol{\phi}}_{j,m}$ and examine its eigenvalue profile over the screening grid $r = 1, \dots, 10$. For each candidate rank r , we then evaluate the common-C null [see (5.4)] using the

Table 5
Portfolio performance under different conditional volatility models
Annualized Sharpe ratio

Table 5 reports annualized Sharpe ratios for one-step-ahead global minimum variance portfolio excess returns relative to the rebalanced equal-weight portfolio. Results are shown for portfolios of 20, 50, and 100 stocks over the out-of-sample period from 2019/01/02 to 2020/12/31. MSV(p) forecasts are obtained using the W-VARMA-MVR estimator with $J = 100$. A dash indicates that the corresponding model did not converge reliably and is therefore not reported. Boldface entries indicate positive Sharpe ratios, while boldface colored entries indicate the best-performing model within each column.

Model	20 assets			50 assets			100 assets		
	2019–2020	2019	2020	2019–2020	2019	2020	2019–2020	2019	2020
RM	-0.773	-0.899	-0.788	-0.254	-1.988	0.380	-0.728	-1.384	-0.513
fDCC	-1.245	-1.047	-1.409	-	-	-	-	-	-
cDCC	-1.217	-1.092	-1.341	-0.947	-1.746	-0.715	-0.675	-0.522	-0.762
GO-GARCH-S	-0.907	-0.762	-1.024	-	-	-	-	-	-
GO-GARCH-GJR	-0.839	-0.686	-0.956	-	-	-	-	-	-
GO-GARCH-S-MANIG	-0.782	-0.756	-0.835	-0.570	-0.569	-0.629	-0.820	-0.465	-1.024
GO-GARCH-GJR-MANIG	-0.886	-0.813	-0.981	-	-	-	-	-	-
FSV	-0.412	-0.492	-0.335	-0.752	-0.897	-0.666	-1.293	-0.770	-1.584
U-MSV(1)	0.453	0.259	0.623	-0.910	0.507	-1.201	-2.004	-3.243	-2.610
U-MSV(2)	1.571	0.889	2.449	0.464	-1.239	1.370	1.807	2.130	1.038
U-MSV(3)	0.776	0.177	1.123	-0.818	-1.714	-0.594	-1.905	1.691	-2.962
U-MSV(4)	-0.738	1.330	-1.282	0.577	-1.187	1.513	1.511	0.661	1.943
U-MSV(5)	-0.029	-0.906	1.359	-0.042	0.802	-1.012	0.735	2.621	0.652
R-MSV(1)	-0.962	-0.958	-0.987	-0.852	-1.500	-0.605	-1.197	-1.433	-1.002
R-MSV(2)	-1.004	-1.228	-0.909	-0.629	-0.831	-0.530	-1.095	-1.205	-0.978
R-MSV(3)	-0.982	-1.121	-0.935	-1.189	-0.656	-1.511	-1.091	-0.974	-1.120
R-MSV(4)	-1.047	-1.373	-0.856	-0.386	-0.138	-0.509	-0.982	-0.872	-0.989
R-MSV(5)	-0.776	-1.276	-0.446	-0.777	-0.629	-0.868	-0.560	-0.745	-0.358

statistic $S_{T,m}^{\text{com}}(r) := T \cdot \max\{0, D_{\text{com},m}(\hat{\phi}; r) - D_{\text{rr},m}(\hat{\phi}; r)\}$, with the LMC procedure as the baseline decision rule and the corresponding MMC p -values reported as a robustness benchmark. The full screening and testing results are reported in Appendix L, where a single table combines the eigenvalue concentration diagnostics with the formal LMC and MMC evidence.

The empirical evidence does not support the common-C restriction. The LMC test rejects the null for every candidate rank in each of the three portfolio dimensions. MMC results lead to the same conclusion, except for a single non-rejection at $(m, r) = (50, 10)$. Since this isolated case occurs only at the upper edge of the tested grid, it does not provide convincing support for a stable low-dimensional common dynamic subspace. We therefore do not retain the common-C restriction in the forecasting exercise. More generally, although the unrestricted volatility dynamics exhibit some eigenvalue concentration, the data do not support the hypothesis that the lag matrices share an exact low-dimensional common right subspace.

Taken together, these findings suggest that the common-C restriction is too restrictive for the equity portfolios considered here. The rejection does not imply the absence of volatility comovement across assets; rather, it indicates that such comovement is not adequately captured by an exact low-dimensional dynamic subspace shared by all volatility-transition matrices. Hence, market-wide and sectoral forces may still contribute to volatility comovement, but the empirical evidence favors a more heterogeneous propagation mechanism with asset-specific dynamics and cross-series volatility spillovers. This interpretation is

consistent with the forecasting exercise, where the unrestricted higher-order MSV specifications perform better than their restricted counterparts.

7.5 Comparing estimated volatilities under MSV and FSV

To assess directly whether the stochastic volatilities implied by MSV and FSV are similar in the empirical application, we compare the fitted series-specific log-variance paths from the unrestricted MSV model with those from the FSV benchmark; detailed discussions and results are reported in Appendix M. The evidence points to limited similarity overall. For $m = 20$ and $m = 50$, the average asset-specific time-series correlations are close to zero, while the standardized path-distance measures are large, indicating substantial differences in the time-series evolution of asset-level volatilities under the two models. For $m = 100$, the average asset-specific correlation increases, and the standardized distance declines somewhat, suggesting modestly greater time-series agreement, but the average date-by-date cross-sectional correlation remains close to zero and its median is negative. Thus, even in the high-dimensional case, the two models do not imply a similar cross-sectional volatility structure. Overall, the empirical results indicate that the stochastic volatilities implied by MSV and FSV are not close substitutes, and any similarity between them is limited and dimension-dependent.

8 Scope, extensions, and practical implementation

The analysis in Sections 2-4 is developed for the stationary $MSV(p)$ specification in (2.1)-(2.2), namely, for an innovation process with zero conditional mean, time-varying log-volatilities, and a time-invariant correlation matrix for standardized shocks. This formulation is adopted for tractability and the development of the core moment-based theory, but it does not exhaust the practical scope of the proposed estimators. More generally, the procedure may be viewed as a scalable *second-moment module* that can be applied to an innovation or pseudo-innovation sequence, provided the transformed second moments satisfy the MSV-type moment restrictions exploited in this paper. Here, pseudo-innovations refer to residuals from a first-stage conditional-mean specification, used as proxies for the unobserved innovations in the second-stage volatility estimation.

This perspective is especially relevant for macroeconomic applications with nontrivial conditional means. In a small- or large-scale VAR, a natural implementation is a two-step procedure. In the first step, one estimates the conditional mean dynamics and constructs fitted residuals or pseudo-innovations. In the second step, one applies the MSV estimator to these innovations in order to recover latent log-volatility states, conditional covariance matrices, and volatility forecasts. The same logic extends to FAVAR and related multi-equation systems. If interactions between the mean and volatility blocks are empirically important, the two stages may be iterated using feasible GLS-type updates. From an asymptotic perspective, such a strategy introduces generated-regressor effects in the distribution of the volatility estimator; in

practice, inference may be handled by an appropriate sandwich variance formula or by joint resampling schemes which re-estimate both stages. Thus, the zero-mean specification used for the core theoretical analysis should be interpreted as a tractable benchmark rather than as a restriction to purely financial settings. In particular, if the innovation sequence from a conditional-mean model is well approximated by a stationary MSV process of the form studied here, then the moment estimator can be applied directly to fitted VAR residuals, thereby providing a practical route to VAR-MSV specifications without requiring full system-wide simulation-based estimation.

At the same time, the proposed method is not intended as a universal estimator for all stochastic-volatility parameterizations. Its direct closed-form applicability relies on the transformed measurement system and weak VARMA representation derived for the $\text{MSV}(p)$ model studied in this paper. For alternative parameterizations – including triangular or Cholesky-type stochastic-volatility models such as [Cogley and Sargent \(2005\)](#) – the method does not automatically apply in the same form, because the relevant transformed moments generally obey a different structure. In such cases, one would need either a suitable mapping to an innovation sequence satisfying MSV-type moment restrictions or a new derivation of model-specific moment equations. Accordingly, our claim is not that the estimator is fully model-agnostic, but rather that it extends naturally to settings in which the object of interest can be reduced to an innovation process with an MSV-type second-moment structure. This modular interpretation is useful in richer macro-financial systems as well. In DFAVAR and related factor-augmented settings, one may first estimate the conditional-mean or factor block and then apply the MSV estimator to the resulting shocks or pseudo-shocks; this is the strategy underlying the DFAVAR-HMSV framework developed in [Ahsan et al. \(2025\)](#). In this framework, the estimated volatility block is used not only for volatility forecasting, but also for uncertainty measurement through macroeconomic uncertainty indices constructed from cross-sectional averages of predicted volatilities. The same strategy extends to regime-switching environments, where one may first estimate the regime-dependent mean or factor structure, and then model the resulting innovations through an MSV block, as in the Markov-switching FAVAR framework studied in [Ahsan et al. \(2026\)](#).

This clarification also helps explain why the MSV specification remains of independent interest. Even with a time-invariant correlation matrix for standardized shocks, the model delivers a fully time-varying conditional covariance matrix through the multivariate dynamics of the latent log-volatility vector. Cross-sectional dependence enters through the unrestricted transition matrices $\{\phi_j\}_{j=1}^p$ and the volatility-innovation covariance matrix Σ_v , so the framework accommodates volatility spillovers, persistent common movements, and rich covariance dynamics in high dimension while remaining computationally feasible. The same modular perspective also points to possible extensions toward richer dependence modelling, for example by using MSV-standardized residuals as inputs to a separate DCC-type correlation procedure. A fully rigorous treatment of such extensions, however, would require additional theoretical development and lies beyond the scope of the present paper. Other extensions, including non-Gaussian

innovations, leverage or cross-leverage effects, and richer factor-based macro-financial systems, remain promising directions for future research.

9 Conclusion

This paper examines the estimation of unrestricted higher-order multivariate stochastic volatility models and proposes a class of computationally simple moment-based estimators built on the weak VARMA representation implied by the transformed MSV(p) model. Asymptotic distributional theory is developed for these estimators, supporting standard large-sample inference while preserving computational tractability. Fast implementation and ease of simulation also make the proposed procedures well suited to simulation-based inference, including Monte Carlo methods with nuisance parameters. Across the simulation designs, the W-VARMA-MVR estimator delivers the best overall performance among the proposed estimators and, relative to Bayesian and QML benchmarks, provides accurate finite-sample estimates at a substantially lower computational cost. Large-scale simulation evidence further indicates that the approach remains feasible in very high dimensions.

Our findings also show that the framework is useful beyond static parameter estimation. In latent log-variance filtering and forecasting exercises, particle-based methods – especially APF – deliver consistent improvements over the Gaussian Kalman benchmark. In comparisons with the FSV model, the proposed MSV procedure provides uniformly stronger latent-state recovery, remains competitive or superior in one-step-ahead volatility forecasting depending on the design, and incurs substantially lower computational cost. These results suggest that the MSV procedure is robust to model misspecification, since it continues to perform well even when the true data-generating process follows an FSV structure rather than an unrestricted MSV specification. In empirical applications to stock returns, unrestricted higher-order MSV specifications achieve the strongest overall covariance-forecasting performance relative to a broad range of competing multivariate volatility models, further underscoring the practical importance of higher-order volatility dynamics. At the same time, the empirical results do not provide convincing support for an exact low-dimensional dynamic subspace shared across lags, and fitted volatility paths implied by MSV and FSV are not close substitutes in the applications considered.

Taken together, the results show that the proposed estimator provides a practical, scalable, and computationally tractable framework for modelling high-dimensional second-moment dynamics. The approach can be combined naturally with first-stage conditional-mean specifications, including VAR, FAVAR, and related systems, while preserving feasibility for volatility forecasting and uncertainty measurement based on predicted volatilities. The framework may also serve as a useful starting point for more flexible specifications of dependence, innovation distributions, leverage or cross-leverage effects, and richer factor-based macro-financial systems in multivariate settings.

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](https://doi.org/10.1080/07350015.2000.10524875).
- Ahn, S. K. and Reinsel, G. C. (1988), ‘Nested reduced-rank autoregressive models for multiple time series’, *Journal of the American Statistical Association* **83**(403), 849–856, doi:[10.1080/01621459.1988.10478673](https://doi.org/10.1080/01621459.1988.10478673).
- 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. (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](https://doi.org/10.1016/j.jeconom.2021.03.008).
- Ahsan, M. N., Dufour, J.-M. and Rodriguez-Rondon, G. (2025), Modèles de volatilité stochastique à haute dimension: applications à l’incertitude macroéconomique au québec et au canada, Technical Report 2025RP-19, CIRANO.
- Ahsan, M. N., Dufour, J.-M. and Rodriguez-Rondon, G. (2026), Markov-switching FAVAR with multivariate stochastic volatility: Theory and inference with applications to uncertainty and growth-at-risk, McGill University working paper.
- 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.
- 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](https://doi.org/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](https://doi.org/10.1016/j.jeconom.2008.12.012).
- 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](https://doi.org/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](https://doi.org/10.1016/0304-4076(86)90063-1).
- Brockwell, P. J. and Davis, R. A. (1996), *Time series: theory and methods*, Springer series in statistics, 2nd edn, Springer, New York.

- 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](https://doi.org/10.1111/j.1467-6419.2004.00232.x).
- Broze, L., Francq, C. and Zakoïan, J.-M. (2001), 'Non-redundancy of high order moment conditions for efficient GMM estimation of weak AR processes', *Economics Letters* **71**(3), 317–322, doi:[10.1016/S0165-1765\(01\)00387-1](https://doi.org/10.1016/S0165-1765(01)00387-1).
- Carnero, M. A., Peña, D. and Ruiz, E. (2004), 'Persistence and kurtosis in GARCH and stochastic volatility models', *Journal of Financial Econometrics* **2**(2), 319–342.
- 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.
- 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](https://doi.org/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.
- Chan, J. C. and Grant, A. L. (2016), 'Modeling energy price dynamics: GARCH versus stochastic volatility', *Energy Economics* **54**, 182–189.
- 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](https://doi.org/10.1137/S0895479896302898).
- 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](https://doi.org/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. (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](https://doi.org/10.1016/S0927-5398(97)00016-9).
- Davidson, J. (1994), *Stochastic limit theory*, Oxford University Press.
- 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](https://doi.org/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](https://doi.org/10.1016/j.jeconom.2005.06.007).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.2307/1912773).
- Engle, R. F. and Kozicki, S. (1993), 'Testing for common features', *Journal of Business & Economic Statistics* **11**(4), 369–380, doi:[10.1080/07350015.1993.10509966](https://doi.org/10.1080/07350015.1993.10509966).
- Engle, R. F. and Susmel, R. (1993), 'Common volatility in international equity markets', *Journal of Business & Economic Statistics* **11**(2), 167–176, doi:[10.1080/07350015.1993.10509945](https://doi.org/10.1080/07350015.1993.10509945).
- Engle, R. F., Ng, V. K. and Rothschild, M. (1990), 'Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills', *Journal of Econometrics* **45**(1–2), 213–237, doi:[10.1016/0304-4076\(90\)90099-F](https://doi.org/10.1016/0304-4076(90)90099-F).

- 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.
- Gourieroux, C. (2006), 'Continuous time wishart process for stochastic risk', *Econometric Reviews* **25**(2-3), 177–217, doi:[10.1080/07474930600713234](https://doi.org/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](https://doi.org/10.1016/j.jeconom.2008.12.016).
- 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](https://doi.org/10.1093/rfs/hhj002).
- 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](https://doi.org/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](https://doi.org/10.1016/j.jeconom.2019.04.028).
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1111/1467-937X.00050).
- 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](https://doi.org/10.1016/j.jempfin.2004.04.009).
- 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](https://doi.org/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](https://doi.org/10.1080/07350015.2020.1713795).
- Phan, T. D., Wachter, J. A., Solomon, E. A. and Kahana, M. J. (2019), 'Multivariate stochastic volatility modeling of neural data', *eLife* **8**, e42950, doi:[10.7554/eLife.42950](https://doi.org/10.7554/eLife.42950).
- Philipov, A. and Glickman, M. E. (2006a), 'Factor multivariate stochastic volatility via Wishart processes', *Econometric Reviews* **25**(2-3), 311–334, doi:[10.1080/07474930600713366](https://doi.org/10.1080/07474930600713366).
- Philipov, A. and Glickman, M. E. (2006b), 'Multivariate stochastic volatility via Wishart processes', *Journal of Business and Economic Statistics* **24**(3), 313–328, doi:[10.1198/073500105000000306](https://doi.org/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.
- Politis, D. N. (2009), 'An algorithm for robust fitting of autoregressive models', *Economics Letters* **102**(2), 128–131, doi:[10.1016/j.econlet.2008.12.004](https://doi.org/10.1016/j.econlet.2008.12.004).
- 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](https://doi.org/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.
- 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](https://doi.org/10.21105/joss.00026).
- 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.
- Shephard, N. and Pitt, M. K. (1997), 'Likelihood analysis of non-gaussian measurement time series', *Biometrika* **84**(3), 653–667.
- 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](https://doi.org/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](https://doi.org/10.1002/for.1123).

- 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.
- Todorov, V. and Tauchen, G. (2012), ‘Inverse realized laplace transforms for nonparametric volatility density estimation in jump-diffusions’, *Journal of the American Statistical Association* **107**(498), 622–635.
- Uhlig, H. (1997), ‘Bayesian vector autoregressions with stochastic volatility’, *Econometrica* **65**(1), 59–73, doi:[10.2307/2171813](https://doi.org/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](https://doi.org/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](https://doi.org/10.1080/09603100110090118).
- 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](https://doi.org/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

Supplementary Appendix

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

August 2026

This online Supplementary Appendix provides additional technical details, implementation material, and complementary empirical and simulation results for the paper.

- A. Proofs
- B. Monte Carlo tests with nuisance parameters
- C. Implicit standard error
- D. MCMC algorithm for MSV models
- E. Composition of portfolios
- F. Forecasting with MSV(p) models
- G. Technical details for common C-space restriction
- H. Additional figures
- I. Additional tables
- J. Supplementary Monte Carlo details and results: MSV versus FSV
- K. Intraday adjustment of high-frequency returns
- L. Empirical common-C diagnostics and test results
- M. Similarity of estimated volatilities under MSV and FSV

[†]Canada Mortgage and Housing Corporation (CMHC), 700 Montreal Road, Ottawa, Ontario K1A 0P7, Canada; e-mail: mahsan@cmhc-schl.gc.ca.

Disclaimer: The views expressed in this article are those of the authors and do not necessarily reflect the position of the Canada Mortgage and Housing Corporation.

[‡]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>

A Proofs

PROOF OF PROPOSITION 3.1 Define the autoregressive matrix polynomial

$$\boldsymbol{\phi}(B) = \mathbf{I}_m - \sum_{j=1}^p \boldsymbol{\phi}_j B^j.$$

From (2.5)-(2.6),

$$\boldsymbol{\phi}(B)(\mathbf{h}_t - \boldsymbol{\mu}) = \mathbf{v}_t, \quad \mathbf{X}_t - \boldsymbol{\mu} = (\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t.$$

Applying $\boldsymbol{\phi}(B)$ to the second equality and using the first gives

$$\begin{aligned} \boldsymbol{\phi}(B)(\mathbf{X}_t - \boldsymbol{\mu}) &= \mathbf{v}_t + \boldsymbol{\phi}(B)\boldsymbol{\epsilon}_t \\ &= \mathbf{v}_t + \boldsymbol{\epsilon}_t - \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\epsilon}_{t-j} =: \mathbf{w}_t. \end{aligned} \tag{A.1}$$

We first characterize the second-order dependence of $\mathbf{w}_t$. By Assumption 2.1, the sequence $\{(\mathbf{u}'_t, \mathbf{v}'_t)'\}$ is i.i.d. Gaussian with block-diagonal covariance matrix. Hence $\mathbf{u}_t$ and $\mathbf{v}_t$ are independent at each date, and the innovation vectors are independent across dates. Since $\boldsymbol{\epsilon}_t$ defined after (2.4) is a measurable function of $\mathbf{u}_t$ with finite second moments, it follows that $\{\boldsymbol{\epsilon}_t\}$ is i.i.d. with mean zero and covariance matrix $\boldsymbol{\Sigma}_\epsilon$, and is independent of $\{\mathbf{v}_t\}$ at all leads and lags. The matrix $\boldsymbol{\Sigma}_\epsilon$ is positive definite. Indeed, if $\mathbf{a}'\boldsymbol{\Sigma}_\epsilon\mathbf{a} = 0$ for some $\mathbf{a} \in \mathbb{R}^m$, then $\sum_{i=1}^m a_i \log(u_{it}^2)$ is almost surely constant. Since $\mathbf{u}_t$ has a density that is strictly positive on $\mathbb{R}^m$ ($\boldsymbol{\Sigma}_u$ being positive definite) and the function $x \mapsto \sum_{i=1}^m a_i \log(x_i^2)$ is continuous on each open orthant, this almost-sure constancy implies that the function is constant on every open orthant. Its gradient $(2a_1/x_1, \dots, 2a_m/x_m)'$ must therefore vanish there, which forces $\mathbf{a} = \mathbf{0}$. In particular,

$$\mathbb{E}(\mathbf{v}_t \mathbf{v}'_s) = \mathbf{0}, \quad \mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}'_s) = \mathbf{0}, \quad t \neq s, \quad \mathbb{E}(\mathbf{v}_t \boldsymbol{\epsilon}'_s) = \mathbf{0}, \quad t, s \in \mathbb{Z}. \tag{A.2}$$

For notational convenience, set

$$\mathbf{A}_0 = \mathbf{I}_m, \quad \mathbf{A}_j = -\boldsymbol{\phi}_j, \quad j = 1, \dots, p.$$

Then

$$\mathbf{w}_t = \mathbf{v}_t + \sum_{j=0}^p \mathbf{A}_j \boldsymbol{\epsilon}_{t-j}.$$

Let

$$\boldsymbol{\Gamma}_{w,k} := \mathbb{E}(\mathbf{w}_t \mathbf{w}'_{t-k}), \quad k \geq 0.$$

Using (A.2), at lag zero we obtain

$$\begin{aligned} \boldsymbol{\Gamma}_{w,0} &= \boldsymbol{\Sigma}_v + \sum_{j=0}^p \mathbf{A}_j \boldsymbol{\Sigma}_\epsilon \mathbf{A}'_j \\ &= \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon + \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\Sigma}_\epsilon \boldsymbol{\phi}'_j. \end{aligned} \tag{A.3}$$

For $1 \leq k \leq p$, a term involving $\boldsymbol{\epsilon}_{t-a}$ in $\mathbf{w}_t$ and a term involving $\boldsymbol{\epsilon}_{t-k-b}$ in $\mathbf{w}_{t-k}$ have nonzero covariance

only when $a = k + b$. Consequently,

$$\begin{aligned}\Gamma_{w,k} &= \sum_{b=0}^{p-k} A_{k+b} \Sigma_\epsilon A_b' \\ &= -\phi_k \Sigma_\epsilon + \sum_{b=1}^{p-k} \phi_{k+b} \Sigma_\epsilon \phi_b', \quad 1 \leq k \leq p,\end{aligned}\tag{A.4}$$

where an empty sum is equal to zero. If $k > p$, the dated innovations entering w_t and w_{t-k} are disjoint, so that

$$\Gamma_{w,k} = \mathbf{0}, \quad k > p.\tag{A.5}$$

Thus $\{w_t\}$ is covariance stationary and second-order p -dependent; indeed, w_t and w_s are independent whenever $|t - s| > p$, because they are measurable functions of disjoint blocks of the i.i.d. innovation sequence in Assumption 2.1.

It remains to establish that w_t admits a finite m -dimensional moving-average representation. This step is a finite matrix spectral-factorization problem; the Wold decomposition alone only guarantees an $\text{MA}(\infty)$ innovations representation. Let $z = e^{-i\lambda}$, $|z| = 1$, and write

$$\phi(z) = I_m - \sum_{j=1}^p \phi_j z^j.$$

Apart from the normalizing factor 2π , the spectral density matrix of w_t is

$$S_w(z) = \Sigma_v + \phi(z) \Sigma_\epsilon \phi(z^{-1})', \quad |z| = 1.\tag{A.6}$$

Because the coefficient matrices are real, $\phi(z^{-1})' = \phi(z)^*$ on the unit circle, where $*$ denotes conjugate transpose. Hence $S_w(z)$ is a Hermitian matrix Laurent polynomial of order at most p . Moreover, Σ_v is positive definite by Assumption 2.1, while Σ_ϵ is positive definite, as established above (although only positive semidefiniteness is needed at this step). Thus, for every nonzero $a \in \mathbb{C}^m$ and every $|z| = 1$,

$$\begin{aligned}a^* S_w(z) a &= a^* \Sigma_v a + [\phi(z)^* a]^* \Sigma_\epsilon [\phi(z)^* a] \\ &> 0.\end{aligned}$$

Therefore $S_w(z)$ is positive definite on the unit circle.

By the matrix Fejér–Riesz theorem [see Ephremidze et al. (2009)], $S_w(z)$ therefore admits an outer polynomial spectral factor of degree at most p . The same theorem implies uniqueness of the outer factor up to right multiplication by a constant unitary matrix. Hence there exists an $m \times m$ matrix polynomial

$$C(z) = C_0 + C_1 z + \cdots + C_p z^p,$$

where some of the highest-order coefficient matrices may be zero, such that

$$S_w(z) = C(z) C(z)^*, \quad |z| = 1.\tag{A.7}$$

The factor $C(z)$ is outer, i.e. with $\det[C(z)] \neq 0$ for $|z| < 1$. We normalize C_0 to be lower triangular with

positive diagonal entries. This normalization is complete: if $\tilde{C}(z) = C(z)Q$ is another outer factor satisfying the same normalization, where Q is constant unitary, then C_0 and $\tilde{C}_0 = C_0Q$ are two lower-triangular positive-diagonal Cholesky factors of the same positive-definite matrix,

$$C_0 C_0^* = \tilde{C}_0 \tilde{C}_0^*.$$

Uniqueness of the Cholesky factor therefore gives $\tilde{C}_0 = C_0$ and hence $Q = I_m$.

Because the coefficient matrices $\Gamma_{w,k}$ of $S_w(z)$ are real, define $D(z) := \overline{C(\bar{z})}$. On the unit circle,

$$D(z)D(z)^* = \overline{S_w(\bar{z})} = S_w(z),$$

so $D(z)$ is also an outer spectral factor. Its constant term $D_0 = \overline{C_0}$ is lower triangular with the same positive diagonal entries, and hence satisfies the same complete normalization. By uniqueness, $D(z) = C(z)$, which implies that $C_0, \dots, C_p$ are real matrices. Consequently, on the unit circle, $C(z)^* = C(z^{-1})'$, so (A.7) can equivalently be written as

$$S_w(z) = C(z)C(z^{-1})', \quad |z| = 1.$$

Since $S_w(z)$ is positive definite on the unit circle, $|\det[C(z)]|^2 = \det[S_w(z)] > 0$ for $|z| = 1$, so that in fact

$$\det[C(z)] \neq 0, \quad |z| \leq 1,$$

and in particular C_0 is nonsingular.

Define

$$\xi_t := C(B)^{-1}w_t.$$

Because $C(z)$ is outer, the inverse filter is causal and stable. From (A.7), the spectral density matrix of ξ_t is

$$\frac{1}{2\pi} C(z)^{-1} S_w(z) [C(z^{-1})']^{-1} = \frac{1}{2\pi} I_m.$$

Thus $\{\xi_t\}$ is an m -dimensional real, mean-zero weak white-noise process with covariance matrix I_m . Set

$$\eta_t := C_0 \xi_t, \quad \Sigma_\eta := C_0 C_0', \quad \theta_j := -C_j C_0^{-1}, \quad j = 1, \dots, p.$$

Then

$$\begin{aligned} w_t &= C_0 \xi_t + \sum_{j=1}^p C_j \xi_{t-j} \\ &= \eta_t - \sum_{j=1}^p \theta_j \eta_{t-j}. \end{aligned} \tag{A.8}$$

Writing

$$\theta(z) = I_m - \sum_{j=1}^p \theta_j z^j,$$

we have

$$\boldsymbol{\theta}(z)C_0 = C(z).$$

The polynomial $\boldsymbol{\theta}(z)$ is invariant to the unitary indeterminacy of the outer factor. Indeed, if another outer factor is $\tilde{C}(z) = C(z)Q$ for a constant unitary matrix Q , then $\tilde{C}_0 = C_0Q$ and

$$\tilde{\boldsymbol{\theta}}(z) = \tilde{C}(z)\tilde{C}_0^{-1} = C(z)Q(C_0Q)^{-1} = C(z)C_0^{-1} = \boldsymbol{\theta}(z).$$

Since C_0 is nonsingular and $\det[C(z)] \neq 0$ for $|z| \leq 1$, it follows that

$$\det[\boldsymbol{\theta}(z)] \neq 0, \quad |z| \leq 1,$$

so the moving-average polynomial is invertible.

Combining (A.1) and (A.8) gives

$$\boldsymbol{\phi}(B)(X_t - \mu) = \boldsymbol{\theta}(B)\eta_t.$$

Expanding both polynomials yields (3.1) and $\eta_t - \sum_{j=1}^p \boldsymbol{\theta}_j \eta_{t-j} = v_t + \epsilon_t - \sum_{j=1}^p \boldsymbol{\phi}_j \epsilon_{t-j}$. By Assumption 2.2, $\det[\boldsymbol{\phi}(z)] \neq 0$ for $|z| \leq 1$, so the autoregressive polynomial is causal. We have already shown that the moving-average polynomial is invertible. Therefore X_t admits a causal and invertible weak VARMA(p, p) representation, with $\Sigma_\eta = C_0 C_0'$ positive definite because C_0 is nonsingular. The corresponding scalar signal-plus-noise result, AR(p) plus white noise yielding an ARMA(p, p) representation, is given by Granger and Morris (1976).

For completeness, the covariance-matching equations for $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$ and Σ_η follow directly from (A.8). At lag zero, comparison with (A.3) gives

$$\Sigma_\eta + \sum_{j=1}^p \boldsymbol{\theta}_j \Sigma_\eta \boldsymbol{\theta}_j' = \Sigma_v + \Sigma_\epsilon + \sum_{j=1}^p \boldsymbol{\phi}_j \Sigma_\epsilon \boldsymbol{\phi}_j'. \quad (\text{A.9})$$

For $k = 1, \dots, p$, comparison with (A.4) gives

$$-\boldsymbol{\theta}_k \Sigma_\eta + \sum_{b=1}^{p-k} \boldsymbol{\theta}_{k+b} \Sigma_\eta \boldsymbol{\theta}_b' = -\boldsymbol{\phi}_k \Sigma_\epsilon + \sum_{b=1}^{p-k} \boldsymbol{\phi}_{k+b} \Sigma_\epsilon \boldsymbol{\phi}_b', \quad k = 1, \dots, p, \quad (\text{A.10})$$

with empty sums equal to zero. In particular, for $k = p$,

$$-\boldsymbol{\theta}_p \Sigma_\eta = -\boldsymbol{\phi}_p \Sigma_\epsilon.$$

The last covariance-matching equation also establishes the exact moving-average order directly. Since both Σ_ϵ and Σ_η are positive definite, they are nonsingular, and therefore

$$\boldsymbol{\theta}_p = \boldsymbol{\phi}_p \Sigma_\epsilon \Sigma_\eta^{-1}. \quad (\text{A.11})$$

Consequently, if the latent VAR is of exact order p , so that $\boldsymbol{\phi}_p \neq \mathbf{0}$, then $\boldsymbol{\theta}_p \neq \mathbf{0}$. Hence the moving-average polynomial is also of exact order p .

We next show that the polynomial pair $[\boldsymbol{\phi}(z), \boldsymbol{\theta}(z)]$ is left coprime. Since $\boldsymbol{\theta}(z)\mathbf{C}_0 = \mathbf{C}(z)$ and $\boldsymbol{\Sigma}_\eta = \mathbf{C}_0\mathbf{C}_0'$,

$$\begin{aligned}\boldsymbol{\theta}(z)\boldsymbol{\Sigma}_\eta\boldsymbol{\theta}(z^{-1})' &= \mathbf{C}(z)\mathbf{C}(z^{-1})' \\ &= \mathbf{S}_w(z), \quad |z| = 1.\end{aligned}$$

Combining this equality with (A.6) gives

$$\boldsymbol{\theta}(z)\boldsymbol{\Sigma}_\eta\boldsymbol{\theta}(z^{-1})' - \boldsymbol{\phi}(z)\boldsymbol{\Sigma}_\epsilon\boldsymbol{\phi}(z^{-1})' = \boldsymbol{\Sigma}_v, \quad |z| = 1. \quad (\text{A.12})$$

Both sides of (A.12) are matrix Laurent polynomials. Since they agree at every point of the unit circle, they are identical as Laurent polynomial matrices; hence (A.12) holds for every $z \neq 0$.

Suppose that a square polynomial matrix $\mathbf{U}(z)$ is a common left divisor, so that

$$\boldsymbol{\phi}(z) = \mathbf{U}(z)\tilde{\boldsymbol{\phi}}(z), \quad \boldsymbol{\theta}(z) = \mathbf{U}(z)\tilde{\boldsymbol{\theta}}(z).$$

Because $\boldsymbol{\phi}(0) = \boldsymbol{\theta}(0) = \mathbf{I}_m$, $\mathbf{U}(0)$ is nonsingular. Substitution into (A.12) gives

$$\mathbf{U}(z)\mathbf{M}(z)\mathbf{U}(z^{-1})' = \boldsymbol{\Sigma}_v,$$

where

$$\mathbf{M}(z) := \tilde{\boldsymbol{\theta}}(z)\boldsymbol{\Sigma}_\eta\tilde{\boldsymbol{\theta}}(z^{-1})' - \tilde{\boldsymbol{\phi}}(z)\boldsymbol{\Sigma}_\epsilon\tilde{\boldsymbol{\phi}}(z^{-1})'$$

is a Laurent polynomial matrix. Since $\boldsymbol{\Sigma}_v$ is nonsingular, the identity above implies that $\mathbf{U}(z)$ is nonsingular for every $z \neq 0$: otherwise its left multiplication would make the left-hand side singular. Moreover, $\mathbf{U}(0)$ is nonsingular because $\boldsymbol{\phi}(0) = \mathbf{I}_m$. Hence the polynomial $\det[\mathbf{U}(z)]$ has no zero in $\mathbb{C}$. By the fundamental theorem of algebra it must therefore be a nonzero constant, so $\mathbf{U}(z)$ is unimodular. Thus $[\boldsymbol{\phi}(z), \boldsymbol{\theta}(z)]$ is left coprime, and cancellation of a nontrivial common left factor is impossible.

The outer spectral factor above selects the invertible solution when the algebraic covariance-matching equations admit more than one solution. $\square$

PROOF OF COROLLARY 3.2 From Proposition 3.1, the observed process $\{\mathbf{X}_t\}$ satisfies the following equation:

$$\mathbf{X}_t - \boldsymbol{\mu} = \sum_{j=1}^p \boldsymbol{\phi}_j(\mathbf{X}_{t-j} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t - \sum_{j=1}^p \boldsymbol{\theta}_j \boldsymbol{\eta}_{t-j} \quad (\text{A.13})$$

or

$$\mathbf{X}_t - \boldsymbol{\mu} = \sum_{j=1}^p \boldsymbol{\phi}_j(\mathbf{X}_{t-j} - \boldsymbol{\mu}) + \mathbf{v}_t + \boldsymbol{\epsilon}_t - \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\epsilon}_{t-j}. \quad (\text{A.14})$$

Multiply both sides of (A.14) by $(\mathbf{X}_{t-k} - \boldsymbol{\mu})'$ and taking expectation we get:

$$\begin{aligned}\boldsymbol{\Gamma}_k &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \mathbb{E}[\mathbf{v}_t(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] \\ &\quad + \mathbb{E}[\boldsymbol{\epsilon}_t(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] - \sum_{j=1}^p \boldsymbol{\phi}_j \mathbb{E}[\boldsymbol{\epsilon}_{t-j}(\mathbf{X}_{t-k} - \boldsymbol{\mu})']. \end{aligned} \quad (\text{A.15})$$

For $k = 0$, we have:

$$\begin{aligned}
\boldsymbol{\Gamma}_0 &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{-p} + \mathbb{E}[\mathbf{v}_t(\mathbf{X}_t - \boldsymbol{\mu})'] \\
&\quad + \mathbb{E}[\boldsymbol{\epsilon}_t(\mathbf{X}_t - \boldsymbol{\mu})'] - \sum_{j=1}^p \boldsymbol{\phi}_j \mathbb{E}[\boldsymbol{\epsilon}_{t-j}(\mathbf{X}_t - \boldsymbol{\mu})'] \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{-p} + \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_\epsilon.
\end{aligned} \tag{A.16}$$

The second equality follows because $\mathbb{E}[\mathbf{v}_t(\mathbf{X}_t - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}_v$, $\mathbb{E}[\boldsymbol{\epsilon}_t(\mathbf{X}_t - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}_\epsilon$, and $\mathbb{E}[\boldsymbol{\epsilon}_{t-j}(\mathbf{X}_t - \boldsymbol{\mu})'] = \mathbf{0}$ for $j \geq 1$.

For $1 \leq k \leq p$, we have:

$$\begin{aligned}
\boldsymbol{\Gamma}_k &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \mathbb{E}[\mathbf{v}_t(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] \\
&\quad + \mathbb{E}[\boldsymbol{\epsilon}_t(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] - \sum_{j=1}^p \boldsymbol{\phi}_j \mathbb{E}[\boldsymbol{\epsilon}_{t-j}(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} - \boldsymbol{\phi}_k \boldsymbol{\Sigma}_\epsilon.
\end{aligned} \tag{A.17}$$

For $k > p$, we have:

$$\begin{aligned}
\boldsymbol{\Gamma}_k &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + \mathbb{E}[\mathbf{v}_t(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] \\
&\quad + \mathbb{E}[\boldsymbol{\epsilon}_t(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] - \sum_{j=1}^p \boldsymbol{\phi}_j \mathbb{E}[\boldsymbol{\epsilon}_{t-j}(\mathbf{X}_{t-k} - \boldsymbol{\mu})'] \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p} + 0 + 0 - 0 \\
&= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{k-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{k-p}.
\end{aligned} \tag{A.18}$$

Combining (A.16)-(A.18), we get the autocovariance structure of the observed process stated in Corollary 3.2. $\square$

PROOF OF COROLLARY 3.3 The closed-form expression for $\boldsymbol{\phi}_{(p)}$ is based on the autocovariance structure of the process $\mathbf{X}_t$. This is the solution of p -system of equations from (3.2) with $k = p+1, \dots, 2p$. So

$$\begin{bmatrix} \boldsymbol{\Gamma}_{p+1} & \boldsymbol{\Gamma}_{p+2} & \cdots & \boldsymbol{\Gamma}_{2p} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_p & \boldsymbol{\Gamma}_{p+1} & \cdots & \boldsymbol{\Gamma}_{2p-1} \\ \boldsymbol{\Gamma}_{p-1} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_{2p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \end{bmatrix}. \tag{A.19}$$

$$\begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_{p+1} & \boldsymbol{\Gamma}_{p+2} & \cdots & \boldsymbol{\Gamma}_{2p} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_p & \boldsymbol{\Gamma}_{p+1} & \cdots & \boldsymbol{\Gamma}_{2p-1} \\ \boldsymbol{\Gamma}_{p-1} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_{2p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \end{bmatrix}^{-1}, \tag{A.20}$$

or

$$\boldsymbol{\phi}_{(p)} = \boldsymbol{\Gamma}_{(p+1)} \boldsymbol{\Gamma}_{[p]}^{-1}. \quad (\text{A.21})$$

where

$$\boldsymbol{\phi}_{(p)} := \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_p \end{bmatrix}, \quad \boldsymbol{\Gamma}_{(p+1)} := \begin{bmatrix} \boldsymbol{\Gamma}_{p+1} & \boldsymbol{\Gamma}_{p+2} & \cdots & \boldsymbol{\Gamma}_{2p} \end{bmatrix}, \quad (\text{A.22})$$

$$\boldsymbol{\Gamma}_{[p]} := \begin{bmatrix} \boldsymbol{\Gamma}_p & \boldsymbol{\Gamma}_{p+1} & \cdots & \boldsymbol{\Gamma}_{2p-1} \\ \boldsymbol{\Gamma}_{p-1} & \boldsymbol{\Gamma}_p & \cdots & \boldsymbol{\Gamma}_{2p-2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 & \cdots & \boldsymbol{\Gamma}_p \end{bmatrix}. \quad (\text{A.23})$$

More generally, taking (3.2) at $k = p + j, \dots, 2p + j - 1$, for any $j \geq 1$, and stacking the resulting equations gives

$$\boldsymbol{\Gamma}_{(p+j)} = \boldsymbol{\phi}_{(p)} \boldsymbol{\Gamma}_{[p+j-1]}. \quad (\text{A.24})$$

Hence,

$$\boldsymbol{\phi}_{(p)} = \boldsymbol{\Gamma}_{(p+j)} \boldsymbol{\Gamma}_{[p+j-1]}^{-1}, \quad j \geq 1. \quad (\text{A.25})$$

where

$$\boldsymbol{\Gamma}_{(p+j)} = \begin{bmatrix} \boldsymbol{\Gamma}_{p+j} & \boldsymbol{\Gamma}_{p+j+1} & \cdots & \boldsymbol{\Gamma}_{2p+j-1} \end{bmatrix}, \quad \boldsymbol{\Gamma}_{[p+j-1]} = \begin{bmatrix} \boldsymbol{\Gamma}_{p+j-1} & \boldsymbol{\Gamma}_{p+j} & \cdots & \boldsymbol{\Gamma}_{2p+j-2} \\ \boldsymbol{\Gamma}_{p+j-2} & \boldsymbol{\Gamma}_{p+j-1} & \cdots & \boldsymbol{\Gamma}_{2p+j-3} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}_j & \boldsymbol{\Gamma}_{j+1} & \cdots & \boldsymbol{\Gamma}_{p+j-1} \end{bmatrix}. \quad (\text{A.26})$$

By covariance stationarity, $\boldsymbol{\Gamma}_{-j} = \boldsymbol{\Gamma}'_j$ for $j \geq 1$, and using (3.2) with $k = 1$, we obtain:

$$\begin{aligned} \boldsymbol{\Gamma}_1 &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_2 \boldsymbol{\Gamma}_{-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{1-p} - \boldsymbol{\phi}_1 \boldsymbol{\Sigma}_\epsilon \\ &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_2 \boldsymbol{\Gamma}'_1 + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}'_{p-1} - \boldsymbol{\phi}_1 \boldsymbol{\Sigma}_\epsilon. \end{aligned} \quad (\text{A.27})$$

hence

$$\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1. \quad (\text{A.28})$$

On substituting the above expression for $\boldsymbol{\Sigma}_\epsilon$ into (3.2) with $k = 0$, we have:

$$\begin{aligned} \boldsymbol{\Gamma}_0 &= \boldsymbol{\phi}_1 \boldsymbol{\Gamma}_{-1} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Gamma}_{-p} + \boldsymbol{\Sigma}_\nu + \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1 \\ &= \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\Gamma}_{-j} + \boldsymbol{\Sigma}_\nu + \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1 \\ &= \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\Gamma}'_j + \boldsymbol{\Sigma}_\nu + \boldsymbol{\Gamma}_0 + \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1. \end{aligned} \quad (\text{A.29})$$

hence

$$\boldsymbol{\Sigma}_\nu = \boldsymbol{\phi}_1^{-1} \boldsymbol{\Gamma}_1 - \boldsymbol{\phi}_1^{-1} \sum_{j=1}^{p-1} \boldsymbol{\phi}_{j+1} \boldsymbol{\Gamma}'_j - \sum_{j=1}^p \boldsymbol{\phi}_j \boldsymbol{\Gamma}'_j. \quad (\text{A.30})$$

Since $\mu = \mathbb{E}(\mathbf{X}_t)$, the proof is completed. $\square$

PROOF OF LEMMA 4.1 Since, under Assumptions 2.1-2.2, the sequence $\{\mathbf{X}_t\}$ is strictly stationary, ergodic and has finite moments of all orders, the Ergodic Theorem [see Theorem 13.12 and Corollary 13.14 in Davidson (1994)] implies that $\frac{1}{T} \sum_{t=1}^T \mathbf{X}_t \xrightarrow{\text{a.s.}} \mathbb{E}(\mathbf{X}_t) = \mu$. Hence, $\hat{\mu} \rightarrow \mu$ almost surely.

Consider now the autocovariance estimators. By definition, $\hat{\Gamma}_k = \frac{1}{T-k} \sum_{t=1}^{T-k} (\mathbf{X}_{t+k} - \hat{\mu})(\mathbf{X}_t - \hat{\mu})'$. Since $\hat{\mu} \rightarrow \mu$ a.s., we can write $\hat{\Gamma}_k - \Gamma_k = \frac{1}{T-k} \sum_{t=1}^{T-k} [(\mathbf{X}_{t+k} - \mu)(\mathbf{X}_t - \mu)' - \Gamma_k] + o(1)$, where $o(1) \xrightarrow{\text{a.s.}} 0$ accounts for replacing $\hat{\mu}$ by μ .

Consider the sequence $\{(\mathbf{X}_{t+k} - \mu, \mathbf{X}_t - \mu)\}_t$. This is a stationary ergodic sequence. By the Ergodic Theorem again, $\frac{1}{T-k} \sum_{t=1}^{T-k} (\mathbf{X}_{t+k} - \mu)(\mathbf{X}_t - \mu)' \xrightarrow{\text{a.s.}} \mathbb{E}[(\mathbf{X}_{t+k} - \mu)(\mathbf{X}_t - \mu)'] = \Gamma_k$. Thus, $\hat{\Gamma}_k \xrightarrow{\text{a.s.}} \Gamma_k$.

Since this holds for each fixed k , it holds simultaneously for $k = 0, 1, \dots, m$. Hence, $\hat{\Lambda}(m) \xrightarrow{\text{a.s.}} \Lambda(m)$. $\square$

PROOF OF LEMMA 4.2 Since $\{\mathbf{X}_t\}$ is strong mixing with geometric decay and has finite fourth moments, standard mixing-based central limit theorems apply (see Brockwell and Davis (1996), Chapter 11 and Davidson (1994), Chapter 14). These results state that if Z_t is a stationary, strong mixing sequence with finite second or fourth moments and mixing coefficients decaying at a geometric rate, then:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T (Z_t - \mathbb{E}[Z_t]) \xrightarrow{d} \mathcal{N}[0, \Sigma_Z] \quad (\text{A.31})$$

for some covariance matrix Σ_Z . We apply this result to each type of empirical moment.

For $\hat{\mu} - \mu$, set $Z_t = \mathbf{X}_t$ and note that $\sum_{t=1}^T (\mathbf{X}_t - \mu) / \sqrt{T}$ converges in distribution to a normal vector. Similarly, each element of $\hat{\Gamma}_k - \Gamma_k$ can be expressed as an average of products $(\mathbf{X}_{t+k} - \mu)(\mathbf{X}_t - \mu)'$ minus their expectation. Since these are also functions of a stationary, strong mixing sequence with finite fourth moments, a joint CLT holds for all autocovariances up to lag m .

By stacking $(\hat{\mu} - \mu, \{\text{vec}(\hat{\Gamma}_k - \Gamma_k)\}_{k=0}^m)$ into one vector, the Cramér-Wold device ensures a joint multivariate normal limit, hence

$$\sqrt{T}[\hat{\Lambda}(m) - \Lambda(m)] \xrightarrow{d} \mathcal{N}[0, \Sigma(\Lambda, m)] \quad (\text{A.32})$$

for some positive semidefinite $\Sigma(\Lambda, m)$. Estimation of $\Sigma(\Lambda, m)$ can be done using HAC estimators which account for dependence. $\square$

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.

From 4.1, $\hat{\Lambda}(2p) \xrightarrow{\text{a.s.}} \Lambda(2p)$, ensuring consistency. From 4.2, we have:

$$\sqrt{T}[\hat{\Lambda}(2p) - \Lambda(2p)] \xrightarrow{d} \mathcal{N}[0, \Sigma(\Lambda, 2p)]. \quad (\text{A.33})$$

Using Assumptions 4.1-4.3, g is continuously differentiable at $\Lambda(2p)$ and $Dg[\Lambda(2p)]$ has full row rank.

Applying the delta method [see [Serfling \(1980\)](#), Chapter 3], we obtain:

$$\sqrt{T}\{g[\hat{\Lambda}(2p)] - g[\Lambda(2p)]\} \xrightarrow{d} \mathcal{N}[0, (Dg[\Lambda(2p)])\Sigma(\Lambda, 2p)(Dg[\Lambda(2p)])'], \quad (\text{A.34})$$

which means that

$$\sqrt{T}(\hat{\Theta} - \Theta) \xrightarrow{d} \mathcal{N}[0, \Omega] \quad (\text{A.35})$$

with $\Omega := (Dg[\Lambda(2p)])\Sigma(\Lambda, 2p)(Dg[\Lambda(2p)])'$. This completes the proof. $\square$

PROOF OF THEOREM 4.4 By extending the results of Lemmas 4.1 and 4.2 to $\hat{\Lambda}(2p+J-1)$, which now includes $[\hat{\mathbf{I}}_0, \hat{\mathbf{I}}_1, \dots, \hat{\mathbf{I}}_{2p+J-1}]$, we have:

$$\hat{\Lambda}(2p+J-1) \xrightarrow{\text{a.s.}} \Lambda(2p+J-1), \quad (\text{A.36})$$

$$\sqrt{T}[\hat{\Lambda}(2p+J-1) - \Lambda(2p+J-1)] \xrightarrow{d} \mathcal{N}[0, \Sigma(\Lambda, 2p+J-1)], \quad (\text{A.37})$$

for some positive definite $\Sigma(\Lambda, 2p+J-1)$.

The winsorized estimator or MVR estimator is obtained from these extended moments via a differentiable mapping g_J :

$$\tilde{\Theta}_J = g_J[\hat{\Lambda}(2p+J-1)]. \quad (\text{A.38})$$

Since g_J is continuously differentiable at $\Lambda(2p+J-1)$ and its Jacobian $Dg_J[\Lambda(2p+J-1)]$ has full row rank, we apply the delta method again:

$$\sqrt{T}(\tilde{\Theta}_J - \Theta) = \sqrt{T}\{g_J[\hat{\Lambda}(2p+J-1)] - g_J[\Lambda(2p+J-1)]\} \xrightarrow{d} \mathcal{N}[0, \Omega_J], \quad (\text{A.39})$$

with $\Omega_J = Dg_J[\Lambda(2p+J-1)]\Sigma(\Lambda, 2p+J-1)Dg_J[\Lambda(2p+J-1)]'$. This establishes consistency and asymptotic normality of the winsorized VARMA-based estimators. $\square$

PROOF OF THEOREM 4.5

Step 1: Baseline ($J = 1$) setup. Consider first the simple VARMA-based estimator which uses exactly one set of moment conditions derived from $\hat{\mathbf{I}}_{(p+1)}$. As shown in 4.3, under the given assumptions:

$$\sqrt{T}(\hat{\Theta}_1 - \Theta) \xrightarrow{d} \mathcal{N}(0, \Omega_1), \quad (\text{A.40})$$

where $\hat{\Theta}_1$ denotes the estimator obtained from these single-lag moment conditions, and Ω_1 is the corresponding asymptotic covariance matrix. Here, Θ is the parameter vector $(\mu', \text{vec}(\Phi_{(p)})', \text{vec}(\Sigma_v)', \text{vec}(\Sigma_e'))'$, as previously defined.

Step 2: Additional moment conditions ($J > 1$). Now, consider the winsorized VARMA-based estimator which uses $J > 1$ sets of moment conditions. In addition to the basic condition from $\hat{\mathbf{I}}_{(p+1)}$, we incorporate further conditions based on $\hat{\mathbf{I}}_{(p+2)}, \hat{\mathbf{I}}_{(p+3)}, \dots, \hat{\mathbf{I}}_{(p+J)}$. Stacking all these conditions into an extended vector $\hat{\Lambda}(p+J)$, we now have more equations than the dimension of Θ , thus an overidentified system. By 4.4, under the same assumptions and the delta method applied to the extended set of moments, the

winsorized estimator $\hat{\Theta}_J$ which incorporates J moment conditions satisfies:

$$\sqrt{T}(\hat{\Theta}_J - \Theta) \xrightarrow{d} \mathcal{N}(0, \Omega_J), \quad (\text{A.41})$$

for some positive definite Ω_J which depends on the extended set of moment conditions.

Step 3: Non-redundancy of additional moment conditions. Non-redundancy means that the additional $J - 1$ moment conditions provide genuinely new identifying information about Θ which is not contained within the original single condition ($J = 1$). Formally, if the original set of moments can be represented by a vector function $g_1(\Lambda(2p))$, and the extended set by $g_J(\Lambda(2p + J - 1))$, then non-redundancy implies that the Jacobian $Dg_J(\Lambda(2p + J - 1))$ has higher rank (or provides additional independent equations) compared to $Dg_1(\Lambda(2p))$. Thus, the parameter vector Θ is more tightly constrained with $J > 1$ conditions than with $J = 1$.

If the additional conditions were redundant, they could be expressed as linear combinations of the original condition. In this case, no improvement in variance could be achieved. Non-redundancy rules this out, ensuring that each new moment condition adds at least one independent piece of information which restricts Θ more effectively.

Step 4: Efficiency gains, the Loewner order and limiting case of $J = 1$. The analogy with GMM estimation is instructive here. In the GMM framework (see Hansen (1982)), adding additional valid and independent moment conditions never increases the asymptotic variance of the estimator; rather, it can stay the same (if conditions are redundant) or decrease (if they are genuinely new).

Since the winsorized estimator for $J > 1$ essentially solves a larger set of moment equations, its asymptotic variance matrix Ω_J is less than or equal to Ω_1 in the Loewner order, i.e., $\Omega_J \preceq \Omega_1$. This ordering means that $\Omega_J - \Omega_1$ is positive semi-definite, reflecting no increase in variance. In typical scenarios, the additional conditions are genuinely informative, making the reduction strict, $\Omega_J < \Omega_1$. Hence, we get a strictly smaller asymptotic variance, which translates into asymptotic efficiency gains.

When $J = 1$, we revert back to the simple VARMA-based estimator with a single moment condition. In this case, Ω_J reduces to Ω_1 , confirming that adding no extra conditions beyond the original one provides no change in variance.

We have shown that under the given assumptions, since the $J - 1$ additional moment conditions are non-redundant, then the asymptotic variance matrix corresponding to the winsorized estimator with J conditions, Ω_J , is smaller than or equal to Ω_1 in the Loewner order. Typically, it is strictly smaller, indicating genuine asymptotic efficiency gains over the simple VARMA-based estimator. This completes the proof. $\square$

B Monte Carlo tests with nuisance parameters

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 which is 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.

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 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 $S_1(\theta), \dots, S_N(\theta)$ 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 $S(N) = (S_1, \dots, S_N)$ under H_0 , independently of S_0 , i.e., $S_0, S_1, \dots, S_N$ be exchangeable.
3. From the simulated samples compute the MC p -value $\hat{p}_N[S] \equiv p_N[S_0; S(N)]$, where

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

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 $S_0, S_1, \dots, S_N$ are all distinct [an event with probability one when the vector $(S_0, S_1, \dots, S_N)'$ 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 $S_0, S_1, \dots, S_N$.

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, $P[\hat{p}_N[S] \leq \alpha] = \alpha$; 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_θ , 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

$$H_0 : \bar{\theta} \in \Omega_0, \quad (\text{B.1})$$

where $\Omega_0 \subset \Omega$, we calculate the relevant test statistic S_0 based on data.

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

$$p_N[x; S(N, \theta)] \equiv \frac{NG_N[x; S(N, \theta)] + 1}{N + 1}. \quad (\text{B.2})$$

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

$$\sup_{\theta \in \Omega_0} \hat{p}_N[S|\theta] \leq \alpha. \quad (\text{B.3})$$

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

$$P[\sup_{\theta \in \Omega_0} \{\hat{p}_N[S|\theta]\} \leq \alpha] \leq \alpha. \quad (\text{B.4})$$

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 (B.3), the test is called a *maximized Monte Carlo* (MMC) test. MMC tests provide valid inference under general regularity conditions such as almost-identified 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 et al. \(1994\)](#)].

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

$$\lim_{T \rightarrow \infty} P[\bar{\theta} \in C_T] = 1 \text{ under } H_0. \quad (\text{B.5})$$

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

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

$$p_{TN}[x; S_T(N, \theta)] \equiv \frac{NG_{TN}[x; S_T(N, \theta)] + 1}{N + 1}. \quad (\text{B.6})$$

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

$$\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha. \quad (\text{B.7})$$

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

$$\lim_{T \rightarrow \infty} P[\sup\{\hat{p}_{TN}[S_T|\theta] : \theta \in C_T\} \leq \alpha] \leq \alpha, \quad (\text{B.8})$$

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

$$C_T = \{\theta \in : \|\hat{\theta}_T - \theta\| < d\} \quad (\text{B.9})$$

with d a fixed positive constant independent of T , satisfies (B.5). 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 which 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 (B.7) is reduced to a single point estimate $\hat{\theta}_T$, i.e. $C_T = \{\hat{\theta}_T\}$, we get a LMC test

$$\hat{p}_{TN}[S_T|\hat{\theta}_T] \leq \alpha, \quad (\text{B.10})$$

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 (B.5) 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 .

For additional discussion on constructing the *implicit standard error* (ISE) of θ from simulation-based confidence intervals, see Appendix C.

C Implicit standard error

In this section, we propose the *implicit standard error* (ISE) of θ which 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 $\mathbb{Y}(N) = (\mathbb{Y}_1, \dots, \mathbb{Y}_N)$ under H_0 , independently of $\mathbb{Y}_0$, i.e., $\mathbb{Y}_0, \mathbb{Y}_1, \dots, \mathbb{Y}_N$ be exchangeable.
3. Calculate $\hat{\theta}(N) = \hat{\theta}_1, \dots, \hat{\theta}_N$ from $\mathbb{Y}(N) = (\mathbb{Y}_1, \dots, \mathbb{Y}_N)$.
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 $\hat{\theta}(0; N) = \hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_N$. Furthermore, if $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_N$ are ordered from smallest to largest such that $\hat{\theta}_1^*, \dots, \hat{\theta}_{N+1}^*$, and $\alpha_L(N+1)$ and $\alpha_H(N+1)$ are integers, then the confidence interval is

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

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

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

6. Finally, a conservative ISE is

$$ISE = \min\{ISE_L, ISE_H\}. \quad (\text{C.3})$$

D MCMC algorithm for MSV models

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 $\mathbf{y}_t$ denote the latent multivariate volatility time-series and $\theta = (\mu, \boldsymbol{\phi}_1, \boldsymbol{\Sigma}_v)$ 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:

$$\log(y_{i,t}^2) = h_{i,t} + \log(u_{i,t}^2). \quad (\text{D.1})$$

Note that (D.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 $\log(u_{i,t}^2) | (d_{i,t} = k) \sim \mathcal{N}(m_k, v_k^2)$. The indicator variable is also estimated in the MCMC sampler. Given the mixture indicator $\mathbf{d}_t$ and the vector parameter θ , the latent volatility series $\mathbf{h}_t$ can be sampled using a forward-filtering and backward-sampling (FFBS) procedure [Carter and Kohn (1994), Frühwirth-Schnatter (1994)]. The mixture indicator $\mathbf{d}_t$ can be sampled from a multinomial distribution

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

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 which 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 which have low persistence.

E 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.
- **Ticker symbols for 100 assets portfolio:** AAPL, ABT, ACN, ADBE, ADI, ADP, ADSK, AMAT, AMD, AMGN, AMT, AMZN, APD, ATVI, AXP, BA, BAC, BDX, BLK, BMY, C, CAT, CCI, CI, CL, CMCSA, CME, COP, COST, CRM, CSCO, CSX, CVS, CVX, D, DE, DHR, DIS, DUK, ECL, EL, EQIX, FDX, FIS, FISV, GE, GILD, GPN, GS, HD, HON, IBM, ICE, ILMN, INTC, INTU, ISRG, ITW, JNJ, JPM, KO, LLY, LOW, LRCX, MA, MCD, MDT, MMC, MMM, MO, MRK, MSFT, MU, NFLX, NKE, NSC, NVDA, ORCL, PEP, PFE, PG, PNC, QCOM, SBUX, SCHW, SHW, SO, SYK, T, TGT, TJX, TMO, UNH, UNP, UPS, USB, VZ, WFC, WMT, XOM.

F Forecasting with MSV(p) models

This section develops s -step-ahead forecasts for the multivariate stochastic volatility model of order p (MSV(p)) introduced in 2.1. Recall that the mean-corrected log-squared transformation yields the measurement equation

$$\mathbf{X}_t = \log \circ (\mathbf{y}_t^2) + (1.2704)\iota_m = \mathbf{h}_t + \epsilon_t, \quad (\text{E1})$$

where ι_m is an $m \times 1$ vector of ones, $\mathbb{E}(\epsilon_t) = \mathbf{0}$ and $\text{Cov}(\epsilon_t) = \Sigma_\epsilon$ is a full $m \times m$ covariance matrix [see (2.7)]. The latent log-volatility follows the VAR(p) dynamics

$$\mathbf{h}_t = \mu + \sum_{j=1}^p \phi_j (\mathbf{h}_{t-j} - \mu) + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}_m(\mathbf{0}, \Sigma_v). \quad (\text{E2})$$

Given $\hat{\mathbf{h}}_{t+s|t}$, the conditional covariance forecast of returns $\mathbf{y}_{t+s}$ is obtained through the MSV decomposition

$$\Sigma_{t+s|t} := \text{Var}(\mathbf{y}_{t+s} | \mathcal{F}_t) = \mathbf{V}_{t+s|t}^{1/2} \Sigma_u \mathbf{V}_{t+s|t}^{1/2}, \quad \mathbf{V}_{t+s|t}^{1/2} := \text{diag}[\exp(\hat{\mathbf{h}}_{t+s|t}/2)] \quad (\text{E3})$$

where $\Sigma_u = (\rho_{ij})$ is the (instantaneous) correlation matrix of return shocks in 2.1.

Forecasting requires filtering $\mathbf{h}_t$ from the non-Gaussian measurement equation in (E1). Three approaches are considered: (i) *Kalman filtering* under the Gaussian approximation $\epsilon_t \approx \mathcal{N}_m(\mathbf{0}, \Sigma_\epsilon)$; (ii) a *bootstrap particle filter* (BPF); and (iii) an *auxiliary particle filter* (APF). In the high-dimensional designs emphasized in this paper, Kalman filtering serves as the default workhorse because it is numerically stable and generally more computationally tractable than particle methods, while remaining coherent with the first- and second-moment characterization of Σ_ϵ implied by (2.7). Particle methods allow one to move beyond the Gaussian approximation for the measurement disturbance. However, with full Σ_ϵ , the exact mul-

tivariate measurement density induced by $\log \circ (u_t^2)$ is not used directly; instead, the particle implementations rely on a *second-moment-compatible* pseudo-likelihood obtained by whitening the residuals and applying a product approximation based on standardized $\log(\chi_1^2)$ marginals, in the spirit of simulation-based likelihood methods for non-Gaussian state-space models; see [Shephard and Pitt \(1997\)](#). When particle filters are employed, the APF is generally preferred to the BPF because its look-ahead weighting and resampling step is designed to mitigate weight degeneracy relative to the BPF; see [Pitt and Shephard \(1999\)](#).

F.1 Forecasting via Kalman filtering

Define the demeaned variables $\tilde{X}_t := X_t - \mu$ and stack the last p lags of the centered log-volatility into the pm -dimensional companion state

$$\tilde{h}_t := \begin{pmatrix} h_t - \mu \\ h_{t-1} - \mu \\ \vdots \\ h_{t-p+1} - \mu \end{pmatrix} \in \mathbb{R}^{pm}. \quad (\text{F.4})$$

The MSV(p) model admits the linear Gaussian state-space representation

$$\begin{aligned} \tilde{h}_{t+1} &= \mathcal{A} \tilde{h}_t + \tilde{v}_{t+1}, \\ \tilde{X}_t &= \mathcal{C} \tilde{h}_t + \epsilon_t, \end{aligned} \quad (\text{F.5})$$

where

$$\mathcal{C} := (\mathbf{I}_m \ \mathbf{0}_{m \times m(p-1)}), \quad (\text{F.6})$$

and the companion-form transition matrix is defined by

$$\mathcal{A} := \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_m & \mathbf{0} \end{pmatrix}. \quad (\text{F.7})$$

Moreover, define

$$\tilde{v}_{t+1} := \begin{pmatrix} v_{t+1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \quad \mathcal{B} := \text{Cov}(\tilde{v}_{t+1}) = \text{diag}(\Sigma_v, \mathbf{0}, \dots, \mathbf{0}). \quad (\text{F.8})$$

For Kalman filtering we employ the Gaussian approximation $\epsilon_t \approx \mathcal{N}_m(\mathbf{0}, \mathcal{D})$ where $\mathcal{D} := \Sigma_\epsilon$.

Let $\hat{\tilde{h}}_{t|t}$ and $\mathbf{P}_{t|t}$ denote the filtered mean and MSE matrix of $\tilde{h}_t$ given $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$. Kalman filter recursions are as follows:

- Initialization:

$$\hat{\tilde{h}}_{1|0} = \mathbf{0}_{(pm \times 1)}, \quad \mathbf{P}_{1|0} = \text{diag}(\Sigma_v, \dots, \Sigma_v)_{(pm \times pm)}. \quad (\text{F.9})$$

- Time update (prediction to t):

$$\hat{\mathbf{h}}_{t|t-1} = \mathcal{A}\hat{\mathbf{h}}_{t-1|t-1}, \quad \mathbf{P}_{t|t-1} = \mathcal{A}\mathbf{P}_{t-1|t-1}\mathcal{A}' + \mathcal{B}. \quad (\text{E.10})$$

- Measurement update (at time t): Define

$$\mathbf{v}_t := \tilde{\mathbf{X}}_t - \mathcal{C}\hat{\mathbf{h}}_{t|t-1}, \quad \mathbf{S}_t := \mathcal{C}\mathbf{P}_{t|t-1}\mathcal{C}' + \mathcal{D}, \quad \mathbf{K}_t := \mathbf{P}_{t|t-1}\mathcal{C}'\mathbf{S}_t^{-1}. \quad (\text{E.11})$$

Then

$$\begin{aligned} \hat{\mathbf{h}}_{t|t} &= \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t\mathbf{v}_t, \\ \mathbf{P}_{t|t} &= (\mathbf{I}_{pm} - \mathbf{K}_t\mathcal{C})\mathbf{P}_{t|t-1}(\mathbf{I}_{pm} - \mathbf{K}_t\mathcal{C})' + \mathbf{K}_t\mathcal{D}\mathbf{K}_t'. \end{aligned} \quad (\text{E.12})$$

- s -step-ahead forecast ($s \geq 1$):

$$\hat{\mathbf{h}}_{t+s|t} = \mathcal{A}^s\hat{\mathbf{h}}_{t|t}, \quad \mathbf{P}_{t+s|t} = \mathcal{A}^s\mathbf{P}_{t|t}(\mathcal{A}')^s + \sum_{k=0}^{s-1} \mathcal{A}^k\mathcal{B}(\mathcal{A}')^k. \quad (\text{E.13})$$

Covariance forecasting is based on the *latent* log-volatility forecast (not on $\hat{\mathbf{X}}_{t+s|t}$), since $\mathbf{X}_t = \mathbf{h}_t + \epsilon_t$. Hence

$$\hat{\mathbf{h}}_{t+s|t} = \mu + \mathcal{C}\hat{\mathbf{h}}_{t+s|t} \quad (\text{E.14})$$

and the plug-in covariance forecast is

$$\hat{\mathbf{V}}_{t+s|t}^{1/2} = \text{diag} \left[\exp(\hat{\mathbf{h}}_{t+s|t}/2) \right], \quad \hat{\Sigma}_{t+s|t} = \hat{\mathbf{V}}_{t+s|t}^{1/2} \Sigma_u \hat{\mathbf{V}}_{t+s|t}^{1/2}. \quad (\text{E.15})$$

F.2 Forecasting via bootstrap particle filtering

We retain the Gaussian transition in (E.5), but treat ϵ_t in (E.1) as non-Gaussian. With full Σ_ϵ , the exact joint density of ϵ_t is typically intractable in high dimensions; we therefore employ a second-moment compatible pseudo-likelihood based on whitening and marginal standardization [Shephard and Pitt (1997)].

Let $\Sigma_\epsilon = \mathbf{L}_\epsilon \mathbf{L}_\epsilon'$ and define the residual and its whitened version

$$\mathbf{e}_t(\tilde{\mathbf{h}}_t) := \tilde{\mathbf{X}}_t - \mathcal{C}\tilde{\mathbf{h}}_t, \quad \mathbf{e}_{w,t}(\tilde{\mathbf{h}}_t) := \mathbf{L}_\epsilon^{-1} \mathbf{e}_t(\tilde{\mathbf{h}}_t). \quad (\text{E.16})$$

Let $\epsilon := \log(\chi_1^2) - \mathbb{E}\{\log(\chi_1^2)\}$ so that $\mathbb{E}(\epsilon) = 0$ and $\text{Var}(\epsilon) = \pi^2/2$. Define $s_\chi := \sqrt{\pi^2/2} = \pi/\sqrt{2}$ and the standardized scalar $Z := \frac{\epsilon}{s_\chi}$, $\mathbb{E}(Z) = 0$, $\text{Var}(Z) = 1$. Denote by f_Z the density of Z . Up to normalizing constants (irrelevant for particle weights), one may write $\log f_Z(z) \propto \frac{1}{2}(s_\chi z - 1.2704) - \frac{1}{2} \exp(s_\chi z - 1.2704)$. We then define the pseudo-likelihood

$$\tilde{f}_\epsilon(\mathbf{e}_t) := \prod_{j=1}^m f_Z((\mathbf{e}_{w,t})_j), \quad (\text{E.17})$$

which is compatible with $\text{Cov}(\mathbf{e}_{w,t}) = \mathbf{I}_m$. The BPF algorithm is given below:

- Initialization ($t = 0$): Let N be the number of particles, and write $\{\tilde{\mathbf{h}}_{t|t}^{(i)}, w_t^{(i)}\}_{i=1}^N$. Draw $\tilde{\mathbf{h}}_{0|0}^{(i)} \sim \mathcal{N}_{pm}(\mathbf{0}, \mathbf{P}_{1|0})$ with $\mathbf{P}_{1|0}$ in (E.9), and set $w_0^{(i)} = 1/N$.

- For $t = 1, \dots, T$ repeat:

- *Propagation (proposal = transition).*

$$\tilde{\mathbf{h}}_{t|t-1}^{(i)} = \mathcal{A}\tilde{\mathbf{h}}_{t-1|t-1}^{(i)} + \tilde{\mathbf{v}}_t^{(i)}, \quad \tilde{\mathbf{v}}_t^{(i)} \sim \mathcal{N}_{pm}(\mathbf{0}, \mathcal{B}). \quad (\text{E.18})$$

- *Weighting (pseudo-likelihood).* Compute $\mathbf{e}_t^{(i)} = \tilde{\mathbf{X}}_t - \mathcal{C}\tilde{\mathbf{h}}_{t|t-1}^{(i)}$ and $\mathbf{e}_{w,t}^{(i)} = \mathbf{L}_\epsilon^{-1} \mathbf{e}_t^{(i)}$, and set

$$\tilde{w}_t^{(i)} = w_{t-1}^{(i)} \times \tilde{f}_\epsilon(\mathbf{e}_t^{(i)}) = w_{t-1}^{(i)} \times \prod_{j=1}^m f_Z((\mathbf{e}_{w,t}^{(i)})_j). \quad (\text{E.19})$$

- *Normalize.*

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{\ell=1}^N \tilde{w}_t^{(\ell)}}. \quad (\text{E.20})$$

- *Resampling (optional).* Compute $\text{ESS} = 1 / \sum_{i=1}^N (w_t^{(i)})^2$; if $\text{ESS} < N/2$, resample $\{\tilde{\mathbf{h}}_{t|t-1}^{(i)}\}$ with probabilities $\{w_t^{(i)}\}$ and reset $w_t^{(i)} = 1/N$.

- *Assignment.* Set $\tilde{\mathbf{h}}_{t|t}^{(i)} = \tilde{\mathbf{h}}_{t|t-1}^{(i)}$.

For $s \geq 1$, generate forward trajectories from the final weighted particle system $\{(\tilde{\mathbf{h}}_{T|T}^{(i)}, w_T^{(i)})\}$:

$$\tilde{\mathbf{h}}_{T+k|T}^{(i)} = \mathcal{A}\tilde{\mathbf{h}}_{T+k-1|T}^{(i)} + \tilde{\mathbf{v}}_{T+k}^{(i)}, \quad \tilde{\mathbf{v}}_{T+k}^{(i)} \sim \mathcal{N}_{pm}(\mathbf{0}, \mathcal{B}), \quad k = 1, \dots, s. \quad (\text{E.21})$$

A natural log-volatility forecast is the weighted particle mean

$$\hat{\mathbf{h}}_{T+s|T} = \mu + \mathcal{C} \left(\sum_{i=1}^N w_T^{(i)} \tilde{\mathbf{h}}_{T+s|T}^{(i)} \right). \quad (\text{E.22})$$

For covariance forecasting, averaging *after* exponentiation accounts for nonlinearities:

$$\hat{\Sigma}_{T+s|T} = \sum_{i=1}^N w_T^{(i)} \mathbf{V}_{T+s|T}^{(i)1/2} \Sigma_u \mathbf{V}_{T+s|T}^{(i)1/2}, \quad \mathbf{V}_{T+s|T}^{(i)1/2} := \text{diag} \left[\exp \left((\mu + \mathcal{C}\tilde{\mathbf{h}}_{T+s|T}^{(i)})/2 \right) \right]. \quad (\text{E.23})$$

Equivalently, letting $\mathbf{d}^{(i)} := \exp \left((\mu + \mathcal{C}\tilde{\mathbf{h}}_{T+s|T}^{(i)})/2 \right) \in \mathbb{R}^m$,

$$\hat{\Sigma}_{T+s|T} = \Sigma_u \circ \left(\sum_{i=1}^N w_T^{(i)} \mathbf{d}^{(i)} \mathbf{d}^{(i)'} \right). \quad (\text{E.24})$$

If a final resampling step is imposed at T (so $w_T^{(i)} = 1/N$), the above expressions reduce to simple averages.

F.3 Forecasting via auxiliary particle filtering

The APF uses the same state-space structure (E.5) and the same pseudo-likelihood (E.17), but introduces a look-ahead resampling step based on predictive fit, which reduces weight degeneracy relative to the BPF [Pitt and Shephard (1999)]. The use of pseudo-likelihood weighting under non-Gaussian measurement errors follows the broader simulation-based likelihood perspective in Shephard and Pitt (1997). The APF algorithm is as follows:

- Look-ahead (first-stage) weights: Maintain particles $\{(\tilde{\mathbf{h}}_{t-1|t-1}^{(j)}, w_{t-1}^{(j)})\}_{j=1}^N$. For each j , compute the transition mean

$$\tilde{\mathbf{h}}_{t|t-1}^{\star(j)} = \mathcal{A}\tilde{\mathbf{h}}_{t-1|t-1}^{(j)}, \quad \mathbf{e}_{w,t}^{\star(j)} = \mathbf{L}_\epsilon^{-1}(\tilde{\mathbf{X}}_t - \mathcal{C}\tilde{\mathbf{h}}_{t|t-1}^{\star(j)}), \quad (\text{E.25})$$

and define

$$g_t^{(j)} := \prod_{r=1}^m f_Z((e_{w,t}^{\star(j)})_r), \quad \alpha_t^{(j)} := \frac{w_{t-1}^{(j)} g_t^{(j)}}{\sum_{\ell=1}^N w_{t-1}^{(\ell)} g_t^{(\ell)}}. \quad (\text{E.26})$$

- Ancestor sampling: Draw $a_i \in \{1, \dots, N\}$ independently with $\Pr(a_i = j) = \alpha_t^{(j)}$.
- Propagation: For each i , simulate

$$\tilde{\mathbf{h}}_{t|t-1}^{(i)} = \mathcal{A}\tilde{\mathbf{h}}_{t-1|t-1}^{(a_i)} + \tilde{\mathbf{v}}_t^{(i)}, \quad \tilde{\mathbf{v}}_t^{(i)} \sim \mathcal{N}_{pm}(\mathbf{0}, \mathcal{B}). \quad (\text{E.27})$$

- Second-stage weights: Compute

$$\mathbf{e}_{w,t}^{(i)} = \mathbf{L}_\epsilon^{-1}(\tilde{\mathbf{X}}_t - \mathcal{C}\tilde{\mathbf{h}}_{t|t-1}^{(i)}) \quad (\text{E.28})$$

and set

$$\tilde{w}_t^{(i)} = \frac{\prod_{r=1}^m f_Z((e_{w,t}^{(i)})_r)}{g_t^{(a_i)}}, \quad w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{\ell=1}^N \tilde{w}_t^{(\ell)}}. \quad (\text{E.29})$$

- Resampling and assignment: If $\text{ESS} < N/2$, resample and reset weights; then set

$$\tilde{\mathbf{h}}_{t|t}^{(i)} = \tilde{\mathbf{h}}_{t|t-1}^{(i)}. \quad (\text{E.30})$$

For $s \geq 1$, forecasting starts from the final weighted particle system

$$\{(\tilde{\mathbf{h}}_{T|T}^{(i)}, w_T^{(i)})\}_{i=1}^N \quad (\text{E.31})$$

and propagates each particle forward under the state transition:

$$\tilde{\mathbf{h}}_{T+k|T}^{(i)} = \mathcal{A}\tilde{\mathbf{h}}_{T+k-1|T}^{(i)} + \tilde{\mathbf{v}}_{T+k}^{(i)}, \quad \tilde{\mathbf{v}}_{T+k}^{(i)} \sim \mathcal{N}_{pm}(\mathbf{0}, \mathcal{B}), \quad k = 1, \dots, s. \quad (\text{E.32})$$

A natural log-volatility forecast is the weighted particle mean

$$\hat{\mathbf{h}}_{T+s|T} = \mu + \mathcal{C} \left(\sum_{i=1}^N w_T^{(i)} \tilde{\mathbf{h}}_{T+s|T}^{(i)} \right). \quad (\text{E.33})$$

Associated with this forecast, one may define the plug-in volatility matrix

$$\hat{\mathbf{V}}_{T+s|T}^{1/2} := \text{diag}[\exp(\hat{h}_{1,T+s|T}/2), \dots, \exp(\hat{h}_{m,T+s|T}/2)]. \quad (\text{E.34})$$

However, covariance forecasting requires conditional expectations of the nonlinear terms $\exp((h_{r,T+s} + h_{c,T+s})/2)$, not merely the exponential of the conditional mean state. For each propagated particle, define

$$\hat{\mathbf{h}}_{T+s|T}^{(i)} := \mu + \mathcal{C}\tilde{\mathbf{h}}_{T+s|T}^{(i)}, \quad i = 1, \dots, N. \quad (\text{E.35})$$

For this reason, the APF naturally delivers the volatility-scale forecast matrix

$$\hat{M}_{T+s|T} = \sum_{i=1}^N w_T^{(i)} \exp \left\{ \frac{1}{2} (\hat{h}_{T+s|T}^{(i)'} \iota_m' + \iota_m \hat{h}_{T+s|T}^{(i)}) \right\}, \quad (\text{E36})$$

where the exponential is taken elementwise. Its (r, c) entry estimates

$$\hat{M}_{rc, T+s|T} = \mathbb{E}[\exp((h_{r, T+s} + h_{c, T+s})/2) | \mathcal{F}_T]. \quad (\text{E37})$$

Equivalently,

$$\hat{M}_{rc, T+s|T} = \mathbb{E}[(\mathbf{V}_{T+s}^{1/2})_{rr} (\mathbf{V}_{T+s}^{1/2})_{cc} | \mathcal{F}_T]. \quad (\text{E38})$$

Thus, $\hat{M}_{T+s|T}$ is the conditional second-moment matrix associated with the diagonal entries of $\mathbf{V}_{T+s}^{1/2}$. In particular, its diagonal entries recover the forecast conditional variances:

$$\hat{M}_{rr, T+s|T} = \mathbb{E}[e^{h_{r, T+s}} | \mathcal{F}_T]. \quad (\text{E39})$$

More generally,

$$\hat{M}_{T+s|T} \neq \hat{\mathbf{V}}_{T+s|T}^{1/2} \iota_m \iota_m' \hat{\mathbf{V}}_{T+s|T}^{1/2} \quad (\text{E40})$$

because the former preserves the nonlinear effect of state uncertainty, whereas the latter is only a plug-in object based on the forecast mean state.

When the correlation matrix of the standardized innovations is constant and equal to Σ_u , the covariance forecast is

$$\hat{\Sigma}_{T+s|T} = \Sigma_u \circ \hat{M}_{T+s|T}. \quad (\text{E41})$$

Equivalently,

$$\hat{\Sigma}_{T+s|T} = \sum_{i=1}^N w_T^{(i)} \mathbf{V}_{T+s|T}^{(i)1/2} \Sigma_u \mathbf{V}_{T+s|T}^{(i)1/2}, \quad \mathbf{V}_{T+s|T}^{(i)1/2} := \text{diag}[\exp((\mu + \mathcal{C} \tilde{h}_{T+s|T}^{(i)})/2)]. \quad (\text{E42})$$

In moderate dimensions, the APF typically achieves a given accuracy with fewer particles than the bootstrap PF; see [Pitt and Shephard \(1999\)](#).

G Technical details for the common C-space restriction

G.1 Proof related to common C-space restriction

PROOF OF LEMMA 5.1 Consider the constrained least-squares problem

$$\hat{\mathbf{C}} := \arg \min_{\mathbf{C} \in \mathbb{R}^{m \times r}: \mathbf{C}'\mathbf{C} = \mathbf{I}_r} \sum_{j=1}^p \|\hat{\phi}_j - \hat{\phi}_j \mathbf{C}\mathbf{C}'\|_F^2. \quad (\text{G.1})$$

Using $\|\mathbf{A} - \mathbf{A}\mathbf{P}\|_F^2 = \text{tr}(\mathbf{A}'\mathbf{A}) - \text{tr}(\mathbf{P}\mathbf{A}'\mathbf{A})$ for any orthogonal projector $\mathbf{P}$ and writing $\mathbf{P} = \mathbf{C}\mathbf{C}'$ with $\mathbf{C}'\mathbf{C} = \mathbf{I}_r$, the objective in (G.1) satisfies

$$\sum_{j=1}^p \|\hat{\boldsymbol{\phi}}_j - \hat{\boldsymbol{\phi}}_j \mathbf{C}\mathbf{C}'\|_F^2 = \sum_{j=1}^p \text{tr}(\hat{\boldsymbol{\phi}}_j' \hat{\boldsymbol{\phi}}_j) - \text{tr}[\mathbf{C}' \left(\sum_{j=1}^p \hat{\boldsymbol{\phi}}_j' \hat{\boldsymbol{\phi}}_j \right) \mathbf{C}] = \text{const} - \text{tr}(\mathbf{C}' \hat{\mathbf{M}}_C \mathbf{C}) \quad (\text{G.2})$$

where $\hat{\mathbf{M}}_C = \sum_{j=1}^p \hat{\boldsymbol{\phi}}_j' \hat{\boldsymbol{\phi}}_j$. Therefore minimizing (G.1) is equivalent to maximizing $\text{tr}(\mathbf{C}' \hat{\mathbf{M}}_C \mathbf{C})$ subject to $\mathbf{C}'\mathbf{C} = \mathbf{I}_r$. The maximizer is obtained by taking $\hat{\mathbf{C}}$ as the r leading eigenvectors of $\hat{\mathbf{M}}_C$. $\square$

G.2 Basis indeterminacy

Under $H_0^{\text{com}}(r)$, $\mathbf{C}$ is identified only up to orthogonal rotation: if $\mathbf{Q} \in \mathbb{R}^{r \times r}$ and $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_r$, then $(\mathbf{B}_j, \mathbf{C})$ and $(\mathbf{B}_j\mathbf{Q}, \mathbf{C}\mathbf{Q})$ generate the same $\boldsymbol{\phi}_j = \mathbf{B}_j\mathbf{C}'$. As a result, inference is naturally expressed in terms of the projector $\mathbf{P}_C = \mathbf{C}\mathbf{C}'$, which is rotation-invariant.

G.3 Delta-method mapping for $\hat{\mathbf{M}}_C$ and $\hat{\mathbf{P}}_C$

The delta-method calculations below take as input the asymptotic distributional theory for the unrestricted estimators developed in Section 4. Let

$$\mathbf{M}_C = \sum_{j=1}^p \boldsymbol{\phi}_j' \boldsymbol{\phi}_j, \quad \hat{\mathbf{M}}_C = \sum_{j=1}^p \hat{\boldsymbol{\phi}}_j' \hat{\boldsymbol{\phi}}_j. \quad (\text{G.3})$$

Define the Jacobian mapping from the stacked vector

$$(\text{vec}(\hat{\boldsymbol{\phi}}_1)', \dots, \text{vec}(\hat{\boldsymbol{\phi}}_p'))' \quad (\text{G.4})$$

to $\text{vec}(\hat{\mathbf{M}}_C)$ as

$$\mathbf{D}_M := \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \dots & \mathbf{D}_p \end{pmatrix}, \quad \mathbf{D}_j := (\mathbf{I}_m \otimes \boldsymbol{\phi}_j') + (\boldsymbol{\phi}_j' \otimes \mathbf{I}_m) \mathbf{K}_{mm}. \quad (\text{G.5})$$

Then, by standard matrix differential calculus,

$$d \text{vec}(\mathbf{M}_C) = \mathbf{D}_M \begin{pmatrix} \text{vec}(d\boldsymbol{\phi}_1) \\ \vdots \\ \text{vec}(d\boldsymbol{\phi}_p) \end{pmatrix} \quad (\text{G.6})$$

so the asymptotic distribution of $\text{vec}(\hat{\mathbf{M}}_C - \mathbf{M}_C)$ follows directly from the delta method.

Under the eigengap condition $\lambda_r(\mathbf{M}_C) > \lambda_{r+1}(\mathbf{M}_C)$, the rank- r spectral projector $\mathbf{P}_C$ is Hadamard differentiable as a function of $\mathbf{M}_C$.¹ Consequently, $\hat{\mathbf{P}}_C = \hat{\mathbf{C}}\hat{\mathbf{C}}'$ is asymptotically normal, and the same holds for the restricted estimators $\hat{\boldsymbol{\phi}}_j \hat{\mathbf{P}}_C$ by the joint delta method applied to the product map.

¹An eigengap condition requires a strict separation between adjacent eigenvalues. Here, $\lambda_r(\mathbf{M}_C) > \lambda_{r+1}(\mathbf{M}_C)$ ensures that the leading rank- r eigenspace is uniquely identified and varies smoothly under small perturbations of $\mathbf{M}_C$.

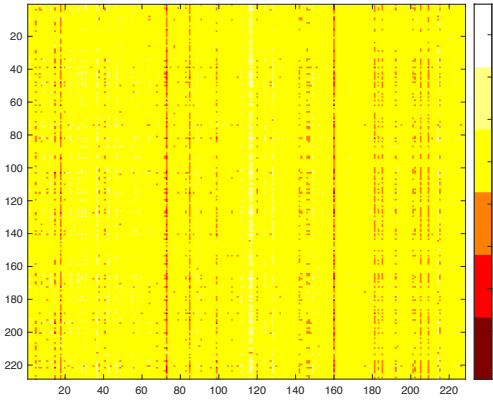
H Additional figures

Figure A1

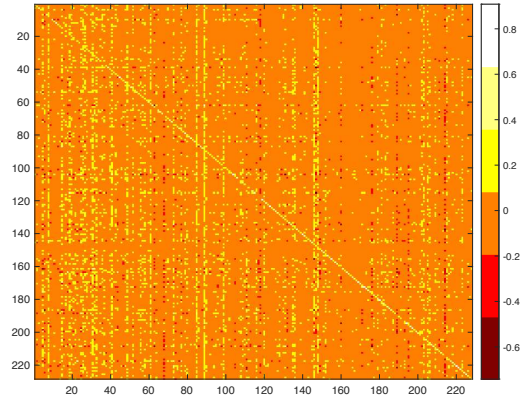
High-dimensional heatmap plots of the estimated first-order volatility persistence matrix with different values of shrinkage control parameter

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

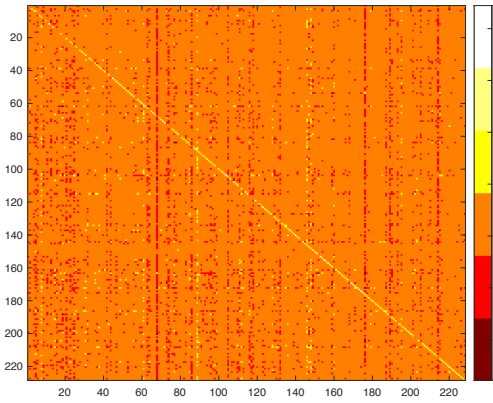
Figure A1 shows the heatmap plots of $\hat{\phi}_{(1)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$). The sample period is from January 01, 2004 to December 31, 2016 and the number of five minute high-frequency observations is $T = 252,021$. The 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 ϕ_1 and diagonal Σ_p . The model has $m = 228$ number of time series and the number of parameters (NP) are $(= m^2 + 2m = 52,440)$ and $\phi_{(1)}$ is a $m \times m$ matrix, which has 51,984 parameters.



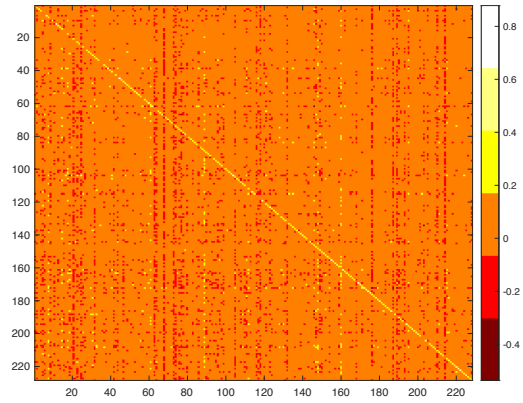
(a) $\hat{\phi}_{(1)}$ with $J = 1$ (unstable)



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



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



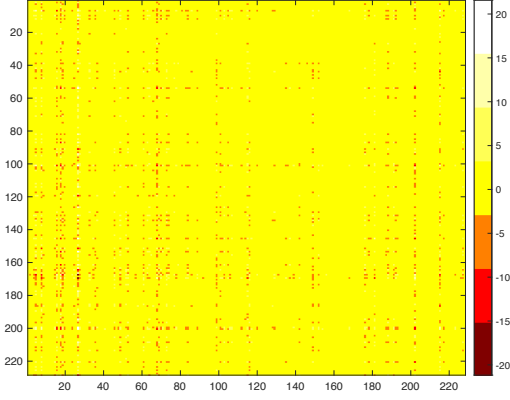
(d) $\hat{\phi}_{(1)}$ with $J = 100$ (stable)

Figure A2

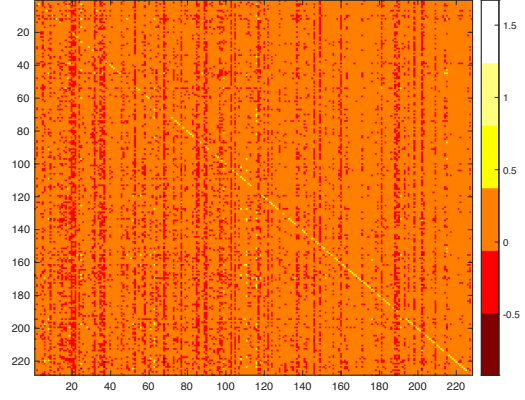
High-dimensional heatmap plots of the estimated first-order volatility persistence matrix with different values of shrinkage control parameter

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

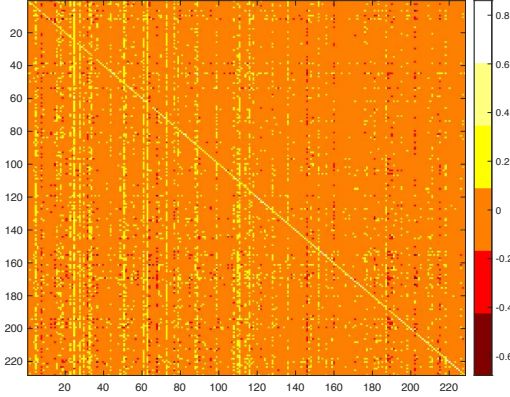
Figure A2 shows the heatmap plots of $\hat{\phi}_{(1)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$). The sample period is from July 01, 2009 to December 31, 2016 and the number of five minute high-frequency observations is $T = 126,011$. The balanced 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 ϕ_1 and diagonal Σ_v . The model has $m = 228$ number of time series and the number of parameters (NP) are $(= m^2 + 2m = 52,440)$ and $\phi_{(1)}$ is a $m \times m$ matrix, which has 51,984 parameters.



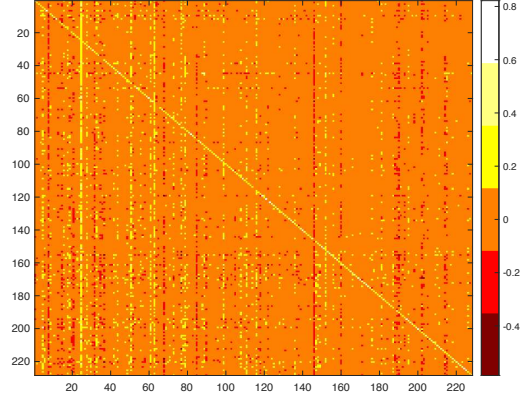
(a) $\hat{\phi}_{(1)}$ with $J = 1$ (unstable)



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



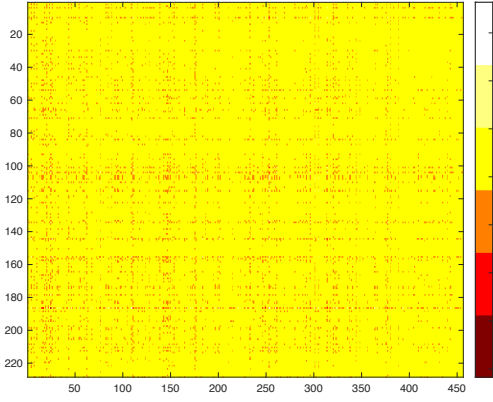
(c) $\hat{\phi}_{(1)}$ with $J = 50$ (stable)



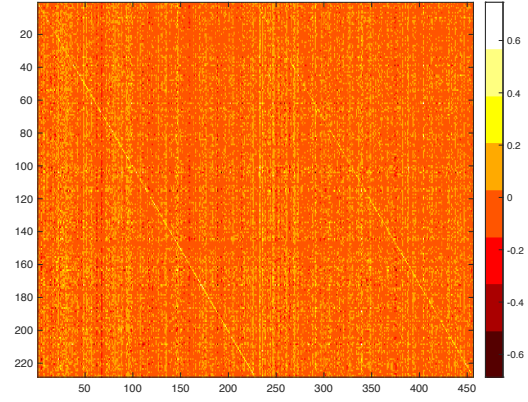
(d) $\hat{\phi}_{(1)}$ with $J = 100$ (stable)

Figure A3
High-dimensional heatmap plots of the estimated second-order volatility persistence matrix with
different values of shrinkage control parameter
Number of asset $m = 228$ and number of high-frequency observations $T = 252,021$

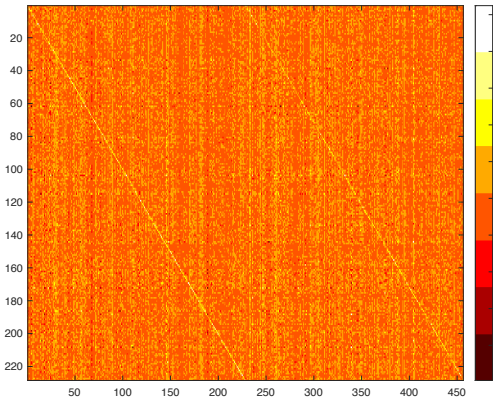
Figure A3 shows the heatmap plots of $\hat{\phi}_{(2)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$). The sample period is from July 01, 2009 to December 31, 2016 and the number of five minute high-frequency observations is $T = 252,021$. The balanced 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 ϕ_1 and ϕ_2 , and diagonal Σ_ν . The model has $m = 228$ number of time series and the number of parameters (NP) are $(= 2m^2 + 2m = 104,424)$ and $\phi_{(2)}$ is a $m \times 2m$ matrix, which has 103,968 parameters.



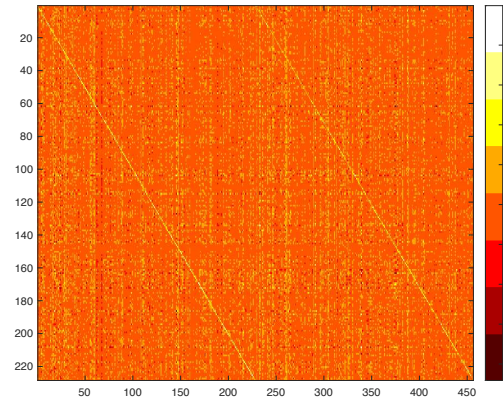
(a) $\hat{\phi}_{(2)}$ with $J = 1$ (unstable)



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



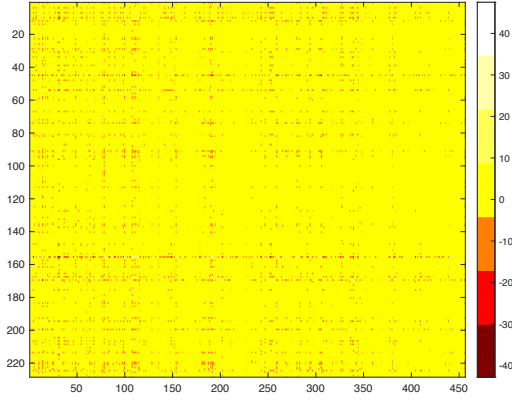
(c) $\hat{\phi}_{(2)}$ with $J = 50$ (stable)



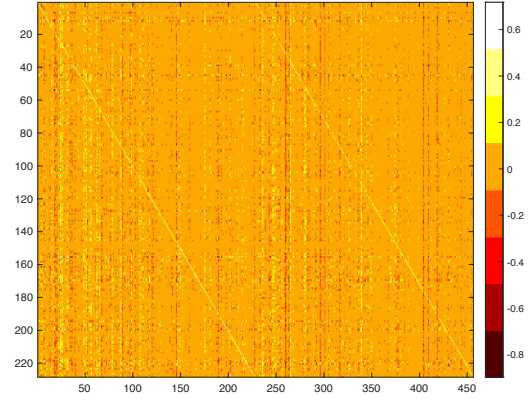
(d) $\hat{\phi}_{(2)}$ with $J = 100$ (stable)

Figure A4
 High-dimensional heatmap plots of the estimated second-order volatility persistence matrix with
 different values of shrinkage control parameter
 Number of asset $m = 228$ and number of high-frequency observations $T = 126,011$

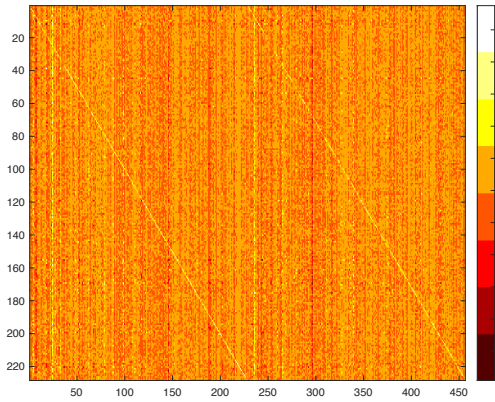
Figure A4 shows the heatmap plots of $\hat{\phi}_{(2)}$ with different values of winsorizing truncation / shrinkage control parameter ($J = 1, 10, 50, 100$). The sample period is from July 01, 2009 to December 31, 2016 and the number of five minute high-frequency observations is $T = 126,011$. The balanced 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 ϕ_1 and ϕ_2 , and diagonal Σ_ν . The model has $m = 228$ number of time series and the number of parameters (NP) are $(= 2m^2 + 2m = 104,424)$ and $\phi_{(2)}$ is a $m \times 2m$ matrix, which has 103,968 parameters.



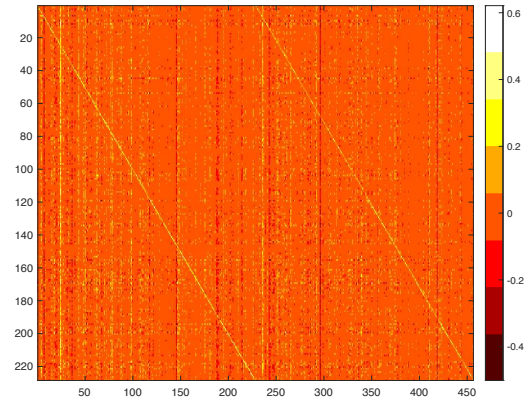
(a) $\hat{\phi}_{(2)}$ with $J = 1$ (unstable)



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



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



(d) $\hat{\phi}_{(2)}$ 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

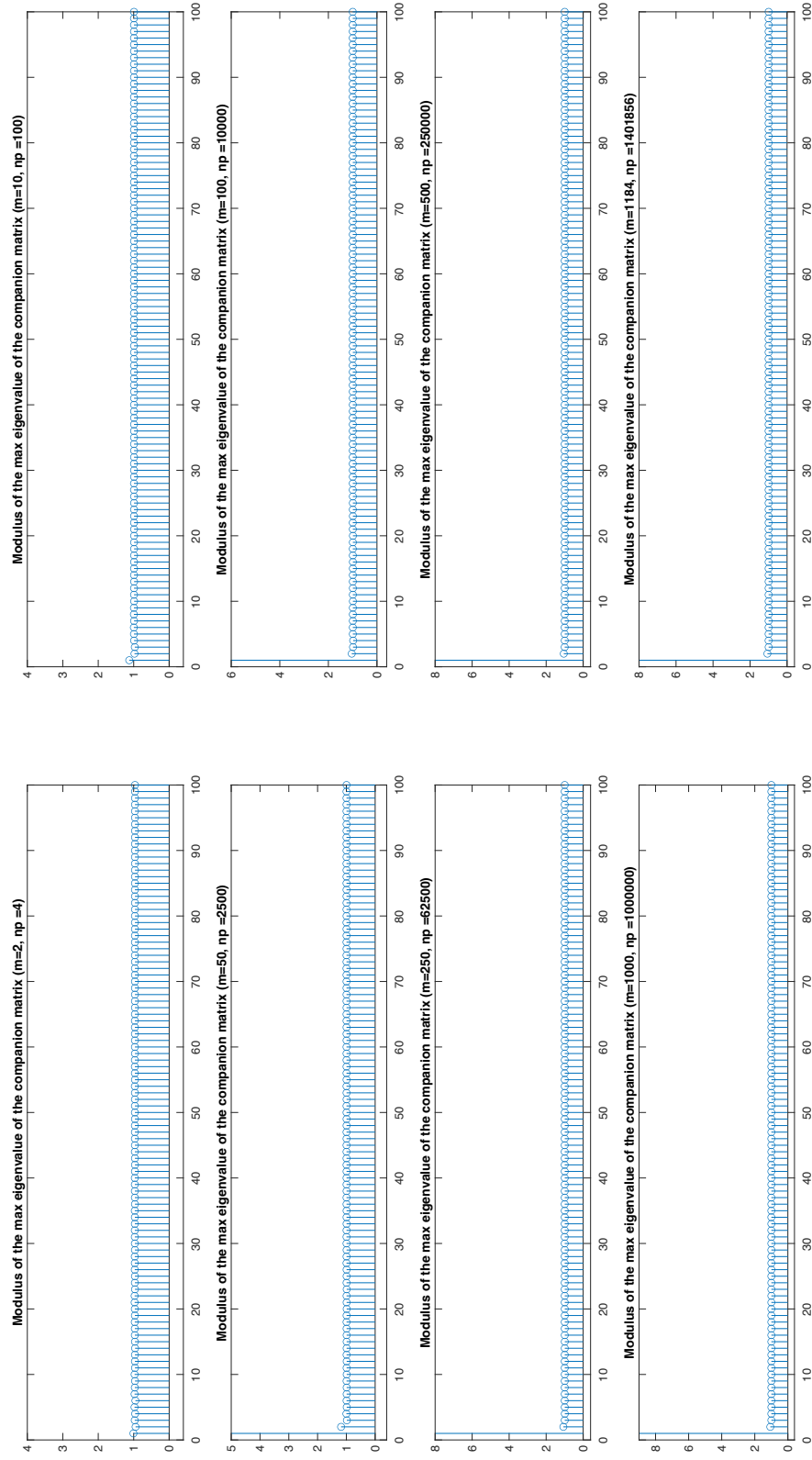


Figure A5 shows the evolution of the modulus of the maximum eigenvalue of the companion matrix associated with a first-order unrestricted MSV model with different numbers of assets $m = (2, 10, 50, 100, 250, 500, 1000, 1184)$ and different values of the shrinkage control parameter $J = (1, \dots, 100)$.

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

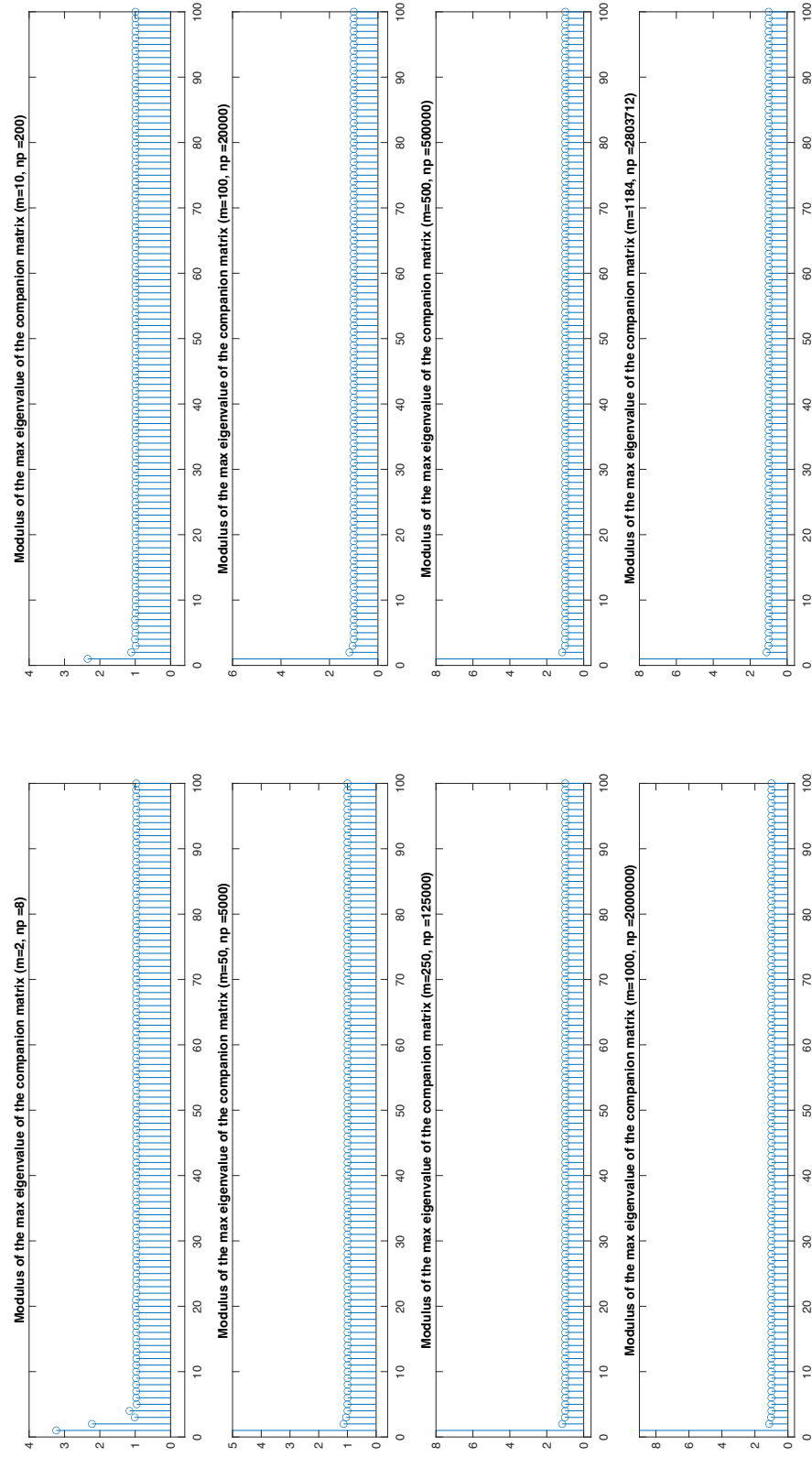


Figure A6 shows the evolution of the modulus of the maximum eigenvalue of the companion matrix associated with a second-order unrestricted MSV model with different numbers of assets $m = (2, 10, 50, 100, 250, 500, 1000, 1184)$ and different values of the shrinkage control parameter $J = (1, \dots, 100)$.

I Additional 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 winsorized estimators reduce to the CF estimator. RMSE is the estimated root mean squared error based on the simulation. NIV denotes the number of inadmissible parameter values produced by the estimators, out of 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

T	Estimators	RCT	NIV	NPD	Bias					RMSE						
					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 A2 reports the performance of QML and W-VARMA-MVR estimators in bi-variate setups. VARMA-type estimators are proposed in Section 3. The 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 $(\phi_{11}, \phi_{22}, \sigma^2_{v,1}, \sigma^2_{v,2}, \sigma_{v,12}, \rho) = (0.98, 0.98, 0.5, 0.5, -0.8)$. We simulate 100 samples from each model. RCT denotes relative computational time with respect to the CF-VARMA estimator. The simulated model has $m = 2$ time series and the number of parameters is 6. The number of inadmissible values (NIV) of ϕ and non-positive definiteness (NPD) of Σ_v are also reported, out of 100. Boldface highlights the smallest bias and RMSE among estimators with no NIV/NPD, and also highlights the estimator with the best overall performance.

Table A3

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

T	Estimators	RCT	NIV	NPD	Bias					RMSE										
					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	$\sigma^2_{v,1}$	$\sigma^2_{v,2}$
1000	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
	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
2000	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
	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
5000	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
	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
5000	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 A3 reports the performance of Bayes and W-VARMA-MVR estimators in bi-variate setups. VARMA-type estimators are proposed in Section 3. The Bayes estimator is based on a Metropolis-within-Gibbs sampler; see Kim et al. (1998), Omori et al. (2007), and Kastner and Frühwirth-Schnatter (2014). Prior distributions, the MCMC algorithm, and computational details are discussed in Section 6.1.3. The posteriors are based on 20,000 draws after discarding 20,000 draws. We consider first-order persistence in latent volatilities. The true DGP is given by (2.1)–(2.2) with $\Sigma_u = \mathbf{I}_m$, $\Sigma_v = 0.4\mathbf{I}_m$, $\mu = -0.5\mathbf{1}$, and stationary ϕ_1 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 model. RCT denotes relative computational time with respect to the CF-VARMA estimator. The simulated model has $m = 2$ time series and the number of parameters is $m^2 + 2m = 8$. NIV and NPD are reported out of 100. Boldface highlights the smallest bias and RMSE among estimators with no NIV/NPD, and also highlights the estimator with the best overall performance.

Table A4
Simulation results for estimated MSV processes: Bayes vs W-VARMA-MVR estimator

Table A4 reports the performance of Bayes and W-VARMA-MVR estimators in bi-variate setups. VARMA-type estimators are proposed in Section 3. The Bayes estimator is based on a Metropolis-within-Gibbs sampler; see Kim et al. (1998), Omori et al. (2007), and Kastner and Frühwirth-Schnatter (2014). Prior distributions, the MCMC algorithm, and computational details are discussed in Section 6.1.3. The posteriors are based on 20,000 draws after discarding 20,000 draws. We consider first-order persistence in latent volatilities. The true DGP is given by (2.1)–(2.2) with $m \in \{2, 5, 10\}$ and $p = 1$. We set $\Sigma_u = I_m$, $\Sigma_v = 0.4I_m$, and $\mu = -0.51$. For $m = 2$, ϕ_1 is non-diagonal with $(\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}) = (0.98, -0.1, 0.12, 0.99)$. For $m = 5$, ϕ_1 is non-diagonal, with diagonal elements $(0.98, 0.985, 0.8, 0.985, 0.98)'$, upper off-diagonal elements equal to -0.15 , and lower off-diagonal elements equal to 0.1 . For $m = 10$, ϕ_1 is non-diagonal, with diagonal elements $(0.95, 0.98, 0.98, 0.99, 0.8, 0.8, 0.99, 0.98, 0.98, 0.95)'$, upper off-diagonal elements equal to -0.11 , and lower off-diagonal elements equal to 0.055 . We simulate 100 samples from each model. RCT denotes relative computational time with respect to the CF-VARMA estimator. The number of parameters is $m^2 + 2m$. NIV and NPD are reported out of 100. Boldface highlights the smallest bias and RMSE among estimators with no NIV/NPD, and also highlights the estimator with the best overall performance. Reported average bias and RMSE are element-wise averages over the relevant vectors and matrices.

$m = 2, p = 1, NP = 8$												
T	Estimators	RCT	NIV	NPD	Average Bias				Average RMSE			
					ϕ	diag[ϕ]	μ	Σ_v	ϕ	diag[ϕ]	μ	Σ_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
$m = 5, p = 1, NP = 35$												
T	Estimators	RCT	NIV	NPD	Average Bias				Average RMSE			
					ϕ	diag[ϕ]	μ	Σ_v	ϕ	diag[ϕ]	μ	Σ_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
$m = 10, p = 1, NP = 120$												
T	Estimators	RCT	NIV	NPD	Average Bias				Average RMSE			
					ϕ	diag[ϕ]	μ	Σ_v	ϕ	diag[ϕ]	μ	Σ_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 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 up to 250 dimensions and first-order persistence in latent volatilities. The true DGP is given by (2.1)–(2.2) with $\Sigma_u = I_m$, $\Sigma_v = I_m$, $\mu = -1$, and stationary ϕ_1 non-diagonal, with diagonal elements equal to 0.9 and off-diagonal elements between -0.01 and 0.01 . We simulate 100 samples per replication is reported in seconds. NP ($= m^2 + 2m$) denotes the number of parameters. NIV and NPD are reported out of 100. Reported average bias and RMSE are element-wise averages over the relevant vectors and matrices.																					
$J = 10$										$J = 100$											
T	m	NP	Time					Average Bias			Average RMSE			Average Bias					Average RMSE		
			Time	NIV	NPD	ϕ	μ	Σ_v	ϕ	μ	Σ_v	Time	NIV	NPD	ϕ	μ	Σ_v	ϕ	μ	Σ_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	
2500	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	
	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	
5000	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	
	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	
10000	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	
	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	
10000	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

J = 10													J = 100							
T	m	NP	Time	NIV	NPD	Average Bias			Average RMSE			Time	NIV	NPD	Average Bias			Average RMSE		
						φ	μ	Σ _v	φ	μ	Σ _v				φ	μ	Σ _v	φ	μ	Σ _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
2500	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
	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
5000	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
	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
10000	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
	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
10000	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 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 up to 250 dimensions and second-order persistence in latent volatilities. The true DGP is given by (2.1)–(2.2) with $\Sigma_u = I_m$, $\Sigma_v = I_m$, $\mu = -I$, and stationary ϕ_1 and ϕ_2 non-diagonal. For ϕ_1 , diagonal elements are equal to 0.9 and off-diagonal elements lie between -0.01 and 0.01 . For ϕ_2 , diagonal elements are equal to -0.85 and off-diagonal elements lie between -0.01 and 0.01 . We simulate 100 samples from each model. Computational time per replication is reported in seconds. NP ($= 2m^2 + 2m$) denotes the number of parameters. NIV and NPD are reported out of 100. Reported average bias and RMSE are element-wise averages over the relevant vectors and matrices.

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 up to 1500 dimensions and second-order persistence in latent volatilities. The true DGP is given by (2.1)–(2.2) with $\Sigma_u = I_m$, $\Sigma_v = I_m$, $\mu = -1$, and stationary ϕ_1 and ϕ_2 non-diagonal. For ϕ_1 , diagonal elements are equal to 0.9 and off-diagonal elements lie between -0.001 and 0.001 . For ϕ_2 , diagonal elements are equal to -0.8 and off-diagonal elements lie between -0.001 and 0.001 . We simulate 100 samples from each model. Computational time per replication is reported in seconds. NP ($= 2m^2 + 2m$) denotes the number of parameters. NIV and NPD are reported out of 100. Reported average bias and RMSE are element-wise averages over the relevant vectors and matrices.

$J = 10$											
T	m	NP	Time	NIV	NPD	Average Bias			Average RMSE		
						ϕ	μ	Σ_v	ϕ	μ	Σ_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
$J = 100$											
T	m	NP	Time	NIV	NPD	Average Bias			Average RMSE		
						ϕ	μ	Σ_v	ϕ	μ	Σ_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
Series-level in-sample estimation for latent log-variance: MSE

Table A9 reports series-level Monte Carlo average filtering losses for the latent log-variance, under the same DGP settings as in Section 6.1.3, for $m \in \{2, 5, 10\}$. These results provide the series-level decomposition of Panel A of Table 3. For each (T, m) and series i , the reported value is the Monte Carlo average over $B = 100$ replications of $T^{-1} \sum_{t=1}^T (h_{t,i} - \hat{h}_{t,i|t})^2$. The column labeled “Avg” reports the cross-series average and coincides with the corresponding entry in Panel A of Table 3. In each Monte Carlo replication, parameters are re-estimated using VARMA-MVR ($J = 10$), while the particle-filter implementations use $N_{\text{pf}} = 10,000$ particles. Boldface identifies, for each fixed (T, m, i) , the lowest loss among KF, BPF and APF.

T	m	Filter	Avg	Series 1	Series 2	Series 3	Series 4	Series 5	Series 6	Series 7	Series 8	Series 9	Series 10
1000	2	KF	1.632	1.428	1.835								
	2	BPF	1.160	1.040	1.280								
	2	APF	1.156	1.035	1.278								
	5	KF	1.314	1.495	1.295	0.995	1.509	1.276					
	5	BPF	0.996	1.129	0.978	0.773	1.137	0.964					
	5	APF	0.993	1.117	0.976	0.770	1.141	0.960					
	10	KF	1.627	2.091	2.273	1.518	1.350	0.999	1.072	3.095	1.534	1.245	1.096
	10	BPF	1.334	1.723	1.895	1.223	1.088	0.857	0.857	2.581	1.219	0.999	0.896
	10	APF	1.323	1.706	1.872	1.210	1.079	0.856	0.849	2.583	1.213	0.989	0.879
2000	2	KF	1.624	1.451	1.797								
	2	BPF	1.145	1.045	1.244								
	2	APF	1.143	1.044	1.243								
	5	KF	1.302	1.481	1.279	0.975	1.500	1.273					
	5	BPF	0.982	1.119	0.968	0.756	1.122	0.943					
	5	APF	0.979	1.117	0.966	0.752	1.119	0.940					
	10	KF	1.545	1.907	2.170	1.447	1.293	0.942	1.015	2.875	1.501	1.234	1.061
	10	BPF	1.233	1.532	1.732	1.141	1.015	0.791	0.807	2.322	1.149	0.971	0.873
	10	APF	1.222	1.520	1.715	1.126	1.002	0.782	0.794	2.325	1.137	0.956	0.859
5000	2	KF	1.645	1.447	1.844								
	2	BPF	1.158	1.043	1.272								
	2	APF	1.158	1.043	1.272								
	5	KF	1.285	1.456	1.271	0.960	1.497	1.239					
	5	BPF	0.963	1.091	0.955	0.738	1.112	0.920					
	5	APF	0.962	1.091	0.955	0.737	1.111	0.919					
	10	KF	1.552	1.986	2.174	1.486	1.278	0.913	1.016	2.855	1.526	1.209	1.073
	10	BPF	1.233	1.599	1.734	1.151	1.005	0.755	0.798	2.302	1.167	0.945	0.873
	10	APF	1.221	1.587	1.727	1.135	0.991	0.747	0.785	2.295	1.154	0.932	0.857

Table A10
Series-level one-step-ahead out-of-sample forecasting for latent log-variance: MSE

Table A10 reports series-level Monte Carlo average one-step-ahead out-of-sample forecasting losses for the latent log-variance, under the same DGP settings as in Section 6.1.3, for $m \in \{2, 5, 10\}$. These results provide the series-level decomposition of Panel B of Table 3. For each (T, m) and series i , the reported value is the Monte Carlo average over $B = 100$ replications of $(h_{T+1,i} - \hat{h}_{T+1|T,i})^2$. The column labeled “Avg” reports the cross-series average and coincides with the corresponding entry in Panel B of Table 3. At each forecast origin, parameters are re-estimated using VARMA-MVR ($J = 10$), while the particle-filter implementations use $N_{\text{pf}} = 10,000$ particles. Boldface identifies, for each fixed (T, m, i) , the lowest loss among KF, BPF, and APF.

T	m	Filter	Avg	Series 1	Series 2	Series 3	Series 4	Series 5	Series 6	Series 7	Series 8	Series 9	Series 10
1000	2	KF	1.737	1.768	1.706								
	2	BPF	1.350	1.394	1.306								
	2	APF	1.355	1.411	1.300								
	5	KF	1.763	2.188	1.532	1.125	1.765	2.206					
	5	BPF	1.462	1.694	1.405	0.903	1.615	1.694					
	5	APF	1.465	1.711	1.393	0.889	1.645	1.685					
	10	KF	2.175	3.759	3.125	1.997	1.869	1.023	1.473	3.291	2.205	1.288	1.722
	10	BPF	1.876	3.483	2.605	1.789	1.676	0.890	1.159	2.559	1.755	1.165	1.685
	10	APF	1.896	3.487	2.612	1.900	1.667	0.920	1.113	2.703	1.721	1.163	1.680
2000	2	KF	2.110	1.429	2.791								
	2	BPF	1.555	1.210	1.900								
	2	APF	1.553	1.207	1.900								
	5	KF	1.705	1.850	1.660	1.135	1.950	1.929					
	5	BPF	1.280	1.401	1.129	0.880	1.587	1.405					
	5	APF	1.244	1.342	1.102	0.859	1.543	1.375					
	10	KF	1.946	3.409	2.270	1.598	1.628	1.381	1.351	2.515	2.181	1.722	1.408
	10	BPF	1.566	3.035	1.645	1.481	1.464	1.126	1.113	1.994	1.416	1.293	1.096
	10	APF	1.545	2.987	1.616	1.324	1.443	1.155	1.092	1.938	1.525	1.207	1.161
5000	2	KF	1.816	1.725	1.907								
	2	BPF	1.402	1.456	1.349								
	2	APF	1.399	1.460	1.338								
	5	KF	1.587	1.532	1.454	1.432	1.757	1.761					
	5	BPF	1.252	1.298	1.195	1.220	1.263	1.284					
	5	APF	1.249	1.282	1.192	1.226	1.230	1.314					
	10	KF	1.840	2.772	2.156	1.612	1.766	1.242	1.093	2.566	2.204	1.405	1.588
	10	BPF	1.528	2.066	1.902	1.315	1.364	0.917	1.086	2.215	2.048	0.991	1.374
	10	APF	1.491	1.988	1.804	1.338	1.426	0.946	1.099	2.094	1.996	0.900	1.319

J Supplementary Monte Carlo details and results: MSV versus FSV

This appendix collects the simulation design, parameter settings, estimation details, prior and hyperparameter information for the FSV benchmark, formal definitions of the performance measures, and the series-specific results underlying Section 6.2.2.

J.1 Data-generating processes and parameter settings for FSV

J.1.1 FSV data-generating process

The FSV data-generating process follows the factor stochastic volatility structure

$$\mathbf{y}_t = \beta + \mathbf{X}\mathbf{f}_t + \mathbf{e}_t, \quad \mathbf{f}_t | \mathbf{h}_t^f \sim \mathcal{N}(\mathbf{0}, \text{diag}(\exp(\mathbf{h}_t^f))), \quad \mathbf{e}_t | \mathbf{h}_t^e \sim \mathcal{N}(\mathbf{0}, \text{diag}(\exp(\mathbf{h}_t^e))), \quad (\text{J.1})$$

where $\mathbf{h}_t^f$ is the vector of factor log-variances and $\mathbf{h}_t^e$ collects idiosyncratic log-variances. The factor log-variance dynamics are

$$\mathbf{h}_t^f = \mu^f + \text{diag}(\phi^f) (\mathbf{h}_{t-1}^f - \mu^f) + \xi_t, \quad \xi_t \sim \mathcal{N}(\mathbf{0}, \mathbf{U}), \quad (\text{J.2})$$

while the idiosyncratic log-variances evolve according to

$$h_{j,t}^e = a_j + \rho_j (h_{j,t-1}^e - a_j) + u_{j,t}, \quad u_{j,t} \sim \mathcal{N}(0, s_j^2), \quad j = 1, \dots, m. \quad (\text{J.3})$$

The implied time-varying covariance matrix is

$$\Sigma_t = \mathbf{X} \text{diag}(\exp(\mathbf{h}_t^f)) \mathbf{X}' + \text{diag}(\exp(\mathbf{h}_t^e)), \quad (\text{J.4})$$

so that the series-specific latent log-variance entering the evaluation is

$$h_{j,t} = \log([\Sigma_t]_{jj}), \quad v_{j,t} = [\Sigma_t]_{jj}. \quad (\text{J.5})$$

The factor dimension is fixed at $k_F = 2$. The simulation code sets

$$\beta = \mathbf{0}, \quad \mu^f = -8 \mathbf{1}_{k_F}, \quad \phi^f = (0.97, 0.93)', \quad \mathbf{U} = 0.5 \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}, \quad (\text{J.6})$$

and, for each series j ,

$$a_j = -8, \quad \rho_j = 0.95, \quad s_j^2 = 0.5. \quad (\text{J.7})$$

The loading matrix $\mathbf{X}$ is generated internally by the FSV simulator under the hierarchical identification.

J.1.2 Estimation settings and hyperparameters

The benchmark FSV specification is estimated on demeaned returns with $k_F = 2$ latent factors. The MCMC algorithm is run for a total of 20,000 iterations, with the first 10,000 discarded as burn-in and a thinning interval of 10, leaving 1,000 retained draws for inference. Hierarchical identification is imposed on the loading matrix, and posterior means are used as point estimates for both the latent log-variance paths

and the one-step-ahead volatility forecasts. Adopting a Bayesian perspective, the specification assigns diffuse Gaussian priors to the intercept vector and the unrestricted factor loadings, namely $\beta \sim \mathcal{N}_m(\mathbf{0}, 10^6 \mathbf{I}_m)$ and $x_{ij} \sim \mathcal{N}(0, 10^6)$ for the free loading entries, subject to the identification restrictions. For the factor log-volatility component, the unconditional mean vector satisfies $\mu \sim \mathcal{N}_{k_F}(\mathbf{0}, 10^2 \mathbf{I}_{k_F})$, the innovation covariance matrix satisfies $\mathbf{U} \sim \mathcal{IW}(k_F + 2, 0.0015 \mathbf{I}_{k_F})$, and the persistence parameters are sampled from diffuse Gaussian posteriors under the restriction $\phi_i \in (0, 1)$ for $i = 1, \dots, k_F$. For the idiosyncratic log-volatility equations, the transformed AR(1) coefficients (c_j, ρ_j) , where $c_j = (1 - \rho_j)a_j$, are assigned diffuse Gaussian priors with covariance matrix $10^6 \mathbf{I}_2$, the innovation variances satisfy $s_j^2 \sim \mathcal{IG}(2.5, 2.5 \times 0.003)$ for $j = 1, \dots, m$, and the persistence parameters are restricted to $\rho_j \in (0, 1)$ to ensure stationary dynamics. The likelihood evaluation further employs the standard seven-component normal-mixture approximation to the $\log(\chi_1^2)$ density used in [Kim et al. \(1998\)](#).

J.2 Performance measures

The first criterion is the pooled RMSE for the latent log-variance paths,

$$\text{RMSE}_h = \left(\frac{1}{mT} \sum_{t=1}^T \sum_{j=1}^m (\hat{h}_{j,t} - h_{j,t})^2 \right)^{1/2}, \quad (\text{J.8})$$

together with the corresponding series-specific measure

$$\text{RMSE}_{h,j} = \left(\frac{1}{T} \sum_{t=1}^T (\hat{h}_{j,t} - h_{j,t})^2 \right)^{1/2}, \quad j = 1, \dots, m. \quad (\text{J.9})$$

The second criterion evaluates one-step-ahead forecasts of the latent conditional volatility. Let

$$\sigma_{j,T+1} = \exp(h_{j,T+1}/2) \quad (\text{J.10})$$

denote the true conditional volatility and let $\hat{\sigma}_{j,T+1|T}$ denote its one-step-ahead forecast. This is a forecast of the conditional volatility, not of the latent log-variance $h_{j,T+1}$. The pooled criterion is

$$\text{RMSE}_\sigma = \left(\frac{1}{m} \sum_{j=1}^m (\hat{\sigma}_{j,T+1|T} - \sigma_{j,T+1})^2 \right)^{1/2}. \quad (\text{J.11})$$

Because there is only one forecast origin per replication, the series-specific quantity reported in the detailed table is the absolute forecast error $|\hat{\sigma}_{j,T+1|T} - \sigma_{j,T+1}|$ for each series j , averaged across replications. Accordingly, the pooled entries in the main table and in the pooled column of the appendix table are Monte Carlo means of the pooled criterion in (J.11); they are not obtained by mechanically aggregating the reported series means.

J.3 Additional discussion of the results

Table A11 complements the pooled evidence reported in the main text by reporting series-specific results together with a pooled column. Under the MSV data-generating process, the advantage of MSV(APF) in

latent path recovery is broad-based across all five series and remains visible at every sample size. The benchmark FSV(MCMC) method is uniformly less accurate for $\text{RMSE}_{h,j}$ in this design.

Under the FSV data-generating process, the path-recovery advantage of MSV(APF) also remains visible at the series level, although it is more moderate than under the MSV design. The series-specific $\text{RMSE}_{h,j}$ values of MSV(APF) range from 0.8869 to 1.0792 across the three sample sizes, whereas the corresponding FSV(MCMC) values range from 1.2417 to 1.3554. Hence, the superiority of MSV(APF) in the pooled RMSE_h metric is not an artifact of aggregation.

The volatility-forecast evidence is more mixed. Under the MSV data-generating process, forecast errors are materially larger and more heterogeneous across series, with neither method uniformly dominating. Under the FSV data-generating process, both procedures yield small series-specific forecast errors, generally around 0.016–0.040, and the ranking varies little across sample sizes. In particular, FSV(MCMC) performs slightly better at $T = 1000$, the two procedures are effectively tied at $T = 2000$ in pooled terms, and MSV(APF) is marginally better at $T = 3000$.

Taken together, the detailed results reinforce the main conclusion. The proposed MSV(APF) procedure is consistently superior for filtering the latent log-variance paths under both data-generating environments. By contrast, one-step-ahead volatility forecasting is more sensitive to specification, but in the present design the differences under the FSV data-generating process remain small rather than dramatic.

Table A11
Series-specific and pooled Monte Carlo RMSE measures

Table A11 reports series-specific Monte Carlo average RMSE measures, together with a pooled column, under the MSV and FSV data-generating processes for $m = 5$ and $T \in \{1000, 2000, 3000\}$. Panel A reports latent log-variance path recovery, and Panel B reports one-step-ahead conditional-volatility forecast errors. Within each panel, results are organized by sample size T . In Panel A, the series-specific measure is $\text{RMSE}_{h,j} = [T^{-1} \sum_{t=1}^T (\hat{h}_{j,t} - h_{j,t})^2]^{1/2}$, while the pooled measure is $\text{RMSE}_h = [(mT)^{-1} \sum_{t=1}^T \sum_{j=1}^m (\hat{h}_{j,t} - h_{j,t})^2]^{1/2}$. In Panel B, the reported series-specific quantity is the absolute one-step-ahead forecast error $|\hat{\sigma}_{j,T+1|T} - \sigma_{j,T+1}|$, while the pooled column reports the Monte Carlo average of the pooled criterion $\text{RMSE}_\sigma = [m^{-1} \sum_{j=1}^m (\hat{\sigma}_{j,T+1|T} - \sigma_{j,T+1})^2]^{1/2}$, where $\sigma_{j,T+1} = \exp(h_{j,T+1}/2)$. All entries are based on $B = 100$ Monte Carlo replications.

T	True DGP	Estimator	Series 1	Series 2	Series 3	Series 4	Series 5	Pooled
Panel A: RMSE_h (latent log-variance path)								
1000	MSV DGP	MSV(APF)	1.0622	0.9951	0.8774	1.0695	0.9775	1.0000
	MSV DGP	FSV(MCMC)	1.7780	1.7062	1.7066	1.7199	1.6428	1.7222
	FSV DGP	MSV(APF)	1.0076	0.8978	1.0668	1.0553	1.0530	1.0217
	FSV DGP	FSV(MCMC)	1.3261	1.2800	1.2517	1.2542	1.2510	1.2750
2000	MSV DGP	MSV(APF)	1.0627	0.9850	0.8702	1.0695	0.9708	0.9950
	MSV DGP	FSV(MCMC)	1.7723	1.7138	1.7207	1.7326	1.6666	1.7300
	FSV DGP	MSV(APF)	0.9927	0.8871	1.0539	1.0792	1.0764	1.0243
	FSV DGP	FSV(MCMC)	1.3554	1.2863	1.2417	1.2518	1.2448	1.2779
3000	MSV DGP	MSV(APF)	1.0502	0.9838	0.8624	1.0687	0.9618	0.9886
	MSV DGP	FSV(MCMC)	1.7201	1.7016	1.7246	1.7000	1.6221	1.6977
	FSV DGP	MSV(APF)	1.0018	0.8869	1.0512	1.0707	1.0684	1.0209
	FSV DGP	FSV(MCMC)	1.3523	1.2923	1.2488	1.2533	1.2518	1.2812
Panel B: RMSE_σ (one-step-ahead volatility forecast)								
1000	MSV DGP	MSV(APF)	3.1895	1.3387	1.2046	2.8227	3.4296	4.5223
	MSV DGP	FSV(MCMC)	3.4569	1.3775	1.5983	2.4697	2.3515	3.7926
	FSV DGP	MSV(APF)	0.0314	0.0277	0.0211	0.0227	0.0260	0.0309
	FSV DGP	FSV(MCMC)	0.0259	0.0209	0.0161	0.0156	0.0204	0.0272
2000	MSV DGP	MSV(APF)	2.6555	1.2784	1.8868	2.8599	2.1084	3.8957
	MSV DGP	FSV(MCMC)	4.2305	1.8034	2.8318	3.3917	2.3666	5.3344
	FSV DGP	MSV(APF)	0.0402	0.0337	0.0217	0.0229	0.0246	0.0364
	FSV DGP	FSV(MCMC)	0.0328	0.0314	0.0205	0.0189	0.0228	0.0364
3000	MSV DGP	MSV(APF)	5.0714	1.6012	1.5249	1.8308	1.2066	3.9462
	MSV DGP	FSV(MCMC)	7.0402	2.4040	2.5404	1.5747	1.2375	5.2462
	FSV DGP	MSV(APF)	0.0294	0.0288	0.0195	0.0213	0.0231	0.0308
	FSV DGP	FSV(MCMC)	0.0300	0.0306	0.0192	0.0199	0.0222	0.0334

K Intraday adjustment of high-frequency returns

This appendix provides details on the intraday adjustment applied to the high-frequency returns in Section 7.1. The adjustment is performed separately for each asset; throughout this appendix, we suppress the asset index. Let P_u denote the continuous-time log-price process. For intraday bin $b = 1, \dots, n^d$ on trading day $\ell = 1, \dots, N^d$, the raw 5-minute return is

$$R_{b,\ell} \equiv \Delta_{b,\ell}^{n^d} P := P_{\ell-1+b/n^d} - P_{\ell-1+(b-1)/n^d}. \quad (\text{K.1})$$

Following [Todorov and Tauchen \(2012\)](#), we represent the spot variance as

$$\tilde{\sigma}_u = \sigma_u d(u - \lfloor u \rfloor), \quad u \in [0, N^d], \quad (\text{K.2})$$

where σ_u denotes the stochastic component of spot variance and $d : [0, 1] \rightarrow (0, \infty)$ is a differentiable deterministic function capturing the systematic within-day volatility pattern. This continuous-time representation is used only to motivate the intraday adjustment.

We estimate the intraday pattern nonparametrically by pooling observations corresponding to the same clock-time bin across trading days. To reduce the influence of jumps and extreme intraday observations, fix a truncation exponent $\varpi \in (0, \frac{1}{2})$; in the empirical implementation, we set $\varpi = 0.49$. For each trading day ℓ , compute the jump-robust bipower variation

$$\text{BV}_\ell = \frac{\pi}{2} \sum_{j=2}^{n^d} \left| \Delta_{j-1,\ell}^{n^d} P \right| \left| \Delta_{j,\ell}^{n^d} P \right|. \quad (\text{K.3})$$

The day-specific truncation level is $\alpha_\ell = 3\sqrt{\text{BV}_\ell}$, and the corresponding threshold is $\tau_\ell = \alpha_\ell (n^d)^{-\varpi}$.

For each intraday bin b , define the pooled truncated second moment as

$$\hat{g}_b = \frac{n^d}{N^d} \sum_{\ell=1}^{N^d} \left| \Delta_{b,\ell}^{n^d} P \right|^2 \mathbf{1}_{\left\{ \left| \Delta_{b,\ell}^{n^d} P \right| \leq \tau_\ell \right\}}, \quad b = 1, \dots, n^d. \quad (\text{K.4})$$

Its intraday average is

$$\hat{g} = \frac{1}{n^d} \sum_{b=1}^{n^d} \hat{g}_b. \quad (\text{K.5})$$

The estimated intraday variance factor for bin b is then

$$\hat{d}_b = \frac{\hat{g}_b}{\hat{g}}, \quad b = 1, \dots, n^d, \quad (\text{K.6})$$

which satisfies the normalization

$$\frac{1}{n^d} \sum_{b=1}^{n^d} \hat{d}_b = 1. \quad (\text{K.7})$$

Finally, the standardized 5-minute return is

$$R_{b,\ell}^{\text{st}} \equiv \frac{R_{b,\ell}}{\sqrt{\hat{d}_b}} = \frac{\Delta_{b,\ell}^{n^d} P}{\sqrt{\hat{d}_b}}, \quad b = 1, \dots, n^d, \quad \ell = 1, \dots, N^d. \quad (\text{K.8})$$

Thus, the adjustment removes the systematic clock-time component of intraday variance while preserving the stochastic volatility dynamics of interest. For each asset, the standardized returns $R_{b,\ell}^{\text{st}}$ are stacked chronologically to form the return series used in the MSV estimation, with $T = n^d N^d$ observations.

L Empirical common-C diagnostics and test results

This appendix reports the empirical screening diagnostics and formal common-C tests for the three sets of return portfolios. Conditional on the unrestricted MSV($\hat{p}_m$) specification selected for each portfolio dimension $m \in \{20, 50, 100\}$, we construct

$$\hat{\mathbf{M}}_{C,m} := \sum_{j=1}^{\hat{p}_m} \hat{\boldsymbol{\phi}}_{j,m}' \hat{\boldsymbol{\phi}}_{j,m}. \quad (\text{L.1})$$

To assess the empirical relevance of a low-dimensional common dynamic subspace, we summarize the eigenvalue concentration of $\hat{\mathbf{M}}_{C,m}$ through the explained dynamic energy

$$\hat{E}_m(r) := \frac{\sum_{k=1}^r \hat{\lambda}_{k,m}}{\sum_{k=1}^m \hat{\lambda}_{k,m}}, \quad r = 1, \dots, 10, \quad (\text{L.2})$$

together with the eigengap ratio

$$\hat{\lambda}_{r,m} / \hat{\lambda}_{r+1,m}. \quad (\text{L.3})$$

Here, $\hat{E}_m(r)$ measures the share of total unrestricted dynamic variation captured by the first r eigendirections of $\hat{\mathbf{M}}_{C,m}$, while the eigengap ratio measures the separation between the r -th and $(r+1)$ -th eigenvalues. Larger values of $\hat{E}_m(r)$ indicate stronger concentration of the unrestricted dynamics in a rank- r approximation, whereas a comparatively large eigengap ratio indicates a sharper break after the first r directions. These diagnostics are descriptive rather than decisive and are used only to assess whether some candidate ranks appear empirically plausible before formal testing.

For each r , we then report the common-C test statistic

$$S_{T,m}^{\text{com}}(r) := T \cdot \max\{0, D_{\text{com},m}(\hat{\boldsymbol{\phi}}; r) - D_{\text{rr},m}(\hat{\boldsymbol{\phi}}; r)\} \quad (\text{L.4})$$

together with the associated LMC and MMC p -values under the null given in (5.4).

Table A12 provides little empirical support for an exact low-dimensional common C-space across the three portfolio dimensions. The screening diagnostics indicate some degree of concentration of the unrestricted dynamics in a few leading directions, especially for $m = 20$, where the explained dynamic energy rises from 0.22 at $r = 1$ to 0.87 at $r = 10$, and for $m = 100$, where a pronounced eigengap appears at $r = 2$. However, these descriptive patterns do not translate into formal evidence in favor of the common-C restriction. For every tested rank $r = 1, \dots, 10$ and every portfolio dimension $m \in \{20, 50, 100\}$, the LMC p -values remain below 0.05, leading to systematic rejection of $H_0^{\text{com}}(r)$. The MMC results are fully consistent with this conclusion, except for a single non-rejection at $(m, r) = (50, 10)$, where the MMC p -value equals 0.096 while the corresponding LMC p -value still leads to rejection. Overall, the table suggests that although the unrestricted volatility dynamics may exhibit moderate low-dimensional concentration, this

concentration is not strong enough to justify an exact common right-subspace representation across lags.

Table A12
Empirical screening and common **C**-space test results across portfolio dimensions

Table A12 reports descriptive screening diagnostics and formal LMC/MMC evidence for the common-**C** restriction

$$H_0^{\text{com}}(r): \exists \mathbf{C} \in \mathbb{R}^{m \times r}, \mathbf{C}'\mathbf{C} = \mathbf{I}_r, \exists \{\mathbf{B}_j\}_{j=1}^p \text{ s.t. } \boldsymbol{\phi}_j = \mathbf{B}_j\mathbf{C}', j = 1, \dots, p,$$

for $m \in \{20, 50, 100\}$ and $r = 1, \dots, 10$. The screening columns $\hat{\lambda}_{r,m}$, $\hat{\lambda}_{r+1,m}$, the eigengap ratio $\hat{\lambda}_{r,m}/\hat{\lambda}_{r+1,m}$, and the explained dynamic energy $\hat{E}_m(r)$ summarize the extent to which the unrestricted volatility dynamics are concentrated in a few dominant directions. In particular, $\hat{E}_m(r)$ measures the share of total unrestricted dynamic variation captured by the first r eigen-directions of $\hat{\mathbf{M}}_{C,m}$, while the eigengap ratio measures the separation between the r -th and $(r+1)$ -th eigenvalues. The remaining columns report the common-**C** test statistic $S_{T,m}^{\text{com}}(r)$ and the associated LMC and MMC p -values. For both Monte Carlo tests, p -values are computed using $N = 999$ replications. The formal decision column is based on the LMC p -value at the nominal level 0.05, whereas the MMC p -values are reported for robustness. The LMC test rejects the common-**C** null for every tested rank. The only MMC non-rejection occurs at $(m, r) = (50, 10)$, which does not alter the overall conclusion that these portfolio dimensions provide no convincing support for an exact low-dimensional common **C**-space.

m	r	$\hat{\lambda}_{r,m}$	$\hat{\lambda}_{r+1,m}$	Eigengap	$\hat{\lambda}_{r,m}/\hat{\lambda}_{r+1,m}$	$\hat{E}_m(r)$	$S_{T,m}^{\text{com}}(r)$	LMC p -value	MMC p -value	Decision
20	1	0.62	0.44		1.401	0.22	426.74	0.001	0.005	Reject
20	2	0.44	0.35		1.274	0.38	681.43	0.001	0.002	Reject
20	3	0.35	0.28		1.229	0.50	789.79	0.001	0.002	Reject
20	4	0.28	0.19		1.457	0.60	825.43	0.001	0.001	Reject
20	5	0.19	0.15		1.286	0.67	797.70	0.001	0.002	Reject
20	6	0.15	0.13		1.172	0.72	807.89	0.001	0.002	Reject
20	7	0.13	0.12		1.089	0.77	770.53	0.002	0.003	Reject
20	8	0.12	0.10		1.215	0.81	730.15	0.003	0.007	Reject
20	9	0.10	0.08		1.144	0.84	681.58	0.009	0.021	Reject
20	10	0.08	0.08		1.056	0.87	627.58	0.012	0.025	Reject
50	1	1.00	0.72		1.394	0.09	4176.71	0.001	0.001	Reject
50	2	0.72	0.69		1.044	0.16	4489.20	0.001	0.001	Reject
50	3	0.69	0.53		1.295	0.22	4505.72	0.001	0.001	Reject
50	4	0.53	0.45		1.191	0.27	4805.17	0.001	0.001	Reject
50	5	0.45	0.37		1.218	0.31	5134.16	0.001	0.001	Reject
50	6	0.37	0.36		1.030	0.35	5544.16	0.001	0.001	Reject
50	7	0.36	0.34		1.047	0.38	5863.35	0.001	0.001	Reject
50	8	0.34	0.33		1.032	0.41	6102.47	0.001	0.001	Reject
50	9	0.33	0.32		1.040	0.44	6281.10	0.001	0.001	Reject
50	10	0.32	0.30		1.043	0.47	6395.32	0.001	0.096	Reject
100	1	2.60	2.24		1.159	0.12	6770.50	0.001	0.001	Reject
100	2	2.24	0.89		2.507	0.23	3435.77	0.001	0.001	Reject
100	3	0.89	0.69		1.303	0.27	3446.70	0.001	0.001	Reject
100	4	0.69	0.58		1.174	0.31	3617.91	0.001	0.001	Reject
100	5	0.58	0.53		1.100	0.33	4000.79	0.001	0.001	Reject
100	6	0.53	0.51		1.040	0.36	4382.37	0.001	0.001	Reject
100	7	0.51	0.50		1.024	0.38	4654.35	0.001	0.001	Reject
100	8	0.50	0.47		1.071	0.41	4900.24	0.001	0.001	Reject
100	9	0.47	0.46		1.020	0.43	5063.82	0.001	0.001	Reject
100	10	0.46	0.42		1.078	0.45	5148.55	0.001	0.001	Reject
Selected ranks		No non-rejected rank under the baseline LMC criterion on the tested grid $r = 1, \dots, 10$ for $m = 20, 50, 100$								

M Similarity of estimated volatilities under MSV and FSV

The simulation comparison in Section 6.2.2 evaluates MSV and FSV through filtering and forecasting accuracy relative to the latent truth. In the empirical application, however, the latent log-volatilities are unobserved, so the relevant issue is whether the two models imply similar fitted volatility paths for the observed assets. To address this question, we compare the model-implied series-specific conditional variances from the unrestricted MSV model and the FSV benchmark for the empirical portfolio dimensions $m \in \{20, 50, 100\}$.

Let $\hat{\Sigma}_t^{\text{MSV}}$ and $\hat{\Sigma}_t^{\text{FSV}}$ denote the fitted conditional covariance matrices under the two models. For each asset $i = 1, \dots, m$, define the fitted log-variance paths by

$$\hat{\ell}_{i,t}^{\text{MSV}} = \log([\text{diag}(\hat{\Sigma}_t^{\text{MSV}})]_i), \quad \hat{\ell}_{i,t}^{\text{FSV}} = \log([\text{diag}(\hat{\Sigma}_t^{\text{FSV}})]_i). \quad (\text{M.1})$$

These quantities are directly comparable across models because they refer to the same asset-specific conditional variances.

We summarize the similarity between the two estimated volatility systems in three ways. First, for each asset i , we compute the time-series correlation $\rho_i = \text{Corr}(\hat{\ell}_{i,t}^{\text{MSV}}, \hat{\ell}_{i,t}^{\text{FSV}})$. Second, after standardizing each fitted path asset by asset, we compute the path-distance measure $d_i = [T^{-1} \sum_{t=1}^T (\tilde{\ell}_{i,t}^{\text{MSV}} - \tilde{\ell}_{i,t}^{\text{FSV}})^2]^{1/2}$, where $\tilde{\ell}_{i,t}^m$ denotes the standardized counterpart of $\hat{\ell}_{i,t}^m$ for $m \in \{\text{MSV}, \text{FSV}\}$. Third, for each date t , we compute the cross-sectional correlation $q_t = \text{Corr}_{i=1, \dots, m}(\hat{\ell}_{i,t}^{\text{MSV}}, \hat{\ell}_{i,t}^{\text{FSV}})$, which captures the extent to which the two models rank assets similarly by conditional volatility at a given date.

Table A13 reports summary statistics for $\{\rho_i\}$, $\{d_i\}$, and $\{q_t\}$ for $m = 20, 50$, and 100 . The results indicate that the similarity between the fitted MSV and FSV volatility paths is generally limited. For $m = 20$ and $m = 50$, the average asset-specific correlations are close to zero and the standardized path distances are large, indicating substantial differences in the dynamic behavior of the fitted asset-level volatilities. For $m = 100$, the average asset-specific correlation increases and the path distance declines somewhat, but the cross-sectional correlation remains centered near zero, implying little systematic agreement in the contemporaneous ranking of assets by volatility. Overall, the empirical comparison suggests that MSV and FSV need not imply similar asset-level volatility dynamics, even when both models are fitted to the same return panel.

Table A13
Similarity of estimated asset-level volatilities under MSV and FSV

Table A13 compares the model-implied series-specific log-variance paths from the unrestricted MSV model and the FSV benchmark for the full-sample empirical applications with $m \in \{20, 50, 100\}$. For each asset i , ρ_i denotes the time-series correlation between the fitted log-variance paths under MSV and FSV. The statistic d_i denotes the root mean squared difference between the two fitted log-variance paths after asset-wise standardization. For each date t , q_t denotes the cross-sectional correlation across the fitted log-variances under the two models. The FSV factor dimension is selected from the data using the [Bai and Ng \(2002\)](#) IC2 criterion applied to the transformed log-square panel, with search range $k \in \{0, \dots, 8\}$. Higher values of ρ_i and q_t , and lower values of d_i , indicate greater similarity between the volatility signals implied by the two models.

Statistic	$m = 20$				$m = 50$				$m = 100$			
	Mean	Median	Min	Max	Mean	Median	Min	Max	Mean	Median	Min	Max
Selected $\hat{k}_{\text{FSV}}$	1				2				2			
Asset-specific time-series correlation ρ_i	0.09	0.09	-0.02	0.18	0.02	0.02	-0.07	0.09	0.30	0.32	0.05	0.50
Standardized path distance d_i	1.32	1.32	1.19	1.44	1.51	1.52	1.42	1.57	1.25	1.24	1.07	1.50
Cross-sectional correlation q_t	0.35	0.38	-0.58	0.90	0.38	0.38	-0.16	0.89	-0.03	-0.07	-0.55	0.81

Appendix References

- Bai, J. and Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.
- 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.
- Brockwell, P. J. and Davis, R. A. (1996), *Time series: theory and methods*, Springer series in statistics, 2nd edn, Springer, New York.
- 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](https://doi.org/10.1093/biomet/81.3.541).
- Davidson, J. (1994), *Stochastic limit theory*, Oxford University Press.
- 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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1016/S0304-4076(02)00108-2).
- Dwass, M. (1957), 'Modified randomization tests for nonparametric hypotheses', *Annals of Mathematical Statistics* **28**, 181–187, doi:[10.1214/aoms/1177707045](https://doi.org/10.1214/aoms/1177707045).
- Ephremidze, L., Janashia, G. and Lagvilava, E. (2009), 'A simple proof of the matrix-valued Fejér–Riesz theorem', *Journal of Fourier Analysis and Applications* **15**(1), 124–127, doi:[10.1007/s00041-008-9051-z](https://doi.org/10.1007/s00041-008-9051-z).
- 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](https://doi.org/10.1111/j.1467-9892.1994.tb00184.x).
- 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](https://doi.org/10.1016/0304-4076(94)90038-8).
- 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](https://doi.org/10.2307/2345178).

- Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**, 1029–1054.
- 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](https://doi.org/10.2307/2297980).
- 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](https://doi.org/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](https://doi.org/10.1111/1467-937X.00050).
- 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](https://doi.org/10.1016/j.jeconom.2006.07.008).
- 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.
- Serfling, R. J. (1980), *Approximation theorems of mathematical statistics*, Wiley Series in Probability and Statistics, 1 edn, Wiley.
- Shephard, N. and Pitt, M. K. (1997), 'Likelihood analysis of non-gaussian measurement time series', *Biometrika* **84**(3), 653–667.
- Todorov, V. and Tauchen, G. (2012), 'Inverse realized laplace transforms for nonparametric volatility density estimation in jump-diffusions', *Journal of the American Statistical Association* **107**(498), 622–635.
- Yu, Y. and Meng, X.-L. (2011), 'To center or not to center: that is not the question—an ancillarity-sufficiency interweaving strategy (Asis) for boosting mcmc efficiency', *Journal of Computational and Graphical Statistics* **20**(3), 531–570, doi:[10.1198/jcgs.2011.203main](https://doi.org/10.1198/jcgs.2011.203main).