AStA Advances in Statistical Analysis (2026) 110:41–63
https://doi.org/10.1007/s10182-025-00536-3

ORIGINAL PAPER

Sequential estimation of multivariate factor stochastic volatility models

Christian Mücher[1,2] · Giorgio Calzolari[3] · Roxana Halbleib[1,2]

Received: 19 April 2024 / Accepted: 12 July 2025 / Published online: 1 August 2025

Abstract

We provide the first "frequentist" method to estimate the parameters of multivariate stochastic volatility models with latent factor structures to capture the time-varying variance–covariance of financial returns. These models alleviate the standard curse of dimensionality, allowing the number of parameters to increase only linearly with the number of series. Although theoretically very appealing, they have only found limited practical application due to huge computational burdens. Our estimation method is simple in implementation as it consists of two steps: first, we estimate the loadings and the unconditional variances by maximum likelihood, and then, we use the efficient method of moments to estimate the parameters of the stochastic volatility structure with the generalised autoregressive conditional heteroskedasticity (GARCH) auxiliary models. In a comprehensive Monte Carlo study, we show the good performance of our method to estimate the parameters of interest accurately. The simulation study and an application to the daily returns on 148 stocks in the cross-sectional dimension provide sound evidence on the computational feasibility of the method proposed and its application.

Keywords Estimation · Efficient method of moments · Multivariate stochastic volatility · Factor models · Curse of dimensionality

✉ Roxana Halbleib
roxana.halbleib@vwl.uni-freiburg.de

Christian Mücher
christian.muecher@vwl.uni-freiburg.de

Giorgio Calzolari
giorgio.calzolari@unifi.it

1 University of Freiburg, Freiburg, Germany

2 Graduate School of Decision Sciences at University of Konstanz, Konstanz, Germany

3 University of Firenze, Florence, Italy

1 Introduction

Modelling and forecasting the multivariate volatility of financial returns is crucial for risk and portfolio management. Most investors hold large, but finite, baskets of financial assets whose risks are time-varying and correlated in time. Dynamic approaches to capture, within a multivariate framework, the unobserved time-varying variation and correlation of the financial returns include the multivariate GARCH models and the multivariate stochastic volatility (MSV) models. While the GARCH specifications are relatively restrictive as they treat the (co-) variance as conditionally deterministic, stochastic volatility models are more flexible as they allow for the variances and the correlations to be stochastic. However, the MSV models are difficult to estimate since the (co-) variances are latent. Moreover, due to the curse of dimensionality, since the number of parameters increases at least quadratically in the dimension of the return vector, MSV models have found so far only limited application.

A solution to the curse of dimensionality of MSV is to impose a factor structure on the vector of underlying returns, where the factors and idiosyncratic errors follow independent autoregressive stochastic volatility (ARSV) processes. This is known as the multivariate factor stochastic volatility (MFSV) model and has been introduced by Harvey et al. (1994) (further developed by Pitt and Shephard (1999b)). The factor representation substantially improves the feasibility of the MSV model in practice by significantly reducing the number of parameters that now increases only linearly with the number of series. However, MFSV still suffers from practical ineffectiveness, even under an exact factor structure, given that the computational burden is enhanced by the latency of the factors and of the idiosyncratic errors, additionally to the one of the variances.

In this paper, we propose a simple method to estimate the parameters of MFSV model of Harvey et al. (1994) that allows it to be easily applied in practice, even to very large, but finite dimensions and at decent computational costs. It consists of two steps: in the first step, we estimate the parameters describing the factor structure, i.e. the loadings and the unconditional variances of the factors and of the idiosyncratic errors, by applying the maximum likelihood (ML) to a static factor representation and by using the parameter convergence results of Anderson and Rubin (1956) and Bai and Li (2012) derived for exact factor models with fixed dimension and arbitrary dynamics. In the second step, we apply the efficient method of moments (EMM) of Bansal et al. (1994) and Gallant and Tauchen (1996) to estimate the ARSV parameters by implementing simple univariate GARCH auxiliary models to each of the extracted static factors and idiosyncratic disturbances from the first step.

This sequential estimation procedure, which has already been successfully applied by Sentana et al. (2008), Aielli et al. (2023), Calzolari and Halbleib (2018) and Calzolari et al. (2021) in other contexts, is computationally feasible and simple in terms of implementation and running time regardless of the dimension of the return vectors. To the best of our knowledge, it is also the first frequentist procedure to estimate MFSV models with exact factor structure, as all existing ones are exclusively Bayesian. Unlike the Bayesian approaches, this paper focuses solely on estimating MFSV models with fixed parameters while for filtration and forecasting of the latent variables of the model we resort to methods already available in the literature. We show that our method can be straightforwardly applied to larger finite dimensions (as well as to more factors) compared to the existing literature without imposing any constraints on the parameters and at very low computational costs.

The existing Bayesian procedures, which simultaneously estimate the parameters and filter the latent variables of MFSV, build on the Markov chain Monte Carlo (MCMC) method as first proposed by Pitt and Shephard (1999b). Although theoretically appealing, this method becomes cumbersome when increasing the dimension of the return vector, as it samples from the entire posterior distribution of the model at each iteration. Chib et al. (2006) partially fix this problem and show that conditionally on the factor loadings and the factors, the problem reduces to sampling from univariate ARSV processes, which significantly decreases the computational burden of the sampling algorithm and allows for efficient sampling from the posterior distribution also for higher-dimensional vectors of returns. Thus, Han (2005) and Nardari and Scruggs (2007) apply the sampling algorithm of Chib et al. (2006) to vectors of around 30 returns and 3 factors and to vectors of 10 returns and 5 factors, respectively. Kastner et al. (2017) improves further the efficiency of the MCMC algorithm proposed by Chib et al. (2006), making it computationally faster, but it still remains burdensome in terms of memory requirements, as it involves storing draws from a very high-dimensional joint density function of all parameters and of all latent variables at each time observation. Thus, the procedure of Kastner et al. (2017) becomes computationally costly for large vector dimensions, many factors and many observations. For this reason, the algorithm is applied to only up to 26 series. The authors do not exploit the potential gains from parallelising their algorithm, since it becomes computationally more intensive in terms of memory requirements. Kastner (2019) makes use of the Bayesian shrinkage to constrain the dimension of the loading parameters in order to apply the procedure to larger dimensions, such as 300 returns.

The multivariate factor stochastic volatility model treated in this paper or in the Bayesian papers mentioned above is a more "traditional" one that models the time-varying variance–covariance matrix of the financial returns specific to portfolios of finite number of assets. The asymptotic results we use in the first estimation step are, thus, for finite number of underlying series N, when the length of time series T tends to infinity. This is in contrast to a more "modern" literature on factor (volatility) models, where both N and T tend to infinity and, thus, motivate the introduction of "approximate" factor models discussed at length most recently by Barigozzi (2023). This literature is very general as it also allows for weakly (cross-) correlated idiosyncratic disturbances (Barigozzi and Hallin 2016, 2017, 2020); however, the way it deals with the volatility specification is structurally different from the approach we deal with in this paper: we work with latent and not approximated volatilities and the factors and the idiosyncratic errors exhibit no autocorrelation, and thus no conditional mean functions, as they inherit the no autocorrelation behaviour of daily financial returns. However, Barigozzi (2023) addresses some serious challenges in using ML estimation when N is fixed in what regards (1) the inference of the loadings being computationally challenging and (2) the factors being not consistently estimated. While our first step auxiliary estimation solves the first problem, for the implementation of our proposed second step, one requires consistent estimates of the parameters in the first step and not of the factors.

We show both in our simulation exercise and in the empirical application that the frequentist sequential method we propose in this paper can estimate the parameters of the MFSV model for very large dimensions of the return vector within minimal computational time and with no memory constraints, especially when parallelising the equation-by-equation estimation of the second step of the procedure. In the simulation exercise, we show that, even without parallelisation, we can obtain accurate estimates for vectors of returns of dimension 10 and 2 factors in less than 13 s if feeding starting values close to the true parameter values and in less than 82 s if the starting values are random. Moreover, we identify the optimal number of simulations in the EMM estimation according to the length of the series and recommend a simple way to choose the starting values that increases the efficiency of the estimates with less computational costs.

In the empirical application, we provide further evidence on the feasibility and speed of our procedure when estimating MFSV for a vector of 148 stock returns and up to 3 factors. Here, we show that, as expected, the computational advantage of our method becomes more pronounced once we use multiple CPU cores to run the second step of the estimation in parallel. With parallelisation enabled, the computational advantage in the real data application increases by a factor larger than 200. While in simulations our method converges in seconds, in the empirical application it may take only up to a few minutes. By implementing some standard particle filtering approaches, we show that the MFSV model outperforms the static model and a comparable factor multivariate GARCH (MGARCH) model when predicting the variance–covariance of future returns one-step and multi-step ahead.

The remainder of this paper is structured as follows. Section 2 formally introduces the MFSV model. In Sect. 3, we present the estimation method proposed in this paper. Section 4 presents the results from a comprehensive simulation study, and in Sect. 5, we present the results from applying our procedure to real data. Section 6 concludes.

2 Multivariate factor stochastic volatility model

Harvey et al. (1994) and Pitt and Shephard (1999b) introduce the factor structure in the context of MSV modelling to capture the common dynamics and the conditional heteroskedasticity of a vector of financial returns of finite dimension as well as to reduce the curse of dimensionality of standard MSV models. This specific type of model is defined for $\boldsymbol{y}_t$, a vector[1] of finite number of return series observed at time t, namely N, such that[2]

$$\boldsymbol{y}_t = B\boldsymbol{f}_t + \boldsymbol{\epsilon}_t \tag{2.1}$$

where $\boldsymbol{f}_t$ is a vector of $k \times 1$ latent factors, $\boldsymbol{f}_t = (f_{t,1}, \dots, f_{t,k})'$ independent of each other, B is of dimension $N \times k$ and contains the factor loadings with $k \leq N$ and rank $(B) = k$, and $\boldsymbol{\epsilon}_t = (\epsilon_{t,1}, \dots, \epsilon_{t,N})'$ is a $N \times 1$ vector of idiosyncratic components that are orthogonal to the factors and orthogonal to each other. Thus, the model presented in Eq. (2.1) is an exact factor model (Watson and Engle 1983; Chamberlain and Rothschild 1983). The model further specifies that each factor and idiosyncratic error follows a univariate ARSV process. For this, we define here $\boldsymbol{x}_t \equiv (f_{t,1}, \dots, f_{t,k}, \epsilon_{t,1}, \dots, \epsilon_{t,N})'$ be the vector of dimension $k + N$ of all factors and idiosyncratic disturbances at time t. Thus, the MFSV model assumes that:

$$x_{t,m} = \sqrt{\sigma^2_{t,m}} u_{t,m} \quad u_{t,m} \overset{i.i.d}{\sim} \mathcal{N}(0,\ 1) \tag{2.2}$$

$$\ln(\sigma^2_{t,m}) = \mu_m + \varphi_m[\ln(\sigma^2_{t-1,m}) - \mu_m] + \sigma_{\eta,m}\eta_{t,m} \quad \eta_{t,m} \overset{i.i.d}{\sim} \mathcal{N}(0, 1), \tag{2.3}$$

where $\eta_{t,m}$ and $u_{t,m}$ are independent of each other and $|\varphi_m| < 1$ that assures covariance stationarity of the process in (2.3) for all $m = 1, \dots, k + N$. The unconditional variances of each of the factors, $\sigma^2_j \equiv \mathbb{V}[f_{j,t}]$ with $j = 1, \dots, k$ and of the idiosyncratic errors, $\sigma^2_{k+i} \equiv \mathbb{V}[\epsilon_{i,t}]$, with $i = 1, \dots, N$ can be computed from Eqs. (2.2) and (2.3) as follows:

$$\sigma^2_m = \exp\left[\mu_m + \frac{\sigma^2_{\eta,m}}{2(1 - \varphi^2_m)}\right] \tag{2.4}$$

for each $m = 1, \dots, k + N$.

To ease the notation, we define for each $m = 1, \dots, k + N$, $h_{t,m} \equiv \ln(\sigma^2_{t,m})$ and $\boldsymbol{h}_t \equiv (h_{t,1}, \dots, h_{t,k+N})'$ be the vector of log variances of the factors and of the idiosyncratic errors, respectively. Thus, $\boldsymbol{f}_t$ and $\boldsymbol{\epsilon}_t$ in Eq. (2.1) follow a conditional joint normal distribution as follows:

$$\begin{pmatrix} \boldsymbol{f}_t \\ \boldsymbol{\epsilon}_t \end{pmatrix} | \boldsymbol{h}_t \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{0}_{k\times1} \\ \boldsymbol{0}_{N\times1} \end{pmatrix}, \begin{pmatrix} \Sigma_{1,t} & 0_{k\times N} \\ 0_{N\times k} & \Sigma_{2,t} \end{pmatrix}\right], \tag{2.5}$$

[1] Throughout the paper, we use uppercase characters to denote matrices, bold face lowercase characters to denote vectors and plain lowercase characters to denote scalars. The only exceptions are: N that indicates cross-sectional dimension, T that indicates the time dimension and H H that is the multiple of the sample length specific to the simulation-based estimation.

[2] Similar to, e.g. Harvey et al. (1994), Yu and Meyer (2006) and Kastner et al. (2017), we restrict the conditional and unconditional mean of the return vector to zero. However, one could easily add a vector of intercepts in Eq. (2.1) to account for nonzero means of the returns or one could demean the series, which we do in the empirical application.

where $\Sigma_{1,t}$ is the diagonal covariance matrix of the k factors with the elements $\sigma^2_{t,1}, \dots, \sigma^2_{t,k}$ on the diagonal and $\Sigma_{2,t}$ is the diagonal covariance matrix of the idiosyncratic errors with the elements $\sigma^2_{t,k+1}, \dots, \sigma^2_{t,k+N}$ on the diagonal. The conditional covariance matrix of the return series at time t can be written as follows:

$$\mathbb{V}[\boldsymbol{y}_t|\boldsymbol{h}_t] = B\Sigma_{1,t}B' + \Sigma_{2,t}. \tag{2.6}$$

For the identification of the parameters of the MFSV model, we choose to restrict B such that $b_{jj} = 1$ and $b_{ij} = 0\,\forall j > i, \quad j = 1, \dots, k$ and $i = 1, \dots, N$, i.e. B is lower triangular, that fixes the rotation indeterminacy (for the same type of identifications see Pitt and Shephard 1999b; Han 2005; Chib et al. 2006; Nardari and Scruggs 2007 among others). Aguilar and West (2000) refer to these set of assumptions as hierarchical structural constraints.[3] In this context, it is realistic to let N, the dimension of the portfolio, to be given (usually chosen by the investor) and, thus, fixed. Thus, the loading parameters left for estimation are b_{ij} with $j = 1, \dots, k$ and $i = j + 1, \dots, N$ and the MFSV model has a total of $Nk - k(k + 1)/2 + 3(k + N)$ parameters. This number of parameters makes the MFSV model parsimonious, as it increases linearly in the number of return series N rather than quadratically, as is the case of the MGARCH and general MSV models. However, the estimation of the parameters of the MFSV is difficult since the factors $(f_{t,1}, \dots, f_{t,k})'$, the errors $(\epsilon_{t,1}, \dots, \epsilon_{t,N})'$ and their stochastic variances $(\sigma^2_{t,1}, \dots \sigma^2_{t,k+N})$ are latent.

The MFSV model specified above can be extended to account for leverage effects in various ways: see Scharth and Li (2022). One practical way for which our estimation procedure below can easily be adapted to is the one of Ishihara and Omori (2017) that allows that $u_{t,m}$ and $\eta_{t,m}$, for $m = 1, \dots, k + N$ are (lagged-) correlated with a different correlation coefficient. Thus, one would have further $k + N$ correlation parameters to estimate. For the estimation, one would replace the GARCH structure of the second estimation step we propose below with a threshold GARCH or exponential GARCH structure. Although very interesting per se, we leave the frequentist estimation of such asymmetric MFSV models for future research.

3 Estimation

In what follows, we present our simple frequentist approach to estimate the parameters of the MFSV model based on two steps.[4] For this reason, we rewrite the vector of the model parameters that need to be estimated as $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', \boldsymbol{\theta}_2'\right)'$, where

[3] Sentana and Fiorentini (2001) show that assuming heteroskedasticity for the factors, one may relax one of these restrictions. However, given that our estimation strategy presented in the next section relies on estimating a static factor model, we need both restrictions on the factor loadings matrix, paired with the restrictions that $\Sigma_{1,t}$ and $\Sigma_{2,t}$ are diagonal (also implying that their unconditional counterparts, $\mathbb{V}[\boldsymbol{f}_t]$, and $\mathbb{V}[\boldsymbol{\epsilon}_t]$ are diagonal) in order to achieve full identification in the static factor model (Bai and Li 2012; Cox 2017; Williams 2020).

[4] Ruiz (1994) and Harvey et al. (1994) propose a "frequentist" QML solution to estimate univariate and multivariate SV, however, without the latent factor structure.

$$\boldsymbol{\theta}_1 = \left(b_{21}, \ldots, b_{N1}, b_{32}, \ldots, b_{N2}, \ldots b_{Nk}, \sigma_1^2, \ldots, \sigma_{k+N}^2\right)' \tag{3.1}$$

is estimated in the first step (to be described in Sect. 3.1) and

$$\boldsymbol{\theta}_2 = \left(\varphi_1, \ldots, \varphi_{k+N}, \sigma_{\eta,1}, \ldots, \sigma_{\eta,k+N}\right)' \tag{3.2}$$

is estimated in the second step (to be described in Sect. 3.2).[5]

3.1 Estimation of the factor model parameters by ML

This section provides the estimates of $\boldsymbol{\theta}_1$. For this, we define a static factor model as follows:

$$\boldsymbol{y}_t = B^* \boldsymbol{g}_t + \boldsymbol{e}_t, \tag{3.3}$$

$$\boldsymbol{g}_t \overset{i.i.d}{\sim} \mathcal{N}\left(\mathbf{0}_{k\times 1}, \Sigma_1^*\right), \tag{3.4}$$

$$\boldsymbol{e}_t \overset{i.i.d}{\sim} \mathcal{N}\left(\mathbf{0}_{N\times 1}, \Sigma_2^*\right), \tag{3.5}$$

with the same assumptions and restrictions on $B^*, \boldsymbol{g}_t, \Sigma_1^*, \Sigma_2^*$ as in the MFSV model presented in Sect. 2, which ensure the full identification of the parameters (see also Williams 2020). Denote the parameter vector of this static factor model by:

$$\boldsymbol{\beta}_1 = (b_{21}^*, \ldots, b_{Nk}^*, \sigma_1^2, \ldots, \sigma_{k+N}^2)', \tag{3.6}$$

i.e. $\boldsymbol{\beta}_1$ gives the factor loadings, the unconditional variances of the factors and the unconditional variances of the idiosyncratic errors of the static factor model defined in Eqs. (3.3)–(3.5). One should note that, as in the true model, N is fixed. Bai and Li (2012) show that, in the exact factor model case, for $T \to \infty$, the (Q)ML estimator of $\boldsymbol{\beta}_1$ from a static factor model, namely $\hat{\boldsymbol{\beta}}_1$,

$$\hat{\boldsymbol{\beta}}_1 \xrightarrow{p} \boldsymbol{\theta}_1, \tag{3.7}$$

i.e. the static factor ML estimators converge to their true parameter counterparts in the dynamic factor model. This result holds if N is fixed, the factors are independent of the error terms, the factor variances are bounded away from zero and the distribution of $(\boldsymbol{f}_t', \boldsymbol{\epsilon}_t')'$ has finite second and fourth moments, as well as finite and summable fourth-order cumulants (Anderson and Rubin 1956; Bai and Li 2012; Anderson 2003), which are the cases for our factor structure. This result simplifies our estimation procedure drastically, especially when applied to vectors of returns and factors of large dimensions.

However, as shown by Bien and Tibshirani (2011) and Bai and Li (2012), the log-likelihood of the static factor model is multi-modal. Thus, classical numerical

[5] Note that $\boldsymbol{\theta}_2$ does not contain the ARSV constants $(\mu_1, \ldots, \mu_{k+N})'$, since we can identify them from the unconditional variances of the ARSV processes described by Eq. (2.4).

optimisation methods are infeasible for higher-dimensional return vectors $\boldsymbol{y}_t$ as they are likely to get stuck at local optima. In this paper, we use the expectation maximisation (EM) algorithm proposed by Bai and Li (2012) to estimate the parameters of the static factor model combined with a modified version of the gradient descent algorithm of Bien and Tibshirani (2011), which considerably reduces the computational time (see also Bai and Liao 2016 and Daniele et al. 2020 for a successful implementation of this combination to estimate other types of factor models). We choose the initial values within this first step estimation by the principal component analysis (PCA). A detailed description of the estimation procedure at this step is presented in Appendix A.

After obtaining $\widehat{B^*}$, $\widehat{\Sigma_1^*}$ and $\widehat{\Sigma_2^*}$, we make use of the asymptotic results of Anderson and Rubin (1956) and Bai and Li (2012) and take them as consistent estimates B, Σ_1 and Σ_2, i.e. of $\boldsymbol{\theta}_1$, by simply setting: $\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\beta}}_1$.

3.2 Estimation of ARSV parameters by EMM

Based on the consistent estimates of $\widehat{B^*}$, $\widehat{\Sigma_1^*}$ and $\widehat{\Sigma_2^*}$ obtained in the first step described in Sect. 3.1, we extract the static factors by the projection formula of Thomson (1954), Lawley and Maxwell (1962)[6]:

$$\hat{\boldsymbol{g}}_t = \underbrace{\left(\widehat{\Sigma_1^*}^{-1} + \widehat{B^*}'\widehat{\Sigma_2^*}^{-1}\widehat{B^*}\right)^{-1}\widehat{B^*}'\widehat{\Sigma_2^*}^{-1}}_{\Pi^*}(\boldsymbol{y}_t - \bar{\boldsymbol{y}}), \tag{3.8}$$

where $\bar{\boldsymbol{y}}$ is the mean of the return series over the whole sample. Define:

$$\hat{\boldsymbol{e}}_t = \boldsymbol{y}_t - \widehat{B^*}\hat{\boldsymbol{g}}_t, \tag{3.9}$$

and let $\hat{\boldsymbol{x}}_t = (\hat{g}_{t,1}, \dots, \hat{g}_{t,k}, \hat{e}_{t,1}, \dots, \hat{e}_{t,N})'$.

At this second step, we estimate the elements of $\boldsymbol{\theta}_2$ by applying the EMM approach of Bansal et al. (1994) and Gallant and Tauchen (1996). Given the independence among and between the factors and the idiosyncratic errors of the factor structure defined in Sect. 2, the EMM approach can be simplified by estimating the parameters of the ARSV models equation-by-equation by applying the appropriate univariate auxiliary model to each component of $\hat{\boldsymbol{x}}_t$. We choose here to apply GARCH(1,1) as the auxiliary model, as it has already been successfully implemented for the EMM estimation of univariate ARSV models by Calzolari et al. (2004) and Monfardini (1998), among others.[7]

[6] One should notice that in this second step of the estimation one does not need the factors to be consistently estimated, which in the factor literature requires that N goes to infinity (Barigozzi 2023), but only the estimates of the parameters.

[7] As an alternative to GARCH, we also implement the autoregressive moving average model with autoregressive and moving average order both equal to one (ARMA(1,1) model) fitted to the log squared transformation of the elements of $\hat{\boldsymbol{x}}_t$ (Monfardini 1998). However, due to the "flatness" in the ARMA(1,1) likelihood for at least one of the elements of $\hat{\boldsymbol{x}}_t$, we were not able to get reliable estimates

In Appendix B, we provide the detailed description of the steps taken to undergo the EMM estimation of $\boldsymbol{\theta}_2$. For this, we define below the structure of the GARCH(1,1) that we apply to each of the series composing $\hat{\boldsymbol{x}}_t$ as follows:

$$\hat{x}_{t,m} = \delta_{t,m}\xi_{t,m}, \tag{3.10}$$

$$\delta^2_{t,m} = (1 - \alpha_{1,m} - \alpha_{2,m})\hat{\sigma}^2_m + \alpha_{1,m}\hat{x}^2_{t-1,m} + \alpha_{2,m}\delta^2_{t-1,m}, \tag{3.11}$$

where $m = 1, \ldots, k, k+1, \ldots, k+N$, $\xi_{t,1}, \ldots, \xi_{t,m}$ are independent white noise processes, $\alpha_{1,m} > 0$, $\alpha_{2,m} > 0$ and $\alpha_{1,m} + \alpha_{2,m} < 1$ $\forall m$ and $\hat{\sigma}^2_m$ is the estimate of m-th unconditional variance contained in $\boldsymbol{\theta}_1$ and estimated in the first step. Here, we follow Sentana and Fiorentini (2001) and Sentana et al. (2008) and identify the GARCH constant through a consistent estimator of the unconditional variance of the respective process. We define $\boldsymbol{\beta}_2 = (\alpha_{1,1}, \ldots, \alpha_{1,k+N}, \alpha_{2,1}, \ldots, , \alpha_{2,k+N})'$ containing all parameters for the $k+N$ univariate GARCH(1,1) models described above that are estimated equation-by-equation by means of quasi-maximum likelihood (QML). To speed up the estimation at this step, we can parallelise the estimation of $\boldsymbol{\theta}_2 = \left(\varphi_1, \ldots, \varphi_{k+N}, \sigma_{\eta,1}, \ldots, \sigma_{\eta,k+N}\right)'$ in $k+N$ EMM estimation algorithms for each of the parameter pairs $\boldsymbol{\theta}_{2,m} = (\varphi_m, \sigma_{\eta,m})'$ with $m = 1, \ldots, k+N$ to which corresponds the auxiliary parameter vector $\boldsymbol{\beta}_{2,m} = (\alpha_{1,m}, \alpha_{2,m})'$. The EMM estimation provides consistent estimators of $\boldsymbol{\theta}_2$ under very general conditions that are fulfilled by our model specification.

Below we provide a short compact version of the two-step estimation algorithm we propose in this paper:

Footnote 7 (Continued)

and standard errors of the MFSV parameters based on this choice of the auxiliary model. The results are, however, available from the authors upon request.

Steps to implement the estimation of MFSV

1 Get the vector $\hat{\boldsymbol{\theta}}_1$ by the EM estimator $\hat{\boldsymbol{\beta}}_1$ of the auxiliary static factor model defined in (3.3) - (3.5).

2 Based on $\hat{\boldsymbol{\theta}}_1$, extract static factors and corresponding idiosyncratic errors, $\hat{\boldsymbol{x}}_t = (\hat{g}_{t,1}, \ldots, \hat{g}_{t,k}, \hat{e}_{t,1}, \ldots, \hat{e}_{t,N})'$ based on equations (3.8) - (3.9).

3 Choose a vector of initial values $\boldsymbol{\theta}_2$, i.e., $\boldsymbol{\theta}_2^0$.

4 Run the following steps for each $m = 1, k+N$. They can be parallelised over the $k+N$ series:

4.1 Get the PML estimates $\hat{\boldsymbol{\beta}}_{2,m}$ of the auxiliary GARCH(1,1) model applied to the m-th element of $\hat{\boldsymbol{x}}_t$ extracted at step 2.

4.2 Compute the auxiliary model score at the m-th element of $\hat{\boldsymbol{x}}_t$ extracted at step 2 and $\hat{\boldsymbol{\beta}}_{2,m}$.

4.3 Run the following steps until the objective function in Equation (B.6) is minimised:

4.3.1 Simulate $T \times H$ observations $\tilde{\boldsymbol{y}}_t(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2^0)$ from the MFSV model given in equations (2.1) -(2.3) and (2.5) by using $\hat{\boldsymbol{\theta}}_1$ obtained at step 1 and $\boldsymbol{\theta}_2^0$ set at step 3.

4.3.2 Compute $\tilde{\boldsymbol{x}}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2^0)$ from equations (3.8) and (3.9), by replacing $\boldsymbol{y}_t$ with its simulated counterpart $\tilde{\boldsymbol{y}}_t(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2^0)$.

4.3.3 Evaluate the auxiliary model score at the m-th element of $\tilde{\boldsymbol{x}}(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2^0)$ extracted at step 4.3.2 and at $\hat{\boldsymbol{\beta}}_{2,m}$.

4.3.4 Update the m-th element of $\boldsymbol{\theta}_2^0$ chosen at step 3, i.e., $\boldsymbol{\theta}_{2,m}^0$ to $\boldsymbol{\theta}_{2,m}^{1m}$, by keeping the rest of the $m-1$ elements of $\boldsymbol{\theta}_2^0$ at the initial values set at step 3. The last updated $\boldsymbol{\theta}_2^{j_m,0}$ that minimises (B.6) is the EMM estimator of $\boldsymbol{\theta}_{2,m}$, i.e., $\hat{\boldsymbol{\theta}}_{2,m}$.

3.3 Standard errors

Both parts of the estimation procedure we present above involve the estimation of an auxiliary model: i.e. in the first part, the static factor model, in the second part, univariate GARCH models on the extracted static factors from the first part. While the estimates of the auxiliary model of the first part $\hat{\boldsymbol{\beta}}_1$ are taken to be the final corresponding estimates (of loadings and unconditional variances) of the "true" dynamic model, the auxiliary estimates of the second part $\hat{\boldsymbol{\beta}}_2$ help at computing the EMM estimates of $\boldsymbol{\theta}_2$. In order to compute the standard errors of the estimators of the

MFSV parameters, we implement the asymptotic variance–covariance matrix of the EMM estimator given by (Gouriéroux et al. 1993):

$$W(H) = \left(1 + \frac{1}{H}\right)\left[\frac{\partial \mathcal{Q}\big(\tilde{\boldsymbol{y}}_t(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\beta}}\big)'}{\partial \boldsymbol{\theta}} \hat{\mathcal{I}}(\hat{\boldsymbol{\beta}})^{-1} \frac{\partial \mathcal{Q}\big(\tilde{\boldsymbol{y}}_t(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\beta}}\big)}{\partial \boldsymbol{\theta}'}\right]^{-1}, \tag{3.12}$$

where H denotes the number of the simulated series in the EMM, $\hat{\mathcal{I}}(\hat{\boldsymbol{\beta}})$ the Fisher information matrix of the auxiliary model and $\mathcal{Q}\big(\tilde{\boldsymbol{y}}_t(\hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\beta}}\big)$ is obtained by stacking vertically together the auxiliary score vectors $\mathcal{Q}(\boldsymbol{y}_t;\hat{\boldsymbol{\beta}}_1)$ and $\mathcal{Q}(\hat{\boldsymbol{x}}_t;\hat{\boldsymbol{\beta}}_2)$, i.e. $\mathcal{Q}(\boldsymbol{y}_t;\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = (\mathcal{Q}(\boldsymbol{y}_t;\hat{\boldsymbol{\beta}}_1)', \mathcal{Q}(\hat{\boldsymbol{x}}_t;\hat{\boldsymbol{\beta}}_2)')'$, where $\mathcal{Q}(\boldsymbol{y}_t;\hat{\boldsymbol{\beta}}_1)$ are obtained by computing the first derivative of the log-likelihood of the static factor model with respect to $\boldsymbol{\beta}_1$ at $\hat{\boldsymbol{\beta}}_1$. The closed-form expression of this score vector is presented in Appendix C. $\mathcal{Q}(\hat{\boldsymbol{x}}_t;\hat{\boldsymbol{\beta}}_2)$ stacks together the $k+N$ GARCH(1,1) scores vectors: $\mathcal{Q}(\hat{\boldsymbol{x}}_t;\hat{\boldsymbol{\beta}}_2) = (\mathcal{Q}_1(\hat{\boldsymbol{x}}_1;\hat{\boldsymbol{\beta}}_{2,m})', \dots, \mathcal{Q}_{k+N}(\hat{\boldsymbol{x}}_{k+N};\hat{\boldsymbol{\beta}}_{2,k+N})')'$ defined in Appendix B.[8] We define $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2')'$.

In Eq. (3.12), the Fisher information matrix of the auxiliary model, $\hat{\mathcal{I}}(\hat{\boldsymbol{\beta}})$, is not available in exact complete form, as it would be if we could estimate all parameters simultaneously, in which case the complete outer product and/or Hessian matrix would be available. Our approach estimates the parameters $\boldsymbol{\beta}$ in two steps. Here, we follow Calzolari et al. (2021) and replace it with a consistent simulation-based estimator: the sample variance–covariance matrix of 1000 independently simulated score vectors of the overall auxiliary model. These score vectors are computed after the last iteration upon convergence. As all components of the auxiliary model score vector are available in closed form, the simulation-based estimation of the Fisher information matrix and of the whole variance–covariance matrix $W(H)$ in Eq. (3.12) are computationally very fast.

As one may see from Eq. (3.12), $W(H)$ decreases with H. As there is no rule on how to choose H, in our empirical application we follow the main stream of the literature and choose H in such a way that the precision of the estimates improves, but not at very high computational costs (usually H is set to 10). However, in our simulation exercise, we find also an optimal choice of H adapted to the number of observations available for the empirical study, T.

In order to get the standard errors of the constant terms of the ARSV processes $\hat{\mu}_1, \dots, \hat{\mu}_{k+N}$ that are computed from the estimates of first step estimation, we use the delta method as described below. From the unconditional variances specified in Sect. 2, for each $m = 1, \dots, k+N$ we get:

$$\mu_m = \ln(\sigma_m^2) - \frac{\sigma_{\eta,m}^2}{2(1-\varphi_m^2)} \tag{3.13}$$

and denote $\boldsymbol{\xi}_m = (\sigma_m^2, \varphi_m, \sigma_{\eta,m})'$. The derivative of μ_m with respect to the parameter vector $\boldsymbol{\xi}_m$ evaluated at $\hat{\sigma}_m^2, \hat{\varphi}_m, \hat{\sigma}_{\eta,m}^2$ is given by:

[8] Note that $\mathcal{Q}(\boldsymbol{y}_t;\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2)$ does not contain $\hat{\boldsymbol{x}}_t$, since it is a function of $\boldsymbol{y}_t$ and $\hat{\boldsymbol{\beta}}_1$

$$\frac{\partial \mu_m}{\partial \xi_m} = \left(\frac{1}{\hat{\sigma}_m^2}, -\frac{\hat{\varphi}_m \hat{\sigma}^2_{\eta,m}}{(1-\hat{\varphi}_m^2)^2}, -\frac{\hat{\sigma}_{\eta,m}}{(1-\hat{\varphi}_m^2)} \right)'. \tag{3.14}$$

The variance of $\hat{\mu}_m$ is then given by:

$$\mathbb{V}[\hat{\mu}_m] = \frac{\partial \mu_m}{\partial \xi_m'} \mathbb{V}[\hat{\xi}_m] \frac{\partial \mu_m}{\partial \xi_m}, \tag{3.15}$$

where $\mathbb{V}[\hat{\xi}_m]$ is extracted correspondingly from $W(H)$ defined above.

4 Monte Carlo simulation

This section provides results on the statistical properties of the estimates obtained with the procedure described in the previous section for various choices of T, N and k that are empirically relevant. We use $N = \{10, 20, 30, 100\}$ simulated return series,[9] $k = \{1, 2, 3\}$ simulated factors to generate $T = \{1000, 4000, 10000\}$ observations. We choose the parameter values similarly to Kastner et al. (2017):

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0.9 & 1 & 0 \\ \vdots & 0.2 & 1 \\ & \vdots & 0.4+\Delta \\ & & \vdots \\ & & 0.7 \\ & & 0.1 \\ \vdots & \vdots & \vdots \\ 0.1 & 0.8 & 0.4 \end{pmatrix}, \quad \begin{aligned} (\varphi_1, \dots, \varphi_N)' &= (0.9, \dots, 0.99)', \\ (\varphi_{N+1}, \varphi_{N+2}, \varphi_{N+3})' &= (0.99, 0.95, 0.91)', \\ (\mu_1, \dots, \mu_N)' &= (-2, \dots, -1.1)', \\ (\mu_{N+1}, \mu_{N+2}, \mu_{N+3})' &= (0, 0, 0)', \\ (\sigma_{\eta,1}, \sigma_{\eta,2}, \sigma_{\eta,3})' &= (0.2, 0.3, 0.4)', \\ (\sigma_{\eta,4}, \dots, \sigma_{\eta,3+N})' &= (0.6, \dots, 0.15)', \end{aligned}$$

where Δ depends on the cross section dimension N and ensures that the fourth element in the third column of B differs from the last element in that column.[10] The dots indicate an evenly spaced grid between the first and the last value. Therefore, we choose the elements of the first column of B such that they are evenly spaced between the values 0.9 and 0.1; the elements of the second column are chosen such that they are evenly spaced between 0.2 and 0.8; the elements of the third column are chosen to be evenly spaced between 0.1 and 0.7. For the simulation cases with $k < 3$, we choose the first k columns of B and the first $k + N$ values from each of the parameter vectors $\boldsymbol{\varphi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_\eta$. For all simulations, we have $R = 1000$ replications.

[9] Following the request of one of the anonymous reviewers, we also undergo a small simulation exercise for $N = 148$, which is the dimension of the series we consider in the empirical application presented in the next section. Given that the results follow the same pattern we present below and are basically equal to the ones of $N = 100$, we refrain from including them in the paper. They can be obtained from the authors upon request.

[10] Readers interested in replicating the Monte Carlo setting should note that we set Δ to $\Delta = (0.7 - 0.1)/(N - 4)$. Thus, for example, for $N = 10$, $\Delta = 0.1$ and for $N = 100$, $\Delta = 0.00625$.

For the step size of the gradient descent method of Bien and Tibshirani (2011), we use the adaptive moments algorithm with $d = 0.005$.[11]

4.1 Choose starting values and *H*

We compare two strategies for selecting the starting values for the parameters of the EMM procedure of the second step. The first strategy is to provide user-specified starting values, where we choose them to be 20% smaller than the true ARSV autoregressive parameters and 20% larger than the ARSV standard deviation parameters. The second strategy is to obtain starting values from the QML approach of Ruiz (1994) applied to the extracted static factors and residuals, i.e. we apply the Kalman filter to their log squared transformations. This second strategy is especially appealing for empirical applications, where the choice of good starting values is not straightforward.

Next, we analyse the effect that H, the number of simulation paths in the EMM, has on the estimation results. We first set $H = 10$ as in most applications presented in the literature (for instance, among many other, by Calzolari et al. (2021)). However, we find that, for small T, increasing H to 100 improves the performance of our procedure, whereas, for large T, the improvement in the performance is negligible while the computational time increases drastically.[12] To balance this behaviour, we follow the idea of Monfardini (1998) and choose, besides fixing it to the value of 10, H as a function of the sample size, T. Specifically, we set H such that $H \cdot T = 10^5$, i.e. $H = 10^5/T$. This is mainly motivated by the reduction in the number of outliers of the EMM estimation procedure, as described below.

4.2 Loss function

We summarise the results of the simulation study by means of the mean squared error (MSE) of the estimated parameters $\hat{\boldsymbol{\theta}}^r$ over the $r = 1, \dots, R$ MC replications. We compute the MSE of $\hat{\boldsymbol{\theta}}$ by

$$MSE(\hat{\boldsymbol{\theta}}) = \frac{1}{(\dim \boldsymbol{\theta})R} \sum_{r=1}^{R} \sum_{i=1}^{\dim \boldsymbol{\theta}} \left(\hat{\theta}_i^r - \theta_i\right)^2, \tag{4.1}$$

i.e. the average squared deviation of the estimated values from the true values, where $\dim \boldsymbol{\theta} = Nk - k(k+1)/2 + 3 * (k + N)$. Analogous MSE values are also computed for subsets of θ, (e.g. the factor loadings, the SV parameters, etc.). We also report results on the ratio between the empirical standard deviations and the asymptotic standard errors of the estimated parameters.

[11] The estimation in this section is conducted on the bwHPC Cluster with multiple jobs, each using 40 CPU cores (two Octa-core Intel Xeon E5-2670 Sandy Bridge Processors with 2.6GHz) in parallel over the MC replications (and not over the second step of the estimation).

[12] We do not report the results for $H = 100$. They are available from the authors upon request.

4.3 Simulation results

Appendices E and F report simulation results for the whole vector of the MFSV parameters, i.e. $\hat{\boldsymbol{\theta}}$ as well as for sub-vectors of it, i.e. for the factor loadings, the unconditional variances and the ARSV parameters, respectively.[13] In Appendix E, we focus on presenting the results on the ratio between the MC standard deviation of the estimated parameters and their asymptotic standard errors, computed as described in Sect. 3, averaged over the components of the corresponding (sub)-vectors. Tables E.1 and E.3 present results for fixed $H = 10$ and Tables E.2 and E.4 report results for $H = 10^5/T$. As expected, increasing H in the EMM estimation procedure reduces the asymptotic variance of the estimated parameters. For $H = 10$ and large T, the average ratio between the empirical standard deviation of the estimated parameter values over the Monte Carlo replications and the estimated asymptotic standard errors is close to 1, regardless of the choice of the starting values. For lower values of T, however, the ratio is smaller than 1 in most cases, indicating that the asymptotic standard errors are too large compared to the Monte Carlo standard deviation. In these cases, the QML-based starting values improve the ratio towards 1. Here again by setting $H = 10^5/T$, also for smaller values of T, the ratios become close to 1.

Appendix F provides detailed numerical results for the MSE of the estimated parameters for the two choices of H and the starting values. From the tables, one may see that, as expected, the MSEs of the parameters estimated in the first step are unaffected by the choice of the EMM starting values of H. However, for the quality of ARSV parameters estimated in the second part of our procedure, both the choice of H and of the starting values are important and exhibit complementary effects for larger Ts. Thus, for T larger than 1000, the procedure provides ARSV estimates with small MSE, also when fixing $H = 10$, by choosing the QML starting values.

Generally, the MSE of the parameter vector $\hat{\boldsymbol{\theta}}$ decreases when T increases and increases when k increases. These effects are particularly pronounced for smaller values of N and reduce when N increases. Similar results are obtained by Calzolari et al. (2021) that show that the burden of adding a new factor to the estimation reduces when the underlying number of returns increases, as it provides more information on their commonalities than when increasing T.

By looking in detail at each set of parameters, one may see that increasing N has a stronger positive effect on the quality of the ARSV estimates than on the quality of the loading and the unconditional variance estimates. While the quality of the loading estimates is mainly improved by increasing T, the quality of the unconditional factor variances also improves by increasing the number of factors. As the constant parameters are identified from the unconditional variance of the error term and the AR and the standard deviation parameters, it inherits their MSE pattern, especially in what regards the dependence on k and N.

Figure 1 plots the MSE values for $H = 10^5/T$ and the QML starting values against the number of observations T for different number of return series N. The

[13] In Appendix D, we provide a short discussion on the occurrence of outliers in the second step of the estimation.

numerical values for this case are detailed in Table F.4 in Appendix F. The figure summarises the MSE results averaged over all parameters. In a nutshell: increasing T significantly reduces the MSE, while increasing k has a detrimental effect on the MSE, especially when T are small and if the portfolio contains only a few assets. Although N is finite, increasing the dimension of the underlying return vectors, thus increasing the dimensionality of the portfolio by adding further assets, also makes the estimation of the parameter more precise, as it provides more information on the commonality in their (co-)variance dynamics.

These effects become even more visible when considering the average MSE for each parameter as displayed in the heat-maps of Appendix G. The unconditional variances of the factors and the loading parameters seem to benefit mostly from increasing T, while the MSE of all estimates reduces when N increases, but increases when k increases.

4.4 Ordering and computational efficiency

Due to the identifying assumptions of the MFSV, the ordering of the returns in the vector $\boldsymbol{y}_t$ matters since the first series only loads on the first factor, the second series only loads on the first two factors, and so on. While the factor loadings matrix, the unconditional variance–covariance matrix of the factors and the constant parameter of the factor ARSV models change by changing the order, the other set of parameters should not change and this is confirmed in a small Monte Carlo simulation

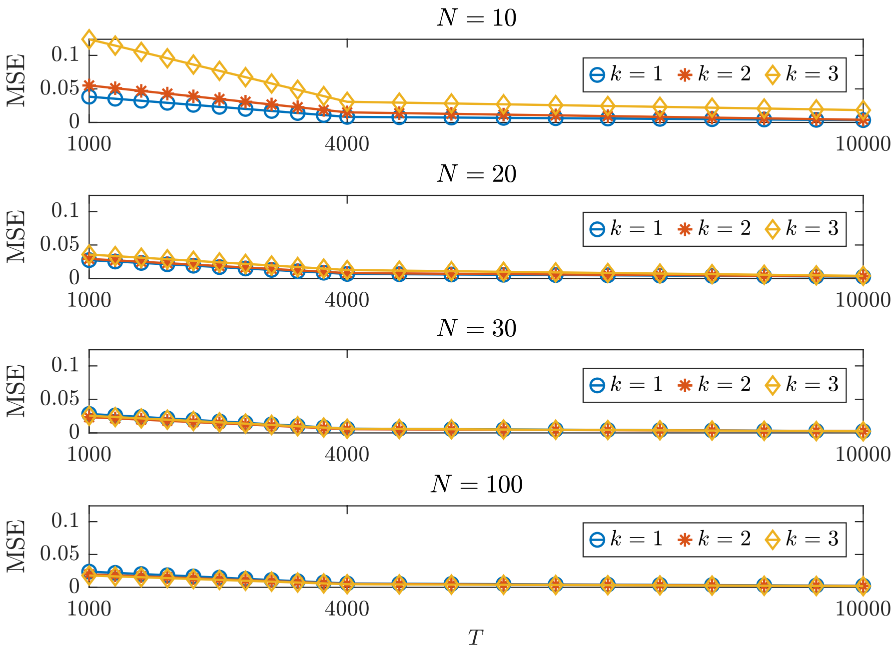

Fig. 1 MSE of $\hat{\theta}$: $H = 10^5/T$ and QML starting values for EMM

study, where we invert the original ordering of the vector of returns. The results, which are computed with $H = 10^5/T$ and the QML starting values, can be obtained from the authors upon request.

One very important advantage of our method is its computational efficiency. Figures 2 and 3 plot the computational time per replication for $H = 10$ and $H = 10^5/T$, respectively, against T for different numbers of factors and return series, however, by implementing the QML starting values for the EMM estimation. The upper panel shows the average computational time over the Monte Carlo replications for $N = 100$ return series for 1, 2 and 3 factors, while the lower panel depicts smaller dimensions, i.e. $N = 10, 20, 30$.

From the two graphs, one may see that the computational time increases linearly in T. Moreover, surprisingly the computational time for $H = 10^5/T \geq 10$ is maximally twice the time needed for $H = 10$. For instance, for $T = 4000$ (i.e. $H = 25$), there are almost no differences in the computational time between the two choices of H. Therefore, one can conclude that adapting the choice of H to T is optimal for any empirical application as it increases the efficiency of the estimates, reduces the number of outliers and is computationally feasible.

In order to present results on the computational time of our procedure, we focus here on the simulation design of Kastner et al. (2017) and choose $N = 10$, $k = 2$ and $T = 1000$ as well as their choice of parameters. Our procedure estimates the parameters of the MFSV model under this setting on average in 13 s when using the QML starting values and setting $H = 10$. Setting $H = 10^5/T$ and with poorer

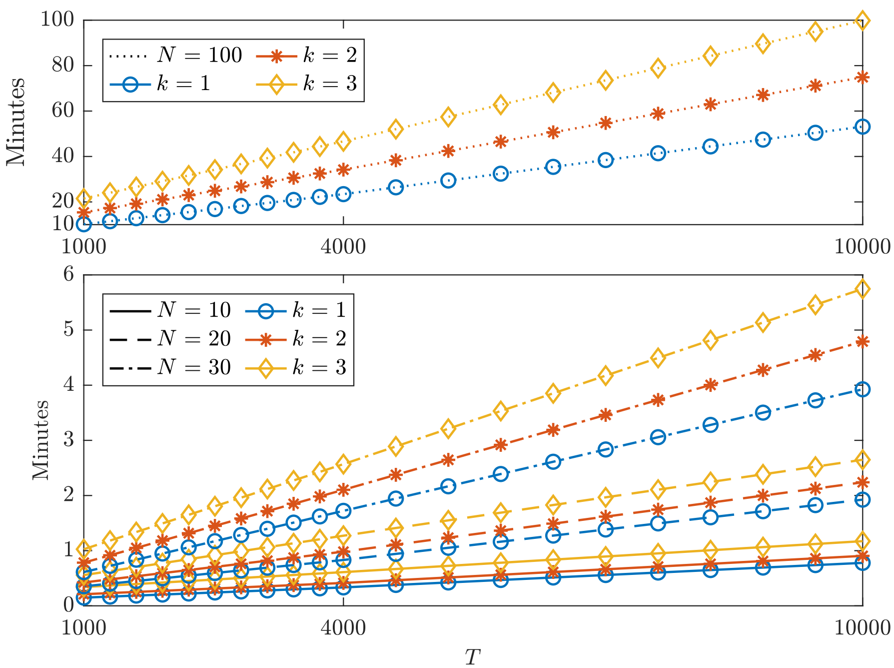

Fig. 2 Average time per Monte Carlo replication: $H = 10$

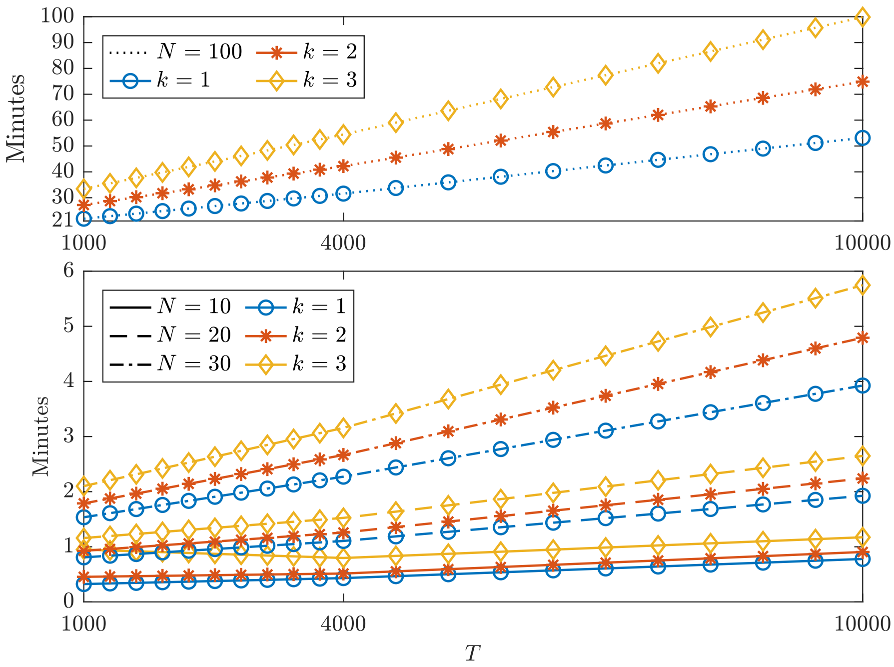

Fig. 3 Average time per Monte Carlo replication: $H = 10^5/T$

starting values, our procedure takes, on average, 82 s (roughly 1.37 min). Therefore, we can conclude that our procedure is very effective and computationally feasible in estimating the parameters of MFSV even for very large vectors of returns as also illustrated in the empirical application presented below.

5 Empirical application

In this section, we apply our proposed algorithm to a vector of $N = 148$ log-returns of large-cap stocks from the S &P500 from 5 January 1981 till 31 December 2018 ($T = 9584$ observations), ordered by market cap.[14][15] We demean and standardise the returns, i.e. for each series subtract its mean and divide by its respective empirical standard deviation, in order to facilitate the estimation. Since we have close to $T = 10000$ observations, we use the adaptive choice of $H = 10^5/T$ and set; thus, $H = 10$.

[14] We provide a list of the ticker symbols and company names ordered by the market capitalisation of the stocks in Table H.1 in Appendix H.

[15] We also implement the procedure on the 26 exchange rate series of Kastner et al. (2017). However, the results both for estimation and forecasting do not drastically differ from the results for the stocks. Therefore, we omit them from this paper and can be obtained from the authors upon request.

As our procedure only regards the estimation of the MFSV model for a pre-specified number of factors, we apply the tests of Onatski (2010) and Ahn and Horenstein (2013) to find the number of dynamic factors in our dataset. Both tests find $k = 1$ number of factors. Figure I.1 in Appendix I shows the number of principal components of the data set we use in this empirical application. The principal components plots we present are in line with applying the procedure presented in Forni et al. (2000) to identify the number of factors for a static factor model. The cut-off between the first and the rest of the principal components is pretty clear, which confirms the results of the tests of Onatski (2010) and Ahn and Horenstein (2013). That is why we present below the empirical results for estimating the MFSV model with $k = 1$ factors, and in Appendix J we discuss the results from estimating the model for $k = 2$ and $k = 3$.

Figures K.1 to K.3 in Appendix K display the estimated parameters of the ARSV for the idiosyncratic error term series. While the estimates of the constant parameter seem to vary less among the 148 idiosyncratic disturbances, the estimates of the AR parameter and of the noise standard deviation display much more variation. Most of the idiosyncratic disturbances have a very persistent stochastic volatility, with the AR parameter being close to 1, while others display less persistency in their dynamic variation, with the AR parameter being around 0.8. Also, the standard deviation of their SVs varies largely among the series from values close to 0 to values larger than 0.3. This mirrors the large heterogeneity in the dynamics of the 148 returns considered in the study.

The AR parameter in the SV structure of the factor is estimated to be around 0.9752, while the constant and the standard deviation are estimated to be -1.5866 and 0.1863, respectively. All parameters of the factor SV are statistically significant at 5% significance level.

Below we discuss the computational time needed for our estimation procedure to estimate the parameters of the MFSV model with 1 factor on the vector of 148 daily returns. When disabling that, the estimation of the second step runs in parallel; the run time is roughly 3,5 h. Parallelising the second step on 40 CPU cores, the run time reduces to approximately 10 min. Because we have no comparison of our estimation time to any other purely estimation algorithm in the literature, we can only present here the running time of the Bayesian procedure of Kastner et al. (2017), which is roughly 59 h. Although it does not make the object of this paper per se, in order to make the comparison complete, we also present some preliminary results from filtering the latent variables based on our parameter estimates. For this, we implement the auxiliary multiple particle filter of Mücher (2023) that aims at solving the curse of dimensionality problem of the bootstrap particle filter method of (Gordon et al. 1993)[16] with 10240 particles per series (i.e. a total of

[16] It is the multiple particle filter of Djuric et al. (2007), which Mücher (2023) adapts by incorporating the auxiliary particle filtering principle of Pitt and Shephard (1999a) in order to get some further computational benefits, i.e. by adding a resampling step prior to the propagation of new particles. Like this, Mücher (2023) obtains a better approximation of the filtering density with the same number of particles, avoiding thus the problem of the hugely dimensionality increase in the number of filters when the dimensionality of the latent series increases (Bengtsson et al. 2008). For further details about the filter, see Djuric et al. (2007) and Mücher (2023).

Table 1 Average Frobenius norm

	No overlap						Overlap					
	MFSV		GARCH		STATIC		MFSV		GARCH		STATIC	
Stocks: All out of sample period												
$h = 1$	**570**	**612**$_{\dagger}$	572	535	596	708***	**570**	**612**$_{\dagger}$	572	535$_{\dagger}$	596	708***
$h = 5$	**2451**	**838**$_{\dagger}$	2483	594	2539	731***	**2646**	**462**$_{\dagger}$	2673	934**	2750	414***
$h = 10$	**4856**	**792**$_{\dagger}$	4961	555	5042	880***	**5056**	**568**$_{\dagger}$	5146	730**	5220	099***
Stocks: Crisis period												
$h = 1$	**1481**	**198**$_{\dagger}$	1491	908$_{\dagger}$	1515	923$^{*}_{\dagger}$	**1481**	**198**$_{\dagger}$	1491	908$_{\dagger}$	1515	923$^{*}_{\dagger}$
$h = 5$	**5657**	**122**$_{\dagger}$	5839	256$_{\dagger}$	5745	258$_{\dagger}$	**6443**	**947**$_{\dagger}$	6617	673$^{**}_{\dagger}$	6457	270$_{\dagger}$
$h = 10$	9865	148$_{\dagger}$	10461	373$_{\dagger}$	**9594**	**818**$_{\dagger}$	12118	313$_{\dagger}$	12619	199$^{*}_{\dagger}$	**11894**	**710**$_{\dagger}$
Exchange rates: All out of sample period												
$h = 1$	**8**	**276**$_{\dagger}$	8	393$_{\dagger}$	8	960***	**8**	**276**$_{\dagger}$	8	393$_{\dagger}$	8	960***
$h = 5$	**37**	**936**$_{\dagger}$	39	380$_{\dagger}$	41	037***	**40**	**972**$_{\dagger}$	41	575$_{\dagger}$	44	004***
$h = 10$	**73**	**091**$_{\dagger}$	78	732$_{\dagger}$	78	547***	**80**	**456**$_{\dagger}$	83	044$_{\dagger}$	85	921***

Bold numbers denote the smallest loss. ***, ** and * denote a rejection of the H0 of the Diebold Mariano test of equal forecasting performance (to the MFSV) at the 1%, 5% and the 10% level, respectively. $_{\dagger}$ denotes models that are in the 90% MCS

1525760 particles), which makes the whole procedure, i.e. the estimation based on the approach of this paper and the filtering roughly three times faster than the Bayesian method of Kastner et al. (2017).[17]

Below we present some results from comparing the forecasting performance of the MFSV model based on the parameter estimates we propose in this paper and two alternatives: the static factor model and the multivariate factor GARCH model of Harvey et al. (1992) on which we apply the EM estimation procedure of Audrino et al. (2016). We apply the filter and obtain the forecasts for the MFSV based on the procedure developed by Mücher (2023).[18] Table 1 presents the Frobenius norm results for 1-step ahead forecasts and for overlapping and non-overlapping 5- and 10-step ahead forecasts during the period 24 April 2006 to 31 December 2018 as well as during the financial crisis period from 2 July 2017 to 31 December 2018. As the table displays, the MFSV performs better than the alternatives providing the smallest loss function, except for the crisis period and 10-step ahead forecasts, for which, however, it enters the model confidence set (MCS) of Hansen et al. (2011). The fact that in the crisis period, almost all models enter MCS is due to the very few number of observations on which the evaluation is done. However, in this case the Diebold Mariano test brings more detailed evidence in favour or MFSV.

[17] We use the *factorstochvol* package provided by the authors for the estimation and, as suggested by them, we use the deep interweaving code. After 20000 burn-in periods, we use 200000 draws from the posterior distribution. The authors use 50000 burn-in and 500000 sampling periods in their paper. Our choice of the burn-in and sampling periods is due to RAM and runtime limitations on our computational device.

[18] We demean and standardise the returns of the rolling window prior to the estimation for each candidate model. The forecasts based on the standardised data are then rescaled using the empirical standard deviation of the respective rolling window.

Additionally to our stock dataset and as a robustness check, we also do the estimation and the forecasting exercise with the exchange rate dataset used by Kastner et al. (2017) of 26 daily series (from 2 April 2005 to 6 August 2015 resulting in 2649 observations, with the forecasting window from 3 January 2012 to 6 August 2015). The crisis forecasting exercise is done only for the stocks, as they were more affected by the crisis than the exchange rates. The results are also presented in Table 1 and confirm the results of the stocks: the MFSV outperforms the alternatives for all choices of h and forecasting window.

6 Conclusion

This paper proposes a two-step procedure to estimate multivariate factor stochastic volatility models with exact factor structure and stochastic volatility dynamics for both the factors and the idiosyncratic disturbances as introduced by Harvey et al. (1994) and Pitt and Shephard (1999b). In the first step, we use the convergence results of Bai and Li (2012) to estimate the factor loadings and the variances of the idiosyncratic errors and of the factors by means of ML applied to a static factor model. In the second step, we estimate the parameters of the stochastic volatility of the factors and of the idiosyncratic errors equation-by-equation by using the EMM procedure of Bansal et al. (1994) and Gallant and Tauchen (1996) and the univariate GARCH as auxiliary models.

Our procedure is the first in the literature to implement "frequentist" techniques to estimate such a model. Due to the two-step procedure and the equation-by-equation estimation in the second step, the procedure we propose is computationally very fast and can easily be applied to very large dimensions of vectors of returns and number of factors. The computational efficiency can be further increased by choosing appropriate starting values and parallelising the EMM estimation in the second step.

In a comprehensive simulation exercise, we show that our estimation provides accurate estimates of the parameters as well as how to choose the optimal number of simulations in the EMM procedure in order to increase the efficiency of the estimation at minimal computational costs. In the empirical application to a large vector of 148 returns, we provide further evidence of the efficacy of our method. While in the simulation exercise we can estimate the model for a vector of 100 returns in a matter of seconds, in the empirical application the estimation converges in about 10 min. We also provide a forecasting evidence in favour of the model we estimate.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10182-025-00536-3.

Acknowledgements We want to thank Matteo Barogozzi, Lukas Bauer, Robin Braun, Ralf Brüggemann, Maurizio Daniele, Gabriele Fiorentini, Giampiero M. Gallo, Lyudmila Grigoryeva, Julie Schnaitmann, Enrique Sentana, Winfried Pohlmeier, the editor and the two anonymous referees for helpful comments. All remaining errors are ours. Christian Mücher acknowledges financial support from the Graduate School of Decision Sciences (GSDS), University of Konstanz, Germany, and from the German federal state of Baden-Württemberg through a Landesgraduiertenstipendium. Roxana Halbleib acknowledges financial support from the German Science Foundation through the project HA 8672/1. We acknowledge computational support by the state of Baden-Württemberg through bwHPC.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Aguilar, O., West, M.: Bayesian dynamic factor models and portfolio allocation. Journal of Business & Economic Statistics **18**, 338–357 (2000)

Ahn, S.C., Horenstein, A.R.: Eigenvalue ratio test for the number of factors. Econometrica **81**, 1203–1227 (2013)

Aielli, G.P., Calzolari, G., Fiorentini, G.: Fast indirect estimation of latent factor models with conditional heteroskedasticity. In Brenatari, E. and Carpita, M., editors, Advances in Latent Variables, Milano. Vita e Pensiero (2013)

Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, 3rd edn. Wiley, Hoboken (2003)

Anderson, T.W., Rubin, H.: Statistical inference in factor analysis. In Neyman, J., editor, In: Proceedings of the third berkeley symposium on mathematical statistics and probability, volume V, pages 111–150, Berkeley. University of California Press (1956)

Audrino, F., Corsi, F., Filipova, K.: Bond risk premia forecasting: a simple approach for extracting macroeconomic information from a panel of indicators. Economet. Rev. **35**, 232–256 (2016)

Bai, J., Li, K.: Statistical analysis of factor models of high dimension. Ann. Stat. **40**, 436–465 (2012)

Bai, J., Liao, Y.: Efficient estimation of approximate factor models via penalized maximum likelihood. J Econom **191**, 1–18 (2016)

Bansal, R., Gallant, A.R., Hussey, R., Tauchen, G.: Computational aspects of nonparametric simulation estimation. In: Belsley, D.A. (ed.) Computational Techniques for Econometrics and Economic Analysis, pp. 3–22. Springer, Dordrecht (1994)

Barigozzi, M.: Quasi maximum likelihood estimation of high-dimensional factor models: a critical review. (2023). arXiv:2303.11777

Barigozzi, M., Hallin, M.: Generalized dynamic factor models and volatilities: recovering the market volatility shocks. Economet. J. **19**, 33–60 (2016)

Barigozzi, M., Hallin, M.: Generalized dynamic factors and volatilities: estimation and forecasting. J Econom **201**, 307–321 (2017)

Barigozzi, M., Hallin, M.: Generalized dynamic factor models and volatilities: consistency, rates, and prediction intervals. J Econom **216**, 4–34 (2020)

Bengtsson, T., Bickel, P., Li, B., et al.: Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. Probability and statistics: essays in honor of David A. Freedman **2**, 316–334 (2008)

Bien, J., Tibshirani, R.J.: Sparse estimation of a covariance matrix. Biometrica **98**, 807–820 (2011)

Broto, C., Ruiz, E.: Estimation methods for stochastic volatility models: a survey. J Econ Surv. **18**, 613–649 (2004)

Calzolari, G., Fiorentini, G., Sentana, E.: Constrained indirect estimation. Rev. Econ. Stud. **71**, 945–973 (2004)

Calzolari, G., Halbleib, R.: Estimating stable latent factor models by indirect inference. J Econom **205**, 280–301 (2018)

Calzolari, G., Halbleib, R., Parrini, A.: Estimating garch-type models with symmetric stable innovations: indirect inference versus maximum likelihood. Comput. Stat. Data Anal. **71**, 945–973 (2014)
Calzolari, G., Halbleib, R., Zagidullina, A.: A latent factor model for forecasting realized variances. J. Financ. Economet. **19**, 860–909 (2021)
Chamberlain, G., Rothschild, M.: Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica **51**(5), 1281–1304 (1983)
Chib, S., Nardari, F., Shephard, N.: Analysis of high dimensional mulitivariate stochastic volatility models. J Econom **134**, 341–371 (2006)
Cox, G.: Weak identification in a class of generically identified models with an application to factor models. Unpublished working paper, Department of economics, Columbia university (2017)
Daniele, M., Pohlmeier, W., Zagidullina, A.: Sparse approximate factor estimation for high-dimensional covariance matrices. (2020). arXiv preprint arXiv:1906.05545
Djuric, P.M., Lu, T., Bugallo, M.F.: Multiple particle filtering. In: 2007 IEEE international conference on acoustics, speech and signal processing-ICASSP'07, volume 3, pages III–1181–III–1184. IEEE (2007)
Forni, M., Hallin, M., Lippi, M., Reichlin, L.: The generalized dynamic-factor model: identification and estimation. Rev. Econ. Stat. **82**, 540–554 (2000)
Gallant, A.R., Tauchen, G.: Which moments to match? Economet. Theor. **12**, 657–681 (1996)
Gordon, N.J., Salmond, D.J., Smith, A.F.: Novel approach to nonlinear/non-gaussian Bayesian state estimation. In: IEE Proceedings F, volume 140, pp. 107–113. IET (1993)
Gouriéroux, C., Monfort, A., Renault, E.: Indirect inference. J. Appl. Economet. **8**, S85–S118 (1993)
Han, Y.: Asset allocation with a high dimensional latent factor stochastic volatility model. Rev Financial Stud **19**, 237–271 (2005)
Hansen, P.R., Lunde, A., Nason, J.M.: The model confidence set. Econometrica **79**, 453–497 (2011)
Harvey, A., Ruiz, E., Sentana, E.: Unobserved component time series models with arch disturbances. J Econom **52**, 129–157 (1992)
Harvey, A., Ruiz, E., Shephard, N.: Multivariate stochastic variance models. Rev. Econ. Stud. **61**, 247–264 (1994)
Ishihara, T., Omori, Y.: Portfolio optimization using dynamic factor and stochastic volatility: evidence on fat-tailed errors and leverage. Jpn. Econ. Rev. **68**, 63–94 (2017)
Kastner, G.: Sparse Bayesian time-varying covariance estimation in many dimensions. J Econom **210**, 98–115 (2019)
Kastner, G., Frühwirth-Schnatter, S., Lopes, H.F.: Efficient Bayesian inference for multivariate factor stochastic volatility models. J. Comput. Graph. Stat. **26**, 905–917 (2017)
Lawley, D.N., Maxwell, A.E.: Factor analysis as a statistical method. J R Stat Soc SerD **12**, 209–229 (1962)
Monfardini, C.: Estimating stochastic volatility models through indirect inference. Economet. J. **1**, 113–128 (1998)
Mücher, C.: Multiple particle filter for the high dimensional multivariate factor stochastic volatility model - beating the curse of dimensionality. (2023). 1013140/RG.2.2.27917.72164
Nardari, F., Scruggs, J.T.: Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and apt pricing restrictions. J Financ Quant Anal **42**, 857–891 (2007)
Onatski, A.: Determining the number of factors from empirical distribution of eigenvalues. Rev. Econ. Stat. **92**, 1004–1016 (2010)
Pitt, M.K., Shephard, N.: Filtering via simulation: auxiliary particle filters. J. Am. Stat. Assoc. **94**, 590–599 (1999)
Pitt, M.K., Shephard, N.: Time-varying covariances: a factor stochastic volatility approach. Bayesian Stat **6**, 547–570 (1999)
Ruiz, E.: Quasi-maximum likelihood estimation of stochastic volatility models. J Econom **63**, 289–306 (1994)
Sahu, S.K., Dey, D.K., Branco, M.D.: A new class of multivariate skew distributions with applications to Bayesian regression models. Can J Stat **31**, 129–150 (2003)
Scharth, M., Li, M.: Leverage, asymmetry, and heavy tails in the high-dimensional factor stochastic volatility model. J Bus Econ Stat **40**, 285–301 (2022)
Sentana, E., Calzolari, G., Fiorentini, G.: Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks. J Econom **146**, 10–25 (2008)

Sentana, E., Fiorentini, G.: Identification, estimation and testing of conditionally heteroskedastic factor modles. J Econom **102**, 143–164 (2001)
Thomson, G.H.: The geometry of mental measurement. London University Press, London (1954)
Watson, M.W., Engle, R.F.: Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. J Econom **23**(3), 385–400 (1983)
Williams, B.: Identification of the linear factor model. Economet. Rev. **39**, 92–109 (2020)
Yu, J., Meyer, R.: Multivariate stochastic volatility models: Bayesian estimation and model comparision. Economet. Rev. **25**, 361–384 (2006)