Forecasting Conditional Covariance Matrices in High-Dimensional Time Series: a General Dynamic Factor Approach

Marc Hallin
ECARES and Département de Mathématique,
Université libre de Bruxelles, Belgium

Luiz K. Hotta
Department of Statistics, University of Campinas, Brazil

João H. G. Mazzeu
Department of Statistics, University of Campinas, Brazil

Carlos Trucíos
São Paulo School of Economics, FGV, Brazil

Pedro L. Valls Pereira
São Paulo School of Economics, FGV, Brazil

Mauricio Zevallos
Department of Statistics, University of Campinas, Brazil

June 2019

ECARES working paper 2019-14

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Forecasting Conditional Covariance Matrices in High-Dimensional Time Series: a General Dynamic Factor Approach*†‡

Carlos Trucíos[1], João H. G. Mazzeu[2], Marc Hallin[3], Luiz K. Hotta[2], Pedro L. Valls Pereira[1], Mauricio Zevallos[2]

[1]São Paulo School of Economics, FGV, Brazil
[2]Department of Statistics, University of Campinas, Brazil
[3]ECARES and Département de Mathématique, Université libre de Bruxelles, Belgium

Abstract

Based on a General Dynamic Factor Model with infinite-dimensional factor space, we develop a new estimation and forecasting procedures for conditional covariance matrices in high-dimensional time series. The performance of our approach is evaluated via Monte Carlo experiments, outperforming many alternative methods. The new procedure is used to construct minimum variance portfolios for a high-dimensional panel of assets. The results are shown to achieve better out-of-sample portfolio performance than alternative existing procedures.

*This paper received the best LACSC 2019 Paper Award at the 4th Latin American Conference for Statistical Computing, held in Guayaquil, Ecuador, May 28-31, 2019.

†Financial support is gratefully acknowledged, from the São Paulo Research Foundation (FAPESP) grants 2016/18599-4 and 2018/03012-3 by the first and fifth authors, from the Coordination for the Improvement of Higher Education Personnel (CAPES) grant 88882.305837/2018-01 by the second author, from the São Paulo Research Foundation (FAPESP) grant 2018/04654-9 by the fourth and sixth authors. All authors acknowledge support from the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS), Centre of Quantitative Studies in Economics and Finance (CEQEF) and European Centre for Advanced Research in Economics and Statistics (ECARES).

‡We thank Mario Forni, Roman Liška, and Matteo Barigozzi for kindly giving access to their Matlab codes. Computational resources have been partially provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) under grant No. 2.5020.11.

Keywords. Dimension reduction, Large panels, High-dimensional time series, Minimum variance portfolio, Volatility, Multivariate GARCH.
JEL classifications. C38, C53, C55, C59, G11.
2010 Mathematics Subject Classification. 62H99, 62M20, 62P20, 91G10.

1 Introduction

Volatility forecasting plays an essential role in a variety of economic and financial applications, such as portfolio allocation, risk management, option pricing, hedging strategies, etc.: see Engle (2009), Hlouskova et al. (2009), Aramonte et al. (2013), Becker et al. (2015), Trucíos et al. (2018) and Engle et al. (2019), to quote only a few.

Several multivariate models have been proposed to model and forecast the conditional covariance matrix of a collection of assets; see Bauwens et al. (2006) or de Almeida et al. (2018) for some reviews. Unfortunately, most of multivariate GARCH (MGARCH) type models badly suffer from the so-called "curse of dimensionality" as the number of assets grows, and cannot be implemented in a high-dimensional context. Therefore, alternative procedures have been proposed, such as Fan et al. (2008), Alessi et al. (2009), Matteson and Tsay (2011), Engle and Kelly (2012), Hu and Tsay (2014), Santos and Moura (2014), Li et al. (2016), Pakel et al. (2017), Chang et al. (2018) and Engle et al. (2019), among others.

Dynamic factor models with high-dimensional asymptotics offer a promising alternative in that context; see the surveys by Barhoumi et al. (2014) and Bai and Wang (2016) for details. Factor models are based on the assumption that prices and volatilities of different assets are driven by a small number of latent factors, which account for their co-movements. They have been used by several authors to model and forecast conditional covariance matrices: see Diebold and Nerlove (1989), Harvey et al. (1992), Aguilar and West (2000), Vrontos et al. (2003), Han (2005), Sentana et al. (2008), Aguilar (2009), Alessi et al. (2009), García-Ferrer et al. (2012), Aramonte et al. (2013) and Dovonon (2013), among others. All these contributions are based on a *static* factor-loading scheme[1] (Bai and Ng, 2002; Stock and Watson,

[1] The latent factors are loaded contemporaneously via some loading matrix, so that the dimension of the factor space reduces to the (finite) number of linearly independent factors.

2002a,b)[2] leading to finite-dimensional factor spaces whose main advantage is to allow for estimation methods based on traditional principal components, which are easy to implement and widely used in practice.

However, as pointed out in Forni and Lippi (2011) and Section 1.1 of Forni et al. (2015), the assumption of a static factor-loading scheme considered in that literature is quite restrictive and rules out some very simple and plausible cross-correlation patterns, leading to infinite-dimensional factor spaces. To overcome this issue, Forni et al. (2000) introduced the so-called *generalized* or *general dynamic factor model* (GDFM), in which factors (equivalently, common shocks) are loaded through filters rather than matrices. As shown in Hallin and Lippi (2013), the GDFM actually follows from a representation result which holds, essentially, without placing any restrictions—beyond second-order stationarity and the existence of a spectrum—on the data-generating process.

The role of traditional principal components in the GDFM is taken over by Brillinger's *dynamic principal components*[3] (Brillinger, 1981), and the estimation method proposed by Forni et al. (2000) naturally relies on this concept. Dynamic principal components, however, involve two-sided filters, producing estimators that are inadequate in forecasting problems. Forni and Lippi (2011) and Forni et al. (2015, 2017)[4] therefore developed an alternative estimation method involving only one-sided filters. Moreover, Monte Carlo simulations indicate that, for estimating impulse-response functions and predicting returns, this one-sided approach outperforms the *static* method of Stock and Watson (2002a,b) and Bai and Ng (2002) even when the actual loading scheme is of the static type (see Section 4 in Forni et al. (2017)).

The Forni et al. (2015, 2017) procedure has been successfully used to forecast inflation and financial returns; see Della Marra (2017), Forni et al. (2018) and Gio-

[2]Similar ideas have been developed also in a non-econometric context, see, e.g., Peña and Box (1987), Stoffer (1999), or Pan and Yao (2008).

[3]Hallin et al. (2018) show that those dynamic principal components, based on the factorization of spectral density matrices, inherit, in a time-series context, the optimality properties that make traditional principal components a successful dimension-reduction device in i.i.d. samples.

[4]The assumptions in those three references yield slight variations; in this paper, unless otherwise stated, we refer to the assumptions in Barigozzi and Hallin (2018).

vannelli et al. (2018). It has also been used in the prediction of conditional variances by (Barigozzi and Hallin, 2016, 2017, 2018), but never, as far as we know, in the prediction of conditional covariance matrices and portfolio optimization.[5] This point constitutes the main goal of this paper.

The rest of the paper is organised as follows. Section 2 briefly describes the GDFM and Section 3 introduces our forecasting procedure. Section 4.1 reports a Monte Carlo study of the finite-sample properties of the proposed procedure. In Section 5, we apply the new procedure in the problem of constructing minimum variance portfolios from a large collection of assets. In Sections 4.1 and 5 we also compare the proposed procedure with other methods. Finally, Section 6 presents the main conclusions and discusses some directions for future research.

2 The general dynamic factor model

In this section, we briefly describe the GDFM to be considered throughout, which basically contains as particular cases all other factor models proposed in the econometric and time series literature, along with the regularity assumptions we need for consistency, which are borrowed, essentially, from Barigozzi and Hallin (2018).

Let $\{\mathbf{X}_t := (X_{1t}\ X_{2t}\dots)',\ t\in\mathbb{Z}\}$, be a double-indexed zero-mean second-order stationary stochastic process, where the first index is cross-sectional and typically refers to assets, while t, as usual, stands for time. The GDFM is based on the decomposition

$$X_{it} = \chi_{it} + \xi_{it}, \qquad i\in\mathbb{N}_0, \quad t\in\mathbb{Z} \tag{1}$$

with

$$\chi_{it} = \sum_{j=1}^{q}\sum_{k=0}^{\infty} b_{ijk}u_{jt-k} = \mathbf{b}_i'(L)\mathbf{u}_t \ \text{ and } \ \xi_{it} = \sum_{k=0}^{\infty} d_{ik}v_{it-k} = d_i(L)v_{it}, \tag{2}$$

where the *common components* χ_{it}, the *idiosyncratic components* ξ_{it}, the *common shocks* or *factors* $\mathbf{u}_t := (u_{1t}\ u_{2t}\ \dots\ u_{qt})'$ driving the common components, and the *idiosyncratic shocks* v_{it} driving the idiosyncratic components all are non-observable.

[5]See, however, Alessi et al. (2009) who assume a factor model decomposition with finite-dimensional factor space on te model of Forni et al. (2005 and 2009).

Letting $\mathbf{X}_n := \{X_{it} | i = 1, ..., n, t \in \mathbb{Z}\}$, $\boldsymbol{\chi}_n := \{\chi_{it} \mid i = 1, ..., n, \ t \in \ \mathbb{Z}\}$, and $\boldsymbol{\xi}_n := \{\xi_{it} | i = 1, ..., n, t \in \mathbb{Z}\}$, equation (2) in vector notation takes the form

$$\mathbf{X}_{nt} = \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} = \mathbf{B}_n(L)\mathbf{u}_t, + \mathbf{D}_n(L)\mathbf{v}_{nt}, \quad n \in \mathbb{N}_0, \quad t \in \mathbb{Z} \tag{3}$$

with $\mathbf{B}_n(L) := (\mathbf{b}_1(L)...\mathbf{b}_n(L))'$, $\mathbf{D}_n(L) := (d_1(L)...d_n(L))'$, and $\mathbf{v}_{nt} := (v_{1t} \dots v_{nt})'$.

On the decomposition (1), we assume the following:

(i) the vector process $\mathbf{u}_t$ is a zero-mean q-dimensional second-order white noise process, with full-rank covariance $\boldsymbol{\Gamma}^{\mathbf{u}}$;

(ii) writing $\mathbf{b}_{ik} := (b_{i1k}...b_{iqk})'$ for the $q \times 1$ coefficient of L^k in $\mathbf{b}_i(L)$, there exists a constant $M_1 > 0$ such that $\sum_{k=0}^{\infty} ||\mathbf{b}_{ik}|| k^{1/2} \leq M_1$ for all $i \in \mathbb{N}$;

(iii) $\mathbf{v}_{nt}$ is a zero-mean second-order stationary process with positive definite covariance $\boldsymbol{\Gamma}_n^v$; moreover, $\mathrm{E}[v_{it}|v_{is}] = 0$ for all $i \in \mathbb{N}$ and $t > s \in \mathbb{Z}$;

(iv) there exists a constant $C_v > 0$ such that $||\boldsymbol{\Gamma}_n^v||_1 \leq C_v$ for all $n \in \mathbb{N}$, and a constant $M_2 > 0$ such that $\sum_{k=0}^{\infty} |d_{ik}| k^{1/2} \leq M_2$ for all $i \in \mathbb{N}$;

(v) $\mathrm{Cov}(u_{jt}, v_{is}) = 0$ for all $i \in \mathbb{N}$, $j = 1, ..., q$, and t, $s \in \mathbb{Z}$;[6]

(vi) there exists a constant $M_3 > 0$ such that, for all j_1, j_2, j_3, j_4,

$$\sum_{k_1,k_2,k_3 \in \mathbb{Z}} |\mathrm{E}(u_{j_1 t} u_{j_2,t-k_1} u_{j_3,t-k_2} u_{j_4,t-k_3})| \leq M_3,$$

and a constant $M_4 > 0$ such that, for all i_1, i_2, i_3, i_4,

$$\sum_{k_1,k_2,k_3 \in \mathbb{Z}} |\mathrm{E}(v_{i_1 t} v_{i_2,t-k_1} v_{i_3,t-k_2} v_{i_4,t-k_3})| \leq M_4;$$

(vii) for all $i \in \mathbb{N}$, $j = 1, \dots, q$ and $z \in \mathbb{C}$, $b_{ij}(z) = \sum_{k=0}^{\infty} b_{ijk} z^k$ has square-summable coefficients, and is the ratio of two finite-order polynomials in z, $b_{ij}(z) = \gamma_{ij}(z)/\delta_{ij}(z)$, where $\gamma_{ij}(z) = \sum_{k=0}^{S_\gamma} \gamma_{ijk} z^k$ and $\delta_{ij}(z) = \sum_{k=0}^{S_\delta} \delta_{ijk} z^k$, with $\delta_{ij}(0) = 1$, have roots outside the closed unit disk only and no common roots, and the orders S_γ and S_δ are independent of i.[7]

Assumption (iii) is the typical assumption of martingale difference innovations used in the GARCH literature. Assumption (vii) entails the existence of a VAR filtering

[6] This implies that the common and idiosyncratic processes are mutually uncorrelated at all leads and lags.

[7] As a consequence, the common components have rational spectral densities; see Assumption (L2) in Barigozzi and Hallin (2018) for more details.

of $\mathbf{X}_n$ satisfying the assumptions of the static factor model where the common shocks $\mathbf{u}_t$ are loaded contemporaneously (see (4) below).

These assumptions also guarantee the existence of the spectral densities $\mathbf{\Sigma}_n^{\chi}(\theta)$, $\mathbf{\Sigma}_n^{\xi}(\theta)$, and $\mathbf{\Sigma}_n^{X}(\theta) = \mathbf{\Sigma}_n^{\chi}(\theta) + \mathbf{\Sigma}_n^{\xi}(\theta)$, $\theta \in [-\pi, \pi]$, of $\boldsymbol{\chi}_n$, $\boldsymbol{\xi}_n$ and $\mathbf{X}_n$, respectively. Then, let $\lambda_{nj}^{\chi}(\theta)$, $\lambda_{nj}^{\xi}(\theta)$ and $\lambda_{nj}^{X}(\theta)$ be the jth eigenvalues (in decreasing order of magnitude) of $\mathbf{\Sigma}_n^{\chi}(\theta)$, $\mathbf{\Sigma}_n^{\xi}(\theta)$ and $\mathbf{\Sigma}_n^{X}(\theta)$, respectively, satisfying the following assumption.

(viii) there exist a positive integer $\bar{n}$ and continuous functions α_j and β_{j-1} from $[-\pi, \pi]$ to $\mathbb{R}$, $j = 1, \ldots, q$, independent of n, and such that, for all $j = 1, \ldots, q$, and all $n > \bar{n}$, $0 < \beta_{j-1}(\theta) < \alpha_j(\theta) \leq \lambda_{nj}^{\chi}(\theta)/n \leq \beta_j(\theta) < \infty$, θ-a.e. in $[-\pi, \pi]$, while $\lambda_{n,q+1}^{\chi}(\theta)$ and $\lambda_{n1}^{\xi}(\theta)$ are bounded, uniformly in $\theta \in [-\pi, \pi]$, as $n \to \infty$.

Hence, as $n \to \infty$, the q idiosyncratic dynamic eigenvalues are exploding linearly (the assumption of factor pervasiveness), while all idiosyncratic eigenvalues are bounded (this is the definition of idiosyncrasy).

The main theoretical result behind the one-sided approach of Forni et al. (2015) is the existence of a block-diagonal VAR filtering of the observations turning the GDFM representation (1) into a static one. More precisely, Forni and Lippi (2011) and Forni et al. (2015) show that, for generic values of the coefficients γ_{ijk} and δ_{ijk} (i.e., except for a subset with Lebesgue measure zero in the $(q+1)(S_\gamma + S_\delta)$-dimensional space of the relevant γ_{ijk} and δ_{ijk} coefficients), any $(q+1)$-dimensional vector $\boldsymbol{\chi}_t^{i_1 \ldots i_{q+1}} := (\chi_{i_1 t}, \ldots, \chi_{i_{q+1} t})'$ with $i_1 < \ldots < i_{q+1}$ admits a VAR representation of the form $\boldsymbol{A}(L)^{i_1 \ldots i_{q+1}} \boldsymbol{\chi}_t^{i_1 \ldots i_{q+1}} = \mathbf{R}^{i_1 \ldots i_{q+1}} \mathbf{u}_t$,[8] where $\boldsymbol{A}(L)^{i_1 \ldots i_{q+1}}$ has degree $S \leq qS_\gamma + q^2 S_\delta$ and the $(q+1) \times q$ matrix $\mathbf{R}^{i_1 \ldots i_{q+1}}$ is of rank q. It follows that *generically*, for any $n = m(q+1)$, partitioning $\boldsymbol{\chi}_{nt} = (\chi_{1t}, \ldots, \chi_{nt})'$ into m subvectors of dimension $(q+1)$, $\boldsymbol{\chi}_{nt}$ admits a block-VAR representation of the form

$$\mathbf{A}_n(L)\boldsymbol{\chi}_{nt} = \begin{bmatrix} \mathbf{A}^1(L) & 0 & \ldots & 0 \\ 0 & \mathbf{A}^2(L) & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \ldots & \mathbf{A}^m(L) \end{bmatrix} \boldsymbol{\chi}_{nt} = \begin{bmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^m \end{bmatrix} \mathbf{u}_t, \quad t \in \mathbb{Z}. \tag{4}$$

[8]See Assumption (L4) in Barigozzi and Hallin (2018) for more details about this VAR representation.

Hence, for $\mathbf{X}_{nt} = (X_{1t}, \ldots, X_{nt})'$, we have

$$\mathbf{A}_n(L)\mathbf{X}_{nt} = \mathbf{A}_n(L)\boldsymbol{\chi}_{nt} + \mathbf{A}_n(L)\boldsymbol{\xi}_{nt} = \mathbf{R}_n\mathbf{u}_t + \boldsymbol{\epsilon}_{nt}, \quad t \in \mathbb{Z} \tag{5}$$

with $\mathbf{R}_n = \left[\mathbf{R}^{1\prime}\,\mathbf{R}^{2\prime}\,\ldots\,\mathbf{R}^{m\prime}\right]'$ and $\boldsymbol{\epsilon}_{nt} = \mathbf{A}_n(L)\boldsymbol{\xi}_{nt}$, where it can be shown that the process $\boldsymbol{\epsilon}_t := \{(\epsilon_{1t}\ \epsilon_{2t} \ldots)',\ t \in \mathbb{Z}\}$ is still idiosyncratic. In other words, using obvious notation

$$\mathbf{A}(L) := \begin{bmatrix} \mathbf{A}^1(L) & 0 & \ldots & 0 & \ldots \\ 0 & \mathbf{A}^2(L) & \ldots & 0 & \ldots \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \ldots & \mathbf{A}^m(L) & \ldots \\ \vdots & \vdots & \ldots & \ldots & \ddots \end{bmatrix} \quad \text{and} \quad \mathbf{R} := \begin{bmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^m \\ \vdots \end{bmatrix}, \tag{6}$$

the filtered process $\mathbf{Y}_t := \mathbf{A}(L)\mathbf{X}_t$ admits a *static* factor model representation

$$\mathbf{Y}_t = \mathbf{R}\mathbf{u}_t + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z} \tag{7}$$

with q-dimensional factor space spanned by $\mathbf{u}_t$. While $\mathbf{R}$ and $\mathbf{u}_t$ are not individually identified, the product $\mathbf{R}\mathbf{u}_t$ is.

The static representation (7), under assumptions *(i)-(ix)*, holds *generically.* Assuming that it holds for the panel under study thus is not a strong requirement; we nevertheless need to make it an assumption:

(ix) For all $n^* \geq q+1$, letting $n = \lfloor n^*/(q+1)\rfloor(q+1)$, there exist block-diagonal filters $\mathbf{A}_n(L)$ and $n \times q$ matrices $\mathbf{R}_n$ such that (5) holds, irrespective of the cross-sectional ordering.

Assumptions *(i)-(ix)* are the main assumptions in Barigozzi and Hallin (2018); on top of these, they also require two less important and more technical ones (Assumptions (L4) and (L5), respectively), which we do not reproduce here. Under those assumptions, Barigozzi and Hallin (2018) show that a consistent reconstruction, based on $\mathbf{X}_t$, $\mathbf{X}_{t-1}$, ..., of the unobserved $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ is possible. It follows that $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ are $\mathcal{F}_t$-measurable, where $\mathcal{F}_t$ denotes the σ-field generated by $\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots$ It is worth noting that, reinforcing the same assumptions (e.g., assuming that $\mathbf{u}_t$ and $\mathbf{v}_{nt}$ are jointly i.i.d., which rules out GARCH-type behaviors), Forni et al. (2017) derive estimators for (1)-(2) and provide a complete asymptotic analysis for the same. On the other hand, Barigozzi and Hallin (2018) do not require i.i.d.ness and, under assumptions that include *(i)-(ix)*, provide consistency and consistency rates for the Forni et al. (2017) estimators. Finally, we assume the following.

(x) The common shocks $\mathbf{u}_t$ and the idiosyncratic shocks v_{it} are stable by aggregation MGARCH and univariate GARCH processes, respectively, and satisfy the conditions for consistent QMLE estimation.

The assumption that the MGARCH driving the common shocks is stable by aggregation is motivated by the fact that $\mathbf{u}_t$ is not fully identifiable (see the remark after (7)): under Assumption *(x)*, any linear transform $\mathbf{R}\mathbf{u}_t$ is driven by a MGARCH model of the same type as $\mathbf{u}_t$ itself. Examples of stable by aggregation MGARCH models are the full VECH (Bollerslev et al., 1988) and full BEKK (Engle and Kroner, 1995) models, which moreover can be consistently estimated via QMLE methods: see Theorems 11.2 and 11.4 in Francq and Zakoian (2010).

3 Predicting the conditional covariance matrix

We present a procedure to predict one-step ahead conditional covariance matrices,[9] i.e, to estimate the conditional covariance matrix $\mathrm{V}(\mathbf{X}_t|\mathcal{F}_{t-1})$ of the observable process $\mathbf{X}_t$. Section 3.1 provides a theoretical expression for that conditional covariance, and Section 3.2 introduces the estimation procedure.

3.1 The conditional covariance matrix

We start with a theoretical expression for the conditional covariance matrix of $\mathbf{X}_t$ in terms of the static representation (7).

Proposition 1. *Let Assumptions (i)-(ix) of Section 2 hold—ensuring the existence of the static representation* (7)*. Assume moreover that $\boldsymbol{u}_t$ and $\boldsymbol{\xi}_t$, conditional on $\mathcal{F}_{t-1}$, are uncorrelated at all leads and lags. Then, the covariance matrix of $\boldsymbol{X}_t$*

[9] The terminology (conditional) covariance *matrix* is used here with a slight abuse: by $\mathrm{V}(\mathbf{X}_t|\mathcal{F}_{t-1})$ we mean the infinite array with (i,j)-element the (conditional) covariance of X_{it} and X_{jt}, $(i,j) \in \mathbb{N}^2$. The same notation $\mathrm{V}(\,.\,|\mathcal{F}_{t-1})$, and the notation $\mathrm{Cov}(\,.,\,.\,|\mathcal{F}_{t-1})$ are used in an obvious fashion for other processes.

conditional on $\mathcal{F}_{t-1}$ *is*

$$\mathrm{V}(\mathbf{X}_t|\mathcal{F}_{t-1}) = \mathbf{R}\mathrm{V}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathbf{R}' + \mathrm{V}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1}). \tag{8}$$

Proof. From (7), we have that

$$\begin{aligned}\mathrm{V}(\mathbf{Y}_t|\mathcal{F}_{t-1}) &= \mathrm{V}(\mathbf{R}\mathbf{u}_t + \boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}) \\ &= \mathbf{R}\mathrm{V}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathbf{R}' + \mathrm{V}(\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}) + \mathrm{Cov}(\mathbf{R}\mathbf{u}_\mathrm{t}, \boldsymbol{\epsilon}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1}) \\ &\qquad +\mathrm{Cov}(\boldsymbol{\epsilon}_\mathrm{t}, \mathbf{R}\mathbf{u}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1}), \quad \mathrm{t} \in \mathbb{Z}.\end{aligned} \tag{9}$$

Without loss of generality we can assume that all VAR filters $\mathbf{A}^k(L)$ in (5) are of the form $\mathbf{A}^k(L) = \mathbf{I}_{q+1} - \boldsymbol{\phi}_1^k L - \cdots - \boldsymbol{\phi}_S^k L^S$ (with $\boldsymbol{\phi}_S^k \neq \mathbf{0}$ for at least one k). Consequently, $\mathbf{A}(L)$ can be written as $\mathbf{A}(L) = \mathbf{I} - \boldsymbol{\Phi}_1 L - \cdots - \boldsymbol{\Phi}_S L^S$. Then, it is easy to check that

$$\begin{aligned}\mathrm{V}(\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}) &= \mathrm{V}(\mathbf{A}(L)\boldsymbol{\xi}_t|\mathcal{F}_{t-1}) = \mathrm{V}\left(\left[\mathbf{I} - \boldsymbol{\Phi}_1 L - \cdots - \boldsymbol{\Phi}_S L^S\right]\boldsymbol{\xi}_t|\mathcal{F}_{t-1}\right) \\ &= \mathrm{V}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1}),\end{aligned} \tag{10}$$

since $\boldsymbol{\xi}_{t-k}$ is $\mathcal{F}_{t-1}$-measurable for $k \geq 1$.

Similarly, we have

$$\mathrm{V}(\mathbf{Y}_t|\mathcal{F}_{t-1}) = \mathrm{V}(\mathbf{A}(L)\mathbf{X}_t|\mathcal{F}_{t-1}) = \mathrm{V}(\mathbf{X}_t|\mathcal{F}_{t-1}). \tag{11}$$

Moreover, since $\mathbf{u}_t$ and $\boldsymbol{\xi}_t$ are conditionally uncorrelated, both $\mathrm{Cov}(\mathbf{R}\mathbf{u}_\mathrm{t}, \boldsymbol{\epsilon}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1})$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_\mathrm{t}, \mathbf{R}\mathbf{u}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1})$ in (9) equal zero. Whence,

$$\mathrm{Cov}(\mathbf{R}\mathbf{u}_\mathrm{t}, \boldsymbol{\epsilon}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1}) = \mathrm{Cov}(\mathbf{R}\mathbf{u}_\mathrm{t}, \mathbf{A}(\mathrm{L})\boldsymbol{\xi}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1}) = \mathbf{R}\mathrm{Cov}(\mathbf{u}_\mathrm{t}, \mathbf{A}(\mathrm{L})\boldsymbol{\xi}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1}).$$

Now,

$$\begin{aligned}&\mathrm{Cov}(\mathbf{u}_\mathrm{t}, \mathbf{A}(\mathrm{L})\boldsymbol{\xi}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1}) = \mathrm{Cov}(\mathbf{u}_\mathrm{t}, \left[\mathrm{I} - \Phi_1\mathrm{L} - ... - \Phi_\mathrm{S}\mathrm{L}^\mathrm{S}\right]\boldsymbol{\xi}_\mathrm{t}|\mathcal{F}_{\mathrm{t}-1}) \\ &\quad = \mathrm{E}(\mathbf{u}_t\left[\boldsymbol{\xi}_t - \boldsymbol{\Phi}_1\boldsymbol{\xi}_{t-1} - ... - \boldsymbol{\Phi}_S\boldsymbol{\xi}_{t-S}\right]'|\mathcal{F}_{t-1}) \\ &\qquad - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathrm{E}([\boldsymbol{\xi}_t - \boldsymbol{\Phi}_1\boldsymbol{\xi}_{t-1} - ... - \boldsymbol{\Phi}_S\boldsymbol{\xi}_{t-S}]'|\mathcal{F}_{t-1}) \\ &\quad = \mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_t'|\mathcal{F}_{t-1}) \\ &\qquad - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathrm{E}(\boldsymbol{\xi}_t'|\mathcal{F}_{t-1}) - \underbrace{\left[\mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_{t-1}'\boldsymbol{\Phi}_1'|\mathcal{F}_{t-1}) - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathrm{E}(\boldsymbol{\xi}_{t-1}'\boldsymbol{\Phi}_1'|\mathcal{F}_{t-1})\right]}_{\mathbf{0}} \\ &\qquad - \ldots - \underbrace{\left[\mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_{t-S}'\boldsymbol{\Phi}_S'|\mathcal{F}_{t-1}) - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathrm{E}(\boldsymbol{\xi}_{t-S}'\boldsymbol{\Phi}_S'|\mathcal{F}_{t-1})\right]}_{\mathbf{0}} \\ &\quad = \mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_t'|\mathcal{F}_{t-1}) - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathrm{E}(\boldsymbol{\xi}_t'|\mathcal{F}_{t-1}) = \mathrm{Cov}(\mathbf{u}_t\boldsymbol{\xi}_t'|\mathcal{F}_{t-1}) = \mathbf{0}\end{aligned}$$

since $\mathrm{Cov}(\mathbf{u}_t \boldsymbol{\xi}'_{t+k} | \mathcal{F}_{t-1}) = \mathbf{0}$ for any k. It then follows from (8)-(11), along with the fact that $\mathrm{Cov}(\boldsymbol{\epsilon}_\mathrm{t}, \mathbf{R}\mathbf{u}_\mathrm{t} | \mathcal{F}_{\mathrm{t}-1}) = 0$, that

$$\mathrm{V}(\mathbf{X}_t | \mathcal{F}_{t-1}) = \mathrm{V}(\mathbf{Y}_t | \mathcal{F}_{t-1}) = \mathbf{R}\mathrm{V}(\mathbf{u}_t | \mathcal{F}_{t-1})\mathbf{R}' + \mathrm{V}(\boldsymbol{\xi}_t | \mathcal{F}_{t-1}),$$

as was to be proved. □

3.2 Estimation

It follows from Proposition 1 that, if $\mathrm{V}(\mathbf{X}_t | \mathcal{F}_{t-1})$ is to be estimated at time $(t-1)$, assumptions have to be made on the dynamics of $\mathrm{V}(\mathbf{u}_t | \mathcal{F}_{t-1})$ and $\mathrm{V}(\boldsymbol{\xi}_t | \mathcal{F}_{t-1})$.

As in Alessi et al. (2009) and Aramonte et al. (2013), we therefore assume that the conditional covariance matrices of the common shocks can be modelled as some q-dimensional MGARCH process. Since q is typically small, this approach escapes the curse of dimensionality. As for the idiosyncratic conditional covariance matrix $\mathrm{V}(\boldsymbol{\xi}_t | \mathcal{F}_{t-1})$, since idiosyncratic cross-correlations are small, it can be approximated by a diagonal matrix where each diagonal element (each marginal conditional variance) is modelled by a univariate GARCH-type model—in the sequel, we use GARCH(1,1) models. In both cases, the MGARCH and the n GARCH(1,1) models are estimated by Gaussian quasi-maximum likelihood (QMLE) (we refer to Francq and Zakoian (2010) for sufficient consistency conditions).

In practice, the actual number of observed series is large, but finite: denote it by N. The estimation of $\mathrm{V}(\mathbf{X}_t | \mathcal{F}_{t-1})$ proceeds as follows.

- **Step 1.** Determine the number q of common shocks, for instance via the Hallin and Liška (2007) criterion.
- **Step 2.** Randomly reorder the N observed series.
- **Step 3.** Compute a consistent[10] estimator

$$\widehat{\boldsymbol{\Sigma}}^X_{NT}(\theta) = \frac{1}{2\pi} \sum_{k=-M_T}^{M_T} e^{-ik\theta} K\left(\frac{k}{B_T}\right) \widehat{\boldsymbol{\Gamma}}^X_k$$

[10]Consistency requires conditions on K, M_T and B_T, for which again we refer to Barigozzi and Hallin (2018).

of the $N \times N$ spectral density matrix of the $\mathbf{X}_t$'s, where $K(\cdot)$ is a kernel function, M_T a truncation parameter, B_T the bandwidth, and $\widehat{\mathbf{\Gamma}}_k^X$ the sample lag-k cross-covariance matrix computed from the observed $N \times T$ panel of $\mathbf{X}_t$ values.

- **Step 4.** Collecting the q normalized column eigenvectors associated with $\widehat{\mathbf{\Sigma}}_{NT}^X(\theta)$'s q largest eigenvalues into the $N \times q$ matrix $\widehat{P}_{NT}^X(\theta)$ (with complex conjugate $\widehat{P}_{NT}^{X*}$) and the corresponding eigenvalues into the $q \times q$ diagonal matrix $\widehat{\mathbf{\Lambda}}_{NT}^X(\theta_h)$, compute
$$\widehat{\mathbf{\Sigma}}_{NT}^{\chi}(\theta) := \widehat{P}_{NT}^X(\theta)\widehat{\mathbf{\Lambda}}_{NT}^X(\theta)\widehat{P}_{NT}^{X*}(\theta)$$
as an estimator of the spectral density matrix of the $\boldsymbol{\chi}_t$'s.

- **Step 5.** Let $N_* := m(q+1)$ with $m := \left\lceil \frac{N}{q+1} \right\rceil$. Dropping the last $N - m(q+1)$ series, denote by $\widehat{\mathbf{\Sigma}}_{N_*T}^{\chi}(\theta)$ the $N_* \times N_*$ spectral density matrix corresponding to the remaining N_* series[11].

- **Step 6.** By inverse Fourier transform of $\widehat{\mathbf{\Sigma}}_{N_*T}^{\chi}(\theta)$, compute the estimated autocovariance matrices $\widehat{\mathbf{\Gamma}}_k^{\chi}$ of the m $(q+1)$-dimensional sub-vectors $\boldsymbol{\chi}_t^k = (\chi_{(k-1)(q+1)+1,t} \cdots \chi_{k(q+1),t})'$, $k = 1, ..., m$. Then, from the latter, obtain, via Akaike order identification and Yule-Walker equations, estimators $\hat{\mathbf{A}}^k(L)$ of the m VAR filters $\mathbf{A}^k(L)$; stacking them into a block-diagonal matrix $\hat{\mathbf{A}}(L)$, compute the estimates $\widehat{\mathbf{Y}}_t := \hat{\mathbf{A}}(L)\mathbf{X}_t$.

- **Step 7.** Obtain the estimates $\widehat{\mathbf{Ru}_t}$ of $\mathbf{Ru}_t$ by computing the first q standard principal components of $\widehat{\mathbf{Y}}_t$; inverting[12] the block-diagonal filters $\hat{\mathbf{A}}(L)$ then using appropriate identification constraints, we obtain the identified quantities $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{u}}_t$, and the corresponding estimates of the impulse-response function $\widehat{\mathbf{B}}_n = [\hat{\mathbf{A}}(L)]^{-1}\widehat{\mathbf{R}}$.

Following Forni et al. (2017) we chose a Cholesky identification scheme to obtain the identification of $\widehat{\mathbf{R}}$ and $\widehat{\mathbf{u}}_t$ (see Section 4.1 of Forni et al. (2017) for more details)—other choices are possible, though.

[11] For the sake of simplicity we keep the same notation for the N_* reordered observed series.

[12] The inverse of $\hat{\mathbf{A}}(L)$ being the block-diagonal filter with $(q+1)\times(q+1)$ diagonal blocks $[\hat{\mathbf{A}}^k(L)]^{-1}$ where q is small; this inversion thus is easily performed.

Steps 1-7 are those described in Forni et al. (2015, 2017) and Barigozzi and Hallin (2018), where we refer to for details. The resulting estimator $\widehat{\boldsymbol{\chi}}_t$, however, depends on the ordering of the panel obtained at Step 2: that ordering indeed determines which elements of $\widehat{\boldsymbol{\Sigma}}^{\chi}_{NT}(\theta)$ are kept in $\widehat{\boldsymbol{\Sigma}}^{\chi}_{N_*T}(\theta)$ and belong to the diagonal blocks of $\widehat{\boldsymbol{\Sigma}}^{\chi}_{N_*T}(\theta)$. Forni et al. (2017) and Barigozzi and Hallin (2018) explain how to deal with this by iterating Steps 2-7 (going back to Step 2, choosing a new random permutation, hence a new N_*-dimensional subpanel, etc.) until numerical stabilization of the averaged (over the permutations) $\widehat{\boldsymbol{\chi}}_t$ values; this typically takes place after few iterations[13].

- **Step 8.** Iterate Steps 2 through 7; average (after obvious reordering of the cross-section) the resulting estimates $\widehat{\mathbf{R}}$, $\widehat{\mathbf{u}}_t$ and $\widehat{\mathbf{B}}_n$. Denote, for the sake of simplicity, the final estimates also by $\widehat{\mathbf{R}}$, $\widehat{\mathbf{u}}_t$ and $\widehat{\mathbf{B}}_n$. Let $\widehat{\boldsymbol{\chi}}_t := \widehat{\mathbf{B}}_n\widehat{\mathbf{u}}_t$ and $\widehat{\boldsymbol{\xi}}_t := \mathbf{X}_t - \widehat{\boldsymbol{\chi}}_t$.

The procedure described so far is the one that has been used in Della Marra (2017), Forni et al. (2018), and Giovannelli et al. (2018) in their forecasting of inflation and financial returns. In order to estimate conditional covariance matrices, we will now exploit the MGARCH and GARCH features of Assumption *(x)*. Thanks to the assumption of stability under aggregation, the choice of identification constraints has no impact, and VECH or BEKK QMLEs can be computed from the $\widehat{\mathbf{u}}_t$'s obtained in Step 8. We then proceed with the following final steps.

- **Step 9a.** Run, over the q-dimensional T-tuple $\widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_T$, a QML estimation procedure for the parameters of the MGARCH model of Assumption *(x)*; this yields an estimator $\widehat{\mathrm{V}}(\mathbf{u}_t|\mathcal{F}_{t-1})$ of $\mathrm{V}(\mathbf{u}_t|\mathcal{F}_{t-1})$.

- **Step 9b.** Similarly run, over each of the N univariate T-tuples $\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_T$, a GARCH QML estimation procedure. This yields N estimators $\widehat{\mathrm{v}}(\xi_{it}|\mathcal{F}^{\xi_i}_{t-1})$ of the variances $\mathrm{v}(\xi_{it}|\mathcal{F}^{\xi_i}_{t-1})$ of the ξ_{it}'s conditional on their past values; the $N \times N$ diagonal matrix $\widehat{\mathrm{V}}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1})$ with diagonal entries $\widehat{\mathrm{v}}(\xi_{it}|\mathcal{F}^{\xi_i}_{t-1})$ is our estimator of $\mathrm{V}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1})$.

[13] Averaging, of course, is performed after rearrangement of the cross-sectional items in the original ordering.

Our estimator $\widehat{\mathrm{V}}(\mathbf{X}_t|\mathcal{F}_{t-1})$ finally is defined as

$$\widehat{\mathrm{V}}(\mathbf{X}_t|\mathcal{F}_{t-1}) := \widehat{\mathbf{R}}\widehat{\mathrm{V}}(\mathbf{u}_t|\mathcal{F}_{t-1})\widehat{\mathbf{R}}' + \widehat{\mathrm{V}}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1}). \tag{12}$$

The following proposition establishes its consistency properties.

Proposition 2. *Assume that $B_T = o(\sqrt{T})$ and $M_T = o(\sqrt{T})$. Under Assumptions (i)-(x) and Assumptions (L4) and (L5) in Barigozzi and Hallin (2018), we have*

$$\widehat{\mathrm{V}}(\mathbf{X}_t|\mathcal{F}_{t-1}) - \mathrm{V}(\mathbf{X}_t|\mathcal{F}_{t-1}) = o_{\mathrm{P}}(1) \tag{13}$$

for any $t \in \mathbb{Z}$ as $n, T \to \infty$ with $n = O(T^c)$ for some finite $c > 0$.

Proof. It follows from Proposition 1 in Barigozzi and Hallin (2018) that, under the assumptions made, letting $\rho_{nT} := \max\left(B_T/\sqrt{T}, 1/B_T, 1/\sqrt{n}\right)$,

$$\frac{1}{\sqrt{n}}\|\widehat{\mathbf{R}} - \mathbf{R}\mathbf{J}\| = \mathrm{O_P}(\rho_{\mathrm{nT}}), \quad \text{and} \quad \max_{\mathrm{t}=1,\ldots,\mathrm{T}} \|\hat{\mathbf{u}}_{\mathrm{t}} - \mathbf{J}\mathbf{u}_{\mathrm{t}}\| = \mathrm{O_P}(\rho_{\mathrm{nT}} \log \mathrm{T}),$$

for some $q \times q$ diagonal matrix $\mathbf{J}$ with entries ± 1, and

$$\max_{1\le i\le n} \max_{1\le t\le T} |\widehat{\xi}_{it} - \xi_{it}| = O_{\mathrm{P}}(\rho_{nT} \log \mathrm{T}).$$

Consequently, $\widehat{\mathbf{R}} - \mathbf{R}\mathbf{J}$, $\hat{\mathbf{u}}_t - \mathbf{J}\mathbf{u}_t$ and $\widehat{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t$ all are $o_{\mathrm{P}}(1)$. The same "two-step estimator" arguments as in Proposition 4 of Alessi et al. (2009) thus apply: since $\hat{\mathbf{u}}_t$ and $\widehat{\xi}_{it}$ consistently estimate $\mathbf{u}_t$ and ξ_{it} in "the first step", computing in "the second step" a maximum likelihood estimator from $\hat{\mathbf{u}}_t$ and $\widehat{\xi}_{it}$ is asymptotically equivalent to computing it from the actual values $\mathbf{u}_t$ and $\boldsymbol{\xi}_t$, and thus leads to consistent estimates of $\mathrm{V}(\mathbf{J}\mathbf{u}_t|\mathcal{F}_{t-1})$ and $\mathrm{V}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1})$, respectively. Now,

$$\mathbf{R}\mathbf{J}\mathrm{V}(\mathbf{J}\mathbf{u}_t|\mathcal{F}_{t-1})\mathbf{J}'\mathbf{R}' = \mathbf{R}\mathbf{J}\mathbf{J}\mathrm{V}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathbf{J}\mathbf{J}\mathbf{R}' = \mathbf{R}\mathrm{V}(\mathbf{u}_t|\mathcal{F}_{t-1})\mathbf{R}',$$

so that

$$\widehat{\mathbf{R}}\widehat{\mathrm{V}}(\mathbf{u}_t|\mathcal{F}_{t-1})\widehat{\mathbf{R}}' + \widehat{\mathrm{V}}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1}) - \mathbf{R}\mathbf{J}\mathrm{V}(\mathbf{J}\mathbf{u}_t|\mathcal{F}_{t-1})\mathbf{J}'\mathbf{R}' - \mathrm{V}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1}) = o_{\mathrm{P}}(1)$$

implies

$$\widehat{\mathbf{R}}\widehat{\mathrm{V}}(\mathbf{u}_t|\mathcal{F}_{t-1})\widehat{\mathbf{R}}' + \widehat{\mathrm{V}}(\boldsymbol{\xi}_t|\mathcal{F}_{t-1}) - \mathrm{V}(\mathbf{X}_t|\mathcal{F}_{t-1}) = o_{\mathrm{P}}(1),$$

as was to be proved. □

In practice, VECH and BEKK QMLEs, however, are numerically quite unstable, and typically strongly depend on the initial values considered in the numerical solution of the likelihood equations. This is a well-documented fact; see, for instance, Lien et al. (2002) and Asai (2015). Rather than VECH or BEKK, we therefore compute DCC QMLEs which are known to be quite robust to missespecification; see Chang et al. (2011), Chevallier (2012), Laurent et al. (2012), Amendola and Candila (2017), or de Almeida et al. (2018). Our Monte Carlo experiments (see Section 4) confirm that, even though the actual data-generating process is BEKK, misspecified DCC QMLEs outperform the correctly specified full BEKK ones.

4 Finite-sample performances

4.1 Monte Carlo experiments

In this section, we investigate the finite-sample performance of the proposed procedure through Monte Carlo simulations.

Simulations were performed from four data generating processes (DGPs). The first two DGPs are static factor models with one and two common factors, respectively; the third and fourth DGPs are dynamic factor models with finite and infinite-dimensional factor spaces, respectively. The common shocks and the idiosyncratic components in all four cases are conditionally heteroscedastic. The first three DGPs are particular cases of the GDFM with static representation and can be consistently estimated by the procedure of Alessi et al. (2009) which, however, cannot consistently estimate the fourth DGP, where the assumption of a finite-dimensional factor space does not hold.

In all DGPs, the idiosyncratic components satisfy $\boldsymbol{\xi}_t|\mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{P}_t)$, where $\mathbf{P}_t$ is an $N \times N$ diagonal matrix containing the conditional variances P_{it} of GARCH(1,1) processes of the form

$$P_{it} = \omega_i + \alpha_i \xi_{it}^2 + \beta_i P_{i,t-1}, \quad i = 1, ..., N,$$

where $\omega_i > 0$, $\alpha_i, \beta_i \geq 0$ and $\alpha_i + \beta_i < 1$; the parameters values α_i and β_i are generated independently from uniform distributions over [0.01, 0.045] and [0.85, 0.95], respectively, and $\omega_i := 1 - \alpha_i - \beta_i$, so that the unconditional variance of ξ_{it}

is $\mathrm{V}(\xi_{it}) = 1$. As for the factors $\mathbf{u}_t$ driving the common components $\boldsymbol{\chi}_t$, they were generated from the following four DGPs.

DGP1. (one common shock; static loadings) One common shock u_t is generated from a univariate GARCH(1,1) model

$$u_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2) \quad \text{with } \sigma_t^2 = 1 + 0.07u_{t-1}^2 + 0.83\sigma_{t-1}^2;$$

here $\boldsymbol{\chi}_t = \mathbf{R}u_t$, where $\mathbf{R}$ is an $N \times 1$ matrix with modulus one randomly generated via the *RandOrthMat* Matlab function.

DGP2. (two common shocks; static loadings) Two common shocks $\mathbf{u}_t = (u_{1t}, u_{2t})'$, generated from a BEKK(1,1,1) model

$$\mathbf{u}_t|\mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{Q}_t) \;\; \text{with } \mathbf{Q}_t = \mathbf{C}_0'\mathbf{C}_0 + \mathbf{C}_1'\mathbf{u}_{t-1}\mathbf{u}_{t-1}'\mathbf{C}_1 + \mathbf{C}_2'\mathbf{Q}_{t-1}\mathbf{C}_2. \qquad (14)$$

In order to guarantee $\mathrm{E}(\mathbf{Q}_t) = \mathrm{E}(\mathbf{u}_{t-1}\mathbf{u}_{t-1}') = \mathbf{I}_q$, we set $\mathbf{C}_0'\mathbf{C}_0 = \mathbf{I}_q - \mathbf{C}_1'\mathbf{C}_1 - \mathbf{C}_2'\mathbf{C}_2$. Parameters of the BEKK are extracted from uniform distributions with ranges as in Alessi et al. (2009): $\mathbf{C}_1$ has diagonal in [0.1,0.5] and off-diagonal elements in [-0.2,0.2], while $\mathbf{C}_2$ has diagonal in [0.8,0.95] and off-diagonal elements in [-0.15,0.15]. At each extraction of the parameters, covariance stationary of the BEKK model has been checked before proceeding. Here, $\boldsymbol{\chi}_t = \mathbf{R}\mathbf{u}_t$ where $\mathbf{R}$ is an $N \times 2$ matrix with orthonormal columns randomly generated via the *RandOrthMat* Matlab function.

DGP3. (four factors driven by $q = 2$ common shocks; static loadings) Four factors $\mathbf{F}_t = (F_{1t}, \dots, F_{4t})'$ driven by $q = 2$ common shocks $\mathbf{u}_t$, yielding a GDFM with finite-dimensional factor space. The shocks are generated from the same BEKK model as in DGP2, the factors are a VAR(4) driven by $\mathbf{u}_t$:

$$\mathbf{F}_t = \boldsymbol{\Phi}\mathbf{F}_{t-1} + \mathbf{K}\mathbf{u}_t \quad \text{and} \quad \mathbf{u}_t|\mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{Q}_t),$$

with $\mathbf{Q}_t$ as in (14), $\boldsymbol{\Lambda}$ is $n \times 4$, $\boldsymbol{\Phi}$ is 4×4 and $\mathbf{K}$ is 4×2. The entries of $\boldsymbol{\Lambda}$ and $\mathbf{K}$ are independently uniformly distributed over $[-1, 1]$. The entries of $\boldsymbol{\Phi}$ are generated as follows: first we generate entries independently uniformly distributed on the interval[-1,1]; second, we divide the resulting matrix by its spectral norm; third, we multiply the resulting matrix by a random variable uniformly distributed on the interval [0.4,0.9] to ensure stationarity while preserving sizeable dynamic responses[14].

DGP4. (two common shocks; dynamic loadings) The two common shocks $\mathbf{u}_t = (u_{1t}, u_{2t})'$ are generated from the same bivariate BEKK model as in (14); the model is a GDFM with infinite-dimensional factor space. Here,

$$\boldsymbol{\chi}_{it} = \begin{pmatrix} a_{i1}(1-\alpha_{i1})^{-1} \\ a_{i2}(1-\alpha_{i2})^{-1} \end{pmatrix} \mathbf{u}_t,$$

where a_{ij} and α_{ij}, $i = 1, ..., n$, $j = 1, 2$ are independent and uniformly distributed over the intervals [-1,1] and [-0.8,0.8], respectively.

For each DGP, we simulated 500 replications of a panel of dimensions N=60 and T=1000. From each replication, the conditional covariance matrix $\boldsymbol{\Sigma}_{T+1|T}$ was estimated using

(a) classical PCA[15] combined with (M)GARCH modelling,

(b) the DCC model with composite likelihood, as described in Pakel et al. (2017),

(c) the procedure of Alessi et al. (2009), and

(d) our method,[16]

denoted as PCA-(M)GARCH, DCC, ABC, and GDFM-CHF, respectively[17]. For simplicity, the correct numbers of factors (for DGP3) and common shocks (for DGPs 1-4) are assumed to be known, since this does not play a role in the comparative performances of procedures (a)-(d). For DGP4, since there are not static factors in its representation, the identification procedure by Bai and Ng (2002) was used in each simulated panel to compute the number of static factors for the estimation of the PCA-(M)GARCH and ABC procedures.[18]

[14]This DGP is similar to the one considered by Alessi et al. (2009).

[15]In the spirit of Diebold and Nerlove (1989) and Van der Weide (2002), static factors are extracted via principal component analysis; an (M)GARCH model then is fitted to the extracted factors. Idiosyncratic components are modelled as independent univariate GARCH processes.

[16]Throughout, we considered 30 cross-sectional permutations and set the order S of the VAR block-diagonal models to one.

[17]GDFM-CHF: General Dynamic Factor Model with Conditionally Heteroscedastic Factors.

[18]In practice, the identification procedures by Bai and Ng (2002) or Alessi et al. (2010) in the static case, by Hallin and Liška (2007) in the GDFM case, should be used prior to the estimation procedure in each replication.

As mentioned in the previous section, estimation of BEKK models is numerically quite unstable and strongly depends on the choice of initial values. For the sake of comparison, for DGPs 2-4 we considered both a DCC(1,1) and a BEKK(1,1,1) estimate of the conditional covariance matrix of the common shocks in the PCA-(M)GARCH, ABC and GDFM-CHF procedures: the DCC-based procedures are denoted as PCA-(M)GARCH-DCC and ABC-DCC and GDFM-CHF-DCC, the BEKK-based ones as PCA-(M)GARCH-BEKK, ABC-BEKK and GDFM-CHF-BEKK, respectively.[19]

In order to compare the performances of those four procedures, we compute, for each simulated panel and each method, a distance between the estimated one-step-ahead conditional covariance matrix $\widehat{\boldsymbol{\Sigma}}_{T+1|T}$ and the theoretical one $\boldsymbol{\Sigma}_{T+1|T}$. Let

$$\mathbf{H}_{T+1|T} := \mathbf{R}\,\mathrm{V}(\mathbf{u}_{\mathrm{T}+1}|\mathcal{F}_{\mathrm{T}})\mathbf{R}' + \mathrm{V}(\boldsymbol{\xi}_{\mathrm{T}+1}|\mathcal{F}_{\mathrm{T}}) \quad \text{for DGP1 and DGP2,}$$

$$\mathbf{H}_{T+1|T} := \boldsymbol{\Lambda}\mathbf{K}\mathrm{V}(\mathbf{u}_{T+1}|\mathcal{F}_{T})\mathbf{K}'\boldsymbol{\Lambda}' + \mathrm{V}(\boldsymbol{\xi}_{T+1}|\mathcal{F}_{T}) \quad \text{for DGP3,}$$

and

$$\mathbf{H}_{T+1|T} = \mathbf{A}\,\mathrm{V}(\mathbf{u}_{T+1}|\mathcal{F}_{T})\mathbf{A}' + \mathrm{V}(\boldsymbol{\xi}_{T+1}|\mathcal{F}_{T}) \quad \text{for DGP4,}$$

where $\mathbf{A}$ is the matrix with elements $a_{i,j}$, $i = 1, ..., N$, $j = 1, 2$. Following Amendola and Candila (2017), we consider four distances, $\mathrm{D}_1, , \ldots, \mathrm{D}_4$, of the form

$$\mathrm{D}(\mathbf{H}_{T+1|T}, \widehat{\boldsymbol{\Sigma}}_{T+1|T}) = \sum_{i=1}^{N}\sum_{j=i}^{N} \omega(i,j)(h_{i,j} - \widehat{\sigma}_{i,j})^2, \tag{15}$$

where $h_{i,j}$ and $\widehat{\sigma}_{i,j}$ are the (i, j) entries of $H_{T+1|T}$ and $\widehat{\boldsymbol{\Sigma}}_{T+1|T}$, respectively, and the weights $\omega(i, j)$ are provided in Table 1.

Distance D_1, which gives equal weights for the variance and covariances, yields a "total" unweighted squared Euclidean distance between $\mathrm{Vech}(\widehat{\boldsymbol{\Sigma}}_{T+1|T})$ and $\mathrm{Vech}(\mathbf{H}_{T+1|T})$; distance D_2 is an unweighted squared Euclidean distance between $\mathrm{Diag}(\widehat{\boldsymbol{\Sigma}}_{T+1|T})$ and $\mathrm{Diag}(\mathbf{H}_{T+1|T})$ (hence disregards the covariances);[20] distance D_3 penalizes negative errors, while D_4 penalizes the positive ones. It is important to note that, in D_3

[19] DCC and BEKK estimations were performed by using the MFE toolbox of Kevin K. Sheppard, freely available at http://www.kevinsheppard.com/MFE_Toolbox.

[20] The classical notation $\mathrm{Vech}(\mathbf{M})$ stands for the vector stacking the upper diagonal entries of a square matrix $\mathbf{M}$, and $\mathrm{Diag}(\mathbf{M})$ for the vector of its diagonal elements.

Table 1: Weights $\omega(i,j)$, $i = 1,\ldots,N$, $j = i,\ldots,N$ in the distances D_1-D_4 in (15).

D_1	$w(i,j) = 1$	for all i and j
D_2	$w(i,j) = 1$	when $i = j$; 0 otherwise
D_3	$w(i,j) = 2$	when $\widehat{\sigma}_{i,j} > h_{i,j}$; 1 otherwise
D_4	$w(i,j) = 2$	when $\widehat{\sigma}_{i,j} < h_{i,j}$; 1 otherwise

and D_4, the weights themselves are data-driven, so that, for a given replication, different methods lead to different weights.

4.2 Simulation results

The results of the Monte Carlo experiments are summarized in Figures 1-4 and Table 2. Figures 1-4 present boxplots of the distances defined in (15), in logarithmic scale and for DGP1, DGP2, DGP3, and DGP4, respectively. Table 2 reports the number of times each estimation procedure achieves the smallest values of the distances for each DGP.

FIGURES 1-4 and TABLE 2 AROUND HERE

Inspection of Figure 1 (DGP1) reveals that ABC and GDFM-CHF perform as well as the simpler PCA-(M)GARCH procedure (with higher variability for GDFM-CHF, though), while DCC is, by far, the worst. According to Figures 2-3, the BEKK-based procedures present much higher variability than the DCC-based ones due, probably, to the numerical instability of BEKK QMLEs. Even when misspecified, DCC-based methods thus are preferable. In Figures 3 (DGP3) and 4 (DGP4), we can observe the good performance of GDFM-CHF-DCC, while ABC-DCC for DGP4, as well as PCA-(M)GARCH-DCC and DCC for DGP3 and DGP4, perform quite poorly.

Due to the high instability of BEKK-based procedures, Table 2 only reports the DCC-based procedures. It appears clearly that, in agreement with the results in Figures 1-4, the DCC method performs worst, except for DGP2. For DGP1 and DGP2, the GDFM-CHF-DCC procedure overperforms PCA-(M)GARCH-DCC and ABC-DCC for all distances but D_2 (where only the conditional variances, not the covariances, are taken into account). In the DGP3 case, the GDFM-CHF-DCC procedure is best for all distances, closely followed by ABC. Finally, for DGP4, the GDFM-CHF-DCC procedure is by far the best for all distances while ABC-DCC performs poorly and PCA-(M)GARCH-DCC even worse. When both conditional variances and covariances are considered (distances D1, D3, and D4), the GDFM-CHF-DCC procedure, irrespective of the DGP, is uniformly best.

Table 2: For each choice of a DGP (DGP1-DGP4) and a distance (D_1-D_4), this table provides the number of times each of the four estimation procedures (PCA-(M)GARCH, DCC, ABC and GDFM-CHF) is the winner across 500 Monte Carlo replications. For DGPs 2-4 we use the DCC-based versions of the PCA-(M)GARCH, ABC, and GDFM-CHF procedures. Highest values are in bold.

	DGP1				DGP2			
Procedure	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4
PCA-(M)GARCH	103	155	114	88	35	75	39	34
DCC	13	38	13	12	45	**214**	45	43
ABC	92	**164**	82	109	59	87	53	62
GDFM-CHF	**292**	143	**291**	**291**	**361**	124	**363**	**361**
	DGP3				DGP4			
Procedure	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4
PCA-(M)GARCH	42	67	41	40	9	1	11	7
DCC	19	7	20	20	3	1	4	3
ABC	211	208	207	219	92	80	91	91
GDFM-CHF	**228**	**218**	**232**	**221**	**396**	**418**	**394**	**399**

5 An application to dynamic portfolio optimization

In this section, we assess our proposal (GDFM-CHF-DCC) in the problem of dynamic portfolio optimisation. The dataset we are considering consists in returns X_{it} from stocks entering the composition of the S&P 500 index, the National Association of Securities Dealers Automated Quotations (NASDAQ-100) and the NYSE

Amex Composite Index (AMEX), on July 27, 2018 and traded from January 2, 2011 through June 29, 2018 (T=1884). It was obtained from *Yahoo Finance* using the R package *quantmod* by Ryan and Ulrich (2017). Because we only considered stocks traded through the whole period, we ended up with $N = 656$ assets.

A window size of 750 days is used for estimation, which represents a concentration ratio of $656/750 = 0.875$; the out-of-sample period was set to 1134 days. An estimator $\widehat{\mathbf{\Sigma}}_{t+1|t}$ of $\mathrm{V}(\boldsymbol{X}_{t+1}|\mathcal{F}_t)$ is computed from the 656×750 subpanel $\{X_{is}|1 \leq i \leq 656,\, t-749 \leq s \leq t\}$ for $t = 750, \ldots, T-1 = 1883$. That estimator is used in the construction, at times $t = 750, \ldots, 1883$ (1134 time points), of a one-step ahead minimal variance portfolio (optimality at time $t+1$)—viz., a vector of weights

$$\widehat{\boldsymbol{\omega}}_{t+1|t} = (\widehat{\omega}_{1;t+1|t}, \ldots, \widehat{\omega}_{656;t+1|t})' = \operatorname*{argmin}_{\boldsymbol{\omega}} \boldsymbol{\omega}' \widehat{\mathbf{\Sigma}}_{t+1|t} \boldsymbol{\omega}$$

where minimisation is with respect to all $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{656})'$ such that $\omega_i \geq 0$ and $\sum_{i=1}^{656} \omega_i = 1$. The resulting (out-of-sample) portfolio return

$$r_{p,t+1} := \sum_{i=1}^{656} \widehat{\omega}_{i;t+1|t} X_{i,t+1}$$

at time $t+1$ then is computed from the observation at time $t+1$.

The minimum-variance portfolio we are proposing is the one based on $\widehat{\mathbf{\Sigma}}_{t+1|t} = \widehat{\mathrm{V}}(\mathbf{X}_{t+1}|\mathcal{F}_t)$, as described in Section 3.2 (but computed from the adequate subpanels), denoted as GDFM-CHF-DCC. For the sake of comparison, we also include the results of the GDFM-CHF-BEKK procedure. We compare its performance with those of (a) the naive equal-weighted portfolio strategy, denoted here by 1/N, (b) the RiskMetrics 2006 methodology (Zumbach, 2007), (c) the OGARCH approach of Alexander and Chibumba (1996), (d) the ABC method of Alessi et al. (2009), (e) the generalized principal volatility components (GPVC)[21] of Li et al. (2016), and (f) the procedure called PCA4TS proposed by Chang et al. (2018), which ex-

[21] A robust version of the GPVC procedure, denoted by RPVC, was proposed by Trucíos et al. (2019). That procedure is based on a robust estimator of the unconditional covariance matrix which can be applied only when the concentration ratio N/T is lower than 0.5. For this reason we did not implement it here. Of course, an adequate robust estimator in an high-dimensional context would be welcome. However, the performance of the RPVC in a $N/T > 0.5$ context has not been analyzed yet.

tends the principal component analysis to second-order stationary vector time series. Those procedures were selected for their feasibility in high-dimensional data.

The GDFM-CHF with DCC or BEKK was implemented with 30 cross-sectional permutations; the order of the VAR block-diagonal models was set to $S = 1$. In practice (when one portfolio is to be estimated at a time), information criteria can be used to determine the order of those VARs. Likewise, following Alessi et al. (2009), the number of static factors, common shocks, volatility components (Li et al., 2016) and groups (Chang et al., 2018) were determined once for all.

The ABC-DCC procedure (Alessi et al., 2009) was implemented with eight static factors and three common shocks determined by the criteria of Bai and Ng (2002) and Hallin and Liška (2007), respectively. The same number of common shocks was used in the GDFM-CHF approach. The GPVC procedure was applied with eight volatility components determined by the criterion of Bai and Ng (2002), the PCA4TS one with 654 groups (two of them with two assets and the remaining ones with only one asset; the groups were obtained following Chang et al. (2018)). The OGARCH procedure was applied as in Becker et al. (2015), that is, with the number of components equal to the number of series.

Following Gambacciani and Paolella (2017), Trucíos et al. (2018), or Engle et al. (2019), among many others, we use annualized performance measures to evaluate out-of-sample portfolio performances. These measures are defined as follows.

(i) Annualized average portfolio (AV):

$$\text{AV} := 252\bar{r}_p = 252\left[\frac{1}{1134}\sum_{t=750}^{1883} r_{p,t+1}\right]$$

(average of the out-of-sample portfolio returns multiplied by 252);

(ii) Annualized standard deviation (SD):

$$\text{SD} = \sqrt{252}\left[\frac{1}{1134}\sum_{t=750}^{1883}(r_{p,t+1} - \bar{r}_p)^2\right]^{1/2}$$

(standard deviation of the out-of-sample portfolio return multiplied by $\sqrt{252}$);

(iii) Annualized information ratio (AV): IR = AV/SD;

(iv) Annualized Sortino's ratio (SR): $\text{SR} = \text{AV}/\left(\text{S}\sqrt{252}\right)$, where

$$S = \left[\frac{1}{1134} \sum_{t=750}^{1883} \min\left(0, r_{p,t+1} - \mathrm{MAR}\right)^2 \right]^{1/2},$$

and the minimal accepted return (MAR) is set to zero.

Because our objective is the selection of a minimum variance portfolio, the most pertinent performance measure should be the SD criterion, as stressed out also by Ledoit and Wolf (2017) and Engle et al. (2019).

The results are reported in Table 3. They reveal that the best performance, for the SD, IR and SR criteria, is achieved by the GDFM-CHF-DCC. The OGARCH model has the second best performance, according to the SD criterion, followed by the ABC-DCC method. The GPVC and the OGARCH procedures exhibit the worst performance according to the AV criterion while ABC has the best performance according to the same criterion, followed by the GDFM-CHF-DCC proposal. The worst out-of-sample performance is obtained by the equal-weight portfolio strategy according to all criteria, but for the AV one. It is worth noting the relative good performance of RM2006, which outperforms GPVC and PCA4TS according to all criteria and loses for OGARCH only through the SD criterium. Finally, note that the results of GDFM-CHF-BEKK are worse than those of GDFM-CHF-DCC, mainly in terms of the SD criterion. This is not surprising since, as mentioned previously, the estimation of the Full BEKK model is hard, unstable and strongly dependent on the initial values, leading to a poor performance (Lien et al., 2002; Laurent et al., 2012; Asai, 2015; Amendola and Candila, 2017; de Almeida et al., 2018).

Taking into account all criteria, the GDFM-CHF-DCC proposal exhibits the best performance, followed by the ABC-DCC procedure.

Table 3: Annualized performance measures: AV, SD, IR and SR stand for the annualized average, standard deviation, information ratio and Sortino's ratio of the out-of-sample portfolio returns, respectively. The dataset is formed by 656 stocks used in the composition of the S&P500, NASDAQ-100 and AMEX indexes and the window size for estimation is equal to 750 days (concentration ratio N/T equal to 0.875). The out-of-sample period goes from January 2, 2014 to June 29, 2018. A ranking of the various methods is provided in parenthesis for each criterion.

	AV	SD	IR	SR
1/N	5.7708 (3)	11.5067 (8)	0.5015 (8)	0.6834 (8)
RM2006	5.5983 (4)	4.5447 (4)	1.2318 (3)	1.7229 (3)
OGARCH	4.9227 (7)	4.4551 (2)	1.1050 (6)	1.5614 (6)
ABC-DCC	6.5267 (1)	4.5313 (3)	1.4404 (2)	1.9677 (2)
GPVC	4.5989 (8)	4.5889 (5)	1.0022 (7)	1.4077 (7)
PCA4TS	5.3677 (6)	4.7255 (6)	1.1359 (5)	1.6024 (5)
GDFM-CHF-DCC	6.2369 (2)	4.0209 (1)	1.5511 (1)	2.2137 (1)
GDFM-CHF-BEKK	5.5834 (5)	4.8954 (7)	1.1405 (4)	1.6287 (4)

6 Conclusions

Based on the one-sided procedure of Forni et al. (2015, 2017) and Barigozzi and Hallin (2018), we propose a forecasting method for the conditional covariance matrix in high-dimensional time series, which we apply to dynamic portfolio optimization.

A Monte Carlo performance comparison of our method with alternative methods is conducted over four different DGPs, using the distance measures proposed in Amendola and Candila (2017). Overall, our method has an excellent performance, and outperforms all its competitors—except, under static factor model DGPs, for the distance D2 which disregards the covariances.

The superiority of our estimator is also empirically established in an application to dynamic portfolio optimisation based on a dataset of 656 assets. Our method achieves the best out-of-sample performance according to the (annualized) standard deviation SD (arguably, the most relevant criterion in the context), information ratio (IR) and Sortino's ratio (SR) criteria, and is second best (after Alessi et al. (2009))

with respect to the (annualized) average criterion.

References

Aguilar, M. (2009). A latent factor model of multivariate conditional heteroscedasticity. *Journal of Financial Econometrics*, 7(4):481–503.

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

Alessi, L., Barigozzi, M., and Capasso, M. (2009). Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors. Working paper series 1115, European Central Bank, Frankfurt am Main, Germany.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate static factor models. *Statistics and Probability Letters*, 80:1806–1813.

Alexander, C. O. and Chibumba, A. (1996). Multivariate orthogonal factor GARCH. *University of Sussex Discussion Papers in Mathematics*.

Amendola, A. and Candila, V. (2017). Comparing multivariate volatility forecasts by direct and indirect approaches. *Journal of Risk*, 19(6):33–57.

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.

Asai, M. (2015). Initial values on quasi-maximum likelihood estimation for BEKK multivariate GARCH models. *Working paper, Soka University*, 44(1):45–52.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Barhoumi, K., Darné, O., and Ferrara, L. (2014). Dynamic factor models: A review of the literature. *OECD Journal of Business Cycle Measurement and Analysis*, 2013(2):73.

Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal*, 19(1):C33–C60.

Barigozzi, M. and Hallin, M. (2017). Generalized dynamic factor models and volatilities: estimation and forecasting. *Journal of Econometrics*, 201(2):307–321.

Barigozzi, M. and Hallin, M. (2018). Generalized dynamic factor models and volatilities: Consistency, rates, and prediction intervals. *arXiv preprint:1811.10045*.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Becker, R., Clements, A., Doolan, M., and Hurn, A. (2015). Selecting volatility forecasting models for portfolio allocation purposes. *International Journal of Forecasting*, 31(3):849–861.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, volume 36. SIAM.

Chang, C.-L., McAleer, M., and Tansuchat, R. (2011). Crude oil hedging strategies using dynamic multivariate GARCH. *Energy Economics*, 33(5):912–923.

Chang, J., Guo, B., and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics*, 46(5):2094–2124.

Chevallier, J. (2012). Time-varying correlations in oil, gas and CO2 prices: an application using BEKK, CCC and DCC-MGARCH models. *Applied Economics*, 44(32):4257–4274.

de Almeida, D., Hotta, L. K., and Ruiz, E. (2018). MGARCH models: Trade-off between feasibility and flexibility. *International Journal of Forecasting*, 34(1):45–63.

Della Marra, F. (2017). A forecasting performance comparison of dynamic factor models based on static and dynamic methods. *Communications in Applied and Industrial Mathematics*, 8(1):43–66.

Diebold, F. X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics*, 4(1):1–21.

Dovonon, P. (2013). Conditionally heteroskedastic factor models with skewness and leverage effects. *Journal of Applied Econometrics*, 28(7):1110–1137.

Engle, R. (2009). *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press, New Jersey.

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

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Fan, J., Wang, M., and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society: series B (Statistical Methodology)*, 70(4):679–702.

Forni, M., Giovannelli, A., Lippi, M., and Soccorsi, S. (2018). Dynamic factor model with infinite-dimensional factor space: forecasting. *Journal of Applied Econometrics*, 33(5):625–642.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185(2):359–371.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: asymptotic analysis. *Journal of Econometrics*, 199(1):74–92.

Forni, M. and Lippi, M. (2011). The general dynamic factor model: One-sided representation results. *Journal of Econometrics*, 163(1):23–28.

Francq, C. and Zakoian, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Gambacciani, M. and Paolella, M. S. (2017). Robust normal mixtures for financial portfolio allocation. *Econometrics and Statistics*, 3:91–111.

García-Ferrer, A., González-Prieto, E., and Peña, D. (2012). A conditionally heteroskedastic independent factor model with an application to financial stock returns. *International Journal of Forecasting*, 28(1):70–93.

Giovannelli, A., Massacci, D., and Soccorsi, S. (2018). Forecasting stock returns with large dimensional factor models. *Working Paper available at SSRN: https://ssrn.com/abstract=2958491*.

Hallin, M., Hörmann, S., and Lippi, M. (2018). Optimal dimension reduction for high-dimensional and functional time series. *Statistical Inference for Stochastic Processes*, 21(2):385–398.

Hallin, M. and Lippi, M. (2013). Factor models in high-dimensional time series—a time-domain approach. *Stochastic Processes and Their Applications*, 123(7):2678–2695.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Han, Y. (2005). Asset allocation with a high dimensional latent factor stochastic volatility model. *The Review of Financial Studies*, 19(1):237–271.

Harvey, A., Ruiz, E., and Sentana, E. (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics*, 52(1-2):129–157.

Hlouskova, J., Schmidheiny, K., and Wagner, M. (2009). Multistage predictions for multivariate GARCH models: Closed form solution and the value for portfolio management. *Journal of Empirical Finance*, 16(2):330–336.

Hu, Y.-P. and Tsay, R. S. (2014). Principal volatility component analysis. *Journal of Business & Economic Statistics*, 32(2):153–164.

Laurent, S., Rombouts, J. V., and Violante, F. (2012). On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics*, 27(6):934–955.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.

Li, W., Gao, J., Li, K., and Yao, Q. (2016). Modeling multivariate volatilities via latent common factors. *Journal of Business & Economic Statistics*, 34(4):564–573.

Lien, D., Tse, Y. K., and Tsui, A. K. (2002). Evaluating the hedging performance of the constant-correlation GARCH model. *Applied Financial Economics*, 12(11):791–798.

Matteson, D. S. and Tsay, R. S. (2011). Dynamic orthogonal components for multivariate time series. *Journal of the American Statistical Association*, 106(496):1450–1463.

Pakel, C., Shephard, N., Sheppard, K., and Engle, R. (2017). Fitting vast dimensional time-varying covariance models. *Manuscript NYU and revision of SSRN, 1354497*.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95:365–379.

Peña, D. and Box, G. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82:836–843.

Ryan, J. A. and Ulrich, J. M. (2017). *quantmod: Quantitative Financial Modelling Framework*. R package version 0.4-12.

Santos, A. A. and Moura, G. V. (2014). Dynamic factor multivariate GARCH model. *Computational Statistics & Data Analysis*, 76:606–617.

Sentana, E., Calzolari, G., and Fiorentini, G. (2008). Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks. *Journal of Econometrics*, 146(1):10–25.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stoffer, D. (1999). Detecting common signals in multiple time series using the spectral envelope. *Journal of the American Statistical Association*, 94:1341–1356.

Trucíos, C., Hotta, L. K., and Pereira, P. L. V. (2019). On the robustness of the principal volatility components. *Journal of Empirical Finance*, 52(1):201–219.

Trucíos, C., Hotta, L. K., and Ruiz, E. (2018). Robust bootstrap densities for dynamic conditional correlations: implications for portfolio selection and Value-at-Risk. *Journal of Statistical Computation and Simulation*, 88(10):1976–2000.

Van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564.

Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003). A full-factor multivariate garch model. *The Econometrics Journal*, 6(2):312–334.

Zumbach, G. O. (2007). A gentle introduction to the RM2006 methodology. *Working Paper available at SSRN: https://ssrn.com/abstract=1420183*.

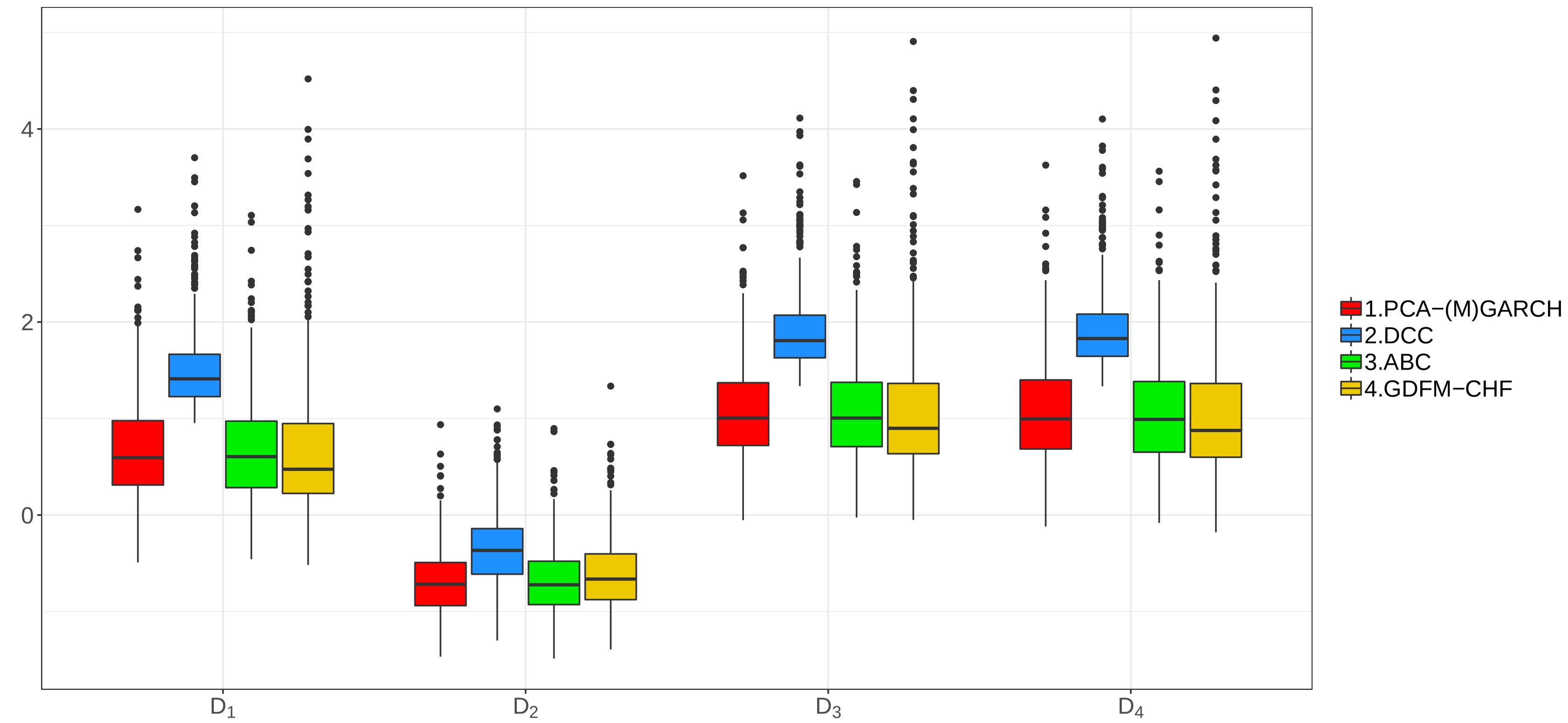

Figure 1: Boxplots of the logarithms of the distances D_1, D_2, D_3, and D_4 for DGP1 across 500 Monte Carlo replications. PCA-(M)GARCH (1), DCC (2), ABC (3) and GDFM-CHF (4) stand for a GARCH model on the common shock with univariate GARCH models on the idiosyncratic components, the DCC with composite likelihood (Pakel et al., 2017), the procedure of Alessi et al. (2009) and our proposal, respectively.

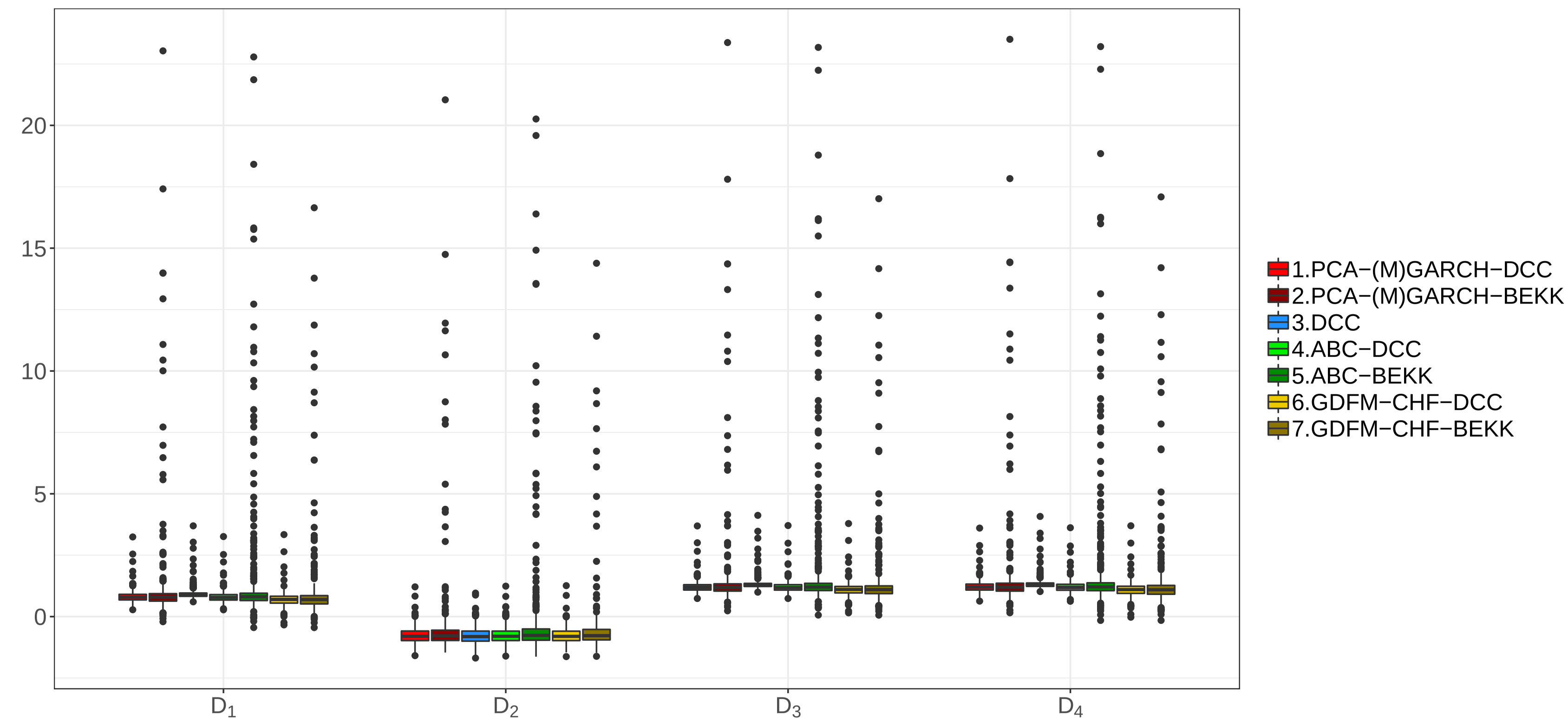

Figure 2: Boxplots of the logarithms of the distances D_1, D_2, D_3, and D_4 for DGP2 across 500 Monte Carlo replications. PCA-(M)GARCH-DCC (1), PCS-(M)GARCH-BEKK (2), DCC (3), ABC-DCC (4), ABC-BEKK (5), GDFM-CHF-DCC (6), GDFM-CHF-BEKK (7) stand for an MGARCH model on the shocks and univariate GARCH models on the idiosyncratic components, the DCC with composite likelihood (Pakel et al., 2017), the procedure of Alessi et al. (2009), and our proposal, respectively.

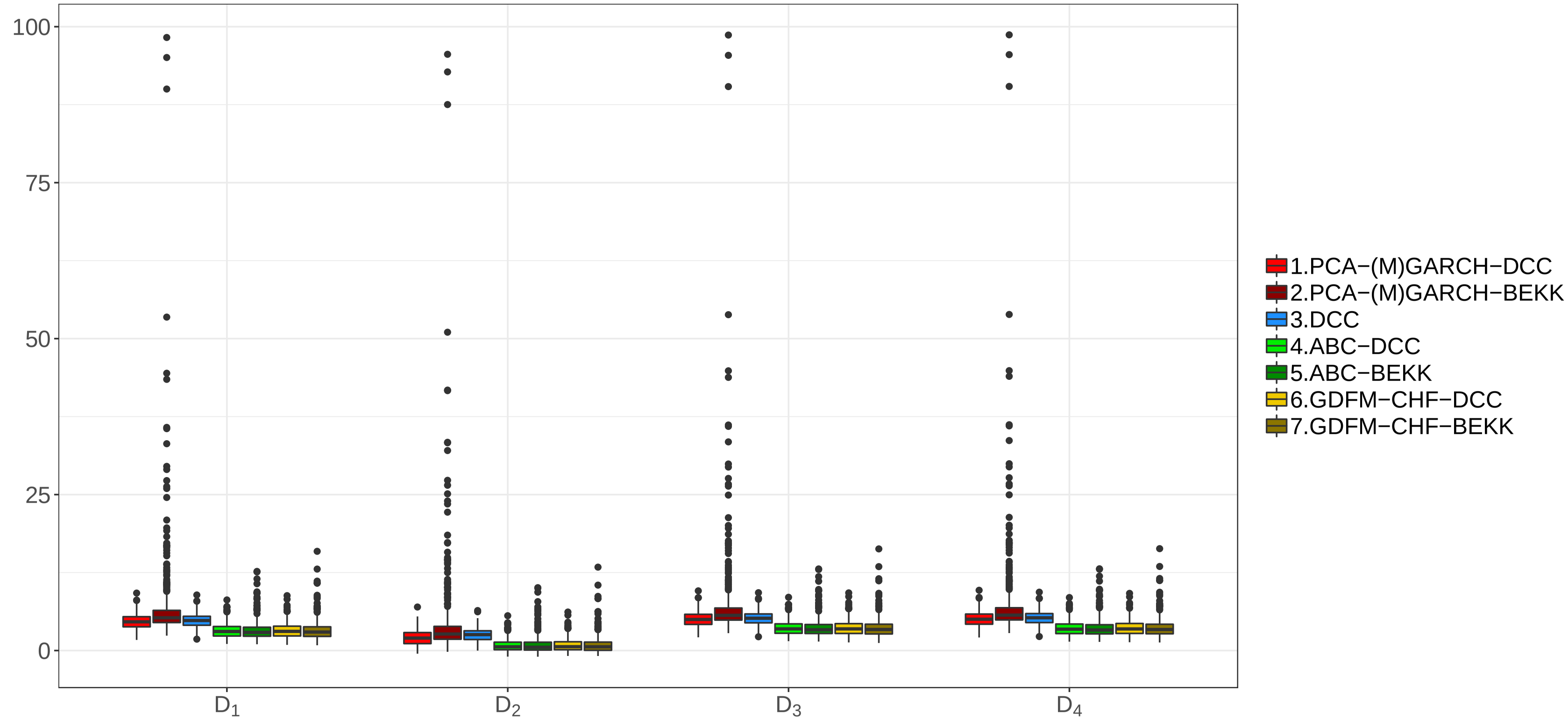

Figure 3: Boxplots of the logarithms of the distances D_1, D_2, D_3, and D_4 for DGP3 across 500 Monte Carlo replications. PCA-(M)GARCH-DCC (1), PCS-(M)GARCH-BEKK (2), DCC (3), ABC-DCC (4), ABC-BEKK (5), GDFM-CHF-DCC (6), GDFM-CHF-BEKK (7) stand for an MGARCH model on the shocks and univariate GARCH models on the idiosyncratic components, the DCC with composite likelihood (Pakel et al., 2017), the procedure of Alessi et al. (2009), and our proposal, respectively.

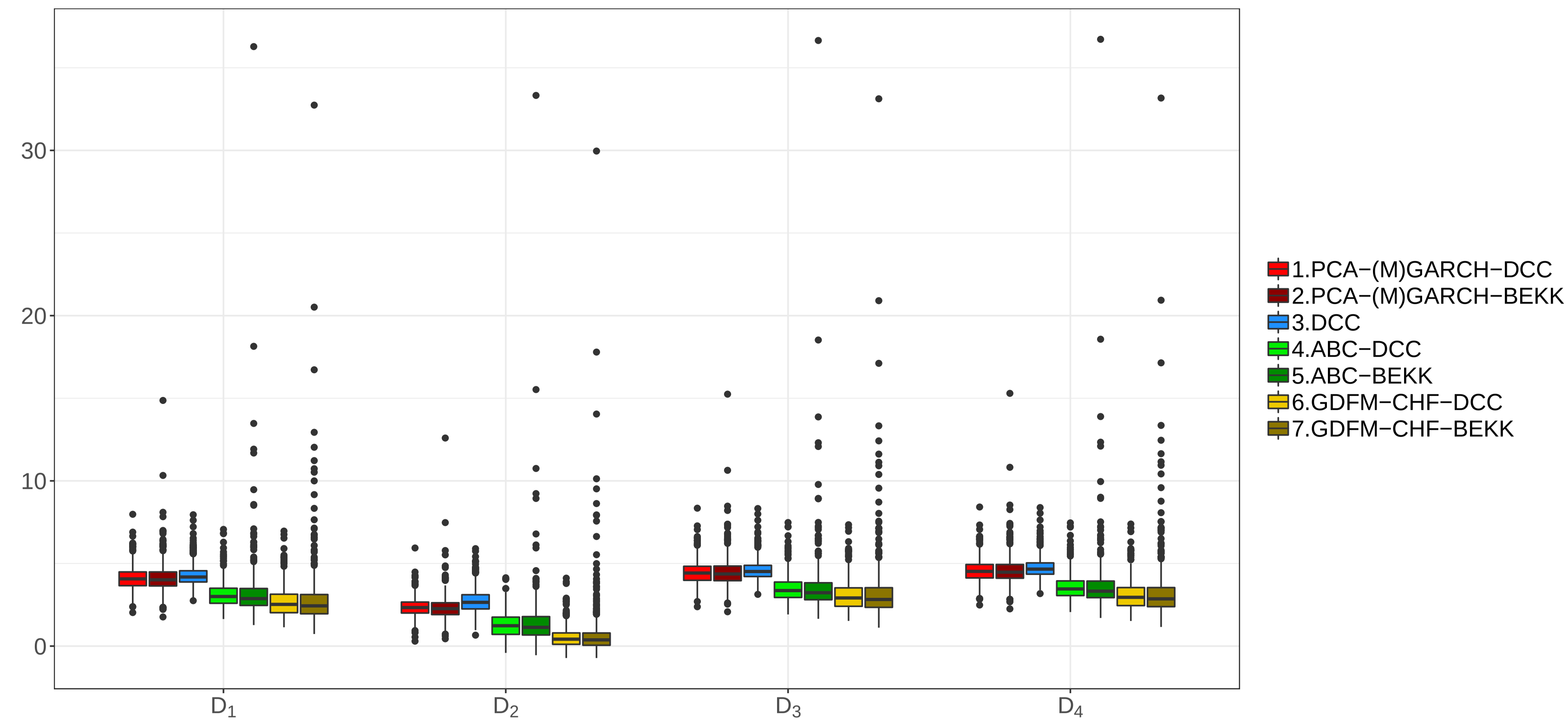

Figure 4: Boxplots of the logarithms of the distances D_1, D_2, D_3, and D_4 for DGP4 across 500 Monte Carlo replications. PCA-(M)GARCH-DCC (1), PCS-(M)GARCH-BEKK (2), DCC (3), ABC-DCC (4), ABC-BEKK (5), GDFM-CHF-DCC (6), GDFM-CHF-BEKK (7) stand for an MGARCH model on the shocks (number of shocks selected via the Bai and Ng (2002) criterion) and univariate GARCH models on the idiosyncratic components, the DCC with composite likelihood (Pakel et al., 2017), the procedure of Alessi et al. (2009), and our proposal, respectively.

Forecasting Conditional Covariance Matrices in High-Dimensional Time Series: a General Dynamic Factor Approach[*,†,‡]

Carlos Trucíos[1], João H. G. Mazzeu[2], Marc Hallin[3], Luiz K. Hotta[2], Pedro L. Valls Pereira[1], Mauricio Zevallos[2]

[1]São Paulo School of Economics, FGV, Brazil
[2]Department of Statistics, University of Campinas, Brazil
[3]ECARES and Département de Mathématique, Université libre de Bruxelles, Belgium

Abstract

Based on a General Dynamic Factor Model with infinite-dimensional factor space and MGARCH common shocks, we develop new estimation and forecasting procedures for conditional covariance matrices in high-dimensional time series. The finite-sample performance of our approach is evaluated via Monte Carlo experiments, outperforming most alternative methods. The new procedure is used to construct one-step-ahead minimum variance portfolios for a high-dimensional panel of assets. The results are shown to achieve better out-of-sample portfolio performance than alternative existing procedures.

Keywords. Dimension reduction, Large panels, High-dimensional time series, Minimum variance portfolio, Volatility, Multivariate GARCH.

[*]This paper received the best LACSC 2019 Paper Award at the 4th Latin American Conference for Statistical Computing, held in Guayaquil, Ecuador, May 28-31, 2019.

[†]Financial support is gratefully acknowledged, from the São Paulo Research Foundation (FAPESP) grants 2016/18599-4 and 2018/03012-3 by the first and fifth authors, from the Coordination for the Improvement of Higher Education Personnel (CAPES) grant 88882.305837/2018-01 by the second author, from the São Paulo Research Foundation (FAPESP) grant 2018/04654-9 by the fourth and sixth authors. All authors acknowledge support from the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS), Centre of Quantitative Studies in Economics and Finance (CEQEF) and European Centre for Advanced Research in Economics and Statistics (ECARES).

[‡]We thank Mario Forni, Roman Liška, and Matteo Barigozzi for kindly giving access to their Matlab codes. Computational resources have been partially provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique (F.R.S.-FNRS) under grant No. 2.5020.11.

JEL classifications. C38, C53, C55, C59, G11.

2010 Mathematics Subject Classification. 62H99, 62M20, 62P20, 91G10.

1 Introduction

Volatility forecasting plays an essential role in a variety of economic and financial applications, such as portfolio allocation, risk management, option pricing, hedging strategies, etc.: see Engle (2009), Hlouskova et al. (2009), Aramonte et al. (2013), Becker et al. (2015), Trucíos et al. (2018), and Engle et al. (2019), to quote only a few.

Several multivariate models have been proposed to model and forecast the conditional covariance matrix of a collection of n assets; see Bauwens et al. (2006) or de Almeida et al. (2018) for reviews. For n small, multivariate GARCH (MGARCH) type models, in that context, constitute fundamental prediction tools. Unfortunately, these models badly suffer from the so-called "curse of dimensionality" as the number n of assets grows, and cannot be implemented in a high-dimensional context. Therefore, alternative procedures have been proposed, see Fan et al. (2008), Alessi et al. (2009), Matteson and Tsay (2011), Engle and Kelly (2012), Hu and Tsay (2014), Santos and Moura (2014), Li et al. (2016), Chang et al. (2018), Engle et al. (2019), Trucíos et al. (2019a) and Pakel et al. (2020), among others.

Dynamic factor models with high-dimensional asymptotics offer a promising approach in that context; see the surveys by Barhoumi et al. (2014) and Bai and Wang (2016) for details. Factor models are based on the assumption that cross-correlations, in a large cross-section of time series data, are accounted by a small number of latent factors or common shocks, which account for their co-movements and have been used by several authors to model and forecast conditional covariance matrices: see Diebold and Nerlove (1989), Harvey et al. (1992), Aguilar and West (2000), Vrontos et al. (2003), Han (2005), Sentana et al. (2008), Aguilar (2009), Alessi et al. (2009), García-Ferrer et al. (2012), Aramonte et al. (2013) and Dovonon (2013), among others. All these contributions are based on a *static* factor-loading scheme[1]

[1] In this static loading scheme, latent factors are loaded contemporaneously via some loading matrix, so that the dimension of the factor space reduces to the (finite) number of linearly independent factors; the number of shocks driving those factors, however, may be strictly less than the

(Bai and Ng, 2002; Stock and Watson, 2002a,b)[2] leading to finite-dimensional factor spaces whose main advantage is to allow for consistent estimation methods based on traditional principal components, which are familiar to most practitioners, easy to implement, and widely used in practice.

However, as pointed out in Forni and Lippi (2011) and Section 1.1 of Forni et al. (2015), the assumption of a static factor-loading scheme considered in that literature is quite restrictive and rules out some very simple and plausible cross-correlation patterns leading to infinite-dimensional factor spaces. To overcome this issue, Forni et al. (2000) introduced the so-called *generalized* or *general dynamic factor model* (GDFM), in which factors (equivalently, common shocks) are loaded through filters rather than matrices; see the monograph by Hallin et al. (2020) for details. As shown in Hallin and Lippi (2013), the GDFM actually follows from a representation result which holds, essentially, without placing any restrictions on the data-generating process—beyond second-order stationarity and the existence of a spectrum.

The role of traditional principal components in the GDFM is taken over by Brillinger's *dynamic principal components*[3] (Brillinger, 1981), and the estimation method proposed by Forni et al. (2000) naturally relies on this concept. Dynamic principal components, however, involve two-sided filters, producing estimators that are inadequate in forecasting problems. Forni and Lippi (2011) and Forni et al. (2015, 2017)[4] therefore developed an alternative consistent estimation method involving one-sided filters only. Monte Carlo simulations indicate that, for estimating impulse-response functions and predicting returns, this one-sided approach outperforms the *static* methods of Stock and Watson (2002a,b) and Bai and Ng (2002) even when the actual loading scheme is of the static type (see Section 4 in Forni et al. (2017)).

The Forni et al. (2015, 2017) procedure has been successfully used to forecast

---

dimension of the factor space.

[2]Similar ideas have been developed also in a non-econometric context, see, e.g., Peña and Box (1987), Stoffer (1999), or Pan and Yao (2008).

[3]Hallin et al. (2018) show that those dynamic principal components, based on the factorization of spectral density matrices, inherit, in a time-series context, the optimality properties that make traditional principal components a successful dimension-reduction device in i.i.d. samples.

[4]The assumptions in those three references yield slight variations; in this paper, unless otherwise stated, we refer to the assumptions in Barigozzi and Hallin (2020).

inflation and financial returns; see Della Marra (2017), Forni et al. (2018) and Giovannelli et al. (2018). It has also been used in the prediction of conditional variances by Barigozzi and Hallin (2016, 2017, 2020), but never, as far as we know, in the prediction of conditional covariance matrices and portfolio optimization.[5] These two points constitute the main goal of this paper.

The rest of the paper is organised as follows. Section 2 briefly describes the GDFM. Section 3 introduces our forecasting procedure and establishes its consistency properties. Section 4 reports a Monte Carlo study of the finite-sample performance of the proposed procedure and their comparison with existing competitors. In Section 5, the new procedure is applied to dynamic portfolio optimization, that is, the problem of constructing, at time T, portfolios with minimum (at time $T+1$) conditional variance from a large collection of assets. In Sections 5 we also compare the proposed procedure with other methods. Section 6 concludes.

2 The general dynamic factor model

In this section, we briefly describe the GDFM to be considered throughout, which basically contains as particular cases all other factor models proposed in the econometric and time series literature, along with the regularity assumptions we need for consistency, which are borrowed, essentially, from Barigozzi and Hallin (2020).

Let $\{\mathbf{X}_t := (X_{1t}\ X_{2t} \dots)', \ t \in \mathbb{Z}\}$, be a double-indexed zero-mean second-order stationary stochastic process, where the first index is cross-sectional and typically refers to assets, while t, as usual, stands for time. The GDFM is based on the decomposition

$$X_{it} = \chi_{it} + \xi_{it}, \qquad i \in \mathbb{N}_0, \quad t \in \mathbb{Z} \tag{1}$$

of X_{it} into two non-observable mutually orthogonal components χ_{it} (the *common components*) and ξ_{it} (the *idiosyncratic components*), where

$$\chi_{it} = \sum_{j=1}^{q} \sum_{k=0}^{\infty} b_{ijk} u_{jt-k} = \mathbf{b}_i'(L)\mathbf{u}_t \ \text{ and } \ \xi_{it} = \sum_{k=0}^{\infty} d_{ik} v_{it-k} = d_i(L) v_{it}; \tag{2}$$

[5] See, however, the unpublished paper by Alessi et al. (2009) who assume a factor model decomposition with finite-dimensional factor space on the model of Forni et al. (2005 and 2009).

the *common shocks* $\mathbf{u}_t := (u_{1t}\ u_{2t}\ ...\ u_{qt})'$ driving the common components, and the *idiosyncratic shocks* v_{it} driving the idiosyncratic components, are also non-observable.

Letting $\mathbf{X}_n := \{X_{it} | i = 1, ..., n, t \in \mathbb{Z}\}$, $\boldsymbol{\chi}_n := \{\chi_{it} \mid i = 1, ..., n,\ t \in \mathbb{Z}\}$, and $\boldsymbol{\xi}_n := \{\xi_{it} | i = 1, ..., n, t \in \mathbb{Z}\}$, equation (2) in vector notation takes the form

$$\mathbf{X}_{nt} = \boldsymbol{\chi}_{nt} + \boldsymbol{\xi}_{nt} = \mathbf{B}_n(L)\mathbf{u}_t, + \mathbf{D}_n(L)\mathbf{v}_{nt}, \quad n \in \mathbb{N}_0, \quad t \in \mathbb{Z} \tag{3}$$

with $\mathbf{B}_n(L) := (\mathbf{b}_1(L)...\mathbf{b}_n(L))'$, $\mathbf{D}_n(L) := \text{diag}(d_1(L)...d_n(L))$, and $\mathbf{v}_{nt} := (v_{1t} \ldots v_{nt})'$.

Let $\|\boldsymbol{A}\|_p$ stand for the L^p norm $\big(\sum_{i,j} A_{ij}^p\big)^{1/p}$ of a real matrix $\boldsymbol{A} = (A_{ij})$ (for $p = 2$, we simply write $\|\boldsymbol{A}\|$). On the GDFM decomposition (1), we assume the following.

Assumption (GDFM) *(i)* the vector process $\mathbf{u}_t$ is a zero-mean q-dimensional second-order white noise process, with full-rank covariance $\boldsymbol{\Gamma}^{\mathbf{u}}$;

(ii) writing $\mathbf{b}_{ik} := (b_{i1k}...b_{iqk})'$ for the $q \times 1$ coefficient of L^k in $\mathbf{b}_i(L)$, there exists a constant $M_1 > 0$ such that $\sum_{k=0}^{\infty} \|\mathbf{b}_{ik}\| k^{1/2} \leq M_1$ for all $i \in \mathbb{N}$;

(iii) $\mathbf{v}_{nt}$ is a zero-mean second-order stationary process with positive definite covariance $\boldsymbol{\Gamma}_n^v$; moreover, $\mathrm{E}[v_{it}|v_{is}] = 0$ for all $i \in \mathbb{N}$ and $t > s \in \mathbb{Z}$;

(iv) there exists a constant $C_v > 0$ such that $\|\boldsymbol{\Gamma}_n^v\|_1 \leq C_v$ for all $n \in \mathbb{N}$, and a constant $M_2 > 0$ such that $\sum_{k=0}^{\infty} |d_{ik}| k^{1/2} \leq M_2$ for all $i \in \mathbb{N}$;

(v) $\mathrm{Cov}(u_{jt}, v_{is}) = 0$ for all $i \in \mathbb{N}$, $j = 1, ..., q$, and t, $s \in \mathbb{Z}$;[6]

(vi) there exists a constant $M_3 > 0$ such that, for all j_1, j_2, j_3, j_4,

$$\sum_{k_1,k_2,k_3 \in \mathbb{Z}} |\mathrm{E}(u_{j_1 t} u_{j_2, t-k_1} u_{j_3, t-k_2} u_{j_4, t-k_3})| \leq M_3,$$

and a constant $M_4 > 0$ such that, for all i_1, i_2, i_3, i_4,

$$\sum_{k_1,k_2,k_3 \in \mathbb{Z}} |\mathrm{E}(v_{i_1 t} v_{i_2, t-k_1} v_{i_3, t-k_2} v_{i_4, t-k_3})| \leq M_4;$$

(vii) for all $i \in \mathbb{N}$ and $j = 1, \ldots, q$, $b_{ij}(z) = \sum_{k=0}^{\infty} b_{ijk} z^k$, $z \in \mathbb{C}$, has square-summable coefficients and is the ratio $\gamma_{ij}(z)/\delta_{ij}(z)$ of two finite-order polynomials in z, $\gamma_{ij}(z) = \sum_{k=0}^{S_\gamma} \gamma_{ijk} z^k$ and $\delta_{ij}(z) = \sum_{k=0}^{S_\delta} \delta_{ijk} z^k$ with roots outside

[6] This implies that the common and idiosyncratic processes are mutually uncorrelated at all leads and lags.

the closed unit disk only, $\delta_{ij}(0) = 1$, and no common roots; the orders S_γ and S_δ, moreover, are independent of i.[7]

Assumption GDFM*(iii)* is the typical assumption of martingale difference innovations used in the GARCH literature. Assumption (vii) entails the existence of a VAR filtering of $\mathbf{X}_n$ satisfying the assumptions of the static factor model where the common shocks $\mathbf{u}_t$ are loaded contemporaneously (see (4) below).

These assumptions also guarantee the existence of the spectral density matrices $\mathbf{\Sigma}_n^\chi(\theta)$, $\mathbf{\Sigma}_n^\xi(\theta)$, and $\mathbf{\Sigma}_n^X(\theta) = \mathbf{\Sigma}_n^\chi(\theta) + \mathbf{\Sigma}_n^\xi(\theta)$, $\theta \in [-\pi, \pi]$, of $\boldsymbol{\chi}_n$, $\boldsymbol{\xi}_n$, and $\mathbf{X}_n$, respectively. Denoting by $\lambda_{nj}^\chi(\theta)$, $\lambda_{nj}^\xi(\theta)$ and $\lambda_{nj}^X(\theta)$ be the jth eigenvalues (in decreasing order of magnitude) of $\mathbf{\Sigma}_n^\chi(\theta)$, $\mathbf{\Sigma}_n^\xi(\theta)$ and $\mathbf{\Sigma}_n^X(\theta)$, respectively, let them satisfy the following assumption.

Assumption (GDFM)*(viii)* There exist an integer $\bar{n} > 0$ and continuous functions α_j and β_{j-1} from $[-\pi, \pi]$ to $\mathbb{R}$, $j = 1, \ldots, q$, independent of n and such that, for all $j = 1, \ldots, q$, and all $n > \bar{n}$,

$$0 < \beta_{j-1}(\theta) < \alpha_j(\theta) \leq \lambda_{nj}^\chi(\theta)/n \leq \beta_j(\theta) < \infty, \quad \theta\text{-a.e. in } [-\pi, \pi],$$

while $\lambda_{n,q+1}^\chi(\theta)$ and $\lambda_{n1}^\xi(\theta)$ are bounded, uniformly in $\theta \in [-\pi, \pi]$, as $n \to \infty$.

Hence, as $n \to \infty$, the q common dynamic eigenvalues are exploding linearly (the assumption of factor *pervasiveness*), while all idiosyncratic eigenvalues are bounded (this is the definition of *idiosyncrasy*).

The main theoretical result behind the one-sided approach of Forni et al. (2015) is the *generic* existence[8] of a block-diagonal VAR filtering of the observations turning the GDFM representation (1) into a static one. More precisely, Forni and Lippi (2011) and Forni et al. (2015) show that, for generic values of the coefficients γ_{ijk} and δ_{ijk} (i.e., except for a subset with Lebesgue measure zero in the $(q+1)(S_\gamma + S_\delta)$-dimensional space of the relevant γ_{ijk} and δ_{ijk} coefficients), any $(q+1)$-dimensional vector $\boldsymbol{\chi}_t^{i_1 \ldots i_{q+1}} := (\chi_{i_1 t}, \ldots, \chi_{i_{q+1} t})'$ with $i_1 < \ldots < i_{q+1}$ admits a VAR representation of the form $\boldsymbol{A}(L)^{i_1 \ldots i_{q+1}} \boldsymbol{\chi}_t^{i_1 \ldots i_{q+1}} = \mathbf{R}^{i_1 \ldots i_{q+1}} \mathbf{u}_t$,[9] where $\boldsymbol{A}(L)^{i_1 \ldots i_{q+1}}$ has

[7] As a consequence, the common components have rational spectral densities; see Assumption (L2) in Barigozzi and Hallin (2020) for more details.

[8] This goes back to results on reduced rank processes: see, e.g., Anderson and Deistler (2008).

[9] See Assumption (L4) in Barigozzi and Hallin (2018) for more details about this VAR representation.

degree $S \leq qS_\gamma + q^2 S_\delta$ and the $(q+1) \times q$ matrix $\mathbf{R}^{i_1 \cdots i_{q+1}}$ is of rank q. It follows that *generically*, for any $n = m(q+1)$, partitioning $\boldsymbol{\chi}_{nt} = (\chi_{1t}, \ldots, \chi_{nt})'$ into m subvectors of dimension $(q+1)$, $\boldsymbol{\chi}_{nt}$ admits a block-VAR representation of the form

$$\mathbf{A}_n(L)\boldsymbol{\chi}_{nt} = \begin{bmatrix} \mathbf{A}^1(L) & 0 & \ldots & 0 \\ 0 & \mathbf{A}^2(L) & \ldots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \ldots & \mathbf{A}^m(L) \end{bmatrix} \boldsymbol{\chi}_{nt} = \begin{bmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^m \end{bmatrix} \mathbf{u}_t, \quad t \in \mathbb{Z}. \qquad (4)$$

Hence, for $\mathbf{X}_{nt} = (X_{1t}, \ldots, X_{nt})'$, we have

$$\mathbf{A}_n(L)\mathbf{X}_{nt} = \mathbf{A}_n(L)\boldsymbol{\chi}_{nt} + \mathbf{A}_n(L)\boldsymbol{\xi}_{nt} = \mathbf{R}_n \mathbf{u}_t + \boldsymbol{\epsilon}_{nt}, \quad t \in \mathbb{Z} \qquad (5)$$

with $\mathbf{R}_n = \left[\mathbf{R}^{1\prime}\, \mathbf{R}^{2\prime} \ldots \mathbf{R}^{m\prime}\right]'$ and $\boldsymbol{\epsilon}_{nt} = \mathbf{A}_n(L)\boldsymbol{\xi}_{nt}$, where it can be shown that the process $\boldsymbol{\epsilon}_t := \{(\epsilon_{1t}\ \epsilon_{2t} \ldots)',\ t \in \mathbb{Z}\}$ is still idiosyncratic. In other words, using obvious notation

$$\mathbf{A}(L) := \begin{bmatrix} \mathbf{A}^1(L) & 0 & \ldots & 0 & \ldots \\ 0 & \mathbf{A}^2(L) & \ldots & 0 & \ldots \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \ldots & \mathbf{A}^m(L) & \ldots \\ \vdots & \vdots & \ldots & \ldots & \ddots \end{bmatrix} \quad \text{and} \quad \mathbf{R} := \begin{bmatrix} \mathbf{R}^1 \\ \mathbf{R}^2 \\ \vdots \\ \mathbf{R}^m \\ \vdots \end{bmatrix}, \qquad (6)$$

the filtered process $\mathbf{Y}_t := \mathbf{A}(L)\mathbf{X}_t$ admits a *static* factor model representation

$$\mathbf{Y}_t = \mathbf{R}\mathbf{u}_t + \boldsymbol{\epsilon}_t, \quad t \in \mathbb{Z} \qquad (7)$$

with q-dimensional factor space spanned by $\mathbf{u}_t$. While $\mathbf{R}$ and $\mathbf{u}_t$ are not individually identified, the product $\mathbf{R}\mathbf{u}_t$ is.

The static representation (7), under assumptions *(i)-(viii)*, holds *generically.* Assuming that it holds for the panel under study this is a very mild requirement; we nevertheless need to make it an assumption:

Assumption (GDFM) *(ix)* For all $n^* \geq q+1$, letting $n = \lfloor n^*/(q+1) \rfloor (q+1)$, there exist block-diagonal filters $\mathbf{A}_n(L)$ and $n \times q$ matrices $\mathbf{R}_n$ such that (5) holds, irrespective of the cross-sectional ordering.

Assumptions (GDFM) *(i)-(ix)* are the main assumptions in Barigozzi and Hallin (2020); on top of these, they also require two less important and more technical ones on the regularity of the VAR operators $\mathbf{A}^m(L)$ (Assumptions (L4) and (L5), respectively), which we do not reproduce here. Under those assumptions, Barigozzi and Hallin (2020) show that a consistent reconstruction, based on $\mathbf{X}_t$, $\mathbf{X}_{t-1}$, ..., of the unobserved $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ is possible. It follows that $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$ are $\mathcal{F}_t$-measurable, where $\mathcal{F}_t$ denotes the σ-field generated by $\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots$ It is worth noting that, reinforcing the same assumptions (e.g., assuming that $\mathbf{u}_t$ and $\mathbf{v}_{nt}$ are jointly i.i.d., which rules out GARCH-type behaviors), Forni et al. (2017) derive estimators for (1)-(2) and provide a complete asymptotic analysis for the same. On the other hand, Barigozzi and Hallin (2020) do not require i.i.d.-ness and, under assumptions that include *(i)-(ix)*, provide consistency and consistency rates for the Forni et al. (2017) estimators.

If, however, $\text{Var}(\mathbf{X}_{nt}|\mathcal{F}_{n;t-1})$ is to be estimated at time $(t-1)$, assumptions have to be made on the dynamics of $\text{Var}(\mathbf{u}_t|\mathcal{F}^{\mathbf{u}}_{t-1})$ and $\text{Var}(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1})$. As in Alessi et al. (2009) and Aramonte et al. (2013), we therefore assume that the conditional covariance matrices of the common shocks can be modelled as some q-dimensional MGARCH process. Since q is typically small, this approach escapes the curse of dimensionality. As for the idiosyncratic conditional covariance matrix $\text{Var}(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1})$, since idiosyncratic cross-correlations are non-pervasive (mild enough that idiosyncratic dynamic eigenvalues remain bounded), it can be approximated by a diagonal matrix where each diagonal element (each marginal conditional variance) is modelled by a univariate GARCH-type model—in the sequel, we use GARCH(1,1) models.[10] In both cases, the MGARCH and the n GARCH(1,1) models are estimated by Gaussian quasi-maximum likelihood (QMLE). We refer to the monograph by Francq and Zakoian (2019) for sufficient QMLE consistency conditions; note, however, that those QMLEs, here, will be computed from the Forni et al. (2017) estimated shocks $\widehat{\mathbf{u}}_t$ and estimated idiosyncratic components $\hat{\boldsymbol{\xi}}_{it}$.

More precisely, we assume the following.

Assumption (GARCH). The common shocks $\mathbf{u}_t$ and the idiosyncratic compo-

[10] From a numerical perspective, this diagonal approximation of idiosyncratic covariances can be seen as a simple regularization device.

nents ξ_{it} are stable by aggregation MGARCH (with parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$) and univariate AR-GARCH (with parameters $\boldsymbol{\vartheta}_i \in \boldsymbol{\Theta}_1$, $i \in \mathbb{N}$) stationary processes, respectively; they are conditionally (on $\mathcal{F}_{n;t-1}$) uncorrelated at all leads and lags; the parameter spaces $\boldsymbol{\Theta}_q$ and $\boldsymbol{\Theta}_1$ are compact; the densities of $\mathbf{u}_t$ and the idiosyncratic shocks v_{it} and the parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$ and $\boldsymbol{\vartheta}_i \in \boldsymbol{\Theta}_1$ jointly satisfy the conditions for consistent QMLE.

The assumption that the MGARCH model generating the common shocks is stable by aggregation is motivated by the fact that $\mathbf{u}_t$ is not fully identified (see the remark after (7)): under Assumption (GARCH), any linear transform $\mathbf{R}\mathbf{u}_t$ is driven by an MGARCH model of the same type as $\mathbf{u}_t$ itself. Examples of stable by aggregation MGARCH models are the full VECH (Bollerslev et al., 1988) and full BEKK (Engle and Kroner, 1995) models, which moreover can be consistently estimated via QMLE methods; see Comte and Lieberman (2003), Hafner and Preminger (2009), and Theorems 10.2 and 10.4 in Francq and Zakoian (2019).

As mentioned before, the idiosyncratic conditional covariance matrix $\mathrm{Var}(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1})$ is approximated by a diagonal matrix where each diagonal element (each marginal conditional variance) is modelled by a univariate GARCH-type model. That approximation, which is justified by the boundedness of idiosyncratic dynamic eigenvalues, is on line with the factor model paradigm, where cross-correlations are essentially accounted for by the common shocks, and the idiosyncratic contribution are negligible. Rather than making the comfortable but unrealistic assumption of mutually orthogonal idiosyncratics, in Section 3.2, we will use the notation $\mathrm{Var}^*(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1})$ and $\mathrm{Var}^*(\mathbf{X}_{nt}|\mathcal{F}_{n;t-1})$ for the approximate conditional covariance matrices of $\boldsymbol{\xi}_{nt}$ and $\mathbf{X}_{nt}$ resulting from neglecting that off-diagonal idiosyncratic contribution.

3 Predicting covariance matrices

In this section, we propose an estimator, based on past observations up to time T, of the covariance matrix of $\mathbf{X}_{n,T+1}$ conditional on $\mathbf{X}_{nT}, \mathbf{X}_{n,T-1}, \ldots$. More precisely, denoting by $\boldsymbol{V}^{\mathbf{X}_n}_{t|t-1}$ the covariance matrix $\mathrm{Var}(\mathbf{X}_{nt}|\mathcal{F}_{n;t-1})$ of $\mathbf{X}_{nt}$ conditional on the σ-field $\mathcal{F}_{n;t-1}$ generated by $\{X_{is}|\, i = 1, \ldots, n; s \leq t-1\}$, we are interested in estimating the $n \times n$ matrix $\boldsymbol{V}^{\mathbf{X}_n}_{T+1|T}$ or some $n_0 \times n_0$ submatrix $\boldsymbol{V}^{\mathbf{X}_{n_0}}_{T+1|T}$ thereof[11] from the observed $n \times T$ panel.[12]

Section 3.1 provides a theoretical expression for $\boldsymbol{V}^{\mathbf{X}_n}_{t|t-1}$ =Var$(\mathbf{X}_{nt}\big|\mathcal{F}_{n;t-1})$; Section 3.2 describes the estimation procedure; Section 3.3 establishes the consistency properties of the estimator.

3.1 The conditional covariance matrix

We start with a theoretical decomposition of the conditional covariance matrix $\boldsymbol{V}^{\mathbf{X}_n}_{t|t-1}$ of $\mathbf{X}_{nt}$ in terms of the elements of the static representation (7). Similar to $\boldsymbol{V}^{\mathbf{X}_n}_{T+1|T}$, the notation $\boldsymbol{V}^{\mathbf{Y}_n}_{T+1|T}$, $\boldsymbol{V}^{\boldsymbol{\chi}_n}_{T+1|T}$, $\boldsymbol{V}^{\boldsymbol{\xi}_n}_{T+1|T}$, $\boldsymbol{V}^{\boldsymbol{\epsilon}_n}_{T+1|T}$, ... is used in an obvious fashion for Var$(\mathbf{Y}_{nt}\big|\mathcal{F}_{n;t-1})$, Var$(\boldsymbol{\chi}_{nt}\big|\mathcal{F}_{n;t-1})$, Var$(\boldsymbol{\xi}_{nt}\big|\mathcal{F}_{n;t-1})$, etc. Note, however, that the $q \times q$ covariance Var$(\mathbf{u}_t\big|\mathcal{F}_{n;t-1})$ of $\mathbf{u}_t$ conditional on $\mathcal{F}_{n;t-1}$, in view of Assumption (GARCH), reduces to Var$(\mathbf{u}_t\big|\mathcal{F}^{\boldsymbol{\chi}}_{n;t-1})$, where $\mathcal{F}^{\boldsymbol{\chi}}_{n;t-1}$ is generated by the past values of $\boldsymbol{\chi}_{nt}$, which in turn, for n large enough, coincides with the σ-field $\mathcal{F}^{\mathbf{u}}_{t-1}$ generated by $\mathbf{u}_t$'s own past. That σ-field no longer involves n—justifying the notation $\boldsymbol{V}^{\mathbf{u}}_{t|t-1}$ or $\boldsymbol{V}^{\mathbf{u}}_{t|t-1}(\mathbf{u}_{t-1}, \mathbf{u}_{t-2}, \ldots)$.

All those conditional covariances can be interpreted as (oracle) predictors, based on observations up to time $t-1$, of the corresponding stochastic covariance the nonobservable realization of which is to take place at time t.

Proposition 1. *Let Assumption (GDFM) (i)-(ix) hold. Then, the covariance matrix $\boldsymbol{V}^{\boldsymbol{X}_n}_{t|t-1}$ of $\boldsymbol{X}_{nt}$ conditional on $\mathcal{F}_{n;t-1}$ decomposes into*

$$\boldsymbol{V}^{\boldsymbol{X}_n}_{t|t-1} = \boldsymbol{R}_n \boldsymbol{V}^{\boldsymbol{u}}_{t|t-1} \boldsymbol{R}'_n + \boldsymbol{V}^{\boldsymbol{\xi}_n}_{t|t-1}. \tag{8}$$

Proof. From (7), we have that

$$\begin{aligned}\text{Var}(\mathbf{Y}_{nt}|\mathcal{F}_{n;t-1}) &= \text{Var}(\mathbf{R}_n\mathbf{u}_t + \boldsymbol{\epsilon}_{nt}|\mathcal{F}_{n;t-1}) \\ &= \mathbf{R}_n\text{Var}(\mathbf{u}_t|\mathcal{F}_{n;t-1})\mathbf{R}'_n + \text{Var}(\boldsymbol{\epsilon}_{nt}|\mathcal{F}_{n;t-1}) + \text{Cov}(\mathbf{R}_n\mathbf{u}_t, \boldsymbol{\epsilon}_{nt}|\mathcal{F}_{n;t-1}) \\ &\quad +\text{Cov}(\boldsymbol{\epsilon}_{nt}, \mathbf{R}_n\mathbf{u}_t|\mathcal{F}_{n;t-1}), \quad t \in \mathbb{Z}. \end{aligned} \tag{9}$$

[11]Without loss of generality, we always consider the $n_0 \times n_0$ left upper corner.

[12]Since the (random) covariance matrix to be estimated is associated with time $T+1$ while observations are limited to time T, this estimator also will be called a *predictor*, although the estimand is never to be observed, which makes this association with time $T+1$ somewhat immaterial.

Without loss of generality we can assume that all VAR filters $\mathbf{A}^k(L)$ in (5) are of the form $\mathbf{A}^k(L) = \mathbf{I}_{q+1} - \boldsymbol{\phi}_1^k L - \cdots - \boldsymbol{\phi}_S^k L^S$ (with $\boldsymbol{\phi}_S^k \neq \mathbf{0}$ for at least one k). Consequently, $\mathbf{A}_n(L)$ can be written as $\mathbf{A}_n(L) = \mathbf{I} - \mathbf{\Phi}_1 L - \cdots - \mathbf{\Phi}_S L^S$. Then, it is easy to check that

$$\begin{aligned}\mathrm{Var}(\boldsymbol{\epsilon}_{nt}|\mathcal{F}_{n;t-1}) &= \mathrm{Var}(\mathbf{A}_n(L)\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}) = \mathrm{Var}\left(\left[\mathbf{I} - \mathbf{\Phi}_1 L - \cdots - \mathbf{\Phi}_S L^S\right]\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}\right) \\ &= \mathrm{Var}(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}), \end{aligned} \tag{10}$$

since $\boldsymbol{\xi}_{n,t-k}$ is $\mathcal{F}_{n;t-1}$-measurable for $k \geq 1$.

Similarly, we have

$$\mathrm{Var}(\mathbf{Y}_{nt}|\mathcal{F}_{n;t-1}) = \mathrm{Var}(\mathbf{A}_n(L)\mathbf{X}_{nt}|\mathcal{F}_{n;t-1}) = \mathrm{Var}(\mathbf{X}_{nt}|\mathcal{F}_{n;t-1}). \tag{11}$$

Moreover, since $\mathbf{u}_t$ and $\boldsymbol{\xi}_{nt}$ are conditionally uncorrelated, both $\mathrm{Cov}(\mathbf{R}_n\mathbf{u}_t, \boldsymbol{\epsilon}_{nt}|\mathcal{F}_{n;t-1})$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_{nt}, \mathbf{R}_n\mathbf{u}_t|\mathcal{F}_{n;t-1})$ in (9) equal zero. Hence,

$$\mathrm{Cov}(\mathbf{R}_n\mathbf{u}_t, \boldsymbol{\epsilon}_{nt}|\mathcal{F}_{n;t-1}) = \mathrm{Cov}(\mathbf{R}_n\mathbf{u}_t, \mathbf{A}_n(L)\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}) = \mathbf{R}_n\mathrm{Cov}(\mathbf{u}_t, \mathbf{A}_n(L)\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}).$$

Now,

$$\begin{aligned}&\mathrm{Cov}(\mathbf{u}_t, \mathbf{A}_n(L)\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}) = \mathrm{Cov}(\mathbf{u}_t, \left[I - \mathbf{\Phi}_1 L - ... - \mathbf{\Phi}_S L^S\right]\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}) \\
&\quad = \mathrm{E}(\mathbf{u}_t\left[\boldsymbol{\xi}_{nt} - \mathbf{\Phi}_1\boldsymbol{\xi}_{n,t-1} - ... - \mathbf{\Phi}_S\boldsymbol{\xi}_{n,t-S}\right]'|\mathcal{F}_{n;t-1}) \\
&\qquad - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{n;t-1})\mathrm{E}(\left[\boldsymbol{\xi}_{nt} - \mathbf{\Phi}_1\boldsymbol{\xi}_{t-1} - ... - \mathbf{\Phi}_S\boldsymbol{\xi}_{n,t-S}\right]'|\mathcal{F}_{n;t-1}) \\
&\quad = \mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_{nt}'|\mathcal{F}_{n;t-1}) - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{n;t-1})\mathrm{E}(\boldsymbol{\xi}_{nt}'|\mathcal{F}_{n;t-1}) \\
&\qquad - \underbrace{\left[\mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_{n,t-1}'\mathbf{\Phi}_1'|\mathcal{F}_{n;t-1}) - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{n;t-1})\mathrm{E}(\boldsymbol{\xi}_{n,t-1}'\mathbf{\Phi}_1'|\mathcal{F}_{n;t-1})\right]}_{\mathbf{0}} \\
&\qquad - \ldots - \underbrace{\left[\mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_{n,t-S}'\mathbf{\Phi}_S'|\mathcal{F}_{n;t-1}) - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{n;t-1})\mathrm{E}(\boldsymbol{\xi}_{n,t-S}'\mathbf{\Phi}_S'|\mathcal{F}_{n;t-1})\right]}_{\mathbf{0}} \\
&\quad = \mathrm{E}(\mathbf{u}_t\boldsymbol{\xi}_{nt}'|\mathcal{F}_{n;t-1}) - \mathrm{E}(\mathbf{u}_t|\mathcal{F}_{n;t-1})\mathrm{E}(\boldsymbol{\xi}_{nt}'|\mathcal{F}_{n;t-1}) = \mathrm{Cov}(\mathbf{u}_t, \boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1}) = \mathbf{0}.\end{aligned}$$

It then follows from (8)-(11), along with the fact that $\mathrm{Cov}(\boldsymbol{\epsilon}_{nt}, \mathbf{R}_n\mathbf{u}_t|\mathcal{F}_{n;t-1}) = \mathbf{0}$, that

$$\mathrm{Var}(\mathbf{X}_{nt}|\mathcal{F}_{n;t-1}) = \mathrm{Var}(\mathbf{Y}_{nt}|\mathcal{F}_{n;t-1}) = \mathbf{R}_n\mathrm{Var}(\mathbf{u}_t|\mathcal{F}_{n;t-1})\mathbf{R}_n' + \mathrm{Var}(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1})$$

with $\mathrm{Var}(\mathbf{u}_t|\mathcal{F}_{n;t-1}) = \boldsymbol{V}_{t|t-1}^{\mathbf{u}}$, as was to be proved. □

The same decomposition (8) applies to the regularized covariances resulting from neglecting idiosyncratic cross-covariances; to avoid overloading notation any further, we do not, however, introduce any formal symbol for the latter.

3.2 Estimation

We start with estimating the GDFM decomposition of the observed $n \times T$ panel.

- **Step 1.** Determine the number q of common shocks, for instance via the Hallin and Liška (2007) criterion.
- **Step 2.** Randomly reorder the n observed series.
- **Step 3.** Compute a consistent[13] estimator

$$\widehat{\mathbf{\Sigma}}^{X}_{nT}(\theta) = \frac{1}{2\pi} \sum_{k=-M_T}^{M_T} e^{-ik\theta} K\left(\frac{k}{B_T}\right) \widehat{\mathbf{\Gamma}}^{X}_{k}$$

  of the $n \times n$ spectral density matrix of the $\mathbf{X}_t$'s, where $K(\cdot)$ is a kernel function, M_T a truncation parameter, B_T the bandwidth, and $\widehat{\mathbf{\Gamma}}^{X}_{k}$ the sample lag-k cross-covariance matrix computed from the observed $n \times T$ panel of $\mathbf{X}_t$ values.

- **Step 4.** Collecting the normalized column eigenvectors associated with $\widehat{\mathbf{\Sigma}}^{X}_{nT}(\theta)$'s q largest eigenvalues into the $n \times q$ matrix $\widehat{\boldsymbol{P}}^{X}_{nT}(\theta)$ (with complex conjugate $\widehat{\boldsymbol{P}}^{X*}_{nT}$) and the corresponding eigenvalues into the $q \times q$ diagonal matrix $\widehat{\mathbf{\Lambda}}^{X}_{nT}(\theta_h)$, compute

$$\widehat{\mathbf{\Sigma}}^{\chi}_{nT}(\theta) := \widehat{\boldsymbol{P}}^{X}_{nT}(\theta) \widehat{\mathbf{\Lambda}}^{X}_{nT}(\theta) \widehat{\boldsymbol{P}}^{X*}_{nT}(\theta)$$

  as an estimator of the spectral density matrix of $\boldsymbol{\chi}_{nt}$.

- **Step 5.** Let $n^* := m(q+1)$ with $m := \left\lceil \frac{n}{q+1} \right\rceil$. Dropping the last $n - m(q+1)$ series, denote by $\widehat{\mathbf{\Sigma}}^{\chi}_{n^*T}(\theta)$ the $n^* \times n^*$ spectral density matrix corresponding to the remaining n^* series[14].
- **Step 6.** By inverse Fourier transform of $\widehat{\mathbf{\Sigma}}^{\chi}_{n^*T}(\theta)$, compute the estimated autocovariance matrices $\widehat{\mathbf{\Gamma}}^{\chi}_{k}$ of the m $(q+1)$-dimensional sub-vectors

$$\boldsymbol{\chi}^{k}_{t} = (\chi_{(k-1)(q+1)+1,t} \cdots \chi_{k(q+1),t})', \quad k = 1, ..., m.$$

[13]Consistency requires conditions on K, M_T and B_T, for which again we refer to Barigozzi and Hallin (2020).

[14]For the sake of simplicity we keep the same notation for the n^* reordered observed series.

Then, from the latter, obtain, via Akaike order identification and Yule-Walker equations, estimators $\hat{\mathbf{A}}^k(L)$ of the m VAR filters $\mathbf{A}^k(L)$; stacking them into a block-diagonal matrix $\hat{\mathbf{A}}_n(L)$, compute the estimates $\widehat{\mathbf{Y}}_{nt} := \hat{\mathbf{A}}_n(L)\mathbf{X}_{nt}$.

- **Step 7.** Obtain the estimates $\widehat{\mathbf{R}_n\mathbf{u}_t}$ of $\mathbf{R}_n\mathbf{u}_t$ by computing the first q standard principal components of $\widehat{\mathbf{Y}}_{nt}$; inverting[15] the block-diagonal filters $\hat{\mathbf{A}}_n(L)$ and then using appropriate identification constraints, we obtain the identified quantities $\widehat{\mathbf{R}}_n$ and $\widehat{\mathbf{u}}_t$, and the corresponding estimates of the impulse-response function $\widehat{\mathbf{B}}_n = [\hat{\mathbf{A}}_n(L)]^{-1}\widehat{\mathbf{R}}_n$.

Following Forni et al. (2017) we adopt a Cholesky identification scheme to obtain the identification of $\widehat{\mathbf{R}}_n$ and $\widehat{\mathbf{u}}_t$ (see Section 4.1 of Forni et al. (2017) for more details)—other choices are possible, though.

Steps 1-7 are those described in Forni et al. (2015, 2017) and Barigozzi and Hallin (2020), where we refer to for details. The resulting estimator $\widehat{\boldsymbol{\chi}}_{nt}$, however, depends on the ordering of the panel obtained at Step 2: that ordering indeed determines which elements of $\widehat{\boldsymbol{\Sigma}}^{\chi}_{nT}(\theta)$ are kept in $\widehat{\boldsymbol{\Sigma}}^{\chi}_{n^*T}(\theta)$ and belong to the diagonal blocks of $\widehat{\boldsymbol{\Sigma}}^{\chi}_{n^*T}(\theta)$. Forni et al. (2017) and Barigozzi and Hallin (2020) explain how to deal with this by iterating Steps 2-7 (going back to Step 2, choosing a new random permutation, hence a new n^*-dimensional subpanel, etc.) until numerical stabilization of the averaged (over the permutations) $\widehat{\boldsymbol{\chi}}_{nt}$ values; this typically takes place after few iterations[16].

- **Step 8.** Iterate Steps 2 through 7; average (after obvious reordering of the cross-section) the resulting estimates $\widehat{\mathbf{R}}_n$, $\widehat{\mathbf{u}}_t$, and $\widehat{\mathbf{B}}_n$. Denote, for the sake of simplicity, the final estimates also as $\widehat{\mathbf{R}}_n$, $\widehat{\mathbf{u}}_t$, and $\widehat{\mathbf{B}}_n$. Let $\widehat{\boldsymbol{\chi}}_{nt} := \widehat{\mathbf{B}}_n\widehat{\mathbf{u}}_t$ and $\hat{\boldsymbol{\xi}}_{nt} := \mathbf{X}_{nt} - \widehat{\boldsymbol{\chi}}_{nt}$.

All these estimators actually are sequences indexed by (n, T). Whenever this is to be emphasized, the notation $\widehat{\mathbf{R}}_n^{(n,T)}$, $\widehat{\mathbf{u}}_t^{(n,T)}$, $\widehat{\boldsymbol{\chi}}_{nt}^{(n,T)}$, $\hat{\boldsymbol{\xi}}_{nt}^{(n,T)}$, and, for n_0-dimensional ($n_0 \leq n$) subvectors, $\widehat{\mathbf{R}}_{n_0}^{(n,T)}$, $\hat{\boldsymbol{\xi}}_{n_0t}^{(n,T)}$, etc. will be adopted.

[15]The inverse of $\hat{\mathbf{A}}_n(L)$ being the block-diagonal filter with $(q+1)\times(q+1)$ diagonal blocks $[\hat{\mathbf{A}}^k(L)]^{-1}$ where q is small, this inversion is easily performed.

[16]Averaging, of course, is performed after rearrangement of the cross-sectional items in the original ordering.

The procedure described so far is the one that has been used in Della Marra (2017), Forni et al. (2018), and Giovannelli et al. (2018) in their forecasting of inflation and financial returns. In order to go one step further and estimate conditional covariance matrices, we will exploit the MGARCH and GARCH features of Assumption (GARCH). Note that, thanks to the assumption of stability under aggregation, the choice of identification constraints has no impact on the validity of Assumption (GARCH), so that VECH or BEKK QMLEs safely can be computed from the $\widehat{\mathbf{u}}_t^{(n,T)}$'s and $\hat{\boldsymbol{\xi}}_{nt}^{(n,T)}$'s obtained in Step 8.

We now proceed with the following final steps. For given $\boldsymbol{\theta}$, the variance of $\mathbf{u}_t$ conditional on $\mathcal{F}_{t-1}$ (equivalently, $\mathcal{F}_{t-1}^{\mathbf{u}}$) is a function $\boldsymbol{V}_{t|t-1;\boldsymbol{\theta}}^{\mathbf{u}}$ of $\mathbf{u}_{t-1}, \mathbf{u}_{t-2}, \ldots$ which, due to stationarity, does not depend on t. Denote by $\boldsymbol{V}_{t,\overline{\tau}|;\boldsymbol{\theta}}^{\mathbf{u}}(\mathbf{v}_1, \ldots, \mathbf{v}_\tau)$ its evaluation at $(\mathbf{v}_1, \ldots, \mathbf{v}_\tau, \mathbf{0}, \mathbf{0}, \ldots)$: then,

$$\boldsymbol{V}_{t|t-1;\boldsymbol{\theta}}^{\mathbf{u}} = \lim_{\tau\to\infty} \boldsymbol{V}_{t,\overline{\tau}|;\boldsymbol{\theta}}^{\mathbf{u}}(\mathbf{u}_{t-1}, \ldots, \mathbf{u}_{t-\tau}) \tag{12}$$

a.s. for any $\boldsymbol{\theta}$ and t. The notation $V_{t|t-1;\boldsymbol{\vartheta}_i}^{\xi_i}$ and $V_{t,\overline{\tau}|;\boldsymbol{\vartheta}_i}^{\xi_i}(v_1, \ldots, v_\tau)$ is used in an obvious similar way for each variable ξ_{it}, with

$$V_{t|t-1;\theta_i}^{\xi_i} = \lim_{\tau\to\infty} V_{t,\overline{\tau}|;\boldsymbol{\vartheta}_i}^{\xi_i}(\xi_{i,t-1}, \ldots, \xi_{i,t-\tau}) \tag{13}$$

a.s. for any $\boldsymbol{\vartheta}_i$, i, and t. Now, since $\boldsymbol{\theta}$ is unknown, denote by $\boldsymbol{\theta}_{(T)}$ its QMLE; more precisely, denote by $\boldsymbol{\theta}_{(T)}$ the mapping from $(\mathbf{v}_T, \ldots, \mathbf{v}_1) \in \mathbb{R}^{qT}$ to the maximizer $\boldsymbol{\theta}_{(T)}(\mathbf{v}_T, \ldots, \mathbf{v}_1) \in \boldsymbol{\Theta}_q$ of the MGARCH likelihood computed at $\mathbf{v}_T, \ldots, \mathbf{v}_1$. The notation $\boldsymbol{\vartheta}_{i;(T)}$ is used in an obvious similar way for each $(\xi_{iT}, \ldots, \xi_{i1})$.

- **Step 9a.** Run, over the q-dimensional T-uple $\widehat{\mathbf{u}}_1^{(n,T)}, \ldots, \widehat{\mathbf{u}}_T^{(n,T)}$, a QML estimation procedure for the parameter $\boldsymbol{\theta}$ of the MGARCH model of Assumption (GARCH); this yields an estimator

$$\hat{\boldsymbol{\theta}}_{(T)}^{(n,T)} := \boldsymbol{\theta}_{(T)}(\widehat{\mathbf{u}}_T^{(n,T)}, \ldots, \widehat{\mathbf{u}}_1^{(n,T)})$$

of $\boldsymbol{\theta}$. Choose a finite lag $\tau < T$ and let[17]

$$\widehat{\boldsymbol{V}}_{T+1,\overline{\tau}|}^{\mathbf{u};(n,T)} := \boldsymbol{V}_{T+1,\overline{\tau}|;\hat{\boldsymbol{\theta}}_{(T)}^{(n,T)}}^{\mathbf{u}}(\widehat{\mathbf{u}}_T^{(n,T)}, \ldots, \widehat{\mathbf{u}}_{T-\tau+1}^{(n,T)}). \tag{14}$$

[17] The subscript $_{(T)}$ indicates that $\hat{\boldsymbol{\theta}}_{(T)}^{(n,T)}$, as a QMLE, is defined over T values of the q-dimensional space of common shocks, that is, is mapping $(\mathbf{v}_T, \ldots, \mathbf{v}_1) \in \mathbb{R}^{qT}$ to $\hat{\boldsymbol{\theta}}_{(T)}^{(n,T)}(\mathbf{v}_T, \ldots, \mathbf{v}_1) \in \boldsymbol{\Theta}_q$; the $\hat{}$ and the $^{(n,T)}$ superscript are the indication that this QMLE $\hat{\boldsymbol{\theta}}_{(T)}(\mathbf{v}_T, \ldots, \mathbf{v}_1)$ is to be computed at $(\mathbf{v}_T, \ldots, \mathbf{v}_1) = (\widehat{\mathbf{u}}_T^{(n,T)}, \ldots, \widehat{\mathbf{u}}_1^{(n,T)})$.

- **Step 9b.** Similarly run, over each of the n univariate T-uples $\hat{\xi}^{(n,T)}_{i1}, \ldots, \hat{\xi}^{(n,T)}_{iT}$, a QML estimation procedure for the parameters $\boldsymbol{\vartheta}_i$, $i = 1, \ldots, n$ of the univariate idiosyncratic AR-GARCH models of Assumption (GARCH); this yields n estimators

$$\hat{\boldsymbol{\vartheta}}^{(n,T)}_{i;(T)} := \boldsymbol{\vartheta}_{i;(T)}(\hat{\xi}^{(n,T)}_{iT}, \ldots, \hat{\xi}^{(n,T)}_{i1}).$$

Let $\widehat{V}^{\xi_i;(n,T)}_{T+1,\overline{\tau}|} := V^{\xi_i}_{T+1,\overline{\tau}|;\hat{\boldsymbol{\vartheta}}^{(n,T)}_{i;(T)}}(\hat{\xi}^{(n,T)}_{iT}, \ldots, \hat{\xi}^{(n,T)}_{i,T-\tau+1})$ and, for $n_0 \le n$, denote by

$$\widehat{\boldsymbol{V}}^{\boldsymbol{\xi}_{n_0};(n,T)}_{T+1,\overline{\tau}|} := \text{diag}\left(\widehat{V}^{\xi_i;(n,T)}_{T+1,\overline{\tau}|}, \ldots, \widehat{V}^{\xi_{n_0};(n,T)}_{T+1,\overline{\tau}|}\right) \tag{15}$$

the $n_0 \times n_0$ diagonal matrix of the predicted (regularized) conditional variances of the idiosyncratic variables $\xi_{1,T+1}, \ldots, \xi_{n_0,T+1}$.

The diagonal matrix (15), however, is neglecting the possible idiosyncratic cross-covariances, which, as explained at the end of Section 2, are mild (non-pervasive) but not nil. As a consequence, (15) yields a predictor of $\text{Var}^*(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1})$ rather than $\text{Var}(\boldsymbol{\xi}_{nt}|\mathcal{F}_{n;t-1})$. Similarly, (16) below is the predictor of $\text{Var}^*(\mathbf{X}_{nt}|\mathcal{F}_{n;t-1})$.

- **Step 9c.** Compute our predictor of the $n_0 \times n_0$ conditional covariance matrix of $(X_{1,T+1}, \ldots, X_{n_0,T+1})$ $(n_0 \le n)$ as

$$\widehat{\boldsymbol{V}}^{\mathbf{X}_{n_0};(n,T)}_{T+1,\overline{\tau}|} := \widehat{\mathbf{R}}^{(n,T)}_{n_0} \widehat{\boldsymbol{V}}^{\mathbf{u};(n,T)}_{T+1,\overline{\tau}|} \widehat{\mathbf{R}}^{(n,T)\prime}_{n_0} + \widehat{\boldsymbol{V}}^{\boldsymbol{\xi}_{n_0};(n,T)}_{T+1,\overline{\tau}|}. \tag{16}$$

3.3 Consistency

Consistency, as well as any other asymptotic property, consists in embedding the actual finite-sample model into a sequence of models indexed by n and T going to infinity. This, however, can be achieved in several ways. Here, we let n_0 denote the (fixed) dimension of the covariance matrix to be predicted and T_0 the point in time where one-step ahead prediction is to be made, while n and T are indexing the sequence of fictitious "future" panels along which asymptotic statements are to be made. As already explained, we are neglecting idiosyncratic cross-covariances; to avoid introducing heavier notation, from now on, we are writing $\boldsymbol{V}^{\mathbf{X}_{n_0}}_{T_0+1|T_0}$ for the resulting conditional covariance matrix $\text{Var}^*(\mathbf{X}_{n_0,T_0+1}\big|\mathcal{F}_{n;T_0})$. With that notation, we are interested in estimating $\boldsymbol{V}^{\mathbf{X}_{n_0}}_{T_0+1|T_0}$ and the estimator (16) we are proposing takes the form

$$\widehat{\boldsymbol{V}}^{\boldsymbol{X}_{n_0};(n_0,T_0)}_{T_0+1,\overline{\tau}|} := \widehat{\mathbf{R}}^{(n_0,T_0)}_{n_0}\widehat{\boldsymbol{V}}^{\mathbf{u};(n_0,T_0)}_{T_0+1,\overline{\tau}|}\widehat{\mathbf{R}}^{(n_0,T_0)\prime}_{n_0} + \widehat{\boldsymbol{V}}^{\boldsymbol{\xi}_{n_0};(n_0,T_0)}_{T_0+1,\overline{\tau}|}. \tag{16$'$}$$

That estimator is to be considered as an element of the (n,T)-indexed sequence

$$\widehat{\boldsymbol{V}}^{\boldsymbol{X}_{n_0};(n,T)}_{T_0+1,\overline{\tau}|} := \widehat{\mathbf{R}}^{(n,T)}_{n_0}\widehat{\boldsymbol{V}}^{\mathbf{u};(n,T)}_{T_0+1,\overline{\tau}|}\widehat{\mathbf{R}}^{(n,T)\prime}_{n_0} + \widehat{\boldsymbol{V}}^{\boldsymbol{\xi}_{n_0};(n,T)}_{T_0+1,\overline{\tau}|} \quad n \geq n_0,\ T \geq T_0$$

based (see (14) and (15)) on the QMLE mappings $\boldsymbol{\theta}_{(T_0)}$ and $\boldsymbol{\vartheta}_{i;(T_0)}$, $i = 1,\ldots,n_0$ involving the T_0 arguments $\widehat{\mathbf{u}}^{(n,T)}_{T_0},\ldots,\widehat{\mathbf{u}}^{(n,T)}_1$ and $\hat{\xi}^{(n,T)}_{iT_0},\ldots,\hat{\xi}^{(n,T)}_{i1}$, respectively.

The following proposition establishes the consistency properties of (16$'$) as n and T tend to infinity (n_0 and T_0 large enough but fixed).

Proposition 2. *Let Assumptions (GDFM)(i)-(ix) and (GARCH), and Assumptions (K), (T), (L4), and (L5) in Barigozzi and Hallin (2020) hold. Then, for any $n_0 \in \mathbb{N}$, any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$ and $\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_{n_0}$ in $\boldsymbol{\Theta}_1$, any $\epsilon > 0$ and $\eta > 0$, there exist $\tau^*(n_0,\boldsymbol{\theta},\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_{n_0};\epsilon,\eta)$ and $T_0^*(n_0,\boldsymbol{\theta},\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_{n_0};\epsilon,\eta)$ and, for any $T_0 \geq T_0^*$, $n^*(n_0,\boldsymbol{\theta},\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_{n_0};T_0;\epsilon,\eta)$, and $T^*(n_0,\boldsymbol{\theta},\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_{n_0};T_0;\epsilon,\eta)$ such that*

$$\mathrm{P}\left[\left\|\widehat{\boldsymbol{V}}^{\boldsymbol{X}_{n_0};(n,T)}_{T_0+1,\overline{\tau}|} - \boldsymbol{V}^{\boldsymbol{X}_{n_0}}_{T_0+1|T_0}\right\| \geq \epsilon\right] \leq \eta \tag{17}$$

for all $n_0 \in \mathbb{N}$, $\tau \geq \tau^*$, $T_0 \geq T_0^*$, $n \geq n^*$, and $T \geq T^*$.

A stronger form of (17), allowing $T_0 = T$, would be

$$\mathrm{P}\left[\left\|\widehat{\boldsymbol{V}}^{\boldsymbol{X}_{n_0};(n,T)}_{T+1,\overline{\tau}|} - \boldsymbol{V}^{\mathbf{X}_{n_0}}_{T+1|T}\right\| \geq \epsilon\right] \leq \eta,$$

for all $n_0 \in \mathbb{N}$, $\tau \geq \tau^*$, $n \geq n^*$, and $T \geq T^*$; this holds true if the values $n^* = n_0$ and $T^* = T_0$ are admissible in Proposition 2; establishing this latter fact, however, would require sharper (namely, sharper than the Barigozzi and Hallin (2020) bound (23) below) results on the magnitude of the differences $\|\widehat{\mathbf{u}}^{(n,T)}_t - \mathbf{u}_t\|$.

The proof of Proposition 2 relies on two lemmas establishing the consistency of $\widehat{\boldsymbol{V}}^{\mathbf{u};(n,T)}_{T+1,\overline{\tau}|}$ and $\widehat{V}^{\xi_i;(n,T)}_{T+1,\overline{\tau}|}$, $i = 1,\ldots,n_0$, respectively.

Lemma 1. *Under the assumptions of Proposition 2, for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$, any $\epsilon_1 > 0$, and any $\eta_1 > 0$, there exist $\tau^\dagger(\epsilon_1,\eta_1,\boldsymbol{\theta})$ and $T_0^\dagger(\epsilon_1,\eta_1;\boldsymbol{\theta})$ and, for any $\boldsymbol{\vartheta}_1,\ldots,\boldsymbol{\vartheta}_{n_0}$ in $\boldsymbol{\Theta}_1$*

and $T_0 \geq T_0^\dagger$, *there exist* $n^\dagger(\epsilon_1, \eta_1; T_0; \boldsymbol{\theta}, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$ *and* $T^\dagger(\epsilon_1, \eta_1; T_0; \boldsymbol{\theta}, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$ *such that*

$$\mathrm{P}\left[\left\| \widehat{\boldsymbol{V}}^{\boldsymbol{u};(n,T)}_{T_0+1,\overline{\tau}|} - \boldsymbol{V}^{\boldsymbol{u}}_{T_0+1|T_0} \right\| \geq \epsilon_1\right] \leq \eta_1 \tag{18}$$

for all $\tau \geq \tau^\dagger$, $T_0 \geq T_0^\dagger$, $n \geq n^\dagger$, *and* $T \geq T^\dagger$.

Lemma 2. *Under the assumptions of Proposition 2, for any* $n_0 \in \mathbb{N}$, *any* $\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0}$ *in* $\boldsymbol{\Theta}_1$, *any* $\epsilon_2 > 0$, *and any* $\eta_2 > 0$, *there exist* $\tau^\ddagger(\epsilon_2, \eta_2; n_0; \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$ *and* $T_0^\ddagger(\epsilon_2, \eta_2; n_0; \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$ *and, for any* $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$ *and* $T_0 \geq T_0^\ddagger$, *there exist* $n^\ddagger(\epsilon_2, \eta_2; n_0, T_0; \boldsymbol{\theta}, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$ *and* $T^\ddagger(\epsilon_2, \eta_2; n_0, T_0; \boldsymbol{\theta}, \boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$ *such that*

$$\max_{1 \leq i \leq n_0} \mathrm{P}\left[\left\| \widehat{V}^{\xi_i;(n,T)}_{T_0+1,\overline{\tau}|} - V^{\xi_i}_{T_0+1|T_0} \right\| \geq \epsilon_2\right] \leq \eta_2 \qquad i = 1, \ldots, n_0 \tag{19}$$

for all $\tau \geq \tau^\ddagger$, $T_0 \geq T_0^\ddagger$, $n \geq n^\ddagger$, *and* $T \geq T^\ddagger$.

These two lemmas rely on a repeated application of the following elementary result.

Lemma 3. *Let* $\epsilon = \epsilon_a + \epsilon_b$ *and* $\eta = \eta_a + \eta_b$ *with* ϵ_a, ϵ_b, η_a, *and* η_b *strictly positive. Denote by* $\boldsymbol{a}$ *and* $\boldsymbol{b}$ *two* d*-dimensional random vectors with unspecified joint distribution such that* $\mathrm{P}[\|\boldsymbol{a}\| \geq \epsilon_a] \leq \eta_a$ *and* $\mathrm{P}[\|\boldsymbol{b}\| \geq \epsilon_b] \leq \eta_b$. *Then,*

$$\mathrm{P}\left[\|\boldsymbol{a} + \boldsymbol{b}\| \geq \epsilon\right] \leq \eta.$$

Proof of Lemma 1. Considering the difference

$$\begin{aligned}
&\boldsymbol{V}^{\mathbf{u}}_{T_0+1|T_0} - \widehat{\boldsymbol{V}}^{\mathbf{u};(n,T)}_{T_0+1,\overline{\tau}|} \\
&\quad = \boldsymbol{V}^{\mathbf{u}}_{T_0+1|T_0}(\mathbf{u}_{T_0}, \mathbf{u}_{T_0-1}, \ldots) - \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}|;\boldsymbol{\theta}_{(T_0)}(\widehat{\mathbf{u}}^{(n,T)}_{T_0}, \ldots, \widehat{\mathbf{u}}^{(n,T)}_1)}(\widehat{\mathbf{u}}^{(n,T)}_{T_0}, \ldots, \widehat{\mathbf{u}}^{(n,T)}_{T_0-\tau+1}),
\end{aligned}$$

decompose it into

$$\begin{aligned}
&\boldsymbol{V}^{\mathbf{u}}_{T_0+1|T_0}(\mathbf{u}_{T_0}, \mathbf{u}_{T_0-1}, \ldots) - \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}|;\boldsymbol{\theta}}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_{T_0-\tau+1}) \\
&+ \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}|;\boldsymbol{\theta}}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_{T_0-\tau+1}) - \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}|;\boldsymbol{\theta}_{(T_0)}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_1)}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_{T_0-\tau+1}) \\
&+ \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}|;\boldsymbol{\theta}_{(T_0)}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_1)}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_{T_0-\tau+1}) - \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}|;\boldsymbol{\theta}_{(T_0)}(\widehat{\mathbf{u}}^{(n,T)}_{T_0}, \ldots, \widehat{\mathbf{u}}^{(n,T)}_1)}(\widehat{\mathbf{u}}^{(n,T)}_{T_0}, \ldots, \widehat{\mathbf{u}}^{(n,T)}_{T_0-\tau+1}) \\
&=: \boldsymbol{E}_1 + \boldsymbol{E}_2 + \boldsymbol{E}_3, \text{ say.}
\end{aligned}$$

The conditions for stationarity in Assumption (GARCH) imply that, uniformly in t, (12) and (13) hold in probability as $\tau \to \infty$. Hence, for all $\epsilon_1 > 0$, $\eta_1 > 0$, and $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$, there exists a $\tau^\dagger$ such that, for all $\tau \geq \tau^\dagger$, all T_0, and, since $\boldsymbol{\Theta}_q$ is compact, all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

$$\mathrm{P}\left[\|\boldsymbol{E}_1\| \geq \epsilon_1/3\right] \leq \eta_1/3. \tag{20}$$

QMLE consistency, on the other hand, implies that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$, $\varepsilon > 0$, and $\eta_1 > 0$, there exists a $T_0^\dagger$ such that, for all $T_0 \geq T_0^\dagger$,

$$\mathrm{P}\left[\left\|\boldsymbol{\theta}_{(T_0)}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_1) - \boldsymbol{\theta}\right\| \geq \varepsilon\right] \leq \eta_1/3. \tag{21}$$

Continuity over a compact implies uniform continuity. Hence, continuity of $\boldsymbol{\theta} \mapsto \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}];\boldsymbol{\theta}}$ entails uniform continuity over $\boldsymbol{\Theta}_q$ and the existence of $\varepsilon > 0$ such that $\left\|\boldsymbol{\theta}_{(T_0)}(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_1) - \boldsymbol{\theta}\right\| \leq \varepsilon$ implies $\|\boldsymbol{E}_2\| \leq \epsilon/3$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$. It follows that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_q$ and $T_0 \geq T_0^\dagger$,

$$\mathrm{P}\left[\|\boldsymbol{E}_2\| \ \geq \epsilon_1/3\right] \leq \eta_1/3. \tag{22}$$

Finally, $\widehat{\mathbf{u}}_t^{(n,T)}$ is uniformly consistent for $\mathbf{u}_t$: Proposition 1 of Barigozzi and Hallin (2020)) entails

$$\max_{1 \leq t \leq T} \|\widehat{\mathbf{u}}_t^{(n,T)} - \mathbf{u}_t\| = O_{\mathrm{P}}\left(\max\left(\frac{B_T}{\sqrt{T}}, \frac{1}{B_T}, \frac{1}{\sqrt{n}}\right)\log T\right), \tag{23}$$

meaning that, for all $\varepsilon > 0$ and $\eta_1 > 0$, any $\boldsymbol{\theta}$, and $(\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$, there exists (n°, T°) such that

$$\mathrm{P}\left[\max_{1 \leq t \leq T}\left\|\widehat{\mathbf{u}}_t^{(n,T)} - \mathbf{u}_t\right\| \geq \varepsilon\right] \leq \eta_1/3 \tag{24}$$

for all $n \geq n^\circ$ and $T \geq T^\circ$. Now, for given T_0, the mapping

$$(\mathbf{v}_{T_0}, \ldots, \mathbf{v}_1) \mapsto \boldsymbol{V}^{\mathbf{u}}_{T_0+1,\overline{\tau}];\boldsymbol{\theta}_{(T_0)}(\mathbf{v}_{T_0},\ldots,\mathbf{v}_1)}(\mathbf{v}_{T_0}, \ldots, \mathbf{v}_{T_0-\tau+1}) \tag{25}$$

is continuous, hence uniformly continuous, over any compact subset C_η of $\mathbb{R}^{qT_0}$ such that $\mathrm{P}\left[(\mathbf{u}_{T_0}, \ldots, \mathbf{u}_1) \in C_\eta\right] \geq 1 - \eta_1/3$. The continuous mapping theorem thus guarantees the existence, for any T_0, any $\boldsymbol{\theta}$ and $(\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_{n_0})$, any $\epsilon_1 > 0$ and $\eta_1 > 0$, of $n^\dagger$ and $T^\dagger$ such that, for $n \geq n^\dagger$ and $T \geq T^\dagger$,

$$\mathrm{P}\left[\|\boldsymbol{E}_3\| \ \geq \epsilon_1/3\right] \leq \eta_1/3. \tag{26}$$

The desired result follows from (20), (22), (26), and Lemma 3 applied to $E_2 + E_3$, then to $E_1 + (E_2 + E_3)$. $\square$

Proof of Lemma 2. The proof, for each $1 \leq i \leq n_0$, hence for a finite collection of n_0 of them, goes along the same lines as for Lemma 1, with a univariate AR-GARCH instead of a q-dimensional MGARCH. □

Proof of Lemma 3. Basic probabilistic operations yield

$$\begin{aligned}\eta_a + \eta_b &\geq \mathrm{P}[\|\mathbf{a}\| \geq \epsilon_a] + \mathrm{P}[\|\mathbf{b}\| \geq \epsilon_b] \\ &\geq \mathrm{P}[\|\mathbf{a}\| \geq \epsilon_a \text{ or } \|\mathbf{b}\| \geq \epsilon_b] \geq \mathrm{P}[\|\mathbf{a}+\mathbf{b}\| \geq \epsilon_a + \epsilon_b = \epsilon].\end{aligned}$$ □

Proof of Proposition 2. The result follows from Lemmas 1 and 2, the consistency, as $n, T \to \infty$, of $\widehat{\mathbf{R}}_{n_0}^{(n,T)}$ as an estimator of $\mathbf{R}_{n_0}$, and an application of Slutsky's Lemma. □

4 Finite-sample performances

In practice, VECH and BEKK QMLEs, however, are reported to be numerically quite unstable, and typically strongly depend on the initial values considered in the numerical solution of the likelihood equations. This is a well-documented fact; see, for instance, Lien et al. (2002) and Manabu (2015). Rather than VECH or BEKK, we therefore compute DCC QMLEs which are known to be quite robust to missespecification; see Chang et al. (2011), Chevallier (2012), Laurent et al. (2012), Amendola and Candila (2017), or de Almeida et al. (2018). Our Monte Carlo experiments confirm that, even though the actual data-generating process is BEKK, misspecified DCC QMLEs outperform the correctly specified full BEKK ones.

4.1 Monte Carlo experiments

In this section, we investigate the finite-sample performance of the proposed procedure through Monte Carlo simulations.

Simulations were performed from three data-generating processes (DGPs). The first DGP is a static factor model with two common factors, the second and third ones are dynamic factor models with finite- and infinite-dimensional factor spaces, respectively. The common shocks and the idiosyncratic components in all four DGPs are conditionally heteroscedastic.

In all DGPs, the idiosyncratic components are defined as $\boldsymbol{\xi}_t := (\xi_{1t}, \dots, \xi_{nt})$ with $\boldsymbol{\xi}_t = \boldsymbol{P}_t^{1/2}\boldsymbol{\zeta}_t$, where $\boldsymbol{P}_t$ is an $n \times n$ diagonal matrix containing the conditional variances $\boldsymbol{P}_{it}$ of ξ_{it}; $\boldsymbol{\zeta}_t := (\zeta_{1t}, \dots, \zeta_{nt})$, where ζ_{it}, $i = 1, \dots, n$, $t = 1, 2, ..., T$ are sequences of i.i.d. innovations generated either from a standard N(0,1) or a centered and standardized Student t_5 distribution. The conditional variances $\boldsymbol{P}_{it}$ follow GARCH(1,1) processes with parameters $\boldsymbol{\vartheta}_i = (\omega_i, \alpha_i, \beta_i)$, of the form

$$\boldsymbol{P}_{it} = \omega_i + \alpha_i \xi_{it}^2 + \beta_i \boldsymbol{P}_{i,t-1}, \quad i = 1, ..., n,$$

where $\omega_i > 0$, $\alpha_i > 0$, $\beta_i \geq 0$, and $\alpha_i + \beta_i < 1$; the parameters values α_i and β_i are independently generated from uniform distributions over [0.01, 0.045] and [0.85, 0.95], respectively, and $\omega_i := 1 - \alpha_i - \beta_i$, so that the unconditional variance of ξ_{it} is $\mathrm{V}(\xi_{it}) = 1$. As for the shocks $\mathbf{u}_t$ driving the common components $\boldsymbol{\chi}_t$, they were generated from the following four DGPs.

DGP1 (two common shocks; static loadings). Two common shocks $\mathbf{u}_t = (u_{1t}, u_{2t})'$, generated from a BEKK(1,1,1) model

$$\mathbf{u}_t = \mathbf{Q}_t^{1/2} \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \quad \text{with } \mathbf{Q}_t = \mathbf{C}_0'\mathbf{C}_0 + \mathbf{C}_1'\mathbf{u}_{t-1}\mathbf{u}_{t-1}'\mathbf{C}_1 + \mathbf{C}_2'\mathbf{Q}_{t-1}\mathbf{C}_2. \tag{27}$$

Here, η_{it}, $i = 1, 2$, are i.i.d. innovations generated by a N(0,1) or a centered and standardized Student t_5 distribution. In order to guarantee $\mathrm{E}(\mathbf{Q}_t) = \mathrm{E}(\mathbf{u}_{t-1}\mathbf{u}_{t-1}') = \mathbf{I}_q$, we set $\mathbf{C}_0'\mathbf{C}_0 = \mathbf{I}_q - \mathbf{C}_1'\mathbf{C}_1 - \mathbf{C}_2'\mathbf{C}_2$. Parameters of the BEKK are extracted from uniform distributions with ranges as in Alessi et al. (2009): $\mathbf{C}_1$ has diagonal elements uniformly distributed over [0.1,0.5] and off-diagonal elements uniformly distributed over [-0.2,0.2], while the diagonal elements of $\mathbf{C}_2$ and the off-diagonal ones are uniformly distributed over [0.8,0.95] and [-0.15,0.15], respectively (all uniforms mutually independent). For each randomly generated set of parameters, the covariance stationary of the resulting BEKK model has been checked before proceeding. Here, $\boldsymbol{\chi}_t = \mathbf{R}\mathbf{u}_t$ where $\mathbf{R}$ is an $n \times 2$ matrix with orthonormal columns randomly generated via the *RandOrthMat* Matlab function.

DGP2 (four factors driven by $q = 2$ common shocks; static loadings). Four factors $\mathbf{F}_t = (F_{1t}, \dots, F_{4t})'$ driven by $q = 2$ common shocks $\mathbf{u}_t$, yielding a GDFM with finite-dimensional factor space. The shocks are generated from the same BEKK model as in DGP2 and the factors are a $\mathbf{u}_t$-driven VAR(4)

$$\mathbf{F}_t = \mathbf{\Phi}\mathbf{F}_{t-1} + \mathbf{K}\mathbf{u}_t \quad \text{and} \quad \mathbf{u}_t = \mathbf{Q}_t^{1/2}\boldsymbol{\eta}_t,$$

with $\mathbf{Q}_t$ as in (27) and $\boldsymbol{\eta}_t$ generated as in DGP2 ($\mathbf{\Phi}$ is 4×4 and $\mathbf{K}$ is 4×2). The entries of $\mathbf{\Lambda}$ and $\mathbf{K}$ are independent and uniformly distributed over $[-1, 1]$. The entries of $\mathbf{\Phi}$ are generated as follows: first we generate independent entries uniformly distributed over the interval[-1,1]; second, we divide the resulting matrix by its spectral norm; third, we multiply the resulting matrix by a random variable uniformly distributed on the interval [0.4,0.9] to ensure stationarity while preserving sizeable dynamic responses[18]. Here, $\boldsymbol{\chi}_t = \mathbf{\Lambda}\mathbf{u}_t$, where $\mathbf{\Lambda}$ is an $n \times 4$ matrix with independent entries uniformly distributed over $[-1, 1]$.

DGP3 (two common shocks; dynamic loadings). The common shocks $\mathbf{u}_t = (u_{1t}, u_{2t})'$ are generated from the same bivariate BEKK model as in (27); the model is a GDFM with infinite-dimensional factor space. Here,

$$\boldsymbol{\chi}_{it} = \begin{pmatrix} a_{i1}(1-\alpha_{i1})^{-1} \\ a_{i2}(1-\alpha_{i2})^{-1} \end{pmatrix} \mathbf{u}_t,$$

where a_{ij} and α_{ij}, $i = 1, ..., n$, $j = 1, 2$ are independent and uniformly distributed over the intervals [-1,1] and [-0.8,0.8], respectively.

For each DGP, we simulated 500 replications of a panel of dimensions n=60 and T=1000 (moderate dimension, $T >> n$) and 500 replications of a high-dimensional panel with n=600 and T=700 ($T \approx n$). From each replication, the covariance matrix $\boldsymbol{V}_{T+1|T}$ of $\mathbf{X}_{T+1}$ conditional on $\mathbf{X}_T, \ldots, \mathbf{X}_1$ was estimated[19] using

(a) classical PCA[20] combined with (M)GARCH modelling,

(b) the DCC model with composite likelihood, as described in Pakel et al. (2020),

(c) the Alessi et al. (2009) model, and

(d) our model[21],

---

[18]This DGP is similar to the one considered by Alessi et al. (2009).

[19]As we are not interested in asymptotics here, we set $n_0 = n$, $T_0 = T$, and $\tau = T - 1$.

[20]In the spirit of Diebold and Nerlove (1989) and Van der Weide (2002), static factors are extracted via principal component analysis; an (M)GARCH model then is fitted to the extracted factors. Idiosyncratic components are modelled as independent univariate GARCH processes.

[21]Throughout, we considered 30 cross-sectional permutations and set the order S of the VAR block-diagonal filters to one.

labeled as PCA, DCC, ABC, and GDFM-CHF, respectively.[22] For simplicity, the correct numbers of factors (for DGP2) and common shocks (for DGPs 1-3) are assumed to be known, since this does not play a role in the comparative performances of procedures (a)-(d). For DGP3, no static factor representation exists and any criterion based on static representation is inappropriate. However, the PCA and ABC methods are based on the surmise that a static factor representation exists. Therefore, before running the PCA and ABC methods, we first determine a (fictitious) number of static factors via the Bai and Ng (2002) procedure.[23]

As mentioned in the previous section, estimation of BEKK models is numerically quite unstable and strongly depends on the choice of initial values. For the sake of comparison, for all DGPs we considered both the DCC(1,1) and BEKK(1,1,1) estimates of the conditional covariance matrix of common shocks in the PCA, ABC and GDFM-CHF models, with lables such as PCA-BEKK, ABC-DCC, etc.[24]

Hereafter, for the sake of simplicity, we denote by $\boldsymbol{V}_{T+1|T}$ the simulated covariance matrix of $\mathbf{X}_{T+1}$ conditional on $\mathbf{X}_T, \ldots, \mathbf{X}_1$ and by $\widehat{\boldsymbol{V}}_{T+1|T}$ its various estimated versions. In order to compare the performances of those various estimators, we compute, for each simulated panel and each method, a distance between $\widehat{\boldsymbol{V}}_{T+1|T}$ and $\boldsymbol{V}_{T+1|T}$. Let

$$\boldsymbol{V}_{T+1|T} := \mathbf{R}\,\mathrm{Var}(\mathbf{u}_{\mathrm{T+1}}|\mathcal{F}_{\mathrm{n,T}})\mathbf{R}' + \mathrm{Var}(\boldsymbol{\xi}_{\mathrm{T+1}}|\mathcal{F}_{\mathrm{n,T}}) \qquad \text{for DGP1}$$
$$\boldsymbol{V}_{T+1|T} := \boldsymbol{\Lambda}\mathbf{K}\mathrm{Var}(\mathbf{u}_{T+1}|\mathcal{F}_{n,T})\mathbf{K}'\boldsymbol{\Lambda}' + \mathrm{Var}(\boldsymbol{\xi}_{T+1}|\mathcal{F}_{n,T}) \qquad \text{for DGP2,}$$

and

$$\boldsymbol{V}_{T+1|T} = \mathbf{A}\,\mathrm{Var}(\mathbf{u}_{n,T+1}|\mathcal{F}_T)\mathbf{A}' + \mathrm{Var}(\boldsymbol{\xi}_{T+1}|\mathcal{F}_{n,T}) \qquad \text{for DGP3,}$$

where $\mathbf{A}$ is the matrix with elements $a_{i,j}$, $i = 1, ..., N$, $j = 1, 2$. Following Amendola

[22] GDFM-CHF stand for General Dynamic Factor Model with Conditionally Heteroscedastic Factors.

[23] In practice, the identification procedures by Bai and Ng (2002) or Alessi et al. (2010) in the static case, by Hallin and Liška (2007) in the GDFM-CHF case, should be used prior to the estimation procedure in each replication.

[24] DCC and BEKK estimations were performed by using the MFE toolbox of Kevin K. Sheppard, freely available at http://www.kevinsheppard.com/MFE_Toolbox.

and Candila (2017), we consider four distances, $\mathrm{D}_1, , \ldots, \mathrm{D}_4$, of the form

$$\mathrm{D}(\boldsymbol{V}_{T+1|T}, \widehat{\boldsymbol{V}}_{T+1|T}) = \sum_{i=1}^{N} \sum_{j=i}^{N} \omega(i,j)(\sigma_{i,j} - \widehat{\sigma}_{i,j})^2, \tag{28}$$

where $\sigma_{i,j}$ and $\widehat{\sigma}_{i,j}$ are the (i, j) entries of $\boldsymbol{V}_{T+1|T}$ and $\widehat{\boldsymbol{V}}_{T+1|T}$, respectively, and the weights $\omega(i, j)$ are provided in Table 1.

Table 1: Weights $\omega(i, j)$, $i = 1, \ldots, n$, $j = i, \ldots, n$ in the distances D_1-D_4 in (28).

D_1	$w(i,j) = 1$	for all i and j
D_2	$w(i,j) = 1$	when $i = j$; 0 otherwise
D_3	$w(i,j) = 2$	when $\widehat{\sigma}_{i,j} > h_{i,j}$; 1 otherwise
D_4	$w(i,j) = 2$	when $\widehat{\sigma}_{i,j} < h_{i,j}$; 1 otherwise

Distance D_1, which gives equal weights for the variance and covariances, yields a "total" unweighted squared Euclidean distance between $\mathrm{Vech}(\widehat{\boldsymbol{V}}_{T+1|T})$ and $\mathrm{Vech}(\boldsymbol{V}_{T+1|T})$; distance D_2 is an unweighted squared Euclidean distance between $\mathrm{Diag}(\widehat{\boldsymbol{V}}_{T+1|T})$ and $\mathrm{Diag}(\boldsymbol{V}_{T+1|T})$ (hence disregards the covariances)[25]; distance D_3 penalizes negative errors, while D_4 penalizes the positive ones. It is important to note that, in D_3 and D_4, the weights themselves are data-driven, so that, for a given replication, different methods lead to different weights.

4.2 Simulation results

The results of the Monte Carlo experiments for moderate and high-dimensional data are summarized in Figure 1 and Table 2 and in Figure 2 and Table 3, respectively. Figures 1 and 2 present the boxplots of the distances defined in (28), in logarithmic scale, and Tables 2 and 3 report the average distances in logarithmic scale and indicate the subset of models with best performance obtained using the Model Confident Set (MCS) approach (Hansen et al., 2011) at 10% level.

For moderate sample size (Figure 1 and Table 2), the conditional covariance of the common shocks were estimated using both BEKK and DCC-based procedures.

[25] The classical notation $\mathrm{Vech}(\mathbf{M})$ stands for the vector stacking the upper diagonal entries of a square matrix $\mathbf{M}$, and $\mathrm{Diag}(\mathbf{M})$ for the vector of its diagonal elements.

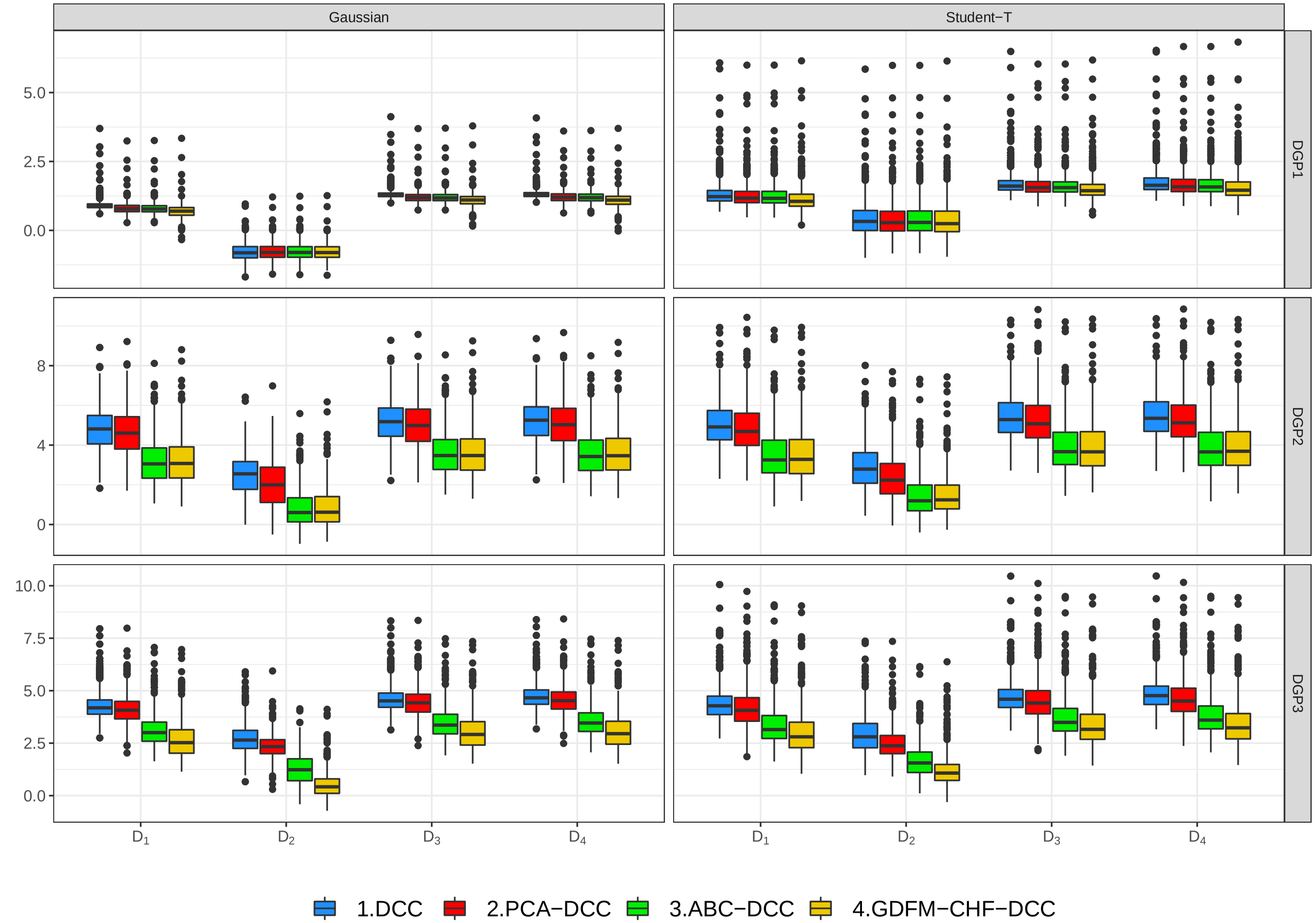

Figure 1: Boxplots of the logarithms of the distances D_1, D_2, D_3, and D_4, for DGP1 (top panel), DGP2 (middle panel), and DGP3 (bottom panel) across 500 Monte Carlo replications using Gaussian innovations (left panel) and Student t_5 innovations (right panel), respectively. Color code: DCC (1), PCA-DCC (2), ABC-DCC (3), and GDFM-CHF-DCC (4); $n = 60$ and $T = 1000$.

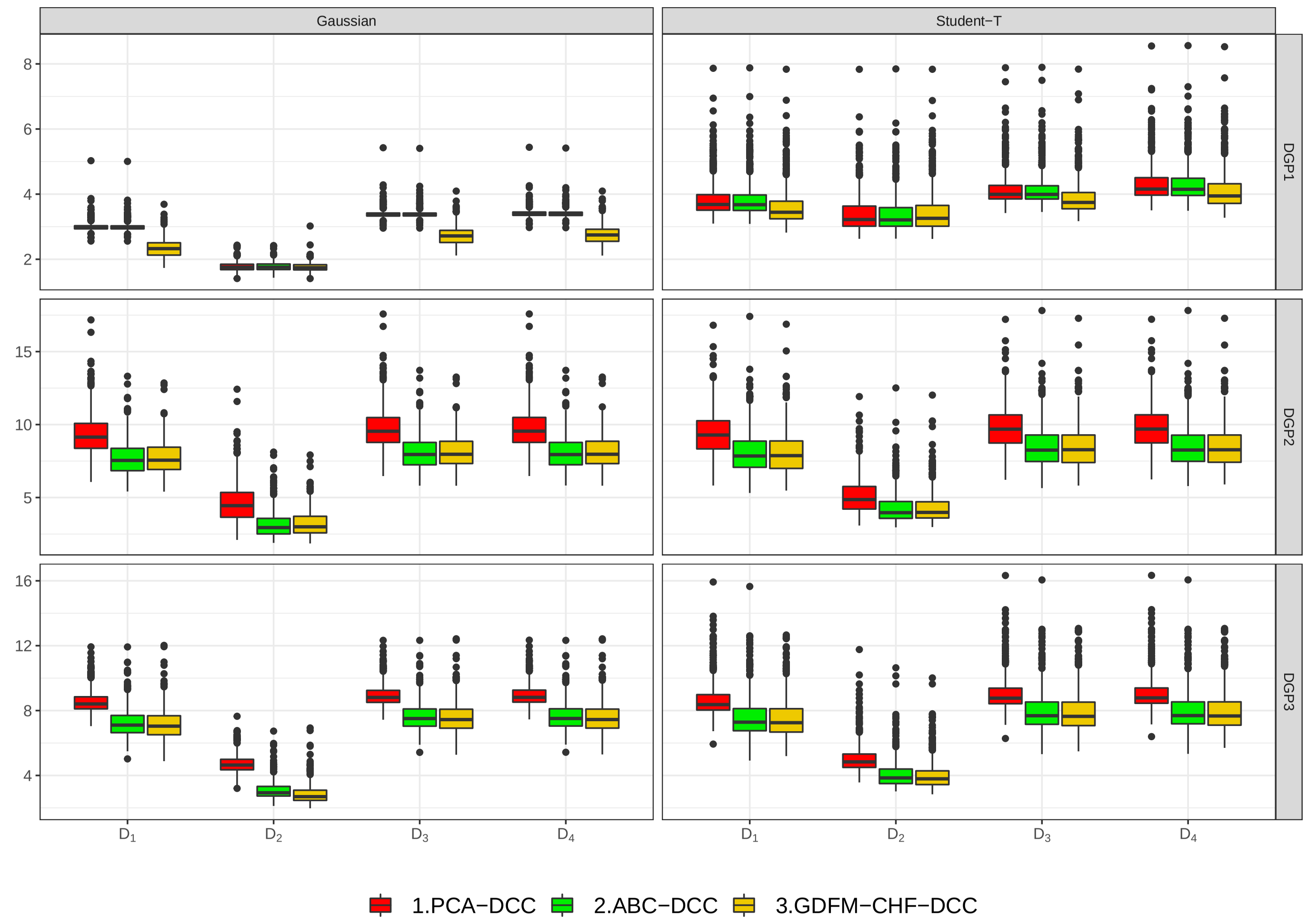

Figure 2: Boxplots of the logarithms of the distances D_1, D_2, D_3, and D_4, for DGP1 (top panel), DGP2 (middle panel), and DGP3 (bottom panel) across 500 Monte Carlo replications using Gaussian innovations (left panel) and Student t_5 innovations (right panel). Color code: PCA-DCC (1), ABC-DCC (2), and GDFM-CHF-DCC (3); $n = 600$ and $T = 700$.

Table 2: Average distances D_1, D_2, D_3, and D_4 (in logarithmic scale) for DGP1 (top panel), DGP2 (middle panel) and DGP3 (bottom panel) across 500 Monte Carlo replications using Gaussian innovations (left panel) and Student t_5 innovations (right panel). Shadowed cells stand for the MCS at 90%. $N = 60$, $T = 1000$

		Gaussian							Student-T						
		DCC	PCA BEKK	PCA DCC	ABC BEKK	ABC DCC	GDFM CHF-BEKK	GDFM CHF-DCC	DCC	PCA BEKK	PCA DCC	ABC BEKK	ABC DCC	GDFM CHF-BEKK	GDFM CHF-DCC
DGP1	D_1	0.9250	1.2738	0.8103	1.3410	0.8031	0.9181	0.6985	1.3651	1.6856	1.2731	1.7581	1.2713	1.4179	1.1645
	D_2	-0.7755	-0.3438	-0.7701	-0.2680	-0.7703	-0.5541	-0.7706	0.4683	0.7801	0.4188	0.8629	0.4228	0.5980	0.4185
	D_3	1.3246	1.6754	1.2118	1.7446	1.2061	1.3288	1.1061	1.7316	2.0607	1.6473	2.1306	1.6461	1.7947	1.5410
	D_4	1.3356	1.6822	1.2190	1.7471	1.2100	1.3171	1.1009	1.8016	2.1139	1.7025	2.1882	1.7001	1.8431	1.5902
DGP2	D_1	4.8267	7.0514	4.6772	3.2381	3.2227	3.2378	3.2209	5.0388	7.4417	4.8741	3.6970	3.5268	3.7246	3.5312
	D_2	2.5417	4.4712	2.0759	0.9182	0.8366	0.9067	0.8577	2.9169	5.0632	2.4392	1.7026	1.4488	1.7243	1.5048
	D_3	5.2040	7.4403	5.0644	3.6585	3.6411	3.6408	3.6227	5.4101	7.8220	5.2578	4.1155	3.9432	4.1215	3.9234
	D_4	5.2591	7.4725	5.1004	3.6274	3.6140	3.6450	3.6292	5.4765	7.8706	5.3002	4.0869	3.9190	4.1369	3.9478
DGP3	D_1	4.2942	4.1107	4.1011	3.2132	3.1201	2.8992	2.6704	4.4337	4.3294	4.1924	3.5553	3.3881	3.1390	3.0276
	D_2	2.7337	2.3416	2.3722	1.3922	1.2670	0.7874	0.5327	2.9380	2.6024	2.5153	1.8506	1.6731	1.3763	1.2453
	D_3	4.6245	4.4536	4.4425	3.5727	3.4790	3.2866	3.0589	4.7603	4.6764	4.5349	3.9178	3.7476	3.5198	3.4051
	D_4	4.7684	4.5739	4.5658	3.6606	3.5682	3.3217	3.0918	4.9106	4.7888	4.6560	3.9997	3.8356	3.5666	3.4585

Table 3: Average distances D_1, D_2, D_3, and D_4 (in logarithmic scale)for DGP1 (top panel), DGP2 (middle panel) and DGP3 (bottom panel) across 500 Monte Carlo replications using Gaussian innovations (left panel) and Student t_5 innovations (right panel). Shadowed cells stand for the MCS at 90%. $N = 600$, $T = 700$.

		Gaussian			Student t_5		
	Dist.	PCA-DCC	ABC-DCC	GDFM-CHF-DCC	PCA-DCC	ABC-DCC	GDFM-CHF-DCC
DGP1	D_1	2.9983	2.9948	2.3326	3.8527	3.8465	3.6217
	D_2	1.7752	1.7749	1.7642	3.4126	3.4079	3.4512
	D_3	3.3913	3.3902	2.7233	4.1567	4.1531	3.8953
	D_4	3.4159	3.4101	2.7519	4.3449	4.3364	4.1365
DGP2	D_1	9.3117	7.7120	7.7699	9.3860	8.1186	8.1406
	D_2	4.6340	3.1794	3.2568	5.0980	4.3034	4.3374
	D_3	9.7148	8.1176	8.1752	9.7856	8.5182	8.5390
	D_4	9.7195	8.1174	8.1755	9.7968	8.5292	8.5523
DGP3	D_1	8.5597	7.2742	7.1764	8.6601	7.6307	7.5554
	D_2	4.7223	3.0885	2.8640	5.0452	4.1039	4.0418
	D_3	8.9581	7.6772	7.5805	9.0551	8.0257	7.9504
	D_4	8.9722	7.6821	7.5831	9.0755	8.0456	7.9704

Considering the six DGPs (counting Gaussian and Student t as distinct models) and four measures of distance, we have a total of 24 comparisons among the models.

The DCC and PCA-DCC models are in the MCS in one case and in two cases, respectively, while the PCA-BEKK does not appear in the MCS. Comparing the estimation of common shocks by BEKK and DCC models, in only one case the BEKK has a slight better performance than DCC in terms of average distance (Gaussian, DGP3, PCA case). In fact, in the majority of cases, the performance is far better using DCC-based models than using the BEKK-based ones. Thus, in Figure 1, we only present the boxplots of the DCC-based models.[26] In general, DCC and PCA procedures achieve the worst performance and in the sequel we concentrate the comparison on the ABC and GDFM-CHF models.

For DGP1, although ABC-DCC and GDFM-CHF-DCC models show similar performance in Figure 1, the ABC-DCC model is included in the MCS only when considering the second distance for both Gaussian and Student t_5 innovations. For DGP2, where ABC models are adequate, the boxplots of ABC-DCC and GDFM-CHF-DCC are very similar. Considering Gaussian innovations, both models belong

[26]The boxplots of the BEKK-based models present much higher variability than those of the DCC-based ones, due, probably, to the numerical instability of BEKK QMLEs as commented in Section 3.2 (figures are available upon request).

to the MCS, as well as their BEKK-based counterparts. For the DGP2 with Student t_5 innovations, ABC-BEKK and GDFM-CHF-BEKK are included in the MSC only for the D3 measure, while GDFM-CHF-DCC is not in the MSC only for the D2 measure. For DGP3, in Figure 1, the GDFM-CHF-DCC model is always performing better than ABC-DCC, and it is the only procedure in the MCS. Finally, we can observe that the distances when the innovations are generated by the Student t_5 distribution are larger than those with Gaussian innovations. Nevertheless, the conclusions in the comparison among the estimated procedures are almost the same for both distributions.

Due to the high instability of BEKK-based procedures, for the high-dimensional data, we only report the results of the DCC-based procedures. We also do not report the results of the DCC model, since it yields the worst performance for n=60 and becomes computationally very expensive for n=600, even when using the composite likelihood method. The results are presented in Figure 2 and Table 3. For DGP1, GDFM-CHF-DCC is among the best procedures in all cases, and in most of them it performs significantly better than all other procedures according to the MCS test, while PCA-DCC has the worst performance. For DGP2, ABC-DCC and GDFM-CHF-DCC are selected as the best procedures when the innovations have Student t_5 distributions, while for Gaussian innovations ABC-DCC is the only procedure in the MCS. Finally, for DGP3, GDFM-CHF-DCC is selected as the only procedure in the MCS, regardless of the distribution of the innovations.

5 An application to dynamic portfolio optimization

In this section, we are applying our (GDFM-CHF-DCC) method in the problem of dynamic portfolio optimisation.

The dataset we are considering consists in returns X_{it} from n = 656 stocks entering the composition of the S&P 500 index, the National Association of Securities Dealers Automated Quotations (NASDAQ-100), and the NYSE Amex Composite Index (AMEX), on July 27, 2018 and traded from January 2, 2011 through June 29, 2018 (T=1884). This dataset was obtained from *Yahoo Finance* using the R package *quantmod* by Ryan and Ulrich (2017). Because we only considered stocks traded through the whole period, we ended up with $n = 656$ assets.

A window size of 750 days is used for estimation, which represents a concentration ratio of $656/750 = 0.875$; the out-of-sample period was set to 1134 days. An estimator $\widehat{\boldsymbol{V}}_{t+1|t}$ of $\mathrm{Var}(\boldsymbol{X}_{n,t+1}|\mathcal{F}_{n,t})$ is computed from the 656×750 subpanels $\{X_{is}|1 \leq i \leq 656,\ t-749 \leq s \leq t\}$ for $t = 750, \ldots, T-1 = 1883$. That estimator is used in the construction, at time $t = 750, \ldots, 1883$ (1134 time points), of a one-step ahead minimal variance portfolio (optimality at time $t+1$)—that is, a vector of weights

$$\widehat{\boldsymbol{\omega}}_{t+1|t} = (\widehat{\omega}_{1;t+1|t}, \ldots, \widehat{\omega}_{656;t+1|t})' := \operatorname*{argmin}_{\boldsymbol{\omega}} \boldsymbol{\omega}' \widehat{\boldsymbol{V}}_{t+1|t} \boldsymbol{\omega}$$

where minimisation is with respect to all $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{656})'$ such that $\omega_i \geq 0$ and $\sum_{i=1}^{656} \omega_i = 1$ and $\widehat{\boldsymbol{V}}_{t+1|t}$ is obtained as in Section 3.2 (with t instead of T_0). The resulting (out-of-sample) portfolio return

$$r_{p,t+1} := \sum_{i=1}^{656} \widehat{\omega}_{i;t+1|t} X_{i,t+1}$$

at time $t+1$ then is computed from the observation at time $t+1$.

For the sake of comparison, we also include the results for the GDFM-CHF-BEKK model and compare them with those of (a) the naive equal-weighted portfolio strategy, denoted here by 1/n, (b) the DCC model with composite likelihood of Pakel et al. (2020), (c) the RiskMetrics 2006 methodology of Zumbach (2007), (d) the OGARCH model of Alexander and Chibumba (1996), (e) the ABC-DCC model of Alessi et al. (2009), (f) the generalized principal volatility components (GPVC) of Li et al. (2016), and (g) the procedure called PCA4TS proposed by Chang et al. (2018), which extends the principal component analysis to second-order stationary vector time series. Those procedures were selected for their feasibility in high-dimensional data.

The GDFM-CHF method with DCC or BEKK models was implemented with 30 cross-sectional permutations; the order of the VAR block-diagonal models was set to $S = 1$. In practice (when one portfolio is to be estimated at a time), information criteria can be used to determine the order of those VARs. Likewise, following Alessi et al. (2009), the number of static factors, common shocks, volatility components (Li et al., 2016) and groups (Chang et al., 2018) were determined once for all.

The ABC-DCC model (Alessi et al., 2009) was implemented with eight static factors and three common shocks determined by the criteria of Bai and Ng (2002) and Hallin and Liška (2007), respectively. The same number of common shocks was used in the GDFM-CHF models. The GPVC procedure was applied with eight volatility components determined by the criterion of Bai and Ng (2002), and the PCA4TS with 654 groups (two of them with two assets and the remaining ones with only one asset; the groups were obtained following Chang et al. (2018)). The OGARCH model was applied as recommended in Becker et al. (2015), that is, with the number of components equal to the number of series.

Following Gambacciani and Paolella (2017), Engle et al. (2019), Trucíos et al. (2019b), among many others, we use various annualized measures to evaluate out-of-sample portfolio performance. These measures are defined as follows:

(i) annualized average portfolio (AV)

$$\mathrm{AV} := 252\bar{r}_p = 252\left[\frac{1}{1134}\sum_{t=750}^{1883} r_{p,t+1}\right]$$

(average of the out-of-sample portfolio returns multiplied by 252);

(ii) annualized standard deviation (SD)

$$\mathrm{SD} := \sqrt{252}\left[\frac{1}{1134}\sum_{t=750}^{1883}\left(r_{p,t+1}-\bar{r}_p\right)^2\right]^{1/2}$$

(standard deviation of the out-of-sample portfolio return multiplied by $\sqrt{252}$);

(iii) annualized information ratio (IR) IR := AV/SD;

(iv) annualized Sortino's ratio (SR) SR := AV/ $\left(\mathrm{S}\sqrt{252}\right)$, where

$$S = \left[\frac{1}{1134}\sum_{t=750}^{1883}\min\left(0, r_{p,t+1}-\mathrm{MAR}\right)^2\right]^{1/2},$$

and the minimal accepted return (MAR) is set to zero.

The results are reported in Table 4. They reveal that the best performance, for the SD, IR and SR criteria, is achieved by the GDFM-CHF-DCC model. The OGARCH model is second best, according to the SD criterion, followed by ABC-DCC. The GPVC and the OGARCH procedures exhibit the worst performance according to the AV criterion while DCC achieves the best one under the same criterion, followed by ABC-DCC. The worst out-of-sample performance is by the equal-weight portfolio strategy according to all criteria but the AV one. It is worth noting the relatively good performance of RM2006, which outperforms GPVC and PCA4TS according to all criteria and loses to DCC and OGARCH models only through the AV and SD criteria, respectively. Finally, note that the results of GDFM-CHF-BEKK are worse than those of GDFM-CHF-DCC, mainly in terms of the SD criterion. This is not surprising since, as mentioned previously, the estimation of the Full BEKK model is hard, unstable and strongly dependent on initial values, leading to a poor performance (Lien et al., 2002; Laurent et al., 2012; Manabu, 2015; Amendola and Candila, 2017; de Almeida et al., 2018). Taking into account all criteria, the GDFM-CHF-DCC proposed model exhibits the best performance, followed by the ABC-DCC model.

In view of our minimum variance objective, the most pertinent performance measure should be the SD criterion, as stressed also by Ledoit and Wolf (2017) and Engle et al. (2019). With that criterion, the GDFM-CHF-DCC methodl is achieving the best performance, followed by the ABC-DCC one.

Table 4: Annualized performance measures: AV, SD, IR, and SR stand for the annualized average, standard deviation, information ratio, and Sortino's ratio of the out-of-sample portfolio returns, respectively. The dataset is formed by 656 stocks used in the composition of the S&P500, NASDAQ-100 and AMEX indexes and the window size for estimation is equal to 750 days (concentration ratio n/T equal to 0.875). The out-of-sample period goes from January 2, 2014 to June 29, 2018. A ranking of the various methods is provided in parenthesis for each criterion.

	AV	SD	IR	SR
1/N	5.7708 (4)	11.5067 (9)	0.5015 (9)	0.6834 (9)
DCC	6.8899 (1)	5.9901 (8)	1.1502 (4)	1.6262 (5)
RM2006	5.6022 (5)	4.5446 (4)	1.2327 (3)	1.7241 (3)
OGARCH	4.9235 (8)	4.4551 (2)	1.1051 (7)	1.5616 (7)
ABC	6.5267 (2)	4.5313 (3)	1.4404 (2)	1.9677 (2)
GPVC	4.5991 (9)	4.5889 (5)	1.0022 (8)	1.4077 (8)
PCA4TS	5.3701 (7)	4.7256 (6)	1.1364 (6)	1.6032 (6)
GDFM-CHF	6.2369 (3)	4.0209 (1)	1.5511 (1)	2.2137 (1)
GDFM-CHF-BEKK	5.5819 (6)	4.8958 (7)	1.1401 (5)	1.6281 (4)

6 Conclusions

Based on the one-sided procedures of Forni et al. (2015, 2017) and Barigozzi and Hallin (2020), we propose a forecasting method for the conditional covariance matrix in high-dimensional time series, which we apply to dynamic portfolio optimization.

A Monte Carlo performance comparison with alternative methods is conducted over three different DGPs, using the distance measures proposed in Amendola and Candila (2017). Overall, our method has an excellent performance, and outperforms all its competitors irrespective of the criterion considered—except, under static factor model DGPs, for the distance D2 which disregards the covariances.

The superiority of our estimator is also empirically established in the context of dynamic portfolio optimisation based on a dataset of 656 assets. Our model, GDFM-CHF-DCC, achieves the best out-of-sample performance according to the (annualized) standard deviation SD (arguably, the most relevant criterion here), information ratio (IR) and Sortino's ratio (SR) criteria, and is third best (after DCC and ABC-DCC models) with respect to the (annualized) average criterion.

References

Aguilar, M. (2009). A latent factor model of multivariate conditional heteroscedasticity. *Journal of Financial Econometrics*, 7(4):481–503.

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

Alessi, L., Barigozzi, M., and Capasso, M. (2009). Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors. Working paper series 1115, European Central Bank, Frankfurt am Main, Germany.

Alessi, L., Barigozzi, M., and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters*, 80(23-24):1806–1813.

Alexander, C. O. and Chibumba, A. (1996). Multivariate orthogonal factor GARCH. *University of Sussex Discussion Papers in Mathematics.*

Amendola, A. and Candila, V. (2017). Comparing multivariate volatility forecasts by direct and indirect approaches. *Journal of Risk*, 19(6):33–57.

Anderson, B. D. and Deistler, M. (2008). Generalized linear dynamic factor models. A structure theory. In *47th IEEE Conference on Decision and Control.*

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.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Wang, P. (2016). Econometric analysis of large factor models. *Annual Review of Economics*, 8:53–80.

Barhoumi, K., Darné, O., and Ferrara, L. (2014). Dynamic factor models: A review of the literature. *Journal of Business Cycle Research*, 2013(2):73.

Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal*, 19(1):C33–C60.

Barigozzi, M. and Hallin, M. (2017). Generalized dynamic factor models and volatilities: estimation and forecasting. *Journal of Econometrics*, 201(2):307–321.

Barigozzi, M. and Hallin, M. (2020). Generalized dynamic factor models and volatilities: Consistency, rates, and prediction intervals. *Journal of Econometrics*, 216:4–34.

Bauwens, L., Laurent, S., and Rombouts, J. V. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21(1):79–109.

Becker, R., Clements, A., Doolan, M., and Hurn, A. (2015). Selecting volatility forecasting models for portfolio allocation purposes. *International Journal of Forecasting*, 31(3):849–861.

Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1):116–131.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, volume 36 of *Classics in Applied Mathematics*. SIAM.

Chang, C.-L., McAleer, M., and Tansuchat, R. (2011). Crude oil hedging strategies using dynamic multivariate GARCH. *Energy Economics*, 33(5):912–923.

Chang, J., Guo, B., and Yao, Q. (2018). Principal component analysis for second-order stationary vector time series. *The Annals of Statistics*, 46(5):2094–2124.

Chevallier, J. (2012). Time-varying correlations in oil, gas and CO^2 prices: an application using BEKK, CCC, and DCC-MGARCH models. *Applied Economics*, 44(32):4257–4274.

Comte, F. and Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, 84(1):61–84.

de Almeida, D., Hotta, L. K., and Ruiz, E. (2018). MGARCH models: Trade-off between feasibility and flexibility. *International Journal of Forecasting*, 34(1):45–63.

Della Marra, F. (2017). A forecasting performance comparison of dynamic factor models based on static and dynamic methods. *Communications in Applied and Industrial Mathematics*, 8(1):43–66.

Diebold, F. X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics*, 4(1):1–21.

Dovonon, P. (2013). Conditionally heteroskedastic factor models with skewness and leverage effects. *Journal of Applied Econometrics*, 28(7):1110–1137.

Engle, R. (2009). *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton University Press, Princeton, New Jersey.

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

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Jounal of Business & Economic Statistics*, 37(2):363–375.

Fan, J., Wang, M., and Yao, Q. (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society Series, B*, 70(4):679–702.

Forni, M., Giovannelli, A., Lippi, M., and Soccorsi, S. (2018). Dynamic factor model with infinite-dimensional factor space: forecasting. *Journal of Applied Econometrics*, 33(5):625–642.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2015). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, 185(2):359–371.

Forni, M., Hallin, M., Lippi, M., and Zaffaroni, P. (2017). Dynamic factor models with infinite-dimensional factor space: asymptotic analysis. *Journal of Econometrics*, 199(1):74–92.

Forni, M. and Lippi, M. (2011). The general dynamic factor model: One-sided representation results. *Journal of Econometrics*, 163(1):23–28.

Francq, C. and Zakoian, J. (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, West Sussex, 2nd edition.

Gambacciani, M. and Paolella, M. S. (2017). Robust normal mixtures for financial portfolio allocation. *Econometrics and Statistics*, 3:91–111.

García-Ferrer, A., González-Prieto, E., and Peña, D. (2012). A conditionally heteroskedastic independent factor model with an application to financial stock returns. *International Journal of Forecasting*, 28(1):70–93.

Giovannelli, A., Massacci, D., and Soccorsi, S. (2018). Forecasting stock returns with large dimensional factor models. available at SSRN 2958491.

Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis*, 100(9):2044–2054.

Hallin, M., Hörmann, S., and Lippi, M. (2018). Optimal dimension reduction for high-dimensional and functional time series. *Statistical Inference for Stochastic Processes*, 21(2):385–398.

Hallin, M. and Lippi, M. (2013). Factor models in high-dimensional time series—a time-domain approach. *Stochastic Processes and Their Applications*, 123(7):2678–2695.

Hallin, M., Lippi, M., Barigozzi, M., Forni, M., and Zaffaroni, P. (2020). *Time Series in High Dimension: the General Dynamic Factor Model*. World Scientific.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Han, Y. (2005). Asset allocation with a high dimensional latent factor stochastic volatility model. *The Review of Financial Studies*, 19(1):237–271.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Harvey, A., Ruiz, E., and Sentana, E. (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics*, 52(1-2):129–157.

Hlouskova, J., Schmidheiny, K., and Wagner, M. (2009). Multistage predictions for multivariate GARCH models: Closed form solution and the value for portfolio management. *Journal of Empirical Finance*, 16(2):330–336.

Hu, Y.-P. and Tsay, R. S. (2014). Principal volatility component analysis. *Journal of Business & Economic Statistics*, 32(2):153–164.

Laurent, S., Rombouts, J. V., and Violante, F. (2012). On the forecasting accuracy of multivariate GARCH models. *Journal of Applied Econometrics*, 27(6):934–955.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.

Li, W., Gao, J., Li, K., and Yao, Q. (2016). Modeling multivariate volatilities via latent common factors. *Journal of Business & Economic Statistics*, 34(4):564–573.

Lien, D., Tse, Y. K., and Tsui, A. K. (2002). Evaluating the hedging performance of the constant-correlation GARCH model. *Applied Financial Economics*, 12(11):791–798.

Manabu, A. (2015). Initial values on quasi-maximum likelihood estimation for BEKK multivariate GARCH models. *Soka Economic Studies Quarterly*, 44(1):45–52.

Matteson, D. S. and Tsay, R. S. (2011). Dynamic orthogonal components for multivariate time series. *Journal of the American Statistical Association*, 106(496):1450–1463.

Pakel, C., Shephard, N., Sheppard, K., and Engle, R. F. (2020). Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics*, (Forthcoming):1–17.

Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379.

Peña, D. and Box, G. E. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82(399):836–843.

Ryan, J. A. and Ulrich, J. M. (2017). *quantmod: Quantitative Financial Modelling Framework*. R package version 0.4-12.

Santos, A. A. and Moura, G. V. (2014). Dynamic factor multivariate GARCH model. *Computational Statistics & Data Analysis*, 76:606–617.

Sentana, E., Calzolari, G., and Fiorentini, G. (2008). Indirect estimation of large conditionally heteroskedastic factor models, with an application to the Dow 30 stocks. *Journal of Econometrics*, 146(1):10–25.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stoffer, D. S. (1999). Detecting common signals in multiple time series using the spectral envelope. *Journal of the American Statistical Association*, 94(448):1341–1356.

Trucíos, C., Hotta, L. K., and Pereira, P. L. V. (2019a). On the robustness of the principal volatility components. *Journal of Empirical Finance*, 52:201–219.

Trucíos, C., Hotta, L. K., and Ruiz, E. (2018). Robust bootstrap densities for dynamic conditional correlations: implications for portfolio selection and value-at-risk. *Journal of Statistical Computation and Simulation*, 88(10):1976–2000.

Trucíos, C., Zevallos, M., Hotta, L. K., and Santos, A. A. (2019b). Covariance prediction in large portfolio allocation. *Econometrics*, 7(2):19.

Van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, 17(5):549–564.

Vrontos, I. D., Dellaportas, P., and Politis, D. N. (2003). A full-factor multivariate garch model. *The Econometrics Journal*, 6(2):312–334.

Zumbach, G. O. (2007). A gentle introduction to the RM2006 methodology. *Working Paper available at SSRN: https://ssrn.com/abstract=1420183.*