Generalized Dynamic Factor Models and Volatilities: Consistency, Rates, and Prediction Intervals

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Generalized Dynamic Factor Models and Volatilities: Consistency, Rates, and Prediction Intervals / Barigozzi M; Hallin M. - In: JOURNAL OF ECONOMETRICS. - ISSN 0304-4076. - STAMPA. - 216:1(2020), pp. 4-34. [10.1016/j.jeconom.2020.01.003]

Availability:
This version is available at: https://hdl.handle.net/11585/722367 since: 2020-06-25

Published:
DOI: http://doi.org/10.1016/j.jeconom.2020.01.003

Generalized Dynamic Factor Models and Volatilities: Consistency, Rates, and Prediction Intervals

MATTEO BARIGOZZI† MARC HALLIN‡

Abstract

Volatilities, in high-dimensional panels of economic time series with a dynamic factor structure on the levels or returns, typically also admit a dynamic factor decomposition. We consider a two-stage dynamic factor model method recovering the common and idiosyncratic components of both levels and log-volatilities. Specifically, in a first estimation step, we extract the common and idiosyncratic shocks for the levels, from which a log-volatility proxy is computed. In a second step, we estimate a dynamic factor model, which is equivalent to a multiplicative factor structure for volatilities, for the log-volatility panel. By exploiting this two-stage factor approach, we build one-step-ahead conditional prediction intervals for large $n \times T$ panels of returns. Those intervals are based on empirical quantiles, not on conditional variances; they can be either equal- or unequal-tailed. We provide uniform consistency and consistency rates results for the proposed estimators as both n and T tend to infinity. We study the finite-sample properties of our estimators by means of Monte Carlo simulations. Finally, we apply our methodology to a panel of asset returns belonging to the S&P100 index in order to compute one-step-ahead conditional prediction intervals for the period 2006-2013. A comparison with the componentwise GARCH benchmark (which does not take advantage of cross-sectional information) demonstrates the superiority of our approach, which is genuinely multivariate (and high-dimensional), nonparametric, and model-free.

JEL Classification: C32, C38, C58.
Keywords: Volatility, Dynamic Factor Models, Prediction intervals, GARCH.

1 Introduction

Data in high dimension unquestionably constitute one of the main challenges of contemporary statistics/econometrics, and have become pervasive in most domains related with data sciences. Time series have not escaped that evolution, and the analysis of high-dimensional time series—equivalently, large cross-sections of univariate time series or panels—today ranks among the most active topics in theoretical and applied econometrics.

The most successful methods so far in the analysis and prediction of high-dimensional time series are based on the so-called factor model approach. That approach, under its various forms, is based on a (non-observed) decomposition of the observation (a large cross-section of time series with complex interrelations) into the sum of two mutually orthogonal (all leads, all lags) components: the *common component*, driven by a small number of *factors* or *common shocks*, and an *idiosyncratic component*, with some variations in the definitions of "common"

† Dipartimento di Scienze Economiche, Università di Bologna, Piazza Scaravilli 2, 40126, Bologna, Italy.
E-Mail: matteo.barigozzi@unibo.it
‡ ECARES, Université libre de Bruxelles CP114/4 B-1050 Bruxelles, Belgium.
E-Mail: mhallin@ulb.ac.be

We thank Christian Brownlees, Christian Francq, and Haeran Cho for helpful comments. This paper was also presented at: "Panel Data Forecasting Conference", University of Southern California, Dornsife, Los Angeles, April, 2019; the "6th Rimini Centre for Economic Analysis (RCEA) Time Series Econometrics Workshop", University of Cyprus, Larnaca, June 2019, the "International Association for Applied Econometrics (IAAE) 2019 Annual Conference", University of Cyprus, Nicosia, June 2019, and the "Workshop on High-Dimensional Data Analysis", Durham University, June 2019.

and "idiosyncratic," and the assumptions made. Regardless of the definition adopted, the common and idiosyncratic components typically are disentangled by means of adequate cross-sectional and/or temporal aggregation of the observed time series.

Those aggregation and factor model approaches are strongly rooted in the multivariate time-series methods developed in the eighties and nineties, of which George Tiao and his collaborators have been most influential and unremittable pioneers: see, for instance, Tiao (1972), Tiao and Hillmer (1978), Tiao and Guttman (1980), Tiao and Box (1981), Tsay and Tiao (1985), Peña and Box (1987), and Tiao and Tsay (1989).

The type of factor model we are considering here is the General or Generalized Dynamic Factor Model (GDFM) introduced by Forni et al. (2000), which, by taking into account all leading and lagging linear dependencies among the data, encompasses most other models, as e.g. the static factor approaches by Bai and Ng (2002), Stock and Watson (2002), and Fan et al. (2013). Moreover, as emphasised in Forni and Lippi (2001) and Hallin and Lippi (2013), beyond the usual assumptions of second-order stationarity and existence of spectral densities, the GDFM decomposition into a common and an idiosyncratic component basically does not place any structural constraints on the data-generating process. In this sense, contrary to static factor approaches, it is canonical, nonparametric and model-free. In this paper, we consider the one-sided GDFM estimation method recently described in Forni et al. (2015, 2017).

Prediction, in classical univariate and moderately multivariate time series analysis, is an obvious and natural objective; it is certainly no less crucial in high dimension. Efficient prediction, however, should exploit the amount of information available, due to the complex cross-dependencies among the many cross-sectional components, in the present and lagged values of the whole cross-section; the larger the cross-section (i.e., the higher the dimension), the more crucial the role of that information, and the more delicate its recovering. Factor models naturally have been used in the construction of *point-predictors*, and quite successfully so: see, e.g., Stock and Watson (2002), Bai and Ng (2008), Forni et al. (2018), to quote only a very few. Those authors, however, are dealing, mostly, with macroeconomic data, while less attention has been given to factor model methods in the analysis and prediction of financial returns: see, e.g. Chamberlain and Rothschild (1983), Connor and Korajczyk (1993), or Aït-Sahalia and Xiu (2017). In particular, when dealing with returns, due to the presence of conditional distribution heterogeneity (of which conditional heteroskedasticity is only a very particular case), conditional volatility phenomenons are essential, and definitely should be taken into account when building conditional prediction limits or conditional prediction intervals.

Most multivariate methods available in the literature for the analysis of conditional heterogeneity are restricted to the study of conditional heteroskedasticity, and rely on parametrisations of the ARCH-GARCH or Stochastic Volatility type: see, for instance, the reviews by Bauwens et al. (2006) and Asai et al. (2006). Because of the curse of dimensionality, however, only the very simplest models can be considered in high-dimensional panels, possibly inducing a nonnegligible loss of efficiency. Among those, the factor GARCH approach is the most popular, see e.g. Diebold and Nerlove (1989), Ng et al. (1992), Harvey et al. (1992), and Sentana et al. (2008). Static factor models directly based on volatilities have also been considered, but these fail to exploit the information contained in the idiosyncratic components of returns, see e.g. Connor et al. (2006) and Fan et al. (2015). For these reasons, Barigozzi and Hallin (2016) introduce a two-step GDFM approach by which the nonparametric and model-free virtues of factor models are used in a joint analysis of returns and volatilities. In Barigozzi and Hallin (2017a), that two-step GDFM is combined with a GARCH strategy in order to produce point-forecasts for volatilities (see also Trucíos et al., 2019 for a recent example), while Barigozzi and Hallin (2017b) and Barigozzi et al. (2019) apply the same methodology in a study of the dynamic interdependencies of US and international financial markets. A two-stage factor approach similar to ours but in a static factor model setting is proposed in Chicheportiche and Bouchaud (2015).

The objective of this paper is to combine the same two-step GDFM approach with a quantile-based construction of conditional confidence limits producing conditional interval predictions rather than point-forecasts for returns. That objective requires nontrivial consistency results on the two-step GDFM estimation method, which are not provided in Barigozzi and Hallin (2016, 2017a,b). The first part of this paper, therefore, is devoted to a careful asymptotic analysis of the two-step GDFM. We then describe the quantile-based construction of conditional confidence limits, which we apply to a dataset of S&P100 daily returns.

The paper is organised as follows. In Section 2, we present the GDFM model for the stochastic processes of returns (levels) and log-volatilities, and give sufficient conditions for its existence and identification. Section 3.1 describes the estimation of the model, and Section 3.2 establishes the consistency properties (with rates) of the proposed estimators. In Section 4, we define the one-step-ahead conditional prediction confidence limits and intervals. In Section 5, we study the finite-sample properties of our estimators via simulations. Section 6 applies our methodology to a panel of daily returns of stocks listed in the S&P100 index and investigates the resulting coverage performance. In Section 7, we conclude. Proofs are postponed to an online Appendix.

Notation

The sub-exponential norm of a scalar random variable X is defined as $\|X\|_{\psi_1} := \sup_{p\geq 1} p^{-1}\mathrm{E}[|X|^p]^{1/p}$ (see e.g. Definition 5.13 in Vershynin, 2012). The transposed complex conjugate of a complex vector $\mathbf{p}$ is denoted as $\mathbf{p}^\dagger$ and $\|\mathbf{p}\| = \mathbf{p}^\dagger\mathbf{p}$. For an hermitian complex $n\times n$ matrix $\mathbf{A}$ with generic (i,j) entry a_{ij} and largest (in modulus) eigenvalue $\mu_1^{\mathbf{A}}$, let $\|\mathbf{A}\|_1 := \max_{j=1,\ldots,n}\sum_{i=1}^n |a_{ij}|$ and $\|\mathbf{A}\| := \mu_1^{\mathbf{A}}$. As usual, L stands for the lag operator, such that, given a stochastic vector process $\{\mathbf{Y}_t|t\in\mathbb{Z}\}$, $L^k\mathbf{Y}_t := \mathbf{Y}_{t-k}$ for any integer k and any $t\in\mathbb{Z}$. Last, we denote by $\mathbb{I}(\mathcal{A})$ the indicator function of an event $\mathcal{A}$.

2 A General Dynamic Factor Model for levels and volatilities

We throughout assume that all stochastic variables in this paper belong to the Hilbert space $L_2(\Omega,\mathcal{F},\mathrm{P})$, where $(\Omega,\mathcal{F},\mathrm{P})$ is some common probability space. We study double-indexed stochastic processes of the form $\mathbf{Y} := \{Y_{it}|i\in\mathbb{N},\ t\in\mathbb{Z}\}$, with n-dimensional sub-processes $\mathbf{Y}_n := \{Y_{it}|i=1,\ldots,n,\ t\in\mathbb{Z}\}$, $n\in\mathbb{N}$. In practice, we deal with the finite observed $n\times T$ realisation

$$\mathbf{Y}_{n,T} := \begin{pmatrix} Y_{11}, & Y_{12}, & \ldots, & Y_{1T} \\ \vdots & \vdots & & \vdots \\ Y_{n1}, & Y_{n2}, & \ldots, & Y_{nT} \end{pmatrix}$$

of $\mathbf{Y}$. In the empirical application of Section 6, the Y_{it}'s are observed values of daily stock returns, and we therefore call $\mathbf{Y}$ the "levels" process. The assumptions in Section 2.1 are mainly taken from Forni et al. (2017), with some modifications, mostly concerning the idiosyncratic components. On the other hand, the assumptions in Section 2.2 are new and are related to the log-volatility proxies originally introduced in Barigozzi and Hallin (2016, 2017a).

2.1 Model and assumptions for levels

The Generalized Dynamic Factor Model (GDFM) for the levels process $\mathbf{Y}$ is a decomposition of Y_{it} into

$$Y_{it} - \mathrm{E}[Y_{it}] = X_{it} + Z_{it}, \quad i\in\mathbb{N},\ t\in\mathbb{Z} \tag{2.1}$$

with

$$X_{it} = \sum_{j=1}^{q}\sum_{k=0}^{\infty} b_{ijk}u_{jt-k} = \mathbf{b}_i'(L)\mathbf{u}_t \quad \text{and} \quad Z_{it} = \sum_{k=0}^{\infty} d_{ik}v_{it-k} = d_i(L)v_{it}, \tag{2.2}$$

where $\mathrm{E}[Y_{it}]$ stands for the expected value of Y_{it} and the processes $\mathbf{u} := \{u_{jt}|j=1,\ldots,q,\ t\in\mathbb{Z}\}$ and $\mathbf{v}_n := \{v_{it}|i=1,\ldots,n,\ t\in\mathbb{Z}\}$ are mutually orthogonal (at all leads and lags) q- and n-dimensional white noises, respectively. Call $\mathbf{u}$ the process of *common factors* or *common shocks* and $\mathbf{v}_n$ the process of *idiosyncratic shocks*; X_{it} and Z_{it} are Y_{it}'s *common* and *idiosyncratic components*, respectively.

Letting $\mathbf{X}_n := \{X_{it}|i=1,\ldots,n,\ t\in\mathbb{Z}\}$ and $\mathbf{Z}_n := \{Z_{it}|i=1,\ldots,n,\ t\in\mathbb{Z}\}$, equations (2.2) in vector notation takes the form

$$\mathbf{X}_{nt} = \mathbf{B}_n(L)\mathbf{u}_t, \qquad \mathbf{Z}_{nt} = \mathbf{D}_n(L)\mathbf{v}_{nt}, \quad n\in\mathbb{N},\ t\in\mathbb{Z}. \tag{2.3}$$

with $\mathbf{B}_n(L) := (\mathbf{b}_1(L)\ldots\mathbf{b}_n(L))'$, and $\mathbf{D}_n(L) := \mathrm{diag}(d_1(L)\ldots d_n(L))$.

More precisely, we assume that (2.1)-(2.2) hold and satisfy the following assumptions:

ASSUMPTION (L1).

(i) *the dimension q of $\mathbf{u}_t$ does not depend on n; the process $\mathbf{u} := \{\mathbf{u}_t | t \in \mathbb{Z}\}$ is second-order white noise, with mean $\mathbf{0}_q$ and diagonal positive definite covariance $\mathbf{\Gamma}^{\mathrm{u}}$;*

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

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

(iv) *there exists a constant $C_{\mathrm{v}} > 0$ such that $\|\mathbf{\Gamma}_n^{\mathrm{v}}\|_1 \leq C_{\mathrm{v}}$ for all $n \in \mathbb{N}$;*

(v) *there exists a constant $M_2 > 0$ such that $\sum_{k=0}^{\infty} |d_{ik}| \, |k| \leq M_2$ for all $i \in \mathbb{N}$;*

(vi) *$\mathrm{Cov}(u_{jt}, v_{is}) = 0$ for all $i \in \mathbb{N}$, $j = 1, \dots, q$, and $t, s \in \mathbb{Z}$;*

(vii) *there exists a constant $M_3 > 0$ such that $\sum_{k_1,k_2,k_3 \in \mathbb{Z}} \left| \mathrm{E}[u_{j_1 t} u_{j_2, t-k_1} u_{j_3, t-k_2} u_{j_4, t-k_3}] \right| \leq M_3$ for all $j_1, j_2, j_3, j_4 = 1, \dots, q$;*

(viii) *there exists a constant $M_4 > 0$ such that $\sum_{k_1,k_2,k_3 \in \mathbb{Z}} \left| \mathrm{E}[v_{i_1 t} v_{i_2, t-k_1} v_{i_3, t-k_2} v_{i_4, t-k_3}] \right| \leq M_4$ for all $i_1, i_2, i_3, i_4 \in \mathbb{N}$.*

These assumptions are standard in the literature with the exception of part *(iv)* which imposes a mild form of sparsity on the covariance matrix of the idiosyncratic innovations. A similar condition can be found in Fan et al. (2013) and is empirically verified by Boivin and Ng (2006) and Bai and Ng (2008) for US macroeconomic data, and by Barigozzi and Hallin (2017b) for stock returns. As a consequence of parts *(iv)* and *(v)*, the idiosyncratic components are allowed to be serially autocorrelated and mildly cross-correlated (see also Lemma 1 below). Moreover, it is easy to check that such assumption is nesting other typical conditions on the cross-sectional dependence of idiosyncratic components (see e.g. Bai and Ng, 2002, and Stock and Watson, 2002, in the static factor model case). Parts *(ii)* and *(v)* imply absolute summability of the autocovariances and therefore the existence of a purely continuous spectral density. Moreover these assumptions and existence of fourth-order moments in parts (vii) and (viii) are classical requirements for consistent estimation of the autocovariances and the spectral density (see e.g. Chapter IV, Theorem 6, in Hannan, 1970, for the autocovariances, and the results in Section 6.2 in Priestley, 2001, and Theorem 5A in Parzen, 1957, for the spectral density). Last, in part *(iii)* we also make the typical assumption of martingale difference innovations used in the GARCH literature (see e.g. Definition 2.1 in Francq and Zakoian, 2011).

It should be insisted, however, that the GDFM is not a *statistical model* in the usual sense, inasmuch as, beyond the requirement of second-order stationarity, the existence of a finite (but unspecified) q, and the existence of a spectrum, it does not really impose any restrictions on the data-generating process: as argued by Forni and Lippi (2001) and Hallin and Lippi (2013), (2.1)-(2.2) indeed constitute a representation result rather than a model equation.

On the filters $\mathbf{b}_i(L)$ and $d_i(L)$ we furthermore impose the following assumptions:

ASSUMPTION (L2).

(i) *$\mathbf{b}_i(L)$ has rational entries, i.e. $b_{ij}(L) = \theta_{ij}(L)\phi_{ij}^{-1}(L)$, where $\phi_{ij}(z)$ and $\theta_{ij}(z)$, for all $i \in \mathbb{N}$ and $j = 1, \dots, q$, are finite-order polynomials;*

(ii) *there exists a constant $\bar{\phi} > 1$ such that $\phi_{ij}(z) \neq 0$ for all $i \in \mathbb{N}$, all $j = 1, \dots, q$, and all $z \in \mathbb{C}$ such that $|z| \leq \bar{\phi}$;*

(iii) *the coefficients θ_{ijk} of $\theta_{ij}(L)$ are such that $|\theta_{ijk}| \leq B^X$ for some positive constant B^X, all $k \in \mathbb{N} \cup \{0\}$, all $i \in \mathbb{N}$, and $j = 1, \dots, q$;*

(iv) *$d_i(L)$ is of the form $c_i^{-1}(L)$ where $c_i(z)$, for all $i \in \mathbb{N}$, is a finite-order polynomial, $c_i(0) = 1$ and $c_i(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.*

This latter assumption is not strictly needed and could be easily relaxed to allow for infinite order autoregressive dynamics—at the expense, however, of heavier notation and longer proofs; see also Section 3.2 for a short discussion. This assumption implies that both the common and idiosyncratic components have a rational spectral density. Rational filters for the common component are also assumed in Forni et al. (2017), while here we also assume that the idiosyncratic component admits a finite autoregressive representation. In particular, using part *(iv)*, we can rewrite the second equation in (2.2) as

$$c_i(L) Z_{it} = v_{it}. \tag{2.4}$$

Let $\mathbf{\Sigma}_n^Y(\theta)$, $\mathbf{\Sigma}_n^X(\theta)$ and $\mathbf{\Sigma}_n^Z(\theta)$, $\theta \in [-\pi, \pi]$, be the $n \times n$ spectral density matrices of the observed panel, the common, and the idiosyncratic components, respectively; the existence of those spectral densities is guaranteed by Assumption (L1). Denote by $\lambda_{nj}^Y(\theta)$, $\lambda_{nj}^X(\theta)$, and $\lambda_{nj}^Z(\theta)$ their respective j-th largest eigenvalues—the panel, common, and idiosyncratic *dynamic eigenvalues*, on which we assume the following. Hereafter, "for all $\theta \in [-\pi, \pi]$" or "$\theta - a.e.$" is to be understood as "for all θ but over a subset of values included in a set with Lebesgue measure zero." Similarly, $\sup_{\theta\in[-\pi,\pi]}$ in the sequel is an *essential* sup, etc.

ASSUMPTION (L3). *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*

$$0 < \beta_{j-1}(\theta) < \alpha_j(\theta) \leq \lambda_{nj}^X(\theta)/n \leq \beta_j(\theta) < \infty \quad \theta\text{-a.e. in } [-\pi, \pi], \text{ all } j = 1, \ldots, q, \text{ and all } n > \bar{n}.$$

Under this assumption, the first q common dynamic eigenvalues, irrespective of the frequency θ (except possibly over a set of measure zero), are diverging linearly as $n \to \infty$. The following results then hold for the idiosyncratic dynamic eigenvalues and those of the panel.

LEMMA 1. *Under Assumptions (L1) and (L3),*

(i) there exists a constant $C^Z > 0$ such that $\sup_{\theta\in[-\pi,\pi]} \lambda_{n1}^Z(\theta) \leq C^Z$ for all $n \in \mathbb{N}$;

(ii) there exist a positive integer $\bar{n}$ and continuous functions α_j^Y and β_{j-1}^Y from $[-\pi, \pi]$ to $\mathbb{R}$, $j = 1, \ldots, q$, independent of n and such that $0 < \beta_{j-1}^Y(\theta) < \alpha_j^Y(\theta) \leq \lambda_{nj}^Y(\theta)/n \leq \beta_j^Y(\theta) < \infty$, θ-a.e. in $[-\pi, \pi]$, all $j = 1, \ldots, q$, and all $n > \bar{n}$;

(iii) there exists a constant $C^Y > 0$ such that $\sup_{\theta\in[-\pi,\pi]} \lambda_{n,q+1}^Y(\theta) \leq C^Y$ for all $n \in \mathbb{N}$.

As a consequence of Lemma 1, identification of the model, i.e., consistently disentangling the unobserved common and idiosyncratic components, is possible, under the assumptions made in the limit, as $n \to \infty$, thanks to the behaviour of the dynamic eigenvalues.

Based on results by Anderson and Deistler (2008) for singular vector processes with a rational spectrum, Forni and Lippi (2011) and Forni et al. (2015) prove that, for generic values of the coefficients of the filters $\mathbf{b}_i(L)$ as defined in Assumption (L2), the space spanned by $u_{j,t-k}$ for $j = 1, \ldots, q$ and $k \geq 0$ is the same as the space spanned by any $(q + 1)$-dimensional subvector of $\mathbf{X}_t$ and its lags; moreover, those subvectors admit an autoregressive representation driven by the common shocks $\mathbf{u}_t$.

More precisely, any $(q + 1)$-dimensional subvector $\mathbf{X}_t^\ddagger$ of $\mathbf{X}_{nt}$ admits an autoregressive representation of the form

$$\mathbf{A}^\ddagger(L)\mathbf{X}_t^\ddagger = \mathbf{H}^\ddagger\mathbf{u}_t, \tag{2.5}$$

where $\mathbf{A}^\ddagger(L)$ is a finite-order VAR operator such that $\mathbf{A}^\ddagger(0) = \mathbf{I}_{q+1}$, $\mathbf{u}_t$ is the vector of common shocks in (2.3), and $\mathbf{H}^\ddagger$ an appropriate $(q + 1) \times q$ matrix. On that representation, we make the following assumptions.

ASSUMPTION (L4). *Let $\mathbf{X}_t^\ddagger$ be an arbitrary $(q + 1)$-dimensional subvector of $\mathbf{X}_{nt}$: the autoregressive representation* (2.5) *is such that*

(i) $\mathbf{A}^\ddagger(L)$ is uniquely defined;

(ii) the degree $S^\ddagger$ of $\mathbf{A}^\ddagger(z)$ is uniformly bounded, that is, $S^\ddagger \leq S$ for some integer $S > 0$ independent of n and the choice of the subvector $\mathbf{X}_t^\ddagger$;

(iii) $\det[\mathbf{A}^\ddagger(z)] \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$;

(iv) $\mathbf{H}^\ddagger$ is $(q + 1) \times q$, with full rank q;

(v) denoting by $\mathbf{\Gamma}_h^{X^\ddagger}$ the lag-h autocovariances of $\mathbf{X}^\ddagger := \{\mathbf{X}_t^\ddagger | t \in \mathbb{Z}\}$ and defining

$$\boldsymbol{\mathcal{C}}^\ddagger := \begin{bmatrix} \mathbf{\Gamma}_0^{X^\ddagger} & \mathbf{\Gamma}_1^{X^\ddagger} & \cdots & \mathbf{\Gamma}_{S-1}^{X^\ddagger} \\ \mathbf{\Gamma}_{-1}^{X^\ddagger} & \mathbf{\Gamma}_0^{X^\ddagger} & \cdots & \mathbf{\Gamma}_{S-2}^{X^\ddagger} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Gamma}_{-S+1}^{X^\ddagger} & \mathbf{\Gamma}_{-S+2}^{X^\ddagger} & \cdots & \mathbf{\Gamma}_0^{X^\ddagger} \end{bmatrix},$$

$\det(\boldsymbol{\mathcal{C}}^\ddagger) > d > 0$, where d is independent of the choice of the subvector $\mathbf{X}_t^\ddagger$.

This assumption allows us to derive an alternative representation of the GDFM (2.2) which is particularly useful for estimation and for the construction, in Section 2.2 below, of a further GDFM for log-volatilities. Without loss of generality, let n factorise into $n = m(q+1)$ for some positive integer m, so that we can partition $\mathbf{X}_n$ into m subprocesses, each of dimension $(q+1)$, of the form $\mathbf{X}_t^{(k)} := (X_{(k-1)(q+1),t} \cdots X_{k(q+1)-1,t})'$, $k = 1, \ldots, m$, with superscript ${}^{(k)}$ substituted for ‡. Each $\mathbf{X}^{(k)}$ satisfies (2.5) and Assumption (L4). Defining the $n \times q$ matrix $\mathbf{H}_n := (\mathbf{H}^{(k)\prime} \cdots \mathbf{H}^{(m)\prime})'$, we thus have the VAR representation

$$\mathbf{A}_n(L)\mathbf{X}_{nt} = \mathbf{H}_n\mathbf{u}_t, \tag{2.6}$$

where $\mathbf{A}_n(L)$ is $n \times n$ block-diagonal with diagonal blocks $\mathbf{A}^{(1)}(L), \ldots, \mathbf{A}^{(m)}(L)$. Moreover, in view of (2.3), we have $[\mathbf{A}_n(L)]^{-1}\mathbf{H}_n = \mathbf{B}_n(L)$ (see Proposition 3 in Forni et al., 2017). Then, the following alternative and equivalent representation of the GDFM holds:

$$\mathbf{A}_n(L)\{\mathbf{Y}_{nt} - \mathrm{E}[\mathbf{Y}_{nt}]\} = \mathbf{H}_n\mathbf{u}_t + \mathbf{A}_n(L)\mathbf{Z}_n. \tag{2.7}$$

The advantage of this representation is that it is "static" in the sense that the common shocks $\mathbf{u}$ now are loaded only contemporaneously and not via filters as in (2.3).

To conclude with, note that the Yule-Walker equations

$$\left(\mathbf{A}_1^{\ddagger} \cdots \mathbf{A}_S^{\ddagger}\right) = \left(\mathbf{\Gamma}_1^{X^{\ddagger}} \cdots \mathbf{\Gamma}_S^{X^{\ddagger}}\right)\left[\boldsymbol{\mathcal{C}}^{\ddagger}\right]^{-1}, \tag{2.8}$$

characterising the S matrix coefficients of $\mathbf{A}^{\ddagger}(L)$ in (2.14) are well defined in view of part *(v)* of Assumption (L4); the same conclusion holds, blockwise, for the n -dimensional VAR (2.6).

For ease of notation, define the filtered processes

$$\mathbf{Y}_n^* := \mathbf{A}_n(L)\{\mathbf{Y}_n - \mathrm{E}[\mathbf{Y}_n]\}, \quad \mathbf{X}_n^* := \mathbf{A}_n(L)\mathbf{X}_n, \quad \text{and} \quad \mathbf{Z}_n^* := \mathbf{A}_n(L)\mathbf{Z}_n$$

with traditional (static) covariance eigenvalues $\mu_{nj}^{Y^*}$, $\mu_{nj}^{X^*}$, and $\mu_{nj}^{Z^*}$, respectively. Since (2.7) is a static factor model, it is natural to make the following assumption on the eigenvalues of the covariance of $\mathbf{X}_n^*$ (see Assumption 4 in Forni et al., 2009 or Assumption 6 in Forni et al., 2017). Unless $q = 1$, indeed, it does not even follow from Assumption (L3) that $\mathrm{E}(\mathbf{X}_n^*\mathbf{X}_n^{*\prime})$ has rank q.

ASSUMPTION (L5). *There exist a positive integer $\bar{n}$ and constants $a_j > b_{j-1}$, $j = 1, \ldots, q$, independent of n such that $0 < a_j \leq \mu_{nj}^{X^*}/n \leq b_j < \infty$ for all $j = 1, \ldots, q$ and all $n > \bar{n}$.*

The following results then hold for the eigenvalues $\mu_{nj}^{Z^*}$ and $\mu_{nj}^{Y^*}$ of the covariance matrices of $\mathbf{Z}_n^*$ and $\mathbf{Y}_n^*$, respectively.

LEMMA 2. *Under Assumptions (L1), (L3), (L4), and (L5),*

(i) there exists a constant $C^{Z^*} > 0$ such that $\mu_{n1}^{Z^*} \leq C^{Z^*}$ for all $n \in \mathbb{N}$;

(ii) there exist a positive integer $\bar{n}$ and constants $a_j^{Y^*} > b_{j-1}^{Y^*}$, $j = 1, \ldots, q$, independent of n such that $0 < a_j^{Y^*} \leq \mu_{nj}^{Y^*}/n \leq b_j^{Y^*} < \infty$ for all $j = 1, \ldots, q$ and all $n > \bar{n}$;

(iii) there exists a constant $C^{Y^*} > 0$ such that $\mu_{n,q+1}^{Y^*} \leq C^{Y^*}$ for all $n \in \mathbb{N}$.

2.2 Model and assumptions for volatilities

We define the vector of common innovations (at time t) as the n-dimensional vector

$$\mathbf{e}_{nt} := (e_{1t}, \ldots, e_{nt})' := \mathbf{H}_n\mathbf{u}_t;$$

for $n > q$, the processes $\mathbf{e}_n := \{\mathbf{e}_{nt} | t \in \mathbb{Z}\}$, $n \in \mathbb{N}$ clearly are singular. Then, letting $s_{it} := e_{it} + v_{it}$, our log-volatility proxy is

$$h_{it} := \log s_{it}^2 = \log(e_{it} + v_{it})^2, \tag{2.9}$$

yielding the double-indexed stochastic process $\mathbf{h} := \{h_{it} | i \in \mathbb{N},\ t \in \mathbb{Z}\}$, with n-dimensional sub-processes $\mathbf{h}_n := \{h_{it} | i = 1, \ldots, n,\ t \in \mathbb{Z}\}$. We call $\mathbf{h}$ the "log-volatilities" process. Similar definitions are used in Engle and Marcucci (2006) and our previous work (Barigozzi and Hallin, 2016, 2017a,b, and Barigozzi et al., 2019). In order for such processes to be well defined we make the following assumption.

ASSUMPTION (V0). *For all $i \in \mathbb{N}$ and $t \in \mathbb{Z}$, $|s_{it}| > 0$ almost surely.*

This assumption makes sure that no cancellation can happen between common and idiosyncratic innovations; it is required, since e_i and v_i, although mutually orthogonal by Assumption (L1.vi), need not be mutually independent (assuming, for instance, that e_i and v_i are absolutely continuous is not sufficient).

Assuming a GDFM with Q factors for the log-volatilities, we obtain

$$h_{it} - \mathrm{E}[h_{it}] = \chi_{it} + \xi_{it} \quad i \in \mathbb{N},\ t \in \mathbb{Z} \tag{2.10}$$

$$\text{with } \chi_{it} = \sum_{j=1}^{Q}\sum_{k=0}^{\infty} f_{ijk}\varepsilon_{jt-k} = \mathbf{f}_i'(L)\boldsymbol{\varepsilon}_t \quad \text{and} \quad \xi_{it} = \sum_{k=0}^{\infty} g_{ik}\nu_{it-k} = g_i(L)\nu_{it}, \tag{2.11}$$

where $\mathrm{E}[h_{it}]$ is h_{it}'s expected value, χ_{it} and ξ_{it} are h_{it}'s *common* and *idiosyncratic* components, and the processes $\boldsymbol{\varepsilon} := \{\varepsilon_{jt}|j = 1,\ldots,Q,\ t \in \mathbb{Z}\}$ and $\boldsymbol{\nu}_n := \{\nu_{it}|i = 1,\ldots,n,\ t \in \mathbb{Z}\}$, $n \in \mathbb{N}$ are mutually orthogonal (at all leads and lags) Q- and n-dimensional white noise, respectively. Note that a GDFM for log-volatilities implies a multiplicative GDFM representation

$$s_{it}^2 = \exp(h_{it}) = \exp(\chi_{it})\exp(\xi_{it})\exp(\mathrm{E}[h_{it}]).$$

for the volatilities themselves. Letting

$$\boldsymbol{\chi}_n := \{\chi_{it}|i = 1,\ldots,n,\ t \in \mathbb{Z}\} \quad \text{and} \quad \boldsymbol{\xi}_n := \{\xi_{it}|i = 1,\ldots,n,\ t \in \mathbb{Z}\},$$

equations (2.11) in vector notation take the form

$$\boldsymbol{\chi}_{nt} = \mathbf{F}_n(L)\boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\xi}_{nt} = \mathbf{G}_n(L)\boldsymbol{\nu}_{nt} \tag{2.12}$$

with $\mathbf{F}_n(L) := (\mathbf{f}_1(L)\ldots\mathbf{f}_n(L))'$ and $\mathbf{G}_n(L) := \mathrm{diag}(g_1(L)\ldots g_n(L))$.

The following assumptions then are the analogues, for log-volatilities and (2.10)-(2.11) , of Assumption (L1).

ASSUMPTION (V1).

(i) The dimension Q of $\boldsymbol{\varepsilon}_t$ does not depend on n; the process $\boldsymbol{\varepsilon} := \{\boldsymbol{\varepsilon}_t|t \in \mathbb{Z}\}$ is second-order white noise, with mean $\mathbf{0}_Q$ and diagonal positive definite covariance $\boldsymbol{\Gamma}^{\varepsilon}$;

(ii) writing $\mathbf{f}_{ik} := (f_{i1k}\ldots f_{iqk})'$ for the $Q \times 1$ coefficient of L^k in $\mathbf{f}_i(L)$, there exists a constant $M_5 > 0$ such that $\sum_{k=0}^{\infty}\|\mathbf{f}_{ik}\|\,|k| \le M_5$ for all $i \in \mathbb{N}$;

(iii) the process $\{\boldsymbol{\nu}_{nt}|t \in \mathbb{Z}\}$ is second-order white noise, with mean $\mathbf{0}_n$ and positive definite covariance $\boldsymbol{\Gamma}_n^{\nu}$; moreover, $\mathrm{E}[\nu_{it}|\nu_{is}] = 0$ for all $i \in \mathbb{N}$ and and $t, s \in \mathbb{Z}$ such that $t > s$;

(iv) there exists a constant $C_\nu > 0$ such that $\|\boldsymbol{\Gamma}_n^{\nu}\|_1 \le C_\nu$ for all $n \in \mathbb{N}$;

(v) there exists a constant $M_6 > 0$ such that $\sum_{k=0}^{\infty}|g_{ik}|\,|k| \le M_6$ for all $i \in \mathbb{N}$;

(vi) $\mathrm{Cov}(\varepsilon_{jt}, \nu_{is}) = 0$ for all $i \in \mathbb{N}$, $j = 1,\ldots,q$, and $t, s \in \mathbb{Z}$;

(vii) there exists a constant $M_7 > 0$ such that $\sum_{k_1,k_2,k_3\in\mathbb{Z}}|\mathrm{E}[\varepsilon_{j_1t-k_1}\varepsilon_{j_2t-k_2}\varepsilon_{j_3t-k_3}\varepsilon_{j_4t}]| \le M_7$ for all $j_1, j_2, j_3, j_4 = 1,\ldots,Q$;

(viii) there exists a constant $M_8 > 0$ such that $\sum_{k_1,k_2,k_3\in\mathbb{Z}}|\mathrm{E}[\nu_{i_1t-k_1}\nu_{i_2t-k_2}\nu_{i_3t-k_3}\nu_{i_4t}]| \le M_8$ for all $i_1, i_2, i_3, i_4 \in \mathbb{N}$.

The same comments made for Assumption (L1) apply here. Moreover, note that all moments of log-transforms of heavy-tailed variables exist and are finite, even for stable distributions (see e.g. Theorem 5.8.1 in Uchaikin and Zolotarev, 2011). Pursuing with assumptions, the following one is the log-volatility counterpart of (L2).

ASSUMPTION (V2).

(i) $\mathbf{f}_i(L)$ has rational entries $f_{ij}(L) = \tilde{\theta}_{ij}(L)\tilde{\phi}_{ij}^{-1}(L)$, where $\tilde{\phi}_{ij}(z)$ and $\tilde{\theta}_{ij}(z)$, for all $i \in \mathbb{N}$ and $j = 1,\ldots,Q$, are finite-order polynomials;

(ii) there exists a constant $\underline{\phi} > 1$ such that $\tilde{\phi}_{ij}(z) \ne 0$ for all $i \in \mathbb{N}$, all $j = 1,\ldots,Q$, and all $z \in \mathbb{C}$ such that $|z| \le \underline{\phi}$;

(iii) *the coefficients $\tilde{\theta}_{ijk}$ of $\tilde{\theta}_{ij}(L)$ are such that $|\tilde{\theta}_{ijk}| \leq B^{\chi}$ for some constant $B^{\chi} > 0$ and all $i \in \mathbb{N}$, $j = 1, \ldots, Q$, and $k \in \mathbb{N} \cup \{0\}$;*
(iv) *$g_i(L)$ is of the form $p_i^{-1}(L)$ where $p_i(z)$, for all $i \in \mathbb{N}$, is a finite-order polynomial, $p_i(0) = 1$ and $p_i(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.*

Assumptions (V2.iv) implies that we can rewrite (2.11) also as

$$p_i(L)Z_{it} = \nu_{it}. \tag{2.13}$$

As in the case of levels, this assumption could be relaxed to allow for an infinite autoregressive order.

Let $\boldsymbol{\Sigma}_n^h(\theta)$, $\boldsymbol{\Sigma}_n^{\chi}(\theta)$, and $\boldsymbol{\Sigma}_n^{\xi}(\theta)$, $\theta \in [-\pi, \pi]$ denote the $n \times n$ spectral density matrices of $\mathbf{h}_n$, its common and its idiosyncratic components, with j-th largest eigenvalues $\lambda_{nj}^h(\theta)$, $\lambda_{nj}^{\chi}(\theta)$ and $\lambda_{nj}^{\xi}(\theta)$, respectively. As in (L3), we assume the following.

ASSUMPTION (V3). *There exist a positive integer $\bar{n}$ and continuous functions $\tilde{\alpha}_j(\theta)$ and $\tilde{\beta}_{j-1}(\theta)$ from $[-\pi, \pi]$ to $\mathbb{R}$, $j = 1, \ldots, Q$, such that $0 < \tilde{\beta}_{j-1}(\theta) < \tilde{\alpha}_j(\theta) \leq \lambda_{nj}^{\chi}(\theta)/n \leq \tilde{\beta}_j(\theta) < \infty$, θ-a.e. in $[-\pi, \pi]$, all $j = 1, \ldots, Q$, and all $n > \bar{n}$.*

Finally, the analogue (V4) of (L4) again is based on the representation results in Forni et al. (2015): any $(Q+1)$- dimensional subvector $\boldsymbol{\chi}_t^{\ddagger}$ of $\boldsymbol{\chi}_{nt}$ admits an autoregressive representation of the form

$$\mathbf{M}^{\ddagger}(L)\boldsymbol{\chi}_t^{\ddagger} = \mathbf{R}^{\ddagger}\boldsymbol{\varepsilon}_t, \tag{2.14}$$

where $\mathbf{M}^{\ddagger}(L)$ is a finite-order VAR operator such that $\mathbf{M}^{\ddagger}(0) = \mathbf{I}_{Q+1}$, $\boldsymbol{\varepsilon}_t$ is the vector of common shocks in (2.12), and $\mathbf{R}^{\ddagger}$ an appropriate $(Q+1) \times Q$ matrix. On that representation, we make the following assumptions:

ASSUMPTION (V4).
(i) *$\mathbf{M}^{\ddagger}(L)$ is uniquely defined;*
(ii) *the degree $\tilde{S}^{\ddagger}$ of $\mathbf{M}^{\ddagger}(z)$ is uniformly bounded, that is, $\tilde{S}^{\ddagger} \leq \tilde{S}$ for some integer $\tilde{S} > 0$ independent of n and the choice of the subvector $\boldsymbol{\chi}_t^{\ddagger}$;*
(iii) $\det[\mathbf{M}^{\ddagger}(z)] \neq 0$ *for all $z \in \mathbb{C}$ such that $|z| \leq 1$.*
(iv) *the $(Q+1) \times Q$ matrix $\mathbf{R}^{\ddagger}$ has full rank Q;*
(v) *denoting by $\boldsymbol{\Gamma}_h^{\chi^{\ddagger}}$ the lag-h autocovariances of $\boldsymbol{\chi}^{\ddagger} := \{\boldsymbol{\chi}_t^{\ddagger}, t \in \mathbb{Z}\}$ and defining $\boldsymbol{\mathcal{V}}^{\ddagger}$ analogously to $\boldsymbol{\mathcal{C}}^{\ddagger}$ in (L4),* $\det(\boldsymbol{\mathcal{V}}^{\ddagger}) > \tilde{d} > 0$, *where $\tilde{d}$ is independent of the choice of the subvector $\boldsymbol{\chi}_t^{\ddagger}$.*

Now, Assumption (V4) implies $[\mathbf{M}_n(L)]^{-1}\mathbf{R}_n = \mathbf{F}_n(L)$, so that, assuming without loss of generality that $n = \bar{m}(Q+1)$ (with $\bar{m} \neq m$ if $Q \neq q$) and defining a block-diagonal autoregressive operator $\mathbf{M}_n(L)$ the way we defined $\mathbf{A}_n(L)$ in the previous section, we can rewrite the GDFM for log-volatilities under the static form

$$\mathbf{M}_n(L)\{\mathbf{h}_{nt} - \mathrm{E}[\mathbf{h}_{nt}]\} = \mathbf{R}_n\boldsymbol{\varepsilon}_t + \mathbf{M}_n(L)\boldsymbol{\xi}_{nt}. \tag{2.15}$$

After defining, with obvious notation, the filtered processes $\mathbf{h}_n^* := \mathbf{M}_n(L)[\mathbf{h}_n - \mathrm{E}[\mathbf{h}_n]]$, $\boldsymbol{\chi}_n^* := \mathbf{M}_n(L)\boldsymbol{\chi}_n$, and $\boldsymbol{\xi}_n^* := \mathbf{M}_n(L)\boldsymbol{\xi}_n$, with (static) spectral eigenvalues $\mu_{nj}^{h^*}$, $\mu_{nj}^{\chi^*}$, and $\mu_{nj}^{\xi^*}$, we conclude with the analogues of (L5) and Lemmas 1 and 2 for the log-volatility panels.

ASSUMPTION (V5). *There exist a positive integer $\bar{n}$ and constants $\tilde{a}_j > \tilde{b}_{j-1} > 0$, $j = 1, \ldots, Q$, independent of n such that $0 < \tilde{a}_j \leq \mu_{nj}^{\chi^*}/n \leq \tilde{b}_j < \infty$ for all $j = 1, \ldots, Q$ and all $n > \bar{n}$.*

We then have the following.

LEMMA 3. *Under Assumptions (V0), (V1), (V3), (V4), and (V5),*
(i) *there exists a constant $C^{\xi} > 0$ such that $\sup_{\theta \in [-\pi, \pi]} \lambda_{n1}^{\xi}(\theta) \leq C^{\xi}$ for all $n \in \mathbb{N}$;*
(ii) *there exist a positive integer $\bar{n}$ and continuous functions $\alpha_j^h(\theta)$ and $\beta_{j-1}^h(\theta)$ from $[-\pi, \pi]$ to $\mathbb{R}$, $j = 1, \ldots, Q$, independent of n and such that $0 < \beta_{j-1}^h(\theta) < \alpha_j^h(\theta) \leq \lambda_{nj}^h(\theta)/n \leq \beta_j^h(\theta) < \infty$, θ-a.e. in $[-\pi, \pi]$, all $j = 1, \ldots, Q$, and all $n > \bar{n}$;*

(iii) *there exists a constant* $C^h > 0$ *such that* $\sup_{\theta \in [-\pi,\pi]} \lambda^h_{n,Q+1}(\theta) \leq C^h$ *for all* $n \in \mathbb{N}$*;*

(iv) *there exists a constant* $C^{\xi^*} > 0$ *such that* $\mu^{\xi^*}_{n1} \leq C^{\xi^*}$ *for all* $n \in \mathbb{N}$*;*

(v) *there exist a positive integer* $\bar{n}$ *and constants* $a^{h^*}_j > b^{h^*}_{j-1}$, $j = 1, \ldots, Q$, *independent of* n *such that* $0 < a^{h^*}_j \leq \mu^{h^*}_{nj}/n \leq b^{h^*}_j < \infty$, *for all* $j = 1, \ldots, Q$ *and all* $n > \bar{n}$*;*

(vi) *there exists a constant* $C^{h^*} > 0$ *such that* $\mu^{h^*}_{n,Q+1} \leq C^{h^*}$ *for all* $n \in \mathbb{N}$.

3 Estimation, consistency, and rates

Hereafter, the terminology "estimation", "estimator", etc. is used, in an orthodox way, for data-driven quantities attempting at evaluating parameters (covariances, spectra, loadings, ...) but also, with a slight abuse, for data-driven quantities attempting at reconstructing unobserved variables (such as common factors, common and idiosyncratic components, ...). All those "estimators", which are $\mathbf{Y}_{n,T}$-measurable random variables (hence depend both on n and T) are carrying "hats".

3.1 Summary of estimation

Estimation proceeds in two parts. The first part deals with the observed $n \times T$ panel $\mathbf{Y}_{n,T}$ of levels, and follows along similar lines as in Forni et al. (2017), yielding estimated log-volatility proxies; the second part consists in repeating the same estimation steps, now based on those estimated log-volatility quantities. Global consistency of the procedure is discussed in the next section, along with further necessary conditions.

To start with, we assume that q and Q are known—an assumption we are relaxing later on. For simplicity of notation, we also assume $\mathbf{Y}_n$ and $\mathbf{h}_n$ to be centred, i.e., to have zero mean; in practice, sample means are to be subtracted in order to obtain centred variables—which has no impact on consistency nor consistency rates.

Here is a detailed list of the steps required for estimation. Further comments on the choice of the quantities needed for estimation and a schematic description of the procedure are given at the end of this section (see also Algorithms 1 and 2).

(*L.i*) To start with, compute the lag-window estimator

$$\widehat{\mathbf{\Sigma}}^Y_n(\theta_h) := \frac{1}{2\pi} \sum_{k=-T+1}^{T-1} \mathrm{K}\left(\frac{k}{B_T}\right) e^{-ik\theta_h} \widehat{\mathbf{\Gamma}}^Y_{nk}, \quad \theta_h = \frac{\pi h}{B_T}, \quad |h| \leq B_T,$$

of the spectral density matrix of returns, where $\widehat{\mathbf{\Gamma}}^Y_{nk} := T^{-1}\sum_{t=|k|+1}^{T} \mathbf{Y}_{nt}\mathbf{Y}'_{nt-|k|}$ is the usual lag-k sample autocovariance matrix of levels and K is a suitable kernel with bandwidth B_T. We here adopt the common choice of a Bartlett kernel

$$\mathrm{K}(x) = \begin{cases} 1 - |x| & \text{if } |x| \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

but other classical kernels are also possible.

(*L.ii*) Collect the q normalised column eigenvectors associated with $\widehat{\mathbf{\Sigma}}^Y_n(\theta_h)$'s q largest eigenvalues into the $n \times q$ matrix $\widehat{\mathbf{P}}^Y_n(\theta_h)$, and collect the corresponding eigenvalues into the $q \times q$ diagonal matrix $\widehat{\mathbf{\Lambda}}^Y_n(\theta_h)$. Take

$$\widehat{\mathbf{\Sigma}}^X_n(\theta_h) := \widehat{\mathbf{P}}^Y_n(\theta_h)\widehat{\mathbf{\Lambda}}^Y_n(\theta_h)\widehat{\mathbf{P}}^{Y\dagger}_n(\theta_h),$$

as an estimate of the spectral density matrix of the level-common component process $\mathbf{X}_n$.

(*L.iii*) By inverse Fourier transform of $\widehat{\mathbf{\Sigma}}^X_n(\theta_h)$, estimate the autocovariance matrices of $\mathbf{X}_n$:

$$\widehat{\mathbf{\Gamma}}^X_{nk} := \frac{\pi}{B_T} \sum_{h=-B_T}^{B_T} e^{ik\theta_h} \widehat{\mathbf{\Sigma}}^X_n(\theta_h), \qquad k \in \mathbb{Z}.$$

(*L.iv*) Assuming, for simplicity, [1] that $n = m(q+1)$, consider the m diagonal $(q+1)\times(q+1)$ blocks of the $\widehat{\mathbf{\Gamma}}^{X}_{nk}$'s. For each block, estimate, via Yule-Walker methods, the coefficients of a $(q+1)$-dimensional VAR model (order determined via AIC or BIC). In other words, compute the sample analogue of (2.8). This yields, for the ℓ-th diagonal block, an estimator $\widehat{\mathbf{A}}^{(\ell)}(L)$ of the autoregressive filter $\mathbf{A}^{(\ell)}(L)$ appearing in Assumption (L4), hence an estimator $\widehat{\mathbf{A}}_n(L)$ of the VAR filter $\mathbf{A}_n(L)$. The resulting estimated filtered process and its estimated covariance matrix are $\widehat{\mathbf{Y}}^*_{nt} := \widehat{\mathbf{A}}_n(L)\mathbf{Y}_{nt}$ and $\widehat{\mathbf{\Gamma}}^{\widehat{Y}^*}_n := T^{-1}\sum_{t=1}^T \widehat{\mathbf{Y}}^*_{nt}\widehat{\mathbf{Y}}^{*\prime}_{nt}$, respectively.

(*L.v*) Collect the q normalised (column) eigenvectors corresponding to $\widehat{\mathbf{\Gamma}}^{\widehat{Y}^*}_n$'s q largest eigenvalues into the $n\times q$ matrix $\widehat{\mathbf{Q}}^{\widehat{Y}^*}_n$. Projecting $\widehat{\mathbf{Y}}^*_{nt}$ onto the space spanned by the columns of $\widehat{\mathbf{Q}}^{\widehat{Y}^*}_n$ provides an estimate $\widehat{\mathbf{e}}_n$ of the innovation process $\mathbf{e}_n$. Taking into account the set of identifying restrictions described in Assumption (I) below, we obtain the estimators

$$\widehat{\mathbf{H}}_n := \sqrt{n}\widehat{\mathbf{Q}}^{\widehat{Y}^*}_n, \qquad \widehat{\mathbf{u}}_t := \frac{1}{n}\widehat{\mathbf{H}}'_n\widehat{\mathbf{Y}}^*_{nt}, \quad \text{and} \quad \widehat{\mathbf{e}}_{nt} := \widehat{\mathbf{H}}_n\widehat{\mathbf{u}}_t = \widehat{\mathbf{Q}}^{\widehat{Y}^*}_n\widehat{\mathbf{Q}}^{\widehat{Y}^{*\prime}}_n\widehat{\mathbf{Y}}^*_{nt}.$$

Our estimator of the dynamic loadings then is $\widehat{\mathbf{B}}_n(L) := \widehat{\mathbf{A}}^{-1}_n(L)\widehat{\mathbf{H}}_n$, where we truncate the filter $\widehat{\mathbf{A}}^{-1}_n(L)$ at some finite lag $\bar{k}_1$. From this we obtain an estimator $\widehat{\mathbf{X}}_{nt} := \widehat{\mathbf{B}}_n(L)\widehat{\mathbf{u}}_t$ of the common component.

(*L.vi*) The resulting estimator of the idiosyncratic component is $\widehat{\mathbf{Z}}_{nt} := \mathbf{Y}_{nt} - \widehat{\mathbf{X}}_{nt}$. Fitting a univariate AR model (order determined via AIC or BIC), either by least squares or via Yule-Walker methods, to each of the n components of $\widehat{\mathbf{Z}}_{nt}$ yields estimators $\widehat{\mathbf{v}}_n$ of the residuals and $\widehat{\mathbf{C}}_n(L)$ of the diagonal matrix of coefficients from which we also obtain $\widehat{\mathbf{D}}_n(L) := \widehat{\mathbf{C}}^{-1}_n(L)$ with $\widehat{\mathbf{C}}^{-1}_n(L)$ truncated at some finite lag $\bar{k}_2$.

(*R*) For all $i = 1,\ldots,n$ and $t = 1,\ldots,T$, let $\widehat{s}_{it} := \widehat{e}_{it} + \widehat{v}_{it}$ and define the estimated log-volatility proxies as capped values of $\log(\widehat{s}^{\,2}_{it})$:

$$\widehat{h}_{it} := \log(\widehat{s}^{\,2}_{it})\,\mathbb{I}(|\widehat{s}_{it}| \geq \kappa_T) + \log(\kappa^2_T)\,\mathbb{I}(|\widehat{s}_{it}| < \kappa_T),$$

where $\kappa_T > 0$ is a sequence of constants to be chosen in order to make our proxy robust to the log-transform. Note that consistency of our estimation procedure requires an adaptive choice of κ_T, depending on the sample size as explained in Assumption (R) below. In particular, κ_T must be strictly positive for consistency to hold.

(*V.i*) Denote by $\widehat{\mathbf{h}}_{nt} := \left(\widehat{h}_{1t}\ldots\widehat{h}_{nt}\right)'$, $t = 1,\ldots,T$ the n-dimensional vector of log-volatility proxies and compute the lag-window estimator

$$\widehat{\mathbf{\Sigma}}^{\widehat{h}}_n(\theta_\ell) := \frac{1}{2\pi}\sum_{k=-T+1}^{T-1}\mathrm{K}\left(\frac{k}{M_T}\right)e^{-ik\theta_\ell}\widehat{\mathbf{\Gamma}}^{\widehat{h}}_{nk}, \quad \theta_\ell = \frac{\pi\ell}{M_T}, \quad |\ell| \leq M_T,$$

of its spectral density matrix, where $\widehat{\mathbf{\Gamma}}^{\widehat{h}}_{nk} := T^{-1}\sum_{t=|k|+1}^T \widehat{\mathbf{h}}_{nt}\widehat{\mathbf{h}}'_{n,t-|k|}$ is the lag-k sample autocovariance matrix of estimated log-volatilities. Here again we adopt the Bartlett kernel, with bandwidth M_T, which could be different from B_T in step (*L.i*).

(*V.ii*)-(*V.vi*) Repeat steps (*L.ii*)-(*L.vi*) for $\widehat{\mathbf{h}}_n$. In particular, steps (*V.ii*)-(*V.v*) yield the estimators $\widehat{\mathbf{M}}_n(L)$ and $\widehat{\mathbf{R}}_n$, from which we compute

$$\widehat{\mathbf{h}}^*_{nt} := \widehat{\mathbf{M}}_n(L)\widehat{\mathbf{h}}_{nt}, \qquad \widehat{\varepsilon}_t := \frac{1}{n}\widehat{\mathbf{R}}'_n\widehat{\mathbf{h}}^*_{nt}, \quad \text{and} \quad \widehat{\mathbf{F}}_n(L) := \widehat{\mathbf{M}}^{-1}_n(L)\widehat{\mathbf{R}}_n,$$

while from (*V.vi*) we obtain $\widehat{\boldsymbol{\nu}}_n$ and $\widehat{\mathbf{P}}_n(L)$, hence $\widehat{\mathbf{G}}_n(L) := \widehat{\mathbf{P}}^{-1}_n(L)$. As before, $\widehat{\mathbf{M}}^{-1}_n(L)$ and $\widehat{\mathbf{P}}^{-1}_n(L)$ are truncated at finite lags $\bar{k}^*_1$ and $\bar{k}^*_2$.

[1] In practice, the last $n - \lfloor n/(q+1)\rfloor(q+1)$ cross-sectional items can be added to the last block in the analysis which will then have size larger than $(q+1)$. Since the arguments in Forni et al. (2017) used in the next section apply to any partition of blocks of size $(q+1)$ or larger, nothing changes in what follows.

Algorithm 1: Estimation of dynamic factor model for levels

Input: data in levels $\mathbf{Y}$ of dimension $n \times T$, number of factors q, bandwidth for estimating spectral density B_T, number of lags for impulse responses $\bar{k}_1$ and $\bar{k}_2$, number of permutations for estimating the common component $nrep$

Output: common component $\widehat{\mathbf{X}}$, idiosyncratic component $\widehat{\mathbf{Z}}$, common shocks $\widehat{\mathbf{u}}$ and $\widehat{\mathbf{e}}$, common impulse responses $\widehat{\mathbf{B}}(L)$, idiosyncratic shocks $\widehat{\mathbf{v}}$, idiosyncratic impulse responses $\widehat{\mathbf{D}}(L)$

Compute autocovariance matrices of data $\widehat{\boldsymbol{\Gamma}}_k^Y$ for $|k| \leq B_T$

Compute the lag-window estimator of the spectral density matrix of data $\widehat{\boldsymbol{\Sigma}}^Y(\theta_h)$ for $\theta_h = \pi h / B_T$ and $|h| \leq B_T$, using $\widehat{\boldsymbol{\Gamma}}_k^Y$ and the Bartlett kernel

for $h \leftarrow -B_T$ **to** B_T **do**

  Compute the q largest eigenvalues $\widehat{\lambda}_1^Y(\theta_h), \dots, \widehat{\lambda}_q^Y(\theta_h)$ of $\widehat{\boldsymbol{\Sigma}}^Y(\theta_h)$ and collect the corresponding eigenvectors into the columns of $\widehat{\mathbf{P}}^Y(\theta_h)$. Let $\widehat{\boldsymbol{\Sigma}}^X(\theta_h) = \widehat{\mathbf{P}}^Y(\theta_h)\text{diag}(\widehat{\lambda}_1^Y(\theta_h), \dots, \widehat{\lambda}_q^Y(\theta_h))\widehat{\mathbf{P}}^{Y\dagger}(\theta_h)$

Compute the autocovariance matrices of the common component $\widehat{\boldsymbol{\Gamma}}_{nk}^X$ for $|k| \leq B_T$ by inverse Fourier transform of $\widehat{\boldsymbol{\Sigma}}^X(\theta_h)$

for $\mathcal{P} \leftarrow 1$ **to** $nrep$ **do**

  Choose a random partition $\mathcal{P}(1), \dots, \mathcal{P}(m(q+1))$ of the n series into $m = \lfloor n/(q+1) \rfloor (q+1)$ blocks such that the first q series are always included and let $\mathbf{Y}_{\mathcal{P}} = (Y_{\mathcal{P}(1)}, \dots, Y_{\mathcal{P}(m(q+1))})'$

  if $m(q+1) < n$ **then** Add the last $n - m$ series to the last block

  for $\ell \leftarrow 1$ **to** m **do**

    Obtain the coefficients $\widehat{\mathbf{A}}_{\mathcal{P}}^{(\ell)}(L)$ fitting a VAR($p_1^{(\ell)}$) on $\mathbf{Y}^{(\ell)} := (Y_{\mathcal{P}((\ell-1)(q+1))} \dots Y_{\mathcal{P}(\ell(q+1)-1)})'$ via Yule Walker equations using $\widehat{\boldsymbol{\Gamma}}_k^X$ for $k = 0, \dots, \ell$, with $p_1^{(\ell)} \leq B_T$ and determined via BIC

  Let $\widehat{\mathbf{A}}_{\mathcal{P}}(L) = \text{diag}(\widehat{\mathbf{A}}_{\mathcal{P}}^{(1)}(L), \dots, \widehat{\mathbf{A}}_{\mathcal{P}}^{(m)}(L))$ and let $\widehat{\mathbf{Y}}_{t,\mathcal{P}}^* = \widehat{\mathbf{A}}_{\mathcal{P}}(L)\mathbf{Y}_{t,\mathcal{P}}$ for $t = 1, \dots, T$

  Compute $\widehat{\mathbf{H}}_{\mathcal{P}}$ as $\sqrt{n}$ times the q leading eigenvectors of the sample covariance matrix of $\widehat{\mathbf{Y}}_{\mathcal{P}}^*$

  Compute $\widetilde{\mathbf{B}}_{\mathcal{P}}(L) = \widehat{\mathbf{A}}_{\mathcal{P}}^{-1}(L)\widehat{\mathbf{H}}_{\mathcal{P}}$ truncating at lag $\bar{k}_1$

  Compute $\widehat{\mathbf{B}}_{\mathcal{P}}(L) = \widetilde{\mathbf{B}}_{\mathcal{P}}(L)\boldsymbol{\mathcal{R}}_{\mathcal{P}}$ with $\boldsymbol{\mathcal{R}}_{\mathcal{P}}$ is $q \times q$ orthogonal and such that the $q \times q$ block of $\widehat{\mathbf{B}}_{\mathcal{P}}(0)$ obtained by isolating the rows corresponding to the first q series in $\mathbf{Y}$ is lower triangular

  Compute $\widehat{\mathbf{u}}_{t,\mathcal{P}} = n^{-1}\boldsymbol{\mathcal{R}}_{\mathcal{P}}'\widehat{\mathbf{H}}_{\mathcal{P}}'\widehat{\mathbf{Y}}_{t,\mathcal{P}}^*$ for $t = 1, \dots, T$

Compute the common shocks as $\widehat{\mathbf{u}}_t = (nrep)^{-1}\sum_{\mathcal{P}=1}^{nrep}\widehat{\mathbf{u}}_{t,\mathcal{P}}$ for $t = 1, \dots, T$

Compute $\widehat{\mathbf{e}}_t = (nrep)^{-1}\sum_{\mathcal{P}=1}^{nrep}\widehat{\mathbf{H}}_{\mathcal{P}}\boldsymbol{\mathcal{R}}_{\mathcal{P}}\widehat{\mathbf{u}}_{t,\mathcal{P}}$

Compute the impulse response functions $\widehat{\mathbf{B}}(L) = (nrep)^{-1}\sum_{\mathcal{P}=1}^{nrep}\widehat{\mathbf{B}}_{\mathcal{P}}(L)$

Compute the common component as $\widehat{\mathbf{X}}_t = \widehat{\mathbf{B}}(L)\widehat{\mathbf{u}}_t$ for $t = 1, \dots, T$

Compute the idiosyncratic component as $\widehat{\mathbf{Z}} = \mathbf{Y} - \widehat{\mathbf{X}}$ such that $\widehat{\mathbf{Z}} = (\widehat{Z}_1 \dots \widehat{Z}_n)'$

for $i \leftarrow 1$ **to** n **do**

  Obtain the coefficients $\widehat{c}_i(L)$ fitting a VAR(s_{1i}) on $\widehat{Z}_i$ via least squares, with s_{1i} determined via BIC

  Let $\widehat{v}_{it} = \widehat{c}_i(L)\widehat{Z}_{it}$ for $t = 1, \dots, T$

Compute the impulse response functions as $\widehat{\mathbf{D}}(L) = \text{diag}(\widehat{c}_1^{-1}(L), \dots, \widehat{c}_n^{-1}(L))$ truncating at lag $\bar{k}_2$

Let the idiosyncratic shocks be $\widehat{\mathbf{v}}_t = (\widehat{v}_{1t} \dots \widehat{v}_{nt})'$ for $t = 1, \dots, T$

Algorithm 2: Estimation of dynamic factor model for log-volatilities

Input: from Algorithm 1: common and idiosyncratic shocks $\widehat{\mathbf{e}}$ and $\widehat{\mathbf{v}}$ both of dimension $n \times T$
number of factors Q, capping constant κ_T, bandwidth for estimating spectral density M_T, number of lags for impulse responses $\bar{k}_1^*$ and $\bar{k}_2^*$, number of permutations for estimating the common component $nrep$

Output: common component $\widehat{\boldsymbol{\chi}}$, idiosyncratic component $\widehat{\boldsymbol{\xi}}$, common shocks $\widehat{\boldsymbol{\varepsilon}}$ and $\widehat{\boldsymbol{\eta}}$, common impulse responses $\widehat{\mathbf{F}}(L)$, idiosyncratic shocks $\widehat{\boldsymbol{\nu}}$, idiosyncratic impulse responses $\widehat{\mathbf{G}}(L)$

for $i \leftarrow 1$ **to** n **do**
    for $t \leftarrow 1$ **to** T **do**
        Compute log-volatility proxy $\widehat{h}_{it}$
        if $|\widehat{e}_{it} + \widehat{v}_{it}| \geq \kappa_T$ **then** $\widehat{h}_{it} = \log(\widehat{e}_{it} + \widehat{v}_{it})^2$
        else $\widehat{h}_{it} = \kappa_T$

Compute autocovariance matrices of log-volatility $\widehat{\boldsymbol{\Gamma}}_k^{\widehat{h}}$ for $|k| \leq M_T$

Compute the lag-window estimator of the spectral density matrix of log-volatility $\widehat{\boldsymbol{\Sigma}}^{\widehat{h}}(\theta_h)$ for $\theta_h = \pi h/M_T$ and $|h| \leq M_T$, using $\widehat{\boldsymbol{\Gamma}}_k^{\widehat{h}}$ and the Bartlett kernel

for $h \leftarrow -M_T$ **to** M_T **do**
    Compute the Q largest eigenvalues $\widehat{\lambda}_1^{\widehat{h}}(\theta_h), \dots, \widehat{\lambda}_Q^{\widehat{h}}(\theta_h)$ of $\widehat{\boldsymbol{\Sigma}}^{\widehat{h}}(\theta_h)$ and collect the corresponding eigenvectors into the columns of $\widehat{\mathbf{P}}^{\widehat{h}}(\theta_h)$. Let $\widehat{\boldsymbol{\Sigma}}^{\widehat{\chi}}(\theta_h) = \widehat{\mathbf{P}}^{\widehat{h}}(\theta_h)\text{diag}(\widehat{\lambda}_1^{\widehat{h}}(\theta_h), \dots, \widehat{\lambda}_Q^{\widehat{h}}(\theta_h))\widehat{\mathbf{P}}^{\widehat{h}\dagger}(\theta_h)$

Compute the autocovariance matrices of the common component $\widehat{\boldsymbol{\Gamma}}_{nk}^{\widehat{\chi}}$ for $|k| \leq M_T$ by inverse Fourier transform of $\widehat{\boldsymbol{\Sigma}}^{\widehat{\chi}}(\theta_h)$

for $\mathcal{P} \leftarrow 1$ **to** $nrep$ **do**
    Choose a random partition $\mathcal{P}(1), \dots, \mathcal{P}(m(Q+1))$ of the n series into $m = \lfloor n/(Q+1) \rfloor (Q+1)$ blocks such that the first Q series are always included and let $\widehat{\mathbf{h}}_{\mathcal{P}} = (\widehat{h}_{\mathcal{P}(1)}, \dots, \widehat{h}_{\mathcal{P}(m(Q+1))})'$
    if $m(Q+1) < n$ **then** Add the last $n-m$ series to the last block
    for $\ell \leftarrow 1$ **to** m **do**
        Obtain the coefficients $\widehat{\mathbf{M}}_{\mathcal{P}}^{(\ell)}(L)$ fitting a VAR($p_2^{(\ell)}$) on $\widehat{\mathbf{h}}^{(\ell)} := (\widehat{h}_{\mathcal{P}((\ell-1)(Q+1))} \dots \widehat{h}_{\mathcal{P}(\ell(Q+1)-1)})'$ via Yule Walker equations using $\widehat{\boldsymbol{\Gamma}}_k^{\widehat{\chi}}$ for $k = 0, \dots, \ell$, with $p_2^{(\ell)} \leq M_T$ and determined via BIC
    Let $\widehat{\mathbf{M}}_{\mathcal{P}}(L) = \text{diag}(\widehat{\mathbf{M}}_{\mathcal{P}}^{(1)}(L), \dots, \widehat{\mathbf{M}}_{\mathcal{P}}^{(m)}(L))$ and let $\widehat{\mathbf{h}}_{t,\mathcal{P}}^* = \widehat{\mathbf{M}}_{\mathcal{P}}(L)\widehat{\mathbf{h}}_{t,\mathcal{P}}$ for $t = 1, \dots, T$
    Compute $\widehat{\mathbf{R}}_{\mathcal{P}}$ as $\sqrt{n}$ times the q leading eigenvectors of the sample covariance matrix of $\widehat{\mathbf{h}}_{\mathcal{P}}^*$
    Compute $\widetilde{\mathbf{F}}_{\mathcal{P}}(L) = \widehat{\mathbf{M}}_{\mathcal{P}}^{-1}(L)\widehat{\mathbf{R}}_{\mathcal{P}}$ truncating at lag $\bar{k}_1^*$
    Compute $\widehat{\mathbf{F}}_{\mathcal{P}}(L) = \widetilde{\mathbf{F}}_{\mathcal{P}}(L)\boldsymbol{\mathcal{R}}_{\mathcal{P}}$ with $\boldsymbol{\mathcal{R}}_{\mathcal{P}}$ is $Q \times Q$ orthogonal and such that the $Q \times Q$ block of $\widehat{\mathbf{M}}_{\mathcal{P}}(0)$ obtained by isolating the rows corresponding to the first Q series in $\widehat{\mathbf{h}}$ is lower triangular
    Compute $\widehat{\varepsilon}_{t,\mathcal{P}} = n^{-1}\boldsymbol{\mathcal{R}}_{\mathcal{P}}'\widehat{\mathbf{R}}_{\mathcal{P}}'\widehat{\mathbf{h}}_{t,\mathcal{P}}^*$ for $t = 1, \dots, T$

Compute the common shocks as $\widehat{\boldsymbol{\varepsilon}}_t = (nrep)^{-1}\sum_{\mathcal{P}=1}^{nrep} \widehat{\varepsilon}_{t,\mathcal{P}}$ for $t = 1, \dots, T$

Compute $\widehat{\boldsymbol{\eta}}_t = (nrep)^{-1}\sum_{\mathcal{P}=1}^{nrep} \widehat{\mathbf{R}}_{\mathcal{P}}\boldsymbol{\mathcal{R}}_{\mathcal{P}}\widehat{\varepsilon}_{t,\mathcal{P}}$

Compute the impulse response functions $\widehat{\mathbf{F}}(L) = (nrep)^{-1}\sum_{\mathcal{P}=1}^{nrep} \widehat{\mathbf{F}}_{\mathcal{P}}(L)$

Compute the common component as $\widehat{\boldsymbol{\chi}}_t = \widehat{\mathbf{F}}(L)\widehat{\varepsilon}_t$ for $t = 1, \dots, T$

Compute the idiosyncratic component as $\widehat{\boldsymbol{\xi}} = \widehat{\mathbf{h}} - \widehat{\boldsymbol{\chi}}$ such that $\widehat{\boldsymbol{\xi}} = (\widehat{\xi}_1 \dots \widehat{\xi}_n)'$

for $i \leftarrow 1$ **to** n **do**
    Obtain the coefficients $\widehat{p}_i(L)$ fitting a VAR(s_{2i}) on $\widehat{\xi}_i$ via least squares, with s_{2i} determined via BIC
    Let $\widehat{\nu}_{it} = \widehat{p}_i(L)\widehat{\xi}_{it}$ for $t = 1, \dots, T$

Compute the impulse response functions as $\widehat{\mathbf{G}}(L) = \text{diag}(\widehat{p}_1^{-1}(L), \dots, \widehat{p}_n^{-1}(L))$ truncating at lag $\bar{k}_2^*$

Let the idiosyncratic shocks be $\widehat{\boldsymbol{\nu}}_t = (\widehat{\nu}_{1t} \dots \widehat{\nu}_{nt})'$ for $t = 1, \dots, T$

An important remark needs to be made here. The cross-sectional ordering of the panel has an impact on the selection of the diagonal blocks in steps (*L.iv*) and (*V.iv*). Each cross-sectional permutation of the panel, thus, would lead to distinct estimators—all sharing the same asymptotic properties. A Rao-Blackwell argument (see Forni et al., 2017 for details) suggests aggregating these estimators into a unique one by simple averaging (after obvious reordering of the cross-section) of the resulting estimated shocks. Although averaging over all $n!$ permutations is clearly unfeasible, as stressed by Forni et al. (2017) and verified empirically also in Forni et al. (2018), a few of them are enough, in practice, to deliver stable averages (which therefore are matching the infeasible average over all $n!$ permutations).

Implementation of the above estimation steps is described in Algorithms 1 and 2. Those algorithms require setting bandwidths B_T and M_T for the estimation of the spectral densities, a capping constant κ_T, and the number of factors q and Q. Concerning the bandwidths and the capping constant, we refer to Section 3.2 for the required asymptotic properties (see Assumptions (K) and (R), respectively), while a numerical assessment of the impact of these quantities is provided in Section 5 on simulated data (see also the results in the online Appendix D) and in Section 6 on real data. Overall, our numerical analysis shows that low levels of capping or even no capping at all are preferable, as they avoid inducing too much bias in the log-volatility distributions. As for the bandwidths, large values of T are required to construct reliable estimates, since they allow setting M_T large enough to capture the high persistence of log-volatility series. Our results are quite insensitive to the choice of B_T, due to the fact that financial returns typically are only weakly autocorrelated.

Finally, we can determine the numbers q and Q of common shocks by means of the information criteria proposed by Hallin and Liška (2007) and applied on the panels $\mathbf{Y}_n$ and $\widehat{\mathbf{h}}_n$, respectively. The resulting data-driven estimators $\widehat{q}$ and $\widehat{Q}$ converge in probability to q and Q, respectively. Since q and Q are integers, this means that, for any $\epsilon > 0$, there exist $n(\epsilon)$ and $T(\epsilon)$ such that, for all $n > n(\epsilon)$ and $T > T(\epsilon)$, $\widehat{q} = q$ and $\widehat{Q} = Q$ with probability larger than $1 - \epsilon$. Hence, in Section 3.2 below, we safely can assume that q and Q are known.

3.2 Consistency and rates

Consistency of the estimators of the GDFM model for levels is proved in Forni et al. (2017). Some differences exist, though, between their approach and ours. First, Forni et al. (2017) make slightly weaker assumptions on idiosyncratic serial dependence and, by exploiting results in Wu and Zaffaroni (2018) on spectral density estimation, they derive their consistency results under the constraint that $B_T \log B_T/T \to 0$ as $T \to \infty$. A more classical approach is adopted here, based on Assumptions (L1) and (V1), which as a consequence requires mildly stronger constraints on the range of admissible values for the bandwidths B_T and M_T. Specifically, we require the following.

ASSUMPTION (K). *As $T \to \infty$, $B_T = o(\sqrt{T})$ and $M_T = o(\sqrt{T})$.*

Note that for $T \simeq 1000$ as in our empirical study, the range of admissible bandwidths is still such that most of the serial dependence in the data is captured when estimating the spectral density (see Section 6 for more details on the choice of the bandwidths).

Second, the results in Forni et al. (2017) hold pointwise in t, which is not sufficient for our needs when it comes to prove consistency in the second part of the estimation procedure. Indeed, we need uniform (over all $t \in \{1, \ldots, T\}$) consistency of the estimators of the common and idiosyncratic components. For this reason, we make additional assumptions on the distribution of common and idiosyncratic components.

ASSUMPTION (T). *There exist constants $K_u > 0$, $K_\varepsilon > 0$, $K_Z > 0$, and $K_\xi > 0$, such that, for any $t = 1, \ldots, T$,*
(i) $\max_{j=1,\ldots,q} \|u_{jt}\|_{\psi_1} \leq K_u$;
(ii) $\max_{j=1,\ldots,Q} \|\varepsilon_{jt}\|_{\psi_1} \leq K_\varepsilon$;
(iii) $\sup_{\boldsymbol{w}_n:\|\boldsymbol{w}_n\|=1} \|\boldsymbol{w}_n' \mathbf{Z}_{nt}\|_{\psi_1} \leq K_Z$, *for all $n \in \mathbb{N}$;*
(iv) $\sup_{\boldsymbol{w}_n:\|\boldsymbol{w}_n\|=1} \|\boldsymbol{w}_n' \boldsymbol{\xi}_{nt}\|_{\psi_1} \leq K_\xi$, *for all $n \in \mathbb{N}$.*

This assumption is equivalent to an assumption of sub-exponential tails of the common factors and the normed linear combinations of idiosyncratic components. Specifically, it can be shown that (T*i*) is equivalent to requiring for any $j = 1, \ldots, q$, that $\mathrm{P}(|u_{jt}| > \epsilon) \leq K_u^* \exp(-\epsilon/K_u^{**})$ for any $\epsilon > 0$ and some finite $K_u^*, K_u^{**} > 0$ (see also

Vershynin, 2012, and Appendix A.3 for details). The same holds also for (T*ii*), (T*iii*), and (T*iv*). See Remark 1 at the end of this section for a discussion of the implications and possible relaxations of this assumption.

Two remarks on (T*iii*) and (T*iv*) are in order here (see Sections 5.2.4 and 5.2.5 in Vershynin, 2012 for details). First, note that by letting $\boldsymbol{w}_n = (0 \dots w_i \dots 0)'$, with $w_i = 1$ for a given i, those assumptions imply that each idiosyncratic component has marginal sub-exponential distribution. Second, an implication of Lemmas 2 and 3 is that vectors of the form $\boldsymbol{w}_n' \mathbf{Z}_n$ and $\boldsymbol{w}_n' \boldsymbol{\xi}_n$ have finite variance for all n, a necessary condition for pointwise consistency. However, (T*iii*) and (T*iv*) are stricter on idiosyncratic cross-sectional dependence, since they control all moments of normed linear combinations of idiosyncratic components. Indeed, since the common components $\mathbf{X}_n$ and $\boldsymbol{\chi}_n$ are recovered by aggregation across the n elements of $\mathbf{Y}_n$ and $\widehat{\mathbf{h}}_n$, respectively, uniform consistency requires limiting the contribution of the tails of the distribution of cross-sectional averages of idiosyncratic components.

Finally, since factors and factor loadings are not separately identified, we can, without loss of generality, impose the following assumptions, which are just identification constraints (see Forni et al., 2009 for similar conditions).

ASSUMPTION (I). *(i) Denoting by* $\mathbf{P}_n^{X^*}$ *the* $n \times q$ *matrix of normalized column eigenvectors corresponding to the* q *largest eigenvalues of the covariance matrix of* $\mathbf{X}_n^*$, *put* $\mathbf{H}_n := \sqrt{n}\mathbf{P}_n^{X^*}$ *and* $\mathbf{u}_t := \mathbf{P}_n^{X^*\prime}\mathbf{X}_n^*/\sqrt{n}$;
(ii) denoting by $\mathbf{P}_n^{\chi^*}$ *the* $n \times Q$ *matrix of normalized eigenvectors corresponding to the* Q *largest eigenvalues of the covariance matrix of* $\boldsymbol{\chi}_n^*$, *put* $\mathbf{R}_n := \sqrt{n}\mathbf{P}_n^{\chi^*}$ *and* $\boldsymbol{\varepsilon}_t := \mathbf{P}_n^{\chi^*\prime}\boldsymbol{\chi}_n^*/\sqrt{n}$.

In other words, Assumption (I) requires the common factors $\mathbf{u}_t$ ($\boldsymbol{\varepsilon}_t$) to be the (non-normalised) principal components of $\mathbf{X}_n^*$ ($\boldsymbol{\chi}_n^*$). Note that, under Assumption (I), both the factors and their loadings depend on n; their product, however, does not, which is particularly convenient and simplifies the proofs. Other identification constraints are commonly used in principal component analysis (see e.g. Fan et al., 2013); they do not affect the results below, but lead to much heavier notation.

The consistency properties of the estimated GDFM for the levels as described in steps (*L.i*)-(*L.vi*) are as follows.

PROPOSITION 1. *Let* $\rho_{nT} := \max\left(B_T/\sqrt{T}, 1/B_T, 1/\sqrt{n}\right)$. *Then, under Assumptions (L1)-(L5), (K), (T), and (I), there exists a* $q \times q$ *diagonal matrix* $\mathbf{J}$ *with entries* ± 1 *such that*
(a) $\max_{i=1,\dots,n} \|\widehat{\mathbf{b}}_{ik}' - \mathbf{b}_{ik}'\mathbf{J}\| = O_{\mathrm{P}}(\rho_{nT})$, *for all* $k \leq \bar{k}_1$;
(b) $\max_{t=1,\dots,T} \|\widehat{\mathbf{u}}_t - \mathbf{J}\mathbf{u}_t\| = O_{\mathrm{P}}(\rho_{nT}\log T)$;
(c) $\max_{i=1,\dots,n} |\widehat{d}_{ik} - d_{ik}| = O_{\mathrm{P}}(\rho_{nT}\log^2 T)$, *for all* $k \leq \bar{k}_2$;
(d) $\max_{i=1,\dots,n}\max_{t=1,\dots,T} |\widehat{v}_{it} - v_{it}| = O_{\mathrm{P}}(\rho_{nT}\log^2 T)$.

The proof of parts *(a)* and *(b)* of Proposition 1 follows directly from Forni et al. (2017) together with Assumptions (T*i*) and (T*iii*). However, parts *(c)* and *(d)* concerning the idiosyncratic components are new results and provide uniform consistency over both time and the cross-section (see also Remark 1 below). In particular, notice that parts *(c)* and *(d)* of Proposition 1 are proved under Assumption (L2*iv*) of a finite-order autoregressive representation for the idiosyncratic component. Relaxing that assumption into possibly infinite-order autoregressive repressentations would require addressing, in the proofs of parts *(c)* and *(d)*, the issue of truncation errors related to finite-order AR fitting. Consistency still could be proved, but with rates depending on the rate of decay of the autocovariances of idiosyncratic components, as shown, for example, in den Haan and Levin (1997). For simplicity, we do not consider this here.

As for the global consistency properties (after the second estimation step), we need a final condition on the choice of the capping sequence κ_T in step (*R*).

ASSUMPTION (R). *The sequence* $\kappa_T > 0$ *is such that the sets* $\mathcal{T}_{i;nT} := \left\{t \in \{1,\dots,T\} \,\middle|\, |\widehat{s}_{it}| < \kappa_T\right\}$ *satisfy* $\max_{i=1,\dots,n} |\mathcal{T}_{i;nT}| = o_{\mathrm{P}}(\sqrt{T})$ *uniformly in* n *as* $T \to \infty$. *Moreover, there exist a positive integer* $\bar{T}$ *and constants* $\varphi > 1$ *and* $0 < \underline{c} \leq \bar{c}$, *independent of* n, *such that* $\underline{c} \leq \kappa_T \log^{\varphi} T \leq \bar{c}$ *for all* $T > \bar{T}$.

The intuition behind this assumption is as follows. As shown in Appendix A.4, an immediate consequence of Proposition 1 is that the volatility proxies are consistently estimated, namely,

$$\max_{i=1,\dots,n}\max_{t=1,\dots,T} |\widehat{s}_{it} - s_{it}| = O_{\mathrm{P}}(\rho_{nT}\log^2 T), \qquad \text{as } n, T \to \infty.$$

Now, setting $\kappa_T = 0$ in step (*R*), then, due to the log-transform, uniform consistency of $\widehat{h}_{it}$ becomes problematic when $\widehat{s}_{it}$ gets "close to zero". For this reason, we need $\kappa_T > 0$. The set $\mathcal{T}_{i;nT}$ is that of all time points $\{1, \ldots, T\}$ at which $\widehat{s}_{it}$ is close to zero, and uniform consistency of $\widehat{h}_{it}$ for $t \in \mathcal{T}^c_{i;nT}$ straightforwardly follows from uniform consistency of $\widehat{s}_{it}$. On the other hand, the sets $\mathcal{T}_{i;nT}$ should not contain too many time points, and have cardinality going to zero at appropriate rate—whence Assumption (R). In particular, we suggest to choose κ_T of the order of $\log^{-\varphi} T$ for all i. Although we do not have theoretical results justifying this choice of κ_T in practice, simulation-based results (see Appendix B) indicate that, the condition on the cardinality of the sets $\mathcal{T}_{i;nT}$ is indeed satisfied for κ_T decreasing logarithmically in T.

Consistency of the estimated GDFM for log-volatilities as described in steps (*R*) and (*V.i*)-(*V.vi*) then follows.

PROPOSITION 2. *Let* $\tau_{nT} := \max\left(B_T M_T/\sqrt{T}, M_T/\sqrt{n}\right)$ *and assume that* $B_T \geq cT^{1/4}$ *for some finite* $c > 0$. *Then, under Assumptions (L1)-(L5), (V1)-(V5), (K), (T), (I), and (R), there exists a* $Q \times Q$ *diagonal matrix* $\mathbf{S}$ *with entries* ± 1 *such that*

(a) $\max_{i=1,\ldots,n} \|\widehat{\mathbf{f}}'_{ik} - \mathbf{f}'_{ik}\mathbf{S}\| = O_{\mathrm{P}}(\tau_{nT} \log^{3+\varphi} T)$ *for all* $k \leq \bar{k}^*_1$*;*
(b) $\max_{t=1,\ldots,T} \|\widehat{\boldsymbol{\varepsilon}}_t - \mathbf{S}\boldsymbol{\varepsilon}_t\| = O_{\mathrm{P}}(\tau_{nT} \log^{4+\varphi} T)$*;*
(c) $\max_{i=1,\ldots,n} |\widehat{g}_{ik} - g_{ik}| = O_{\mathrm{P}}(\tau_{nT} \log^{5+\varphi} T)$ *for all* $k \leq \bar{k}^*_2$*;*
(d) $\max_{i=1,\ldots,n} \max_{t=1,\ldots,T} |\widehat{\nu}_{it} - \nu_{it}| = O_{\mathrm{P}}(\tau_{nT} \log^{5+\varphi} T)$.

This result, which is new, provides the theoretical foundation for the consistency of the estimators used in Barigozzi and Hallin (2016, 2017a,b) and in this paper. Note that parts *(c)* and *(d)*, just as parts *(c)* and *(d)* of Proposition 1, are proved under Assumption (V2*iv*) of a finite-order autoregressive representation for the idiosyncratic components; the same comments as for Proposition 1 apply.

Our results show that, up to logarithmic factors and the bandwidth-related ones, the rates of consistency of our estimators are of order $\min(\sqrt{T}, \sqrt{n})$ as in classical one-step factor models. The following three technical remarks discuss how our assumptions, in particular Assumptions (T) and (K), affect the consistency rates, and how the effect of those logarithmic and bandwidth-related factors could be controlled further if we were willing to make additional assumptions.

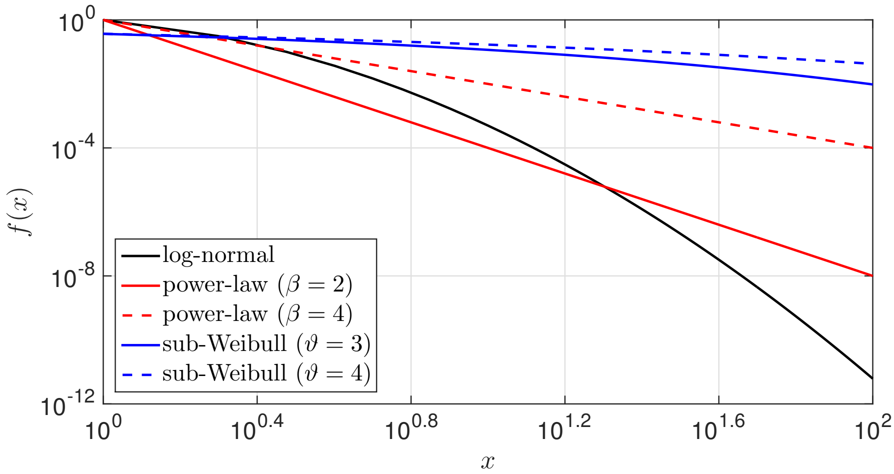
REMARK 1 (Serial dependence of idiosyncratic components). Inspection of the proof of part *(c)* of Proposition 1 shows that the extra (with respect to part *(b)*) $\log T$ factor there is due to terms of the type $T^{-1}\sum_{t=1}^T Z_{it}$. Now, while the cross-sectional dependence of idiosyncratic components is controlled via Assumption (T*iii*), we do not impose (beyond weak stationarity) any specific assumption on their serial dependence. However, it is worth noting that, if we made some mild additional mixing assumption controlling that serial dependence, then those terms could be bounded by a Bernstein-type inequality, as for example in Theorem 1 by Merlevède et al. (2011). Similar comments apply to Proposition 2 and bounds on the idiosyncratic sums $T^{-1}\sum_{t=1}^T \xi_{it}$. If such additional assumptions were made, the rates in Proposition 1 parts *(c)* and *(d)* would change to $O_{\mathrm{P}}(\rho_{nT} \log T)$, those in Proposition 2 part *(a)* to $O_{\mathrm{P}}(\tau_{nT} \log^{1+\varphi} T)$, those in part *(b)* to $O_{\mathrm{P}}(\tau_{nT} \log^{2+\varphi} T)$, and those in parts *(c)* and *(d)* to $O_{\mathrm{P}}(\tau_{nT} \log^{2+\varphi} T)$.

REMARK 2 (Tail behavior). In Section 6, we analyze a panel of stock returns, and it is therefore worth discussing how our assumptions relate to the distributional properties of financial data. First, let us stress that it is common, in the financial econometrics literature, to assume Gaussianity of log-volatility proxies (see e.g. Alizadeh et al., 2002). This is in agreement with the tail Assumptions (T*ii*) and (T*iv*) since sub-Gaussians tails are lighter than sub-exponentials. In the Gaussian case, the rates in Proposition 2 part *(a)* would change to $O_{\mathrm{P}}(\tau_{nT} \log^{5/2+\varphi} T)$, those in part *(b)* to $O_{\mathrm{P}}(\tau_{nT} \log^{3+\varphi} T)$, and those in parts *(c)*, and *(d)* to $O_{\mathrm{P}}(\tau_{nT} \log^{7/2+\varphi} T)$.

Second, Assumption (T*i*) straightforwardly generalizes to more general classes of distributions such that, for some finite constants $K^*_u > 0$, $K^{**}_u > 0$, and $\vartheta > 0$, $\mathrm{P}(|u_{jt}| > \epsilon) \leq K^*_u \exp\left(-\epsilon^{\vartheta}/K^{**}_u\right)$ for any $\epsilon > 0$ and $j = 1, \ldots, q$; (T*iii*) can be generalized similarly for level idiosyncratic components. These distributions are studied in the literature under the name of *sub-Weibull distributions* (Kuchibhotla and Chakrabortty, 2018, and Vladimirova and Arbel, 2019) or *semi-exponential* (Borovkov, 2000).[2] By letting $\vartheta < 1$, we could allow for

[2]Note that the assumption of a sub-Weibull tail decay is equivalent to the moment condition $(\mathrm{E}[|u_{jt}|^k])^{1/k} \leq Ck^{1/\vartheta}$ for all $k \geq 1$ and some finite $C > 0$ (see Vladimirova and Arbel, 2019, Theorem 2.1); fourth-order moments in that case always exist.

FIGURE 1: *Comparison of the tails of log-normal (black), power-law (red) and sub-Weibull (blue) probability density functions $f(x)$ (log-scales on both axis).*



tails, which, although still exponentially decaying, could be heavier than assumed in Assumption (T), thus accounting for moderately extreme events. Following the same steps as in Appendix A.3, it is easily seen that in this case the rates in Proposition 1 part *(b)* would change to $O_{\mathrm{P}}(\rho_{nT}\log^{1/\vartheta} T)$ and those in parts *(c)* and *(d)*) to $O_{\mathrm{P}}(\rho_{nT}\log^{2/\vartheta} T)$. As for Proposition 2, would we assume a sub-Weibull distribution also in (T*ii*) and (T*iv*) (with the same value of ϑ), then rates would change to $O_{\mathrm{P}}(\tau_{nT}\log^{3/\vartheta+\varphi} T)$ in part *(a)*, to $O_{\mathrm{P}}(\tau_{nT}\log^{4/\vartheta+\varphi} T)$ in part *(b)*, and to $O_{\mathrm{P}}(\tau_{nT}\log^{5/\vartheta+\varphi} T)$ in parts *(c)* and *(d)*. To conclude, assuming sub-Gaussian tails in (T*ii*) and (T*iv*) modifies the rates in part *(a)* of Proposition 2 into $O_{\mathrm{P}}(\tau_{nT}\log^{2/\vartheta+1/2+\varphi} T)$, those in part *(b)* into $O_{\mathrm{P}}(\tau_{nT}\log^{2/\vartheta+1+\varphi} T)$, and those in parts *(c)* and *(d)* into $O_{\mathrm{P}}(\tau_{nT}\log^{2/\vartheta+3/2+\varphi} T)$.

Finally, in principle, we also could assume *power-law decay*—that is, the existence of finite constants $K_u^* > 0$ and $\beta > 0$ such that $\mathrm{P}(|u_{jt}| > \epsilon) \leq K_u^* \epsilon^{-\beta}$ for any $j = 1, \ldots, q$ and $\epsilon > 0$; we similarly could generalize (T*iii*) for level idiosyncratic components. We do not explore this possibility in detail, but we notice that , in order to have consistency under this setting, we would need at least $\beta > 2$; moreover, the smaller β, the smaller the range of admissible choices for the bandwidths B_T and M_T. Notice however that, in practice, determining the actual values of ϑ and β is very tricky, and that small values of ϑ can generate a tail behavior which is comparable to the power-law behavior (see Figure 1).

REMARK 3 (Bandwidths and estimation of spectral densities). The results in Propositions 1 and 2 require uniform consistency of the estimated spectral density over all frequencies. For this reason, we have stronger than usual asymptotic constraints on the bandwidths. These could be relaxed if we made stronger assumptions on the shocks. First, notice that in our setting the level shocks are just uncorrelated (see Assumptions (L1*i*) and (L1*iii*)), and are by no means independent. However, if we are willing to assume the existence, for level shocks, of moments of all orders, then we could apply Theorem 7.7.4 in Brillinger (2001), which would allow us to replace B_T with $B_T^{\epsilon}\sqrt{B_T}$ for any $\epsilon > 0$ in the definition of ρ_{nT} in Proposition 1. Second, we could, in principle, allow for independent shocks on log-volatilities (e.g. assuming Gaussianity, see Remark 1) and therefore make use of Theorem 4 and Section 4.2 in Wu and Zaffaroni (2018), which would allow us to replace M_T with $\sqrt{M_T \log M_T}$ in the definition of τ_{nT} in Proposition 2.

4 Conditional prediction intervals

Before describing our prediction intervals, let us summarise here the main notation developed in the previous sections. Given an observed dataset of size $n \times T$, we have, for the levels,

$$
\begin{aligned}
&Y_{it} = X_{it} + Z_{it} + \mathrm{E}[Y_{it}], \qquad (4.1)\\
&X_{it} = \mathbf{b}'_{i0}\mathbf{u}_t + \sum_{k=1}^{\infty}\mathbf{b}'_{ik}\mathbf{u}_{t-k} := e_{it} + X_{it|t-1}, \quad Z_{it} = d_{i0}v_{it} + \sum_{k=1}^{\infty} d_{ik}v_{it-k} := v_{it} + Z_{it|t-1},\\
&s_{it} := e_{it} + v_{it}, \;\; i = 1, \dots, n, \; t = 1, \dots, T
\end{aligned}
$$

where $d_{i0} = 1$ because of (2.4) and, for the log-volatilities,

$$
\begin{aligned}
&h_{it} := \log s_{it}^2 = \chi_{it} + \xi_{it} + \mathrm{E}[h_{it}], \qquad (4.2)\\
&\chi_{it} = \mathbf{f}'_{i0}\boldsymbol{\varepsilon}_t + \sum_{k=1}^{\infty}\mathbf{f}'_{ik}\boldsymbol{\varepsilon}_{t-k} := \eta_{it} + \chi_{it|t-1}, \quad \xi_{it} = g_{i0}\nu_{it} + \sum_{k=1}^{\infty} g_{ik}\nu_{it-k} := \nu_{it} + \xi_{it|t-1},\\
&\omega_{it} := \eta_{it} + \nu_{it}, \;\; i = 1, \dots, n, \; t = 1, \dots, T
\end{aligned}
$$

where $g_{i0} = 1$ because of (2.13).

The optimal one-step-ahead linear predictors of level Y_{it} and log-volatility h_{it} are thus

$$
Y_{it|t-1} := X_{it|t-1} + Z_{it|t-1} + \mathrm{E}[Y_{it}] \quad \text{and} \quad h_{it|t-1} := \chi_{it|t-1} + \xi_{it|t-1} + \mathrm{E}[h_{it}], \qquad (4.3)
$$

with innovations s_{it} and ω_{it}, respectively. As a consequence, the level innovations are

$$
s_{it} = \exp\big(h_{it}/2\big)\mathrm{sign}(s_{it}) = \exp\big(h_{it|t-1}/2\big)\exp\big(\omega_{it}/2\big)\mathrm{sign}(s_{it}).
$$

We therefore define a one-step-ahead predictor of the volatilities as

$$
s_{it|t-1} := \exp\big(h_{it|t-1}/2\big),
$$

with associated "multiplicative innovations"

$$
w_{it} := \exp\big(\omega_{it}/2\big)\mathrm{sign}(s_{it}).
$$

Note, however, that, due to the nonlinear nature of the exponential transformation from h_i to s_i, this multiplicative decomposition of volatilities into a predictor and an "innovation" does not enjoy (in the space of volatilities) the traditional L^2 optimality properties, which only hold for their logarithms (in the space of log-volatilities). This, however, will not be a concern in the quantile-based construction we now describe, due to the fact that the coverage probabilities of a interquantile interval are invariant under continuous monotone transformations: the quantile of w_{it} .

Denoting by $q(\alpha; w_i)$ the (unconditional) α-quantile of $w_i := \{w_{it} | t = 1, \dots, T\}$, $i = 1, \dots, n$ (which, by stationarity, does not depend on t), theoretical lower and upper prediction bounds with confidence level $(1-\alpha)$ and $\alpha \in (0, 1)$ are

$$
\mathcal{L}_{it|t-1}(\alpha) := Y_{it|t-1} + s_{it|t-1}\, q(\alpha; w_i) \;\text{ and }\; \mathcal{U}_{it|t-1}(\alpha) := Y_{it|t-1} + s_{it|t-1}\, q(1-\alpha; w_i), \qquad (4.4)
$$

respectively. Note that $Y_{it|t-1}$ lies above $\mathcal{L}_{it|t-1}(\alpha)$ for $\alpha < \mathrm{P}[w_{it} \le 0]$ and lies below $\mathcal{U}_{it|t-1}(\alpha)$ for $\alpha < 1 - \mathrm{P}[w_{it} \le 0]$. Prediction intervals with coverage probability $(1-\alpha)$ can be constructed as

$$
\mathcal{I}_{it|t-1}(\alpha) := \big[\, \mathcal{L}_{it|t-1}(\alpha^-), \mathcal{U}_{it|t-1}(\alpha^+)\, \big] \qquad (4.5)
$$

with $\alpha^\pm < 1/2$ and $\alpha^- + \alpha^+ = \alpha$, covering $Y_{it|t-1}$ (see (4.8)) provided that

$$
\alpha^- < \mathrm{P}[w_{it} \le 0] \quad \text{and} \quad \alpha^+ < 1 - \mathrm{P}[w_{it} \le 0]. \qquad (4.6)
$$

Clearly, the lower bound $\mathcal{L}_{it|t-1}(\alpha)$ provides a measure of the Value-at-Risk of level α at time t, which we denote as $\mathrm{VaR}_{it}(\alpha) := -\mathcal{L}_{it|t-1}(\alpha)$ (see Section 12.3.1 in Francq and Zakoian, 2011 for a review).[3]

The advantage of quantile-based prediction intervals of the form (4.5) over their conditional heteroskedasticity-based competitors stems from the fact that, irrespective of the way w_i has been obtained, the conditional α-quantiles of Y_{it} (conditional on $Y_{i,t-1}, Y_{i,t-2}, \ldots$) *are* of the form (4.4). This quantile-based approach moreover allows for unequal tails ($\alpha^- \neq \alpha^+$ in (4.6)—hence, distinct attitudes towards losses and gains) and automatically takes into account the typical skewness of financial data distributions.

In practice, the model is estimated from a $n \times T$ observed panel; the empirical counterparts of $Y_{i,T+1|T}$ and $h_{i,T+1|T}$ for $i = 1, \ldots, n$ are

$$\widehat{Y}_{i,T+1|T} = \widehat{X}_{i,T+1|T} + \widehat{Z}_{i,T+1|T} + \frac{1}{T}\sum_{t=1}^{T} Y_{it} = \sum_{k=1}^{\bar{k}_1} \widehat{\mathbf{b}}'_{ik}\widehat{\mathbf{u}}_{T-k+1} + \sum_{k=1}^{\bar{k}_2} \widehat{d}_{ik}\widehat{v}_{i,T-k+1} + \frac{1}{T}\sum_{t=1}^{T} Y_{it}$$

and

$$\widehat{h}_{i,T+1|T} = \widehat{\chi}_{i,T+1|T} + \widehat{\xi}_{i,T+1|T} + \frac{1}{T}\sum_{t=1}^{T} \widehat{h}_{it} = \sum_{k=1}^{\bar{k}_1^*} \widehat{\mathbf{f}}'_{ik}\widehat{\boldsymbol{\varepsilon}}_{T-k+1} + \sum_{k=1}^{\bar{k}_2^*} \widehat{g}_{ik}\widehat{\nu}_{i,T-k+1} + \frac{1}{T}\sum_{t=1}^{T} \widehat{h}_{it},$$

and we accordingly define $\widehat{s}_{i,T+1|T} := \exp\big(\widehat{h}_{i,T+1|T}/2\big)$; based on the estimates $\widehat{s}_{it}$ and $\widehat{\omega}_{it}$ of s_{it} and ω_{it}, let $\widehat{w}_{it} := \exp\big(\widehat{\omega}_{it}/2\big)\mathrm{sign}(\widehat{s}_{it})$.

For any i, denote by $\widehat{w}_{i(1)}, \ldots, \widehat{w}_{i(T)}$ the order statistic of $\widehat{w}_{i1}, \ldots, \widehat{w}_{iT}$; the empirical quantile $w_{i(\lceil T\alpha\rceil)}$ then can be used as an estimator of $q(\alpha; w_i)$. Empirical versions of the prediction limits and intervals (4.4) and (4.5) are

$$\widehat{\mathcal{L}}_{i,T+1|T}(\alpha) := \widehat{Y}_{i,T+1|T} + \widehat{s}_{i,T+1|T}\, \widehat{w}_{i(\lceil T\alpha\rceil)}, \quad \widehat{\mathcal{U}}_{i,T+1|T}(\alpha) := \widehat{Y}_{i,T+1|T} + \widehat{s}_{i,T+1|T}\, \widehat{w}_{i(\lceil T(1-\alpha)\rceil)}$$

and

$$\widehat{\mathcal{I}}_{i,T+1|T}(\alpha) := \big[\widehat{\mathcal{L}}_{i,T+1|T}(\alpha^-), \widehat{\mathcal{U}}_{i,T+1|T}(\alpha^+)\big] \tag{4.7}$$

with $\alpha^\pm < 1/2$ and $\alpha^- + \alpha^+ = \alpha \in (0,1)$. A schematic description of this procedure is given in Algorithm 3.

If the w_{it}'s were i.i.d. instead of weak white noise, the convergence (for given α^- and α^+, without rates) of (4.7) to (4.5) would follow from the fact that, as a consequence of the consistent estimation of the GDFMs for levels and volatilities, for any n_0 and T_0, $\max_{1\leq i\leq n_0}\max_{1\leq t\leq T_0} |\widehat{w}_{it} - w_{it}|$ converges to zero as n and T tend to infinity.

Then, the difference between the empirical quantile of order α computed from $\{\widehat{w}_{1t}, \ldots, \widehat{w}_{iT_0}\}$ and the empirical quantile of order α computed from the unobservable $\{w_{i1}, \ldots, w_{iT_0}\}$ is $o_{\mathrm{P}}(1)$ for given $1 \leq i \leq n_0$ as n and T tend to infinity. Now, for given i, were the w_{it}'s i.i.d., the empirical α-quantile computed from $\{w_{i1}, \ldots, w_{iT_0}\}$ is, for T_0 large enough, arbitrarily close to its theoretical counterpart $q(\alpha; w_i)$ with probability arbitrarily close to one. The same conclusion extends to the present case where the w_{it}'s are stationary and uncorrelated provided that they satisfy some additional mild ergodicity or mixing assumption. The literature on Glivenko-Cantelli and quantile consistency under ergodicity and mixing is abundant, and we will not proceed with imposing any specific mixing conditions here which anyway hardly can be checked from the data. The reader may like to refer to Theorem 3.1 in Francq and Zakoian (2019) for details.

Once prediction regions have been constructed, it is important to evaluate their actual coverage performance. For this, it is useful to define the conditional coverage indicators—namely, for prediction intervals $\widehat{\mathcal{I}}_{i,T+1|T}(\alpha)$,

$$\widehat{\mathcal{H}}_{i,T+1|T}(\alpha) := \mathbb{I}\big(Y_{i,T+1} \in \widehat{\mathcal{I}}_{i,T+1|T}(\alpha)\big). \tag{4.8}$$

For a given i, we say that $\widehat{\mathcal{I}}_{i,T+1|T}(\alpha)$ provides the correct coverage if

$$\mathrm{P}(Y_{i,T+1} \in \widehat{\mathcal{I}}_{i,T+1|T}(\alpha)|Y_{i,T}, \ldots, Y_{i1}) = \mathrm{E}[\widehat{\mathcal{H}}_{i,T+1|T}(\alpha)|Y_{i,T}, \ldots, Y_{i1}] = (1-\alpha),$$

[3]Usually, a Value-at-Risk is reported as a positive quantity. That will be the case with $\mathrm{VaR}_{it}(\alpha)$ for α small enough. Positive values of $\mathcal{L}_{it|t-1}(\alpha)$ are possible, though: in such cases, $\mathrm{VaR}_{it}(\alpha)$ is defined to be zero by convention (see Francq and Zakoian, 2011, Definition 12.1).

Algorithm 3: Estimation of conditional prediction intervals

Input: data in levels $\mathbf{Y}$, $\alpha^- \in [0,1/2]$ and $\alpha^+ \in [0,1/2]$ such that the confidence level is $\alpha = \alpha^+ + \alpha^- \in (0,1)$
from Algorithm 1: common level shocks $\widehat{\mathbf{u}}$ of size $q \times T$ and $\widehat{\mathbf{e}}$ of size $n \times T$, idiosyncratic level shocks $\widehat{\mathbf{v}}$ of size $n \times T$, common level impulse responses $\widehat{\mathbf{B}}(L)$ of size $n \times q \times \bar{k}_1$, idiosyncratic level impulse responses $\widehat{\mathbf{D}}(L)$ of size $n \times n \times \bar{k}_2$
from Algorithm 2: log-volatility proxy $\mathbf{h}$ of size $n \times T$, common log-volatility shocks $\widehat{\boldsymbol{\varepsilon}}$ of size $Q \times T$ and $\widehat{\boldsymbol{\eta}}$ of size $n \times T$, idiosyncratic log-volatility shocks $\widehat{\boldsymbol{\nu}}$ of size $n \times T$, common level impulse responses $\widehat{\mathbf{F}}(L)$ of size $n \times Q \times \bar{k}_1^*$, idiosyncratic level impulse responses $\widehat{\mathbf{G}}(L)$ of size $n \times n \times \bar{k}_2^*$
Output: lower bounds of conditional prediction interval $\widehat{\mathcal{L}}_{1,T+1|T}(\alpha^-), \ldots, \widehat{\mathcal{L}}_{n,T+1|T}(\alpha^-)$
upper bounds of conditional prediction interval $\widehat{\mathcal{U}}_{1,T+1|T}(\alpha^+), \ldots, \widehat{\mathcal{U}}_{n,T+1|T}(\alpha^+)$

```
Compute Ȳ the sample mean of levels Y
Compute the one-step-ahead prediction of common and idiosyncratic components of levels X̂_{T+1|T} = Σ_{k=1}^{k̄_1} B̂_k û_{T-k+1}
Compute the one-step-ahead prediction of idiosyncratic component of levels Ẑ_{T+1|T} = Σ_{k=1}^{k̄_2} D̂_k v̂_{T-k+1}
Compute the one-step-ahead prediction of levels Ŷ_{T+1|T} = X̂_{T+1|T} + Ẑ_{T+1|T} + Ȳ such that
    Ŷ_{T+1|T} = (Ŷ_{1,T+1|T} ... Ŷ_{n,T+1|T})'
Compute h̄̂ the sample mean of log-volatilities ĥ
Compute the one-step-ahead prediction of common component of log-volatilities χ̂_{T+1|T} = Σ_{k=1}^{k̄_1^*} F̂_k ε̂_{T-k+1}
Compute the one-step-ahead prediction of idiosyncratic component of log-volatilities ξ̂_{T+1|T} = Σ_{k=1}^{k̄_2^*} Ĝ_k ν̂_{T-k+1}
Compute the one-step-ahead prediction of log-volatilities ĥ_{T+1|T} = χ̂_{T+1|T} + ξ̂_{T+1|T} + h̄̂
Compute the one-step-ahead prediction of volatilities ŝ_{T+1|T} = exp(ĥ_{T+1|T}/2) such that ŝ_{T+1|T} = (ŝ_{1,T+1|T} ... ŝ_{n,T+1|T})'
Compute the log-volatility innovations ω̂_t = η̂_t + ν̂_t for t = 1, ..., T
Compute the volatility proxy ŝ_t = exp(h_t/2) or equivalently ŝ_t = ê_t + v̂_t for t = 1, ..., T
Compute the volatility innovations ŵ_t = exp(ω̂_t/2)sign(ŝ_t) such that ŵ_t = (ŵ_{1t} ... ŵ_{nt})' for t = 1, ..., T
for i ← 1 to n do
    Compute the order statistics w_{i(⌈Tα^-⌉)} and w_{i(⌈T(1-α^+)⌉)} of w_i
    Compute the lower bound L̂_{i,T+1|T}(α^-) = Ŷ_{i,T+1|T} + ŝ_{i,T+1|T} ŵ_{i(⌈Tα^-⌉)}
    Compute the upper bound Û_{i,T+1|T}(α^+) = Ŷ_{i,T+1|T} + ŝ_{i,T+1|T} ŵ_{i(⌈T(1-α^+)⌉)}
```

which is equivalent (see e.g. Lemma 1 in Christoffersen, 1998) to the hypothesis that

$$\widehat{\mathcal{H}}_{i,T+1|T}(\alpha) \overset{iid}{\sim} \text{Bernoulli}(1-\alpha). \tag{4.9}$$

That hypothesis can be tested against alternatives of insufficient coverage probability values, against non-sharp prediction limits, or against alternatives of serial dependence. We refer to Section 6.3 for details and implementation.

5 Simulation study

5.1 Setup

To study the performance of our estimator on finite samples, we simulate data ($\mathcal{M}$ replications) according to the model described in (4.1)-(4.2).

For each Monte Carlo replication $m = 1, \ldots, \mathcal{M}$ and for given values of n, T, q, and Q, we first simulate a multiplicative factor model for the volatilities which in turn implies a factor structure also for the levels. The common component of the log-volatilities is generated as

$$\boldsymbol{\chi}_{nt,m} := (\mathbf{M}_{n,m}(L))^{-1}\mathbf{R}_{n,m}\boldsymbol{\varepsilon}_{t,m}, \quad t = 1, \ldots, T,$$

where $\boldsymbol{\varepsilon}_{t,m} \overset{iid}{\sim} N(\mathbf{0}_Q, \mathbf{I}_Q)$, $\mathbf{R}_{n,m}$ is $n \times Q$ with entries $[\mathbf{R}_{n,m}]_{ij} \overset{iid}{\sim} N(0,1)$ and rescaled such that $\mathbf{R}_{n,m}'\mathbf{R}_{n,m} = n$, and $\mathbf{M}_{n,m}(L) = \mathbf{I}_n - \sum_{k=1}^{3} \mathbf{M}_{kn,m} L^k$ where the coefficients $\mathbf{M}_{kn,m}$ are diagonal $n \times n$ matrices with entries $[\mathbf{M}_{kn,m}]_{ij} \stackrel{iid}{\sim} N(0,1)$ and rescaled in such a way that $\det(\mathbf{M}_{n,m}(z)) \neq 0$ for $|z| \leq 1$.[4] Then, we generate the process

$$\boldsymbol{\xi}^*_{nt,m} := (\mathbf{P}^*_{n,m}(L))^{-1}\boldsymbol{\nu}^*_{nt,m}, \quad t = 1, \dots, T,$$

where $\boldsymbol{\nu}^*_{nt,m} \stackrel{iid}{\sim} N(\mathbf{0}_n, \boldsymbol{\Sigma}_{n,m})$, with $\boldsymbol{\Sigma}_{n,m}$ a Toeplitz matrix with entries $[\boldsymbol{\Sigma}_{n,m}]_{ij} := 0.5^{|i-j|}$, if $|i-j| \leq 2$ and zero otherwise, and $\mathbf{P}^*_{n,m}(L)$ generated in the same way as $\mathbf{M}_{n,m}(L)$. Denoting by $\xi^*_{it,m}$ the ith element of $\boldsymbol{\xi}^*_{it,m}$, we rescale it into $\xi^{**}_{it,m} := \xi^*_{it,m}[\text{Var}(\chi_{it,m})/\{2\text{Var}(\xi^*_{it,m})\}]^{1/2}$ so that the signal-to-noise ratio is 2.

Define

$$e^*_{it,m} := \exp(\chi_{it,m}/2)\ \pi_{it,m}, \ \text{ and } \ v^*_{it,m} := \exp(\chi_{it,m}/2)\ \exp(\xi^{**}_{it,m}/2)\pi_{it,m}, \quad t = 1, \dots, T,\ i = 1, \dots, n,$$

where $\pi_{it,m} = \pm 1$ with equal probabilities 0.5 and $\chi_{it,m}$ is the ith element of $\boldsymbol{\chi}_{t,m}$. The volatility and log-volatility proxies then are

$$\begin{aligned} s^2_{it,m} &:= (e^*_{it,m} + v^*_{it,m})^2 = \exp(\chi_{it,m}) \left[1 + \exp(\xi^{**}_{it,m}) + 2\exp(\xi^{**}_{it,m}/2)\right], \\ h_{it,m} &:= \log(s^2_{it,m}) = \chi_{it,m} + \log\left[1 + \exp(\xi^{**}_{it,m}) + 2\exp(\xi^{**}_{it,m}/2)\right], \ t = 1, \dots, T,\ i = 1, \dots, n, \end{aligned}$$

from which we see that, since each $\chi_{i,m}$ is driven by the Q-dimensional vector of shocks $\boldsymbol{\varepsilon}_m$, it has the role of common log-volatility, while the n shocks $\boldsymbol{\nu}^*_{n,m}$ have only an idiosyncratic role.

Letting $\mathbf{V}$ be the q normalized eigenvectors corresponding to the q largest eigenvalues of the sample covariance of the vector $\mathbf{e}^*_{nt,m} := (e^*_{1t,m} \cdots e^*_{nt,m})'$, we build the level shocks as

$$\mathbf{e}_{nt,m} := \mathbf{V}\mathbf{V}'\mathbf{e}^*_{nt,m}, \ \text{ and } \ \mathbf{v}_{nt,m} := \mathbf{V}_\perp\mathbf{V}_\perp'\mathbf{e}^*_{nt,m} + \mathbf{v}^*_{nt,m}, \quad t = 1, \dots, T,$$

where $\mathbf{V}_\perp$ is $n \times (n-q)$ such that $\mathbf{V}_\perp'\mathbf{V} = \mathbf{0}_{(n-q)\times q}$, and $\mathbf{v}^*_{nt,m} := (v^*_{1t,m} \cdots v^*_{nt,m})'$. Note that, by construction, the elements $e_{it,m}$ and $v_{it,m}$ of the vectors $\mathbf{e}_{nt,m}$ and $\mathbf{v}_{nt,m}$ are such that $(e_{it,m}+v_{it,m}) = (e^*_{it,m}+v^*_{it,m})$: therefore, we can also write $s^2_{it} = (e_{it,m} + v_{it,m})^2$.

Finally, we generate the vectors of common and idiosyncratic components of the levels as

$$\mathbf{X}_{nt,m} := (\mathbf{I}_n - \mathbf{A}_{n,m}L)^{-1}\mathbf{e}_{nt,m}, \ \text{ and } \ \mathbf{Z}_{nt,m} := (\mathbf{I}_n - \mathbf{C}_{n,m}L)^{-1}\mathbf{v}_{nt,m}, \quad t = 1, \dots, T$$

where $\mathbf{A}_{n,m}$ is a diagonal $n \times n$ matrix with entries $[\mathbf{A}_{n,m}]_{ij} \stackrel{iid}{\sim} U[-0.3, 0.7]$, and $\mathbf{C}_{n,m}$ is generated in the same way but with entries from a uniform distribution over $[\mathbf{C}_{n,m}]_{ij} \stackrel{iid}{\sim} U[-0.5, 0.5]$; since these matrices are diagonal, the autoregressive models for $\mathbf{X}_{n,m}$ and $\mathbf{Z}_{n,m}$ are causal. The panel of levels then is generated as $\mathbf{Y}_{nt,m} := \mathbf{X}_{nt,m} + \mathbf{Z}_{nt,m}$.

In our numerical study, we let $n \in \{100, 200\}$, $T \in \{200, 500, 1000\}$, and either $q = 1$ and $Q = 1$, $q = 3$ and $Q = 2$ (as in the empirical application of the next section), or $q = 2$ and $Q = 3$. For each configuration considered, we simulate and estimate the model $\mathcal{M} = 200$ times.

It has to be noticed that the data-generating process we are considering is similar to a stochastic volatility model. To illustrate the properties of the generated data, we report in Table 1 the autocorrelations up to lag 10 of $h_{i,m}$, $s_{i,m}$, $e_{i,m}$, $v_{i,m}$, $X_{i,m}$, $X^2_{i,m}$, $Z_{i,m}$, $Z^2_{i,m}$, $Y_{i,m}$, and $Y^2_{i,m}$, averaged over all $\mathcal{M}$ replications and over all n series, and when $n = 200$, $T = 1000$. It can be seen that log-volatilities $h_{i,m}$ and volatilities $s_{i,m}$ have high persistence, while, due to the way they are generated, the shocks $e_{i,m}$ and $v_{i,m}$ display no linear serial dependence, i.e. are weak white noises. Turning to the kurtosis of the level shocks reported in the left panel of Table 2, these display heavy tails (especially the common ones) for the case $q = 1$ and $Q = 1$, while the kurtosis tends to decrease when increasing Q, possibly due to the aggregation of shocks in generating the common components of the log-volatility $\chi_{i,m}$. Similar comments apply to the absolute values of skewness reported in the right panel of Table 2: especially in the case $q = 1$ and $Q = 1$, the common shocks display a high degree of asymmetry. Because of these features of the simulated data the case $q = 1$ and $Q = 1$ is particularly interesting to study to assess the performance of our estimators when dealing with heavy-tailed and skewed data.

[4] In particular, when looking at simulated data 25% of the total $3n^2$ roots are found to be in the range $(0.7, 1)$, thus accounting for high persistence in log-volatilities, see also Table 1 below.

TABLE 1: *Autocorrelations of simulated variables. Average values over all n series and all $\mathcal{M}$ replications for $n = 200$, $T = 1000$, and $\mathcal{M} = 200$. Values outside the $[\pm 1.96/\sqrt{T}] = [\pm 0.0620]$ interval are starred.*

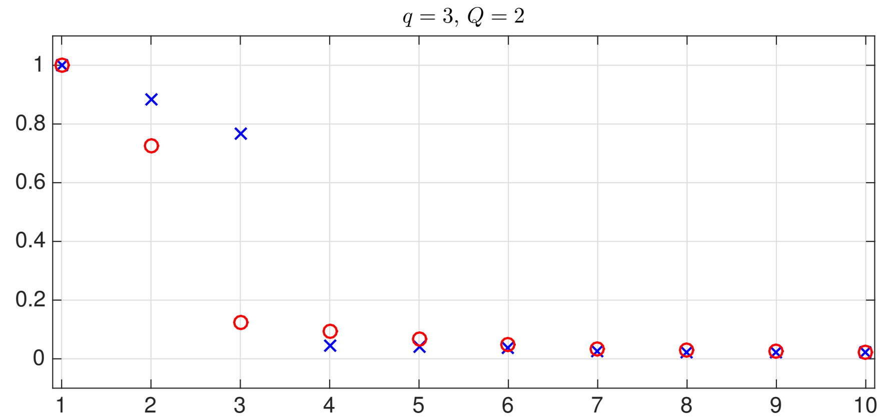
$q=1$, $Q=1$	lag 1	2	3	4	5	6
$h_{i,m}$	0.2967*	0.2856*	0.1178*	0.1691*	0.0528	0.1142*
$s_{i,m}$	0.3082*	0.2743*	0.1201*	0.1426*	0.0552	0.0758*
$e_{i,m}$	-0.0378	0.0696*	-0.0565	0.0089	-0.0099	-0.0283
$e^2_{i,m}$	0.0326	0.0495	0.0020	0.0035	0.0031	0.0042
$v_{i,m}$	-0.0023	-0.0017	-0.0035	0.0005	-0.0015	-0.0028
$v^2_{i,m}$	0.2961*	0.2674*	0.1196*	0.1445*	0.0565	0.0771*
$X_{i,m}$	0.1753*	0.1770*	0.0096	0.0302	-0.0032	-0.0262
$X^2_{i,m}$	0.0989*	0.0791*	0.0068	0.0023	0.0016	-0.0037
$Z_{i,m}$	0.0020	0.0783*	-0.0033	0.0118	-0.0018	-0.0002
$Z^2_{i,m}$	0.2877*	0.2260*	0.1054*	0.1200*	0.0522	0.0622*
$Y_{i,m}$	0.1174*	0.1439*	0.0051	0.0219	-0.0031	-0.0172
$Y^2_{i,m}$	0.0994*	0.0798*	0.0073	0.0023	0.0018	0.0029
$q=3$, $Q=2$	lag 1	2	3	4	5	6
$h_{i,m}$	0.2654*	0.2826*	0.1183*	0.1611*	0.0508	0.1272*
$s_{i,m}$	0.2757*	0.2607*	0.1237*	0.1305*	0.0511	0.0913*
$e_{i,m}$	-0.0089	-0.0690*	-0.0028	0.0304	-0.0353	-0.0071
$e^2_{i,m}$	0.1939*	0.0721*	0.0034	0.0166	0.0077	0.0113
$v_{i,m}$	-0.0010	-0.0011	-0.0030	0.0002	-0.0038	-0.0030
$v^2_{i,m}$	0.2635*	0.2515*	0.1212*	0.1283*	0.0503	0.0907*
$X_{i,m}$	0.1977*	0.0592	0.0465	0.0492	-0.0142	0.0002
$X^2_{i,m}$	0.2545*	0.0822*	0.0161	0.0222	0.0073	0.0101
$Z_{i,m}$	-0.0172	0.0814*	-0.0052	0.0119	-0.0045	-0.0013
$Z^2_{i,m}$	0.2708*	0.2133*	0.1076*	0.1059*	0.0502	0.0764*
$Y_{i,m}$	0.1227*	0.0671*	0.0285	0.0374	-0.0123	-0.0003
$Y^2_{i,m}$	0.2166*	0.0909*	0.0237	0.0242	0.0079	0.0009
$q=2$, $Q=3$	lag 1	2	3	4	5	6
$h_{i,m}$	0.2730*	0.2692*	0.1254*	0.1717*	0.0494	0.1293*
$s_{i,m}$	0.2375*	0.2348*	0.1036*	0.1481*	0.0238	0.0885*
$e_{i,m}$	-0.0221	0.0127	-0.0204	-0.0001	-0.0071	-0.0067
$e^2_{i,m}$	0.0025	0.0640*	0.0222	0.0616	0.0001	0.0134
$v_{i,m}$	-0.0026	-0.0002	0.0000	0.0026	-0.0011	-0.0035
$v^2_{i,m}$	0.2330*	0.2313*	0.1015*	0.1456*	0.0229	0.0879*
$X_{i,m}$	0.1835*	0.1254*	0.0364	0.0277	0.0079	0.0008
$X^2_{i,m}$	0.1159*	0.1020*	0.0388	0.0377	0.0131	0.0057
$Z_{i,m}$	0.0317	0.0872*	0.0032	0.0156	-0.0008	-0.0007
$Z^2_{i,m}$	0.2509*	0.1918*	0.0969*	0.1227*	0.0293	0.0711*
$Y_{i,m}$	0.1313*	0.1121*	0.0253	0.0220	0.0040	0.0012
$Y^2_{i,m}$	0.1047*	0.0869*	0.0426	0.0425	0.0043	0.0133

TABLE 2: *Kurtosis and absolute value of skewness of simulated common level shocks $e_{i,m}$ and idiosyncratic level shocks $v_{i,m}$. Maximum and average values over all n series and all $\mathcal{M}$ replications for $n = 200$, $T = 1000$, and $\mathcal{M} = 200$.*

	kurtosis						skewness					
	$q=1, Q=1$		$q=3, Q=2$		$q=2, Q=3$		$q=1, Q=1$		$q=3, Q=2$		$q=2, Q=3$	
	max.	aver.	max.	aver.	max.	aver.	max.	aver.	max.	aver.	max.	aver.
$e_{i,m}$	161.40	83.53	67.60	10.94	31.86	5.60	7.88	0.28	4.08	0.03	2.53	0.02
$v_{i,m}$	15.03	3.02	12.68	3.02	9.36	3.01	1.14	0.01	1.07	0.01	0.85	0.01

Furthermore, notice that $\mathbf{e}_{n,m}$, by construction, is a singular vector (as it should be) and has the role of a common level innovation. Moreover, the elements of $\mathbf{v}_{n,m}$, in general, are cross-sectionally dependent. As a consequence, both $\mathbf{Y}_{n,m}$ and $\mathbf{h}_{n,m}$ have an approximate dynamic factor structure. In Figure 2 we show scree-plots with the ten largest eigenvalues of the zero-frequency sample spectral density matrices of $\mathbf{Y}_{n,m}$ (blue crosses),

FIGURE 2: *Normalized eigenvalues of the zero-frequency spectral densities of simulated data for $n = 200$ and $T = 1000$. Blue crosses: levels, $\mathbf{Y}_{n,m}$; red circles: log-volatilities, $\mathbf{h}_{n,m}$.*



and $\mathbf{h}_{n,m}$ (red circles), normalized by the largest zero-frequency eigenvalue, averaged over all $\mathcal{M}$ realisations, when $n = 200$, $T = 1000$, $q = 3$, and $Q = 2$.

5.2 Results

For each replication, we estimate the model as described in Section 3. The capping constants κ_T and the bandwidths B_T and M_T involved in the estimation of the spectral density are chosen as in the empirical analysis of the next section. Specifically, we let $\kappa_T \in \{0, 0.2, 0.4\}$, while the bandwidths values are $B_T = 2$ and $M_T = 10$ for $T = 200$, $B_T = 2$ and $M_T = 15$ for $T = 500$, $B_T = 2$ and $M_T = 20$ for $T = 1000$ (see Appendix D for results based on other values). Once we obtain estimated common components $\widehat{X}_{i,m}$ for the levels and $\widehat{\chi}_{i,m}$ for the log-volatilities, we compute the global error measures

$$MSE^X = \frac{1}{\mathcal{M}nT}\sum_{m=1}^{\mathcal{M}}\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it,m} - \widehat{X}_{it,m})^2, \quad MSE^\chi = \frac{1}{\mathcal{M}nT}\sum_{m=1}^{\mathcal{M}}\sum_{i=1}^{n}\sum_{t=1}^{T}(\chi_{it,m} - \widehat{\chi}_{it,m})^2,$$
$$MAD^X = \frac{1}{\mathcal{M}nT}\sum_{m=1}^{\mathcal{M}}\sum_{i=1}^{n}\sum_{t=1}^{T}|X_{it,m} - \widehat{X}_{it,m}|, \quad MAD^\chi = \frac{1}{\mathcal{M}nT}\sum_{m=1}^{\mathcal{M}}\sum_{i=1}^{n}\sum_{t=1}^{T}|\chi_{it,m} - \widehat{\chi}_{it,m}|$$

and the maximal errors over all realizations:

$$MAX^X = \max_{i=1,\dots,n}\max_{t=1,\dots,T}\max_{m=1,\dots,\mathcal{M}}|X_{it,m} - \widehat{X}_{it,m}|,$$
$$MAX^\chi = \max_{i=1,\dots,n}\max_{t=1,\dots,T}\max_{m=1,\dots,\mathcal{M}}|\chi_{it,m} - \widehat{\chi}_{it,m}|.$$

Notice that the error in the estimation of the common component $X_{i,m}$ of the levels (first step of the estimation procedure) has already been studied in Forni et al. (2017) and Forni et al. (2018). We therefore consider it as the benchmark error with respect to which the performance of the second estimation step, which is the novelty of this paper, is to be compared. Results are provided in Table 3. We note that MSE and MAD in the second step tend to be about 1.5 times higher than in the first step, which is not unexpected as first- and second- step errors typically cumulate in a two-stage procedure. However, when turning to MAX, this is no longer the case, since levels in our data-generating process display heavier tails than log-volatilities—in line with the typical behavior of daily stock returns and their volatilities. Increasing n and T improves the performance of all estimators; the role of n, in that respect, seems to be the main one—a manifestation of the “blessing of dimensionality". On the other hand increasing Q the number of common log-volatility shocks, tends to make estimation of the second step harder, but still results are in line with the case $Q = 1$. Capping has an effect in controlling the maximum error but does not affect the MSE and MAD results much. To illustrate the good performances of our method,

TABLE 3: *Simulation results. MSEs and MADs for common components. Bandwidths are* $B_T = 2$ *and* $M_T = 10$ *for* $T = 200$*;* $B_T = 2$ *and* $M_T = 15$ *for* $T = 500$*;* $B_T = 2$ *and* $M_T = 20$ *for* $T = 1000$.

$q = 1, Q = 1$

		$T = 200$		$T = 500$		$T = 1000$	
		$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
MSE^X		0.215	0.219	0.168	0.164	0.125	0.154
MSE^{χ}	$\kappa_T = 0$	0.368	0.321	0.302	0.247	0.251	0.241
MSE^{χ}	$\kappa_T = 0.2$	0.369	0.324	0.276	0.247	0.240	0.230
MSE^{χ}	$\kappa_T = 0.4$	0.377	0.334	0.279	0.238	0.238	0.230
MAD^X		0.278	0.245	0.255	0.226	0.234	0.228
MAD^{χ}	$\kappa_T = 0$	0.442	0.401	0.394	0.346	0.360	0.342
MAD^{χ}	$\kappa_T = 0.2$	0.436	0.395	0.367	0.346	0.344	0.323
MAD^{χ}	$\kappa_T = 0.4$	0.437	0.397	0.364	0.324	0.337	0.317
MAX^X		10.797	13.560	17.113	16.352	14.405	19.757
MAX^{χ}	$\kappa_T = 0$	6.587	6.709	8.996	6.270	7.462	7.912
MAX^{χ}	$\kappa_T = 0.2$	8.259	7.247	7.967	6.270	9.328	8.431
MAX^{χ}	$\kappa_T = 0.4$	8.987	8.402	8.295	10.223	9.821	8.960

$q = 3, Q = 2$

		$T = 200$		$T = 500$		$T = 1000$	
		$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
MSE^X		0.143	0.155	0.101	0.112	0.086	0.085
MSE^{χ}	$\kappa_T = 0$	0.291	0.284	0.237	0.216	0.209	0.185
MSE^{χ}	$\kappa_T = 0.2$	0.262	0.261	0.197	0.199	0.179	0.163
MSE^{χ}	$\kappa_T = 0.4$	0.250	0.252	0.182	0.176	0.162	0.141
MAD^X		0.265	0.261	0.227	0.228	0.210	0.205
MAD^{χ}	$\kappa_T = 0$	0.412	0.402	0.370	0.348	0.346	0.322
MAD^{χ}	$\kappa_T = 0.2$	0.389	0.381	0.336	0.327	0.317	0.299
MAD^{χ}	$\kappa_T = 0.4$	0.378	0.372	0.321	0.307	0.300	0.275
MAX^X		6.634	8.693	9.682	12.606	10.943	19.277
MAX^{χ}	$\kappa_T = 0$	5.532	5.624	5.128	6.725	4.488	5.209
MAX^{χ}	$\kappa_T = 0.2$	5.688	5.800	5.561	6.204	4.813	4.675
MAX^{χ}	$\kappa_T = 0.4$	5.713	6.180	5.630	7.317	4.978	5.235

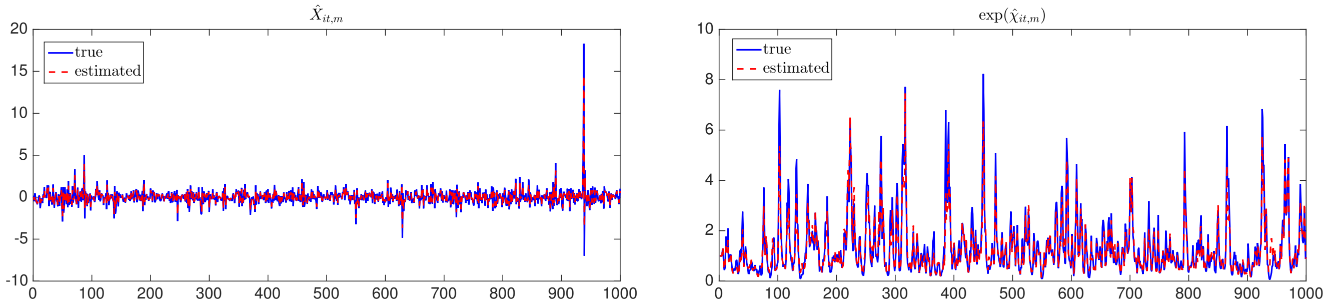
$q = 2, Q = 3$

		$T = 200$		$T = 500$		$T = 1000$	
		$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
MSE^X		0.151	0.161	0.112	0.129	0.082	0.085
MSE^{χ}	$\kappa_T = 0$	0.356	0.323	0.290	0.277	0.247	0.224
MSE^{χ}	$\kappa_T = 0.2$	0.324	0.299	0.259	0.247	0.210	0.190
MSE^{χ}	$\kappa_T = 0.4$	0.311	0.287	0.248	0.231	0.193	0.173
MAD^X		0.279	0.271	0.241	0.248	0.212	0.213
MAD^{χ}	$\kappa_T = 0$	0.461	0.434	0.414	0.398	0.382	0.360
MAD^{χ}	$\kappa_T = 0.2$	0.437	0.415	0.387	0.372	0.350	0.327
MAD^{χ}	$\kappa_T = 0.4$	0.426	0.404	0.376	0.356	0.334	0.311
MAX^X		5.395	9.180	5.900	8.936	6.210	11.411
MAX^{χ}	$\kappa_T = 0$	4.833	5.066	5.009	5.654	5.111	5.571
MAX^{χ}	$\kappa_T = 0.2$	4.988	5.325	5.269	5.998	5.355	5.654
MAX^{χ}	$\kappa_T = 0.4$	5.058	5.748	5.660	6.208	5.599	5.413

in Figure 3 we show, for one replication, the estimated (in red) and simulated (in blue) common components of levels, and of volatilities, respectively, for $n = 200$, $T = 1000$, $q = 1$, and $Q = 1$ (which is the case exhibiting the heaviest tails), setting $\kappa_T = 0.2$. The choice of bandwidths adopted seems to work quite well, and, comparing to alternative choices considered in Appendix D, it can be shown that M_T must be large enough to capture the persistence in log-volatilities, while lower values of B_T are enough for levels and do not affect much the second step of estimation.

Finally, for $T = 1000$, we estimated the model using the first 900 observations, then ran a recursive pseudo-out-of-sample forecasting exercise constructing one-step-ahead prediction intervals for the remaining 100 obser-

FIGURE 3: *Simulation results. True (blue) and estimated (red) common components of levels,* $\widehat{X}_{it,m}$*, and of volatilties,* $\exp(\widehat{\chi}_{it,m})$*, when* $n = 200$*,* $T = 1000$*,* $q = 1$*,* $Q = 1$*, and* $\kappa_T = 0.2$*. One series and one realisation.*



vations (from 901 to 1000), as described in Section 4. The $\alpha/2$-upper and $\alpha/2$-lower bounds $\widehat{\mathcal{U}}_{i,\tau+1|\tau,m}(\alpha/2)$ and $\widehat{\mathcal{L}}_{i,\tau+1|\tau,m}(\alpha/2)$ of prediction intervals with coverage probability $(1-\alpha)$ are then computed for each series and replication and each out-of-sample observation. From the latter, we compute the observed coverage frequencies across all series and replications

$$C(\alpha) := \frac{1}{\mathcal{M}n100} \sum_{m=1}^{\mathcal{M}} \sum_{i=1}^{n} \sum_{\tau=900}^{999} \mathbb{I}\Big(\widehat{\mathcal{L}}_{i,\tau+1|\tau,m}(\alpha/2) \leq Y_{i,\tau+1,m} \leq \widehat{\mathcal{U}}_{i,\tau+1|\tau,m}(\alpha/2)\Big)$$

and the proportions of coverage violations in the upper and lower tails,

$$V_+(\alpha/2) := \frac{1}{\mathcal{M}n100} \sum_{m=1}^{\mathcal{M}} \sum_{i=1}^{n} \sum_{\tau=900}^{999} \mathbb{I}\Big(Y_{i,\tau+1,m} > \widehat{\mathcal{U}}_{i,\tau+1|\tau,m}(\alpha/2)\Big),$$

and

$$V_-(\alpha/2) := \frac{1}{\mathcal{M}n100} \sum_{m=1}^{\mathcal{M}} \sum_{i=1}^{n} \sum_{\tau=900}^{999} \mathbb{I}\Big(Y_{i,\tau+1,m} < \widehat{\mathcal{L}}_{i,\tau+1|\tau,m}(\alpha/2)\Big),$$

respectively. Results are shown in Table 4. Overall performances look reasonably good—the larger n and T, the better. We note that capping has a clear effect on the empirical coverage; too much capping seems to affect mostly the cases in which $\alpha = 0.32$ and 0.2. No capping at all works quite well in practice, despite the fact that theoretical results require $\kappa_T > 0$. Moreover, the same comments apply to empirical coverage as for the choice of bandwidths, with the additional finding that higher values of B_T yield more reliable prediction performances (see Appendix D).

6 Interval prediction for S&P100 returns

In this section, we apply our methodology to a panel of $n = 90$ daily returns of stocks from the Standard & Poor's 100 Index. Data are observed from January 4, 2000 through September 30, 2013, for a total of $T = 3456$ observations. We run a pseudo-out-of-sample forecasting exercise by estimating the model using data over the period $t = 1, \ldots, \tau$, with $\tau = (T-M), \ldots, (T-1)$ and $M = 1948$, corresponding to an evaluation period running from January 3, 2006 through September 27, 2013. For each value of τ, we estimate the $n = 90$ one-step-ahead prediction intervals as defined in (4.7). The data cover the following sectors (in parentheses, the number of series in each sector): Consumer Discretionary (11), Consumer Staples (10), Energy (12), Financials (13), Health Care (11), Industrials (14), Information Technology (12), Materials (3), Telecommunications Services (2), Utilities (2) (see Appendix C for the names of individual stocks).

Although we should, in principle, fully re-estimate the whole model at each of the M iterations, some quantities were kept fixed throughout the exercise. In particular, when applied to the full $n \times T$ panel, the Hallin and

TABLE 4: *Simulation results. Empirical coverage and frequencies of prediction bounds violations, averaged over all n series and all $\mathcal{M}$ replications, for $T = 1000$ and $\mathcal{M} = 200$. Bandwidths values: $B_T = 2$ and $M_T = 10$ for $T = 200$; $B_T = 2$ and $M_T = 15$ for $T = 500$; $B_T = 2$ and $M_T = 20$ for $T = 1000$.*

$q = 1, Q = 1$

		$n = 100$					$n = 200$				
		α					α				
		0.32	0.2	0.1	0.05	0.01	0.32	0.2	0.1	0.05	0.01
$C(\alpha)$	$\kappa_T = 0$	0.6409	0.7637	0.8667	0.9312	0.9869	0.6765	0.7992	0.9082	0.9573	0.9926
$V_+(\alpha/2)$		0.1810	0.1195	0.0674	0.0342	0.0057	0.1635	0.1026	0.0470	0.0221	0.0040
$V_-(\alpha/2)$		0.1781	0.1168	0.0659	0.0346	0.0074	0.1601	0.0983	0.0449	0.0206	0.0035
$C(\alpha)$	$\kappa_T = 0.2$	0.6691	0.7769	0.8685	0.9197	0.9681	0.7201	0.8285	0.9226	0.9628	0.9934
$V_+(\alpha/2)$		0.1636	0.1124	0.0682	0.0403	0.0153	0.1422	0.0876	0.0392	0.0194	0.0034
$V_-(\alpha/2)$		0.1673	0.1107	0.0633	0.0400	0.0166	0.1378	0.0840	0.0383	0.0179	0.0033
$C(\alpha)$	$\kappa_T = 0.4$	0.7072	0.7987	0.8799	0.9238	0.9699	0.7119	0.7957	0.8763	0.9257	0.9703
$V_+(\alpha/2)$		0.1453	0.1023	0.0617	0.0384	0.0145	0.1429	0.1007	0.0600	0.0360	0.0150
$V_-(\alpha/2)$		0.1475	0.0990	0.0584	0.0378	0.0156	0.1453	0.1037	0.0638	0.0383	0.0147

$q = 3, Q = 2$

		$n = 100$					$n = 200$				
		α					α				
		0.32	0.2	0.1	0.05	0.01	0.32	0.2	0.1	0.05	0.01
$C(\alpha)$	$\kappa_T = 0$	0.6350	0.7523	0.8543	0.9131	0.9688	0.6718	0.7917	0.8929	0.9457	0.9895
$V_+(\alpha/2)$		0.1810	0.1252	0.0749	0.0449	0.0157	0.1619	0.1037	0.0530	0.0277	0.0053
$V_-(\alpha/2)$		0.1840	0.1225	0.0708	0.0420	0.0155	0.1664	0.1047	0.0542	0.0266	0.0053
$C(\alpha)$	$\kappa_T = 0.2$	0.6767	0.7775	0.8665	0.9206	0.9701	0.7031	0.8081	0.8986	0.9469	0.9916
$V_+(\alpha/2)$		0.1601	0.1145	0.0691	0.0422	0.0162	0.1458	0.0935	0.0503	0.0250	0.0039
$V_-(\alpha/2)$		0.1632	0.1080	0.0644	0.0372	0.0137	0.1512	0.0985	0.0512	0.0282	0.0046
$C(\alpha)$	$\kappa_T = 0.4$	0.7129	0.7993	0.8803	0.9267	0.9724	0.7565	0.8447	0.9222	0.9610	0.9923
$V_+(\alpha/2)$		0.1445	0.1033	0.0614	0.0384	0.0147	0.1209	0.0780	0.0382	0.0196	0.0043
$V_-(\alpha/2)$		0.1426	0.09740	0.0583	0.0349	0.0129	0.1227	0.0774	0.0397	0.0195	0.0035

$q = 2, Q = 3$

		$n = 100$					$n = 200$				
		α					α				
		0.32	0.2	0.1	0.05	0.01	0.32	0.2	0.1	0.05	0.01
$C(\alpha)$	$\kappa_T = 0$	0.6888	0.8045	0.8981	0.9500	0.9874	0.6391	0.7563	0.8623	0.9215	0.9784
$V_+(\alpha/2)$		0.1568	0.0995	0.0527	0.0258	0.0065	0.1786	0.1212	0.0678	0.0387	0.0107
$V_-(\alpha/2)$		0.1544	0.0960	0.0492	0.0242	0.0061	0.1824	0.1226	0.0700	0.0399	0.0110
$C(\alpha)$	$\kappa_T = 0.2$	0.7335	0.8290	0.9105	0.9539	0.9890	0.6770	0.7814	0.8752	0.9277	0.9791
$V_+(\alpha/2)$		0.1345	0.0863	0.0459	0.0239	0.0056	0.1602	0.1077	0.0613	0.0355	0.0103
$V_-(\alpha/2)$		0.1320	0.0847	0.0436	0.0222	0.0054	0.1629	0.1110	0.0636	0.0368	0.0106
$C(\alpha)$	$\kappa_T = 0.4$	0.7733	0.8539	0.9213	0.9595	0.9898	0.7167	0.8066	0.8879	0.9352	0.9809
$V_+(\alpha/2)$		0.1154	0.0747	0.0408	0.0207	0.0055	0.1395	0.0953	0.0551	0.0326	0.0093
$V_-(\alpha/2)$		0.1113	0.0714	0.0379	0.0198	0.0047	0.1439	0.0981	0.0570	0.0322	0.0099

Liška (2007) criterion returns $\widehat{q} = 3$ common factors for the level panel and $\widehat{Q} = 2$ common factors for log-volatility panel: those values are used in all subsequent analyzes. We also choose the bandwidths by minimizing, over a grid of possible bandwidth values, the mean-squared errors

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(Y_{it} - \widehat{X}_{it|t-1})^2 \quad \text{and} \quad \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(\widehat{h}_{it} - \widehat{\chi}_{it|t-1})^2,$$

respectively, leading to possibly distinct bandwidths for $\widehat{X}_{it|t-1}$ and $\widehat{\chi}_{it|t-1}$. More precisely, we first determine B_T and then determine M_T using the chosen B_T to compute $\widehat{h}_{it}$. As a result we throughout use $B_T = 2$ and $M_T = 17$. The VAR orders and the orders of their truncated inverse MA representations needed to compute impulse responses are set as follows: (i) deg$[\mathbf{A}_n(L)] = 1$, with inverse MA truncated at lag $\bar{k}_1 = 20$; (ii) deg$[\mathbf{C}_n(L)] = 1$, with inverse MA truncated at lag $\bar{k}_2 = 20$; (iii) deg$[\mathbf{M}_n(L)] = 5$, with inverse MA truncated at lag $\bar{k}_1^* = 100$;

(iv) $\deg[\mathbf{P}_n(L)] = 1$, with inverse MA truncated at lag $\bar{k}_2^* = 100$. The estimation of the GDFM is based on 10 cross-sectional permutations, as explained at the end of Section 3.1. Finally, regarding the choice of the capping constant κ_T, we choose $\kappa_T \in \{0,\, 0.1,\, 0.25,\, 0.5\}$ irrespective of i; note that, with reference to Assumption (R), we have $\log^{-1} T = 0.12$. Also note that, on the average across the M iterations, 6%, out of the total $n\tau$ observations, are capped when $\kappa_T = 0.1$, 14% when $\kappa_T = 0.25$, and 27% when $\kappa_T = 0.5$.

For any given sample size τ, we compute the quantiles of $\widehat{w}_i$ using $(\widehat{w}_{i,\tau-\ell+1}, \ldots, \widehat{w}_{i,\tau})$, where we set $\ell \in \{126, 252, 504, \tau\}$, hence using either the past six months, one year, or two years of available data, or using all available past observations. Denoting by $\widehat{\mathbf{w}}_i^{(\ell)}$ the vector of the most recent ℓ observations (so that $\widehat{\mathbf{w}}_i^{(\tau)}$ coincides with $\widehat{\mathbf{w}}_i$), for levels $\alpha \in \{0.32, 0.2, 0.1, 0.05, 0.01\}$ and window sizes ℓ, and for $\tau = (T-M), \ldots, (T-1)$, we obtain the estimates

$$\widehat{\mathcal{U}}^{(\ell)}_{i,\tau+1|\tau}(\alpha) := \widehat{Y}_{i,\tau+1|\tau} + \widehat{s}_{i,\tau+1|\tau}\, \widehat{w}^{(\ell)}_{i(\lceil \ell(1-\alpha)\rceil)}, \qquad \widehat{\mathcal{L}}^{(\ell)}_{i,\tau+1|\tau}(\alpha) := \widehat{Y}_{i,\tau+1|\tau} + \widehat{s}_{i,\tau+1|\tau}\, \widehat{w}^{(\ell)}_{i(\lceil \ell\alpha\rceil)},$$

$$\widehat{\mathcal{I}}^{(\ell)}_{i,\tau+1|\tau}(\alpha) := \big[\widehat{\mathcal{L}}^{(\ell)}_{i,\tau+1|\tau}(\alpha^-), \widehat{\mathcal{U}}^{(\ell)}_{i,\tau+1|\tau}(\alpha^+)\big], \quad \text{and} \quad \widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha) := \mathbb{I}\big(Y_{i,\tau+1} \in \widehat{\mathcal{I}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)\big).$$

6.1 Coverage performance: qualitative analysis

For each of the $n = 90$ series considered we compute the coverage frequency

$$C_i^{(\ell)}(\alpha) := \frac{1}{M} \sum_{\tau=T-M}^{T-1} \widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha) = \frac{1}{M} \sum_{\tau=T-M}^{T-1} \mathbb{I}\Big(\widehat{\mathcal{L}}^{(\ell)}_{i,\tau+1|\tau}(\alpha^-) \le Y_{i,\tau+1} \le \widehat{\mathcal{U}}^{(\ell)}_{i,\tau+1|\tau}(\alpha^+)\Big),$$

the proportions

$$V_{i,+}^{(\ell)}(\alpha^+) := \frac{1}{M} \sum_{\tau=T-M}^{T-1} \mathbb{I}\Big(Y_{i,\tau+1} > \widehat{\mathcal{U}}^{(\ell)}_{i,\tau+1|\tau}(\alpha^+)\Big) \text{ and } V_{i,-}^{(\ell)}(\alpha^-) := \frac{1}{M} \sum_{\tau=T-M}^{T-1} \mathbb{I}\Big(Y_{i,\tau+1} < \widehat{\mathcal{L}}^{(\ell)}_{i,\tau+1|\tau}(\alpha^-)\Big)$$

of coverage violations in the upper and lower tails, and the average interval length

$$L_i^{(\ell)}(\alpha) := \frac{1}{M} \sum_{\tau=T-M}^{T-1} \Big(\widehat{\mathcal{U}}^{(\ell)}_{i,\tau+1|\tau}(\alpha) - \widehat{\mathcal{L}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)\Big).$$

Table 5 reports, for $\alpha^+ = \alpha^- = \alpha/2$ with $\alpha \in \{0.32, 0.2, 0.1, 0.05, 0.01\}$ (corresponding to coverage levels 68%, 80%, 90%, 95% and 99%) and $\kappa_T \in \{0, 0.1, 0.25, 0.5\}$, the cross-sectional average $C^{(\ell)}(\alpha)$ of the empirical coverage frequencies $C_i^{(\ell)}(\alpha)$, the cross-sectional averages $V_+^{(\ell)}(\alpha/2)$ and $V_-^{(\ell)}(\alpha/2)$ of the proportions of coverage violations $V_{i,+}^{(\ell)}(\alpha^+)$ and $V_{i,-}^{(\ell)}(\alpha^-)$, and the cross-sectional average $L^{(\ell)}(\alpha)$ of the average interval lengths $L_i^{(\ell)}(\alpha)$.

Inspection of the table reveals that $C^{(\ell)}(\alpha) \simeq (1-\alpha)$ and $V_+^{(\ell)}(\alpha/2) \simeq V_-^{(\ell)}(\alpha/2) \simeq \alpha/2$, which is a qualitative confirmation of the validity of our methodology (see Section 6.3 for more formal validation). Three remarks emerge from these results. First, regarding the sensitivity of our procedure to capping, lower values of κ_T, in general, provide better results when α is higher, while larger values of κ_T provide better results for lower values of α; in all cases, $\kappa_T = 0.5$ yields a mostly conservative coverage frequency higher than $(1-\alpha)$. In particular, note that the choice of $\kappa_T = 0$ (no capping at all), although ruled out by Assumption (R), still provides very good results. Second, setting $\ell = \tau$, that is, considering the entire past history to compute quantiles apparently is not the best strategy, and shorter horizons ℓ seem preferable. This finding is possibly related to some time variation in the distribution of the innovations of log-volatilities at horizons longer than one year. Third, for any given α, shorter intervals are obtained when setting $\ell = 252$ or 504 regardless of the choice of κ_T. Overall, choosing $\kappa_T = 0.1$ and $\ell = 252$ or 504 works best for $\alpha = 0.32$ and 0.2, while $\kappa_T = 0.25$ and $\ell = 126$ or 252 works best for $\alpha = 0.1, 0.05$, and 0.01.

In Figures 4 and 5, we set $\kappa_T = 0.25$ and $\ell = 252$ and we show (in grey) $Y_{i,\tau+1}$ for some selected individual stocks, together with (in red) the estimated upper and lower bounds of the 90% one-step-ahead prediction interval, i.e. $\widehat{\mathcal{U}}^{(252)}_{i,\tau+1|\tau}(0.05)$ and $\widehat{\mathcal{L}}^{(252)}_{i,\tau+1|\tau}(0.05)$, respectively. Figure 4 shows results for six of the most volatiles stocks in

TABLE 5: *Standard & Poor's 100 Index data ($n = 90$ daily returns). Empirical coverage, frequency of prediction bounds violations, and average length of prediction intervals for GDFM, averaged over the cross-section.*

	$\kappa_T = 0$					$\kappa_T = 0.1$				
	α					α				
	0.32	0.2	0.1	0.05	0.01	0.32	0.2	0.1	0.05	0.01
$C^{(126)}(\alpha)$	0.6709	0.7894	0.8887	0.9400	0.9812	0.6874	0.7985	0.8931	0.9416	0.9813
$V_+^{(126)}(\alpha/2)$	0.1641	0.1048	0.0552	0.0299	0.0094	0.1559	0.1002	0.0533	0.0291	0.0095
$V_-^{(126)}(\alpha/2)$	0.1650	0.1058	0.0561	0.0301	0.0094	0.1566	0.1013	0.0536	0.0292	0.0091
$L^{(126)}(\alpha)$	3.3934	4.5156	6.1305	7.7681	12.4174	3.4726	4.5726	6.1553	7.7698	12.3130
$C^{(252)}(\alpha)$	0.6708	0.7903	0.8902	0.9415	0.9848	0.6882	0.7999	0.8940	0.9424	0.9846
$V_+^{(252)}(\alpha/2)$	0.1647	0.1044	0.0544	0.0289	0.0077	0.1560	0.0998	0.0526	0.0287	0.0078
$V_-^{(252)}(\alpha/2)$	0.1644	0.1053	0.0554	0.0296	0.0075	0.1558	0.1003	0.0534	0.0289	0.0076
$L^{(252)}(\alpha)$	3.3621	4.4794	6.0949	7.7240	12.5008	3.4351	4.5290	6.1078	7.7074	12.3767
$C^{(504)}(\alpha)$	0.6711	0.7895	0.8895	0.9412	0.9846	0.6886	0.7995	0.8929	0.9419	0.9843
$V_+^{(504)}(\alpha/2)$	0.1651	0.1057	0.0551	0.0290	0.0078	0.1561	0.1005	0.0536	0.0288	0.0081
$V_-^{(504)}(\alpha/2)$	0.1638	0.1047	0.0554	0.0298	0.0076	0.1553	0.1000	0.0535	0.0292	0.0076
$L^{(504)}(\alpha)$	3.3034	4.4179	6.0266	7.6643	12.1190	3.3786	4.4708	6.0439	7.6539	12.0462
$C^{(\tau)}(\alpha)$	0.7010	0.8142	0.9049	0.9506	0.9881	0.7187	0.8244	0.9096	0.9523	0.9881
$V_+^{(\tau)}(\alpha/2)$	0.1516	0.0933	0.0474	0.0247	0.0061	0.1424	0.0879	0.0452	0.0237	0.0062
$V_-^{(\tau)}(\alpha/2)$	0.1474	0.0925	0.0477	0.0248	0.0058	0.1389	0.0877	0.0452	0.0239	0.0057
$L^{(\tau)}(\alpha)$	3.4523	4.6632	6.4305	8.2802	13.3895	3.5562	4.7560	6.5201	8.3747	13.5115
	$\kappa_T = 0.25$					$\kappa_T = 0.5$				
	α					α				
	0.32	0.2	0.1	0.05	0.01	0.32	0.2	0.1	0.05	0.01
$C^{(126)}(\alpha)$	0.7126	0.8141	0.8997	0.9452	0.9821	0.7552	0.8391	0.9119	0.9507	0.9836
$V_+^{(126)}(\alpha/2)$	0.1435	0.0926	0.0500	0.0274	0.0091	0.1222	0.0800	0.0436	0.0243	0.0082
$V_-^{(126)}(\alpha/2)$	0.1439	0.0932	0.0504	0.0274	0.0088	0.1226	0.0809	0.0446	0.0251	0.0081
$L^{(126)}(\alpha)$	3.6203	4.6949	6.2419	7.8371	12.3547	3.9076	4.9426	6.4443	8.0189	12.5330
$C^{(252)}(\alpha)$	0.7138	0.8143	0.9009	0.9452	0.9851	0.7556	0.8398	0.9127	0.9512	0.9866
$V_+^{(252)}(\alpha/2)$	0.1433	0.0928	0.0491	0.0271	0.0077	0.1224	0.0798	0.0432	0.0242	0.0069
$V_-^{(252)}(\alpha/2)$	0.1428	0.0929	0.0500	0.0277	0.0072	0.1219	0.0804	0.0440	0.0246	0.0065
$L^{(252)}(\alpha)$	3.5796	4.6500	6.1923	7.7700	12.4405	3.8737	4.9105	6.4108	7.9706	12.7086
$C^{(504)}(\alpha)$	0.7149	0.8150	0.9002	0.9449	0.9846	0.7588	0.8422	0.9132	0.9514	0.9861
$V_+^{(504)}(\alpha/2)$	0.1430	0.0927	0.0495	0.0274	0.0080	0.1204	0.0782	0.0426	0.0236	0.0072
$V_-^{(504)}(\alpha/2)$	0.1420	0.0923	0.0502	0.0277	0.0074	0.1208	0.0795	0.0442	0.0250	0.0067
$L^{(504)}(\alpha)$	3.5336	4.6035	6.1513	7.7434	12.1898	3.8584	4.9070	6.4371	8.0357	12.6302
$C^{(\tau)}(\alpha)$	0.7430	0.8387	0.9162	0.9551	0.9886	0.7824	0.8633	0.9283	0.9613	0.9900
$V_+^{(\tau)}(\alpha/2)$	0.1301	0.0808	0.0415	0.0221	0.0061	0.1091	0.0680	0.0351	0.0189	0.0053
$V_-^{(\tau)}(\alpha/2)$	0.1269	0.0805	0.0422	0.0228	0.0054	0.1085	0.0687	0.0366	0.0198	0.0047
$L^{(\tau)}(\alpha)$	3.7420	4.9317	6.6982	8.5677	13.7991	4.1045	5.2913	7.0734	8.9901	14.4295

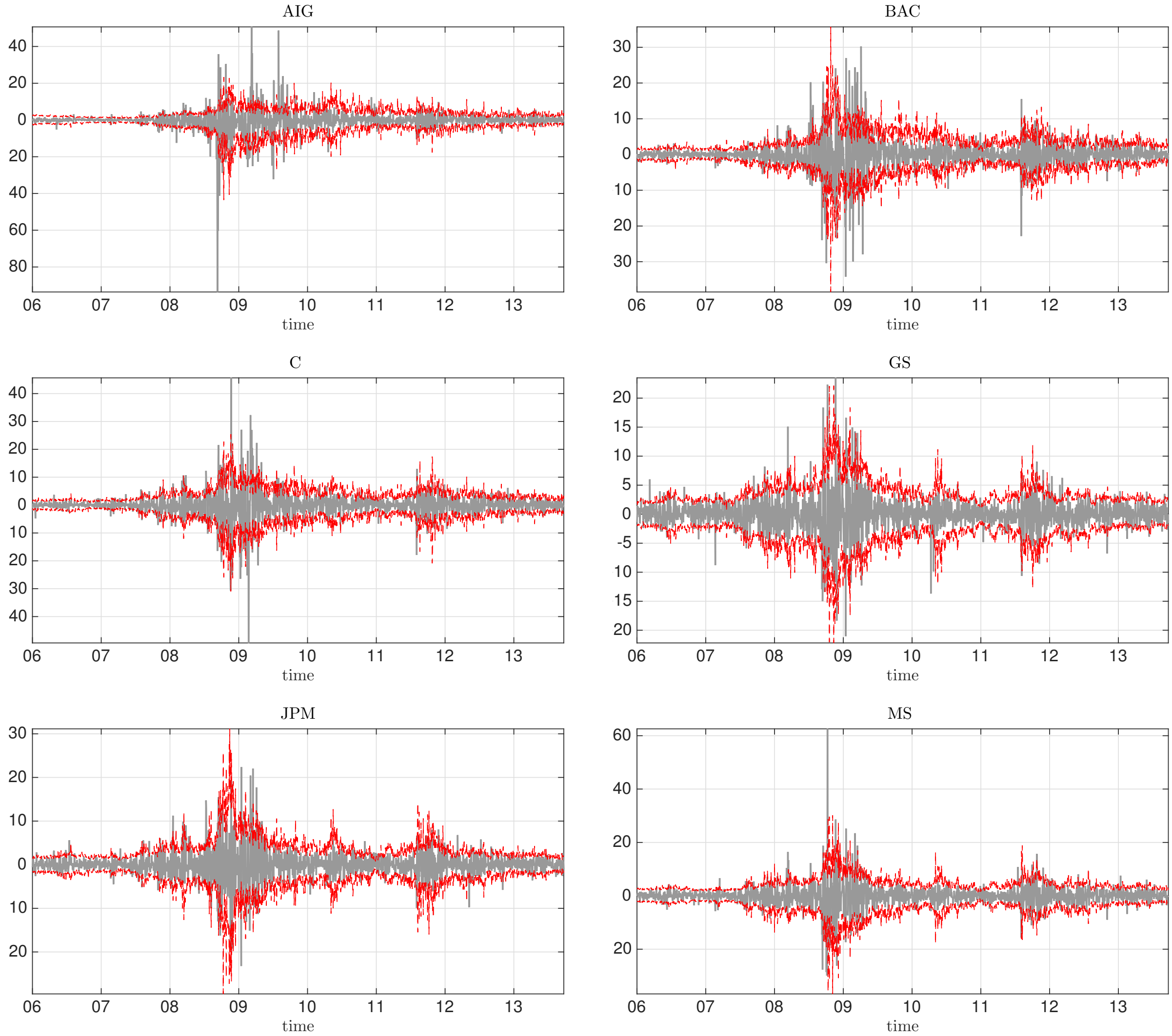
our dataset, all belonging to the financial sector: America International Group (AIG), Bank of America (BAC), Citigroup (C), Goldman Sachs (GS), JPMorgan Chase (JPM), Morgan Stanley (MS). Figure 5 provides the same results for eight relevant non-financial stocks: Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), Wallgreens (WAG), Exxon Mobil (XOM), Johnson & Johnson (JNJ), Boeing (BA), General Electric (GE). Volatilities, in those series, which were the most seriously affected by the great financial crisis, are notoriously hard to predict.

6.2 Coverage: comparison with GARCH

The novelty of our prediction intervals is that they are exploiting the information contained in the available cross-section of $n = 90$ stocks. This is in sharp contrast with the usual GARCH approach, which is strongly univariate, and disregards cross-sectional information by analyzing the n series one by one. Moreover, estimating 90 univariate GARCH models requires much more computing time than estimating our model. GARCH nevertheless constitute the more common practice in this context, and serves as a natural benchmark.

We therefore compare our prediction intervals with those obtained by fitting, via quasi-maximum likelihood,

FIGURE 4: *One-step-ahead 90% conditional prediction intervals (in red; $\ell = 252$): America International Group (AIG), Bank of America (BAC), Citigroup (C), Goldman Sachs (GS), JPMorgan Chase (JPM), Morgan Stanley (MS).*



univariate GARCH(1,1) models to all series in our panel. Specifically, for each series i, we estimate the model

$$Y_{it} = \mathrm{E}[y_{it}] + \sigma_{it}\epsilon_{it}, \quad \epsilon_{it} \overset{iid}{\sim} (0,1), \qquad t = 1, \ldots, \tau,$$
$$\sigma^2_{it} = \omega_i + \gamma_i Y^2_{it-1} + \beta_i \sigma^2_{it-1}, \quad \omega_i > 0,\ \gamma_i, \beta_i \geq 0,\ \gamma_i + \beta_i < 1.$$

For given $\tau = (T-M), \ldots, (T-1)$, we obtain estimated parameters $\widehat{\omega}_i, \widehat{\gamma}_i$ and $\widehat{\beta}_i$, from which we compute the estimated volatilities $\widehat{\sigma}^2_{it}$ and the innovation values $\widehat{\epsilon}_{it} = Y_{it}/\widehat{\sigma}_{it}$, $t = 1, \ldots, \tau$. Innovation quantiles are computed from $(\widehat{\epsilon}_{i,\tau-\ell+1}, \ldots, \widehat{\epsilon}_{i,\tau})$, where as before we set $\ell \in \{126, 252, 504, \tau\}$. Then, for any given level α and window size ℓ, and for $\tau = (T-M), \ldots, (T-1)$, given the one-step-ahead volatility predictor $\widehat{\sigma}^2_{i,\tau+1|\tau} = \widehat{\omega}_i + \widehat{\gamma}_i Y^2_{i,\tau} + \widehat{\beta}_i \widehat{\sigma}^2_{i,\tau}$, we compute the the upper and lower confidence bounds

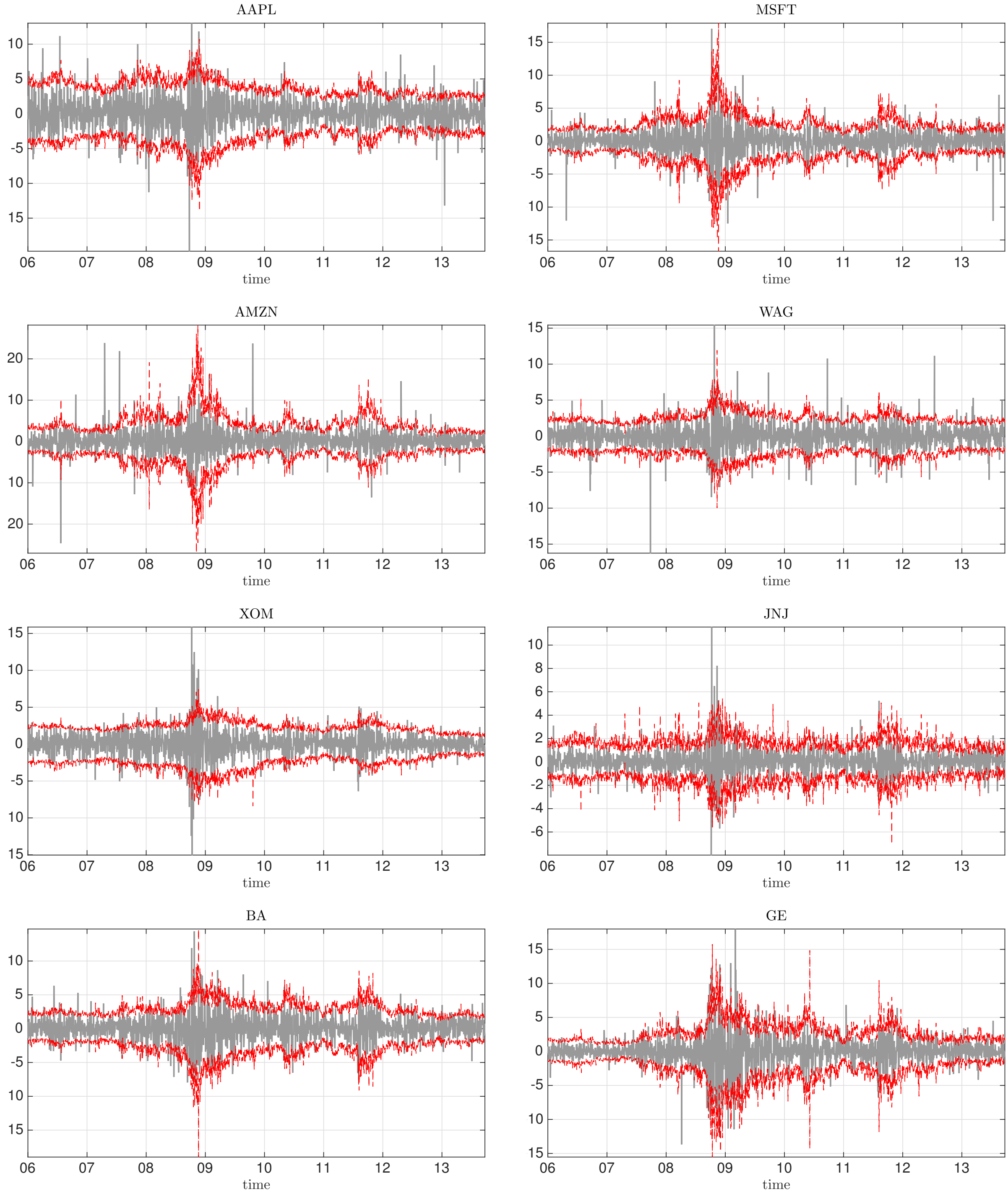
$$\widehat{\mathcal{U}}^{(\ell)\mathrm{GARCH}}_{i,\tau+1|\tau}(\alpha) := \bar{Y}_i + \widehat{\sigma}_{i,\tau+1|\tau}\,\widehat{\epsilon}^{(\ell)}_{i(\lceil \ell(1-\alpha)\rceil)} \quad \text{and} \quad \widehat{\mathcal{L}}^{(\ell)\mathrm{GARCH}}_{i,\tau+1|\tau}(\alpha) := \bar{Y}_i + \widehat{\sigma}_{i,\tau+1|\tau}\,\widehat{\epsilon}^{(\ell)}_{i(\lceil \ell\alpha\rceil)},$$

yielding the one-step-ahead prediction intervals

$$\widehat{\mathcal{I}}^{(\ell)\mathrm{GARCH}}_{i,\tau+1|\tau}(\alpha) := \big[\widehat{\mathcal{L}}^{(\ell)\mathrm{GARCH}}_{i,\tau+1|\tau}(\alpha/2), \widehat{\mathcal{U}}^{(\ell)\mathrm{GARCH}}_{i,\tau+1|\tau}(\alpha/2)\big]$$

and the indicators of correct interval prediction $\widehat{\mathcal{H}}^{(\ell)\mathrm{GARCH}}_{i,\tau+1|\tau}(\alpha) := \mathbb{I}(Y_{i,\tau+1} \in \widehat{\mathcal{I}}^{(\ell)\mathrm{GARCH}}_{i,\tau+1|\tau}(\alpha))$. Based on these quantities, we then compute, for $\alpha \in \{0.32, 0.2, 0.1, 0.05, 0.01\}$, the empirical coverage frequency, denoted

FIGURE 5: *One-step-ahead 90% conditional prediction intervals (in red; $\ell = 252$): Apple (AAPL), Microsoft (MSFT), Amazon (AMZN), Wallgreens (WAG), Exxon Mobil (XOM), Johnson & Johnson (JNJ), Boeing (BA), General Electric (GE).*



as $C_i^{(\ell)\mathrm{GARCH}}(\alpha)$, the proportions of coverage violations in the upper and lower tail, denoted as $V_{i,+}^{(\ell)\mathrm{GARCH}}(\alpha/2)$ and $V_{i,-}^{(\ell)\mathrm{GARCH}}(\alpha/2)$, respectively, and the average interval length, denoted as $L_i^{(\ell)\mathrm{GARCH}}(\alpha)$. Averages of these quantities over the n series under study are shown in Table 6. Inspection of this table reveals that the GDFM performances are slightly better than the GARCH ones in terms of coverage frequencies, based on similar interval lengths. This, however, is mainly a descriptive and, due to cross-sectional dependence, somewhat misleading assessment, which ideally should be reinforced into a more formal testing analysis.

A formal comparison between the GDFM and GARCH(1,1) coverage performances should take into account

TABLE 6: *Standard & Poor's 100 Index data ($n = 90$ daily returns). Empirical coverage, frequency of prediction bounds violations, and average length of prediction intervals for GARCH, averaged over the cross-section.*

	α				
	0.32	0.2	0.1	0.05	0.01
$C^{(126)\mathrm{GARCH}}(\alpha)$	0.6755	0.7947	0.8933	0.9429	0.9834
$V_+^{(126)\mathrm{GARCH}}(\alpha/2)$	0.1576	0.0991	0.0507	0.0267	0.0076
$V_-^{(126)\mathrm{GARCH}}(\alpha/2)$	0.1669	0.1062	0.0560	0.0304	0.0090
$L^{(126)\mathrm{GARCH}}(\alpha)$	3.4401	4.5562	6.1207	7.7282	12.3986
$C^{(252)\mathrm{GARCH}}(\alpha)$	0.6786	0.7981	0.8968	0.9460	0.9871
$V_+^{(252)\mathrm{GARCH}}(\alpha/2)$	0.1567	0.0978	0.0491	0.0255	0.0060
$V_-^{(252)\mathrm{GARCH}}(\alpha/2)$	0.1647	0.1041	0.0541	0.0285	0.0069
$L^{(252)\mathrm{GARCH}}(\alpha)$	3.4142	4.5235	6.0755	7.6329	12.2536
$C^{(504)\mathrm{GARCH}}(\alpha)$	0.6807	0.7994	0.8983	0.9479	0.9878
$V_+^{(504)\mathrm{GARCH}}(\alpha/2)$	0.1560	0.0975	0.0488	0.0248	0.0056
$V_-^{(504)\mathrm{GARCH}}(\alpha/2)$	0.1633	0.1031	0.0529	0.0274	0.0066
$L^{(504)\mathrm{GARCH}}(\alpha)$	3.3822	4.4801	6.0220	7.5581	11.7469
$C^{(\tau)\mathrm{GARCH}}(\alpha)$	0.6920	0.8077	0.9036	0.9510	0.9897
$V_+^{(\tau)\mathrm{GARCH}}(\alpha/2)$	0.1520	0.0935	0.0458	0.0228	0.0048
$V_-^{(\tau)\mathrm{GARCH}}(\alpha/2)$	0.1560	0.0988	0.0505	0.0262	0.0055
$L^{(\tau)\mathrm{GARCH}}(\alpha)$	3.4139	4.4942	6.0156	7.5369	11.6268

the fact that the coverage results of the two methods, for given i and τ, are not independent. The situation is quite similar to that of comparing paired proportions, where tests are to be carried out on the basis of the traditional McNemar (1947) test. For given α and ℓ, consider, for all i, the events (discordant GDFM and GARCH coverage results)

$$\mathcal{A}_{i,\tau+1|\tau}^{(\ell)}(\alpha) := \left\{Y_{i,\tau+1} \in \widehat{\mathcal{I}}_{i,\tau+1|\tau}^{(\ell)}(\alpha) \cap Y_{i,\tau+1} \notin \widehat{\mathcal{I}}_{i,\tau+1|\tau}^{(\ell)\mathrm{GARCH}}(\alpha)\right\}$$
$$\mathcal{B}_{i,\tau+1|\tau}^{(\ell)}(\alpha) := \left\{Y_{i,\tau+1} \notin \widehat{\mathcal{I}}_{i,\tau+1|\tau}^{(\ell)}(\alpha) \cap Y_{i,\tau+1} \in \widehat{\mathcal{I}}_{i,\tau+1|\tau}^{(\ell)\mathrm{GARCH}}(\alpha)\right\},$$

and define

$$n_{12i}^{(\ell)}(\alpha) := \sum_{\tau=T-M}^{T-1} \mathbb{I}\left(\mathcal{A}_{i,\tau+1|\tau}^{(\ell)}(\alpha)\right) \quad \text{and} \quad n_{21i}^{(\ell)}(\alpha) := \sum_{\tau=T-M}^{T-1} \mathbb{I}\left(\mathcal{B}_{i,\tau+1|\tau}^{(\ell)}(\alpha)\right).$$

Consider the null hypothesis under which the indicators of a successful interval prediction in both methods are i.i.d. Bernoulli, with identical (but otherwise unspecified) coverage probabilities. The McNemar test of that hypothesis is conditioning on the sum $n_{\mathrm{disc},i}^{(\ell)}(\alpha) := n_{12i}^{(\ell)}(\alpha) + n_{21i}^{(\ell)}(\alpha)$ of discordant coverage results: concordant results indeed carry no information on a difference between coverage probabilities. Conditional on $n_{\mathrm{disc},i}^{(\ell)}(\alpha)$, the null distribution of $n_{12i}^{(\ell)}(\alpha)$ is binomial $\mathrm{Bin}(n_{\mathrm{disc},i}^{(\ell)}(\alpha), 0.5)$. At probability level δ, the test rejects in favour of a better GDFM coverage for "large values" of $n_{12i}^{(\ell)}(\alpha)$, in favour of a better GARCH coverage for "small values" of the same (equivalently, "large values" of $n_{21i}^{(\ell)}$), with critical values the $(1-\delta)$ and δ binomial quantiles, respectively.

Table 7 reports the McNemar empirical rejection frequencies (over the $n = 90$ series)—in favour of a better GDFM coverage in the left-hand panel, in favour of a better GARCH coverage in the right-hand one. We consider the cases in which $\alpha = 0.1$ or 0.05, $\ell = 126$ or 252, $\kappa_T = 0.25$ (for the GDFM); testing was performed at significance levels $\delta = 0.1$, 0.05, and 0.01. Irrespective of ℓ and α, the GDFM approach appears to outperform, quite consistently and significantly, the GARCH one.

6.3 Coverage: backtesting

As explained in Section 4, a formal assessment of the validity of our approach can be based on the backtesting procedure proposed by Christoffersen (1998). The idea consists in testing the null hypothesis (4.9) under which

TABLE 7: *Standard & Poor's 100 Index data ($n = 90$ daily returns). Proportions of McNemar rejections in favour of a better GDFM coverage (left-hand panel), in favour of a better GARCH coverage (right-hand panel).*

	better GDFM coverage			better GARCH coverage		
$\alpha = 0.1$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
$\ell = 126$	0.6000	0.5444	0.3667	0.1000	0.0778	0.0667
$\ell = 252$	0.5333	0.4556	0.2889	0.1333	0.1222	0.0889
$\alpha = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
$\ell = 126$	0.4111	0.2778	0.1556	0.1111	0.0778	0.0556
$\ell = 252$	0.2556	0.2000	0.0556	0.1778	0.1222	0.1111

the $\widehat{\mathcal{H}}_{i,\tau+1|\tau}(\alpha)$'s (the indicators of a successful interval prediction) are i.i.d. Bernoulli$(1-\alpha)$. Depending on the objectives, several alternatives can be considered. One can be interested (Section 6.3.1) in the validity of interval prediction or the sharpness of the nominal coverage level. Else, one may consider (Section 6.3.2) alternatives of serial dependence. Or, those two issues can be combined (Section 6.3.3) by merging the corresponding alternatives.

Irrespective of the alternative, however, it should be insisted that all those tests—one for each cross-sectional item—are intrinsically univariate. When simultaneously performing several or all of them, one should be extremely cautious with the interpretation of the results. The tables we are providing below are reporting empirical rejection frequencies (over the $n = 90$ series). Those n tests, however, are not functionally interrelated (as they would be if the prediction intervals were based on the quantiles of common shocks only); hence, they are not about testing the validity of *joint prediction intervals* with global asymptotic coverage level $(1-\alpha)$. Neither are they mildly interrelated (as they would be if the prediction intervals were exclusively based on idiosyncratic quantiles), providing joint prediction intervals with global asymptotic coverage level of the order of $(1-\alpha)^n$. High rejection frequencies across the n series thus do not imply bad forecasting properties, but can result from complex cross-sectional dependencies. A standard attitude would consist in adopting a Bonferroni or a Šidák correction; for $n = 90$, and for a global testing level of 1%, this would lead to implementing the $n = 90$ individual tests at an overly conservative level $\delta \approx 0.0001 = 10^{-4}$—a level at which none of the null hypotheses under study is rejected.

All tests below are performed for $\kappa_T = 0.25$, $\alpha = 0.1$ or 0.05, $\ell = 126$ or 252; testing significance levels are $\delta = 0.1$, 0.05, and 0.01.

6.3.1 Testing for valid or sharp conditional coverage probabilities

If we are interested in the validity of interval prediction, the relevant testing problems are (one-sided)

$$H_{0i} : \mathrm{E}[\widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)] \geq (1-\alpha) \quad \text{versus} \quad H_{1i} : \mathrm{E}[\widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)] < (1-\alpha). \tag{6.1}$$

If instead we are interested in testing whether $(1-\alpha)$, as a nominal confidence level, is sharp, the testing problems are (still one-sided)

$$H_{0i} : \mathrm{E}[\widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)] \leq (1-\alpha) \quad \text{versus} \quad H_{1i} : \mathrm{E}[\widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)] > (1-\alpha). \tag{6.2}$$

Both testing problems (6.1) and (6.2), admit a level-δ uniformly most powerful solution, rejecting H_{0i} whenever the test statistic

$$n^{(\ell)}_{1i}(\alpha) := \sum_{\tau=T-M}^{T-1} \widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)$$

falls below the binomial Bin$(M, 1-\alpha)$ quantile of order δ when testing (6.1), or above the Bin$(M, 1-\alpha)$ quantile of order $(1-\delta)$ when testing (6.2). Since M is large, the same tests are well approximated by rejecting H_{0i} whenever the proportion $n^{(\ell)}_{1i}(\alpha)/M$ of correct coverage is smaller than $(1-\alpha) - z_\delta\sqrt{\alpha(1-\alpha)}$ when testing (6.1), or larger than $(1-\alpha) + z_\delta\sqrt{\alpha(1-\alpha)}$ when testing (6.2), where z_δ stands for the $(1-\delta)$ standard normal quantile. A two-sided coverage test can also be computed

$$LR^{(\ell)}_{\text{cover},i}(\alpha) := \big(n^{(\ell)}_{1i}(\alpha) - M(1-\alpha)\big)^2 / M\alpha(1-\alpha), \tag{6.3}$$

TABLE 8: *Standard & Poor's 100 Index data* ($n = 90$ *daily returns). Proportion of rejections when testing for valid nominal coverage* (6.1) *(left-hand panel) and for sharp nominal coverage* (6.2) *(middle panel), and when considering the two-sided test* (6.3) *(right-hand panel)*

	valid nominal coverage test			sharp nominal coverage test			two-sided coverage test		
$\alpha = 0.1$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
$\ell = 126$	0.1444	0.1222	0.0889	0.2444	0.1444	0.0333	0.2667	0.1889	0.0778
$\ell = 252$	0.1556	0.1333	0.0778	0.3111	0.2111	0.0889	0.3444	0.2556	0.1333
$\alpha = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
$\ell = 126$	0.2889	0.1889	0.1333	0.0111	0.0000	0.0000	0.1889	0.1556	0.1000
$\ell = 252$	0.3000	0.2000	0.1444	0.0556	0.0111	0.0000	0.2111	0.1556	0.1333

with asymptotic $\chi^2_{(1)}$ null distribution (as $M \to \infty$).[5]

Table 8 reports the empirical rejection frequencies (over $n = 90$ series) when testing (6.1) (left-hand panel) and (6.2) (right-hand panel), respectively and using the normal approximation of the binomial. The general comments above apply when interpreting those tables: the only valid global conclusions are those resulting from Bonferroni or Šidák corrections, which do not lead to any rejections.

6.3.2 Testing against serial dependence

If the alternative of interest is serial dependence among coverage indicators, we propose considering, for each individual stock i, alternatives of binary first-order Markov dependence. More precisely, defining the transition probabilities

$$p_{hk,i}(\alpha) = \mathrm{P}\Big(\widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha) = k \,\Big|\, \widehat{\mathcal{H}}^{(\ell)}_{i,\tau|\tau-1}(\alpha) = h\Big), \quad h, k = 1, 0,$$

we consider the testing problem (with unspecified unconditional probability $p_i(\alpha)$ of correct coverage)

$$H_{0i} : p_{01,i}(\alpha) = p_{11,i}(\alpha) =: p_i(\alpha) \qquad \text{versus} \qquad H_{1i} : p_{01,i}(\alpha) \neq p_{11,i}(\alpha); \tag{6.4}$$

note that $p_{01,i}(\alpha) = p_{11,i}(\alpha)$ automatically implies $p_{00,i}(\alpha) = p_{10,i}(\alpha)$. Defining

$$n^{(\ell)}_{11i}(\alpha) := \sum_{\tau=T-M+1}^{T-1} \widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)\widehat{\mathcal{H}}^{(\ell)}_{i,\tau|\tau-1}(\alpha), \qquad n^{(\ell)}_{10i}(\alpha) := n^{(\ell)}_{1i}(\alpha) - n^{(\ell)}_{11i}(\alpha),$$

$$n^{(\ell)}_{01i}(\alpha) := \sum_{\tau=T-M+1}^{T-1} \widehat{\mathcal{H}}^{(\ell)}_{i,\tau+1|\tau}(\alpha)\big(1 - \widehat{\mathcal{H}}^{(\ell)}_{i,\tau|\tau-1}(\alpha)\big), \qquad n^{(\ell)}_{00i}(\alpha) := n^{(\ell)}_{0i}(\alpha) - n^{(\ell)}_{01i}(\alpha),$$

the statistics

$$\pi^{(\ell)}_i(\alpha) := \big(n^{(\ell)}_{01i}(\alpha) + n^{(\ell)}_{11i}(\alpha)\big)/M$$

are estimators of the $p_i(\alpha)$'s under the null, while

$$\pi^{(\ell)}_{11i}(\alpha) := n^{(\ell)}_{11i}(\alpha)/n^{(\ell)}_{1i}(\alpha), \qquad \pi^{(\ell)}_{10i}(\alpha) := 1 - \pi^{(\ell)}_{11i}(\alpha),$$
$$\pi^{(\ell)}_{01i}(\alpha) := n^{(\ell)}_{01i}(\alpha)/\big(M - n^{(\ell)}_{1i}(\alpha)\big), \text{ and } \pi^{(\ell)}_{00i}(\alpha) := 1 - \pi^{(\ell)}_{01i}(\alpha)$$

are estimating the transition probabilities $p_{hk,i}(\alpha)$ under the alternative. Log-likelihoods under the null and the alternative are

$$L_{0i}(\alpha) = (n^{(\ell)}_{00i}(\alpha) + n^{(\ell)}_{10i}(\alpha)) \log[1 - \pi^{(\ell)}_i(\alpha)] + n^{(\ell)}_{01i}(\alpha) + n^{(\ell)}_{11i}(\alpha) \log[\pi^{(\ell)}_i(\alpha)],$$

and

$$L^{(\ell)}_{1i}(\alpha) = n^{(\ell)}_{00i}(\alpha) \log[1 - \pi^{(\ell)}_{01i}(\alpha)] + n^{(\ell)}_{01i}(\alpha) \log[\pi^{(\ell)}_{01i}(\alpha)] \\ + n^{(\ell)}_{10i}(\alpha) \log[1 - \pi^{(\ell)}_{11i}(\alpha)] + n^{(\ell)}_{11i}(\alpha) \log[\pi^{(\ell)}_{11i}(\alpha)],$$

[5]It is easily seen that $LR^{(\ell)}_{\text{cover},i}(\alpha)$ is equivalent, up to a constant term, to the so-called "unconditional coverage" likelihood ratio test statistic proposed in Section 3.1 of Christoffersen (1998) which therefore yields the same results.

TABLE 9: *Standard & Poor's 100 Index data ($n = 90$ daily returns). Proportion of rejections when testing against serial dependence* (6.4) *(left-hand panel) and in the combined problem* (6.5) *(right-hand panel).*

$\alpha = 0.1$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
$\ell = 126$	0.3222	0.2222	0.0778	0.3667	0.2222	0.1222
$\ell = 252$	0.4000	0.3556	0.1889	0.4889	0.4222	0.2556
$\alpha = 0.05$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
$\ell = 126$	0.2778	0.2000	0.0556	0.2778	0.2111	0.1222
$\ell = 252$	0.3667	0.2667	0.1667	0.3556	0.2667	0.2222

respectively. For any given i, α and ℓ, thus, we can construct a likelihood-ratio test for (6.4), based on the asymptotically $\chi^2_{(1)}$ null distribution (as $M \to \infty$) of $LR^{(\ell)}_{\text{ind},i}(\alpha) := 2\big[L^{(\ell)}_{1i}(\alpha) - L_{0i}(\alpha)\big]$ (see also Section 3.2 in Christoffersen, 1998). More general alternatives, involving higher-order serial dependencies, could be considered as well, based on the tests proposed by Dufour et al. (1998).

In Table 9 (left-hand panel), we report the proportions of rejections (over the n series) when testing (6.4). The same remarks apply as in the interpretation of Table 8.

6.3.3 Combined test

Combining the above tests, a likelihood ratio test (given i, α, and ℓ) for

$$H_{0i} : p_{01,i}(\alpha) = p_{11,i}(\alpha) = (1-\alpha) \;\text{ versus }\; H_{1i} : p_{01,i}(\alpha) \neq p_{11,i}(\alpha) \;\text{ or }\; p_{01,i}(\alpha) = p_{11,i}(\alpha) \neq (1-\alpha) \quad (6.5)$$

can be based on the asymptotically $\chi^2_{(2)}$ (as $M \to \infty$) null distribution of

$$LR^{(\ell)}_{i}(\alpha) = LR^{(\ell)}_{\text{cover},i}(\alpha) + LR^{(\ell)}_{\text{ind},i}(\alpha)$$

(see also Section 3.3 in Christoffersen, 1998). The fraction of rejections (over n series) when testing (6.5) is reported in Table 9 (right-hand panel). The same remarks as in Table 8 still apply.

6.4 Discussion

In Table 10, we report (four panels, according to the values of α and ℓ) the ten individual series for which the four tests above return the most significant rejections. Rejecting in (6.1) the null hypothesis of a valid coverage ("small" values of $n^{(\ell)}_{1i}/M$) means that the approximations we are making in the construction of the intervals lead to a loss of prediction accuracy for that specific series: the intervals for that series are not wide enough—equivalently, their actual coverage probability is less than the nominal $(1-\alpha)$ level. The series listed in the first column of each panel thus are "hardest to predict". Among them are stocks belonging to the Financial sector, as America International Group (AIG), Bank of America (BAC), and Citigroup (C). These series, in particular, were among those mostly affected by the great financial crisis. Rejecting in (6.2) the null hypothesis of a sharp coverage ("large" values of $n^{(\ell)}_{1i}/M$) also means that the approximations we are making in the construction of the intervals lead to a loss of prediction accuracy for that specific series, now in the sense that we could do better: the intervals for that series are too wide—their actual coverage probability is more than the nominal $(1-\alpha)$ level. The series listed in the second column of each panel thus are "easiest to predict". Among them, stocks belonging to the Energy and Consumers sectors, as Exxon Mobil (XOM), Cisco Systems (CSCO), and McDonalds (MCD).

When testing against serial dependence, rejection ("large" values of $LR^{(\ell)}_{\text{ind},i}(\alpha)$) indicates that the predictive information available in past observations has not been fully exploited in the construction of the prediction intervals. This could be the case, for example, if some informative idiosyncratic cross-correlation is available: idiosyncratic cross-correlations indeed are not captured by our univariate autoregressive modelling of idiosyncratic components. Alternative multivariate models for idiosyncratic components, such as sparse VAR, are likely to improve on this (see e.g. the approach proposed in Barigozzi and Hallin, 2017b), and could be incorporated into our two-step GDFM approach. We do not explore this any further in this paper, though. Such dependencies could be related to sectoral co-movements which, being specific to some restricted sector, are not captured by the market-wide factors. This seems to be the case especially for Financial and Energy stocks. A symptom of that phenomenon is the fact that the explained variance of the common component of the Financial stock returns is about 30% less than the variance explained by the common component of all other stock returns. The importance of this idiosyncratic variation, which is not accounted for by our approach, may explain why combined tests of correct coverage and independence exhibit, for Financial stock returns, high rejection frequencies.

TABLE 10: *Standard & Poor's 100 Index data ($n = 90$ daily returns). Series tickers for which the null hypotheses considered in Section 5.3 are rejected most significantly.*

$\alpha = 0.1$	smallest $n_{1i}^{(\ell)}(\alpha)/M$	largest $n_{1i}^{(\ell)}(\alpha)/M$	largest $LR_{\text{ind},i}^{(\ell)}(\alpha)$	largest $LR_i^{(\ell)}(\alpha)$	$\alpha = 0.05$	smallest $n_{1i}^{(\ell)}(\alpha)/M$	largest $n_{1i}^{(\ell)}(\alpha)/M$	largest $LR_{\text{ind},i}^{(\ell)}(\alpha)$	largest $LR_i^{(\ell)}(\alpha)$
$\ell = 126$	BAC	MCD	AIG	BAC	$\ell = 126$	SPG	COST	AIG	AIG
	SPG	CSCO	BRK.B	SPG		BAC	MCD	BRK.B	BAC
	C	CVX	AMGN	AIG		C	TGT	AMGN	SPG
	AIG	GILD	SPG	C		AIG	EMC	DVN	C
	WFC	MO	BAC	WFC		WFC	WMT	MRK	WFC
	USB	TXN	COP	BRK.B		BRK.B	CVX	BAC	COP
	JPM	WMT	SO	AMGN		SO	GILD	XOM	BRK.B
	COF	XOM	AAPL	USB		USB	T	CVS	DVN
	BRK.B	EMC	APC	COP		MS	COP	COP	APC
	MS	SLB	JNJ	JPM		COF	CSCO	EXC	USB
$\alpha = 0.1$	smallest $n_{1i}^{(\ell)}(\alpha)/M$	largest $n_{1i}^{(\ell)}(\alpha)/M$	largest $LR_{\text{ind},i}^{(\ell)}(\alpha)$	largest $LR_i^{(\ell)}(\alpha)$	$\alpha = 0.05$	smallest $n_{1i}^{(\ell)}(\alpha)/M$	largest $n_{1i}^{(\ell)}(\alpha)/M$	largest $LR_{\text{ind},i}^{(\ell)}(\alpha)$	largest $LR_i^{(\ell)}(\alpha)$
$\ell = 252$	BAC	GILD	AIG	BAC	$\ell = 252$	SPG	MCD	AIG	AIG
	SPG	MCD	SPG	SPG		BAC	GILD	BRK.B	BAC
	C	XOM	COP	AIG		C	ORCL	COP	SPG
	AIG	TXN	BRK.B	C		AIG	QCOM	DVN	C
	WFC	CSCO	DVN	WFC		WFC	CVX	EXC	WFC
	USB	CVX	APC	BRK.B		BRK.B	CSCO	OXY	BRK.B
	MS	EBAY	LLY	USB		JPM	WMT	ALL	COP
	BRK.B	EMC	BAC	MS		USB	COST	BAC	SO
	JPM	TGT	UNH	SO		FCX	EMC	SPG	OXY
	COF	MO	C	AMGN		TWX	TGT	KO	ALL

7 Conclusions

In this paper, we consider a two-step GDFM approach for jointly modelling stock returns and their volatilities in order to build conditional prediction intervals. A careful study of the consistency properties (as the cross-sectional dimension n and the sample size T both tend to infinity) of the resulting estimators is conducted. Those results are the theoretical foundation of (Barigozzi and Hallin, 2016, 2017a,b, and Barigozzi et al., 2019); here, we are using them in the construction of one-step-ahead prediction intervals.

We then apply our methodology to a panel of 90 daily returns of stocks listed in the S&P100. Through a recursive exercise, we show that we are able to obtain one-step-ahead prediction intervals which are in general more accurate than univariate GARCH methods.

Many extensions of this work are possible, which are left for future research. First, our empirical results indicate that, by exploiting also the cross-sectional lagged dependencies among idiosyncratic components, we could achieve better coverage especially for those series belonging to the Financial sector, which remains strongly interconnected even after controlling for common factors. This could be achieved by computing predictions of idiosyncratic components by fitting multivariate models such as sparse VARs. Second, our methodology immediately allows us to consider bivariate or multivariate prediction intervals. Third, asymmetric prediction intervals can also be considered. In particular, Value-at-Risk indicators are readily computable; moreover, by considering many values of the coverage, we can approximate the whole conditional distribution of returns. Last, another possible application consists in the construction od prediction intervals for macroeconomic variables as GDP or inflation taking into account, in a way similar to Jurado et al. (2015), the uncertainty related to the business cycle.

References

Aït-Sahalia, Y. and Xiu, D. (2017). Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *Journal of Econometrics*, 201:384–399.

Alizadeh, S., Brandt, M. W., and Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, 57:1047–1091.

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

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

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

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146:304–317.

Barigozzi, M. and Hallin, M. (2016). General dynamic factors and volatilities: Recovering the market volatility shocks. *The Econometrics Journal*, 19:C33–C60.

Barigozzi, M. and Hallin, M. (2017a). General dynamic factors and volatilities: Estimation and forecasting. *Journal of Econometrics*, 201:307–321.

Barigozzi, M. and Hallin, M. (2017b). Networks, dynamic factors, and the volatility analysis of high-dimensional financial series. *Journal of the Royal Statistical Society, Series C*, 66:581–605.

Barigozzi, M., Hallin, M., and Soccorsi, S. (2019). Identification of global and local shocks in international financial markets via general dynamic factor models. *Journal of Financial Econometrics*, 33:625–642.

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

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 127:169–194.

Borovkov, A. A. (2000). Large deviation probabilities for random walks with semiexponential distributions. *Siberian Mathematical Journal*, 41:1061–1093.

Brillinger, D. (2001). *Time Series: Data Analysis and Theory*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean–variance analysis on large asset markets. *Econometrica*, 51:1281–304.

Chicheportiche, R. and Bouchaud, J.-P. (2015). A nested factor model for non-linear dependencies in stock returns. *Quantitative Finance*, 15:1789–1804.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39:841–862.

Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *the Journal of Finance*, 48:1263–1291.

Connor, G., Korajczyk, R. A., and Linton, O. (2006). The common and specific components of dynamic volatility. *Journal of Econometrics*, 132:231–255.

den Haan, W. J. and Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust Inference*, volume 15 of *Handbook of Statistics*. Elsevier.

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

Dufour, J., Hallin, M., and Mizera, I. (1998). Generalized run tests for heteroscedastic time series. *Journal of Nonparametric Statistics*, 9:39–86.

Engle, R. F. and Marcucci, J. (2006). A long–run pure variance common features model for the common volatilities of the Dow Jones. *Journal of Econometrics*, 132:7–42.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society, Series B*, 75:603–680.

Fan, J., Liao, Y., and Shi, X. (2015). Risks of large portfolios. *Journal of Econometrics*, 186:367–387.

Forni, M., Giannone, D., Lippi, M., and Reichlin, L. (2009). Opening the black box: Structural factor models versus structural VARs. *Econometric Theory*, 25:1319–1347.

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

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

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

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

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory*, 17:1113–1141.

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

Francq, C. and Zakoian, J.-M. (2011). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons.

Francq, C. and Zakoian, J.-M. (2019). Virtual Historical Simulation for estimating the conditional VaR of large portfolios. mimeo.

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

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

Hannan, E. J. (1970). *Multiple Time Series*. John Wiley & Sons.

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

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

Kuchibhotla, A. K. and Chakrabortty, A. (2018). Moving beyond sub-gaussianity in high dimensional statistics: Applications in covariance matrix estimation and linear regressions. arXiv preprint arXiv:1804.02605.

McNemar, Q. M. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12:153–157.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151:435–474.

Ng, V., Engle, R. F., and Rothschild, M. (1992). A multi-dynamic-factor model for stock returns. *Journal of Econometrics*, 52:245–266.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *The Annals of Mathematical Statistics*, 28:329–348.

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

Priestley, M. B. (2001). *Spectral Analysis and Time Series*. Elsevier Academic Press.

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*, 146:10–25.

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

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika*, 59:525–531.

Tiao, G. C. and Box, G. E. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76:802–816.

Tiao, G. C. and Guttman, I. (1980). Forecasting contemporal aggregates of multiple time series. *Journal of Econometrics*, 12:219–230.

Tiao, G. C. and Hillmer, S. C. (1978). Some consideration of decomposition of a time series. *Biometrika*, 65:497–502.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51:157–195.

Trucíos, . C., Mazzeu, J., Hallin, M., Zevallos, M., Hotta, L., and Pereira, P. V. (2019). Forecasting conditional covariance matrices in high-dimensional time series with application to dynamic portfolio optimization: a general dynamic factor approach. Technical Report 2019-32, ECARES.

Tsay, R. S. and Tiao, G. C. (1985). Use of canonical analysis in time series model identification. *Biometrika*, 72:299–315.

Uchaikin, V. V. and Zolotarev, V. M. (2011). *Chance and Stability: Stable Distributions and their Applications*. Walter de Gruyter.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In Eldar, Y. and Kutyniok, G., editors, *Compressed Sensing. Theory and Applications.* Cambridge University Press.

Vladimirova, M. and Arbel, J. (2019). Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. arXiv preprint arXiv:1905.04955.

Wu, W. B. and Zaffaroni, P. (2018). Asymptotic theory for spectral density estimates of general multivariate time series. *Econometric Theory*, 34:1–22.