Forecasting Value-at-Risk and Expected Shortfall in Large Portfolios: a General Dynamic Factor Approach

Marc Hallin
Department of Mathematics and ECARES, Université libre de Bruxelles

Carlos Trucios
Federal University of Rio de Janeiro

December 2020

ECARES working paper 2020-50
Forecasting Value-at-Risk and Expected Shortfall in Large Portfolios: a General Dynamic Factor Model Approach

Marc Hallin¹ and Carlos Trucios*²,³

¹Department of Mathematics and ECARES
Université libre de Bruxelles, Belgium

²Center for Mathematics, Computing and Cognition
Federal University of ABC (UFABC), Brazil
³Faculty of Business Administration and Accounting
Federal University of Rio de Janeiro (UFRJ), Brazil

December 15, 2020

Abstract

Beyond their importance from a regulatory policy point of view, Value-at-Risk (VaR) and Expected Shortfall (ES) play an important role in risk management, portfolio allocation, capital level requirements, trading systems, and hedging strategies. Unfortunately, due to the curse of dimensionality, their accurate estimation in large portfolios is quite a challenge. To tackle this problem, we propose a filtered historical simulation method in which high-dimensional conditional covariance matrices are estimated via a general dynamic factor model with infinite-dimensional factor space and conditionally heteroscedastic factors. The procedure is applied to a panel with concentration ratio close to one. Backtesting and scoring results indicate that both VaR and ES are accurately estimated under our method, which outperforms alternative approaches available in the literature.

Keywords: conditional covariance, high-dimensional time series, large panels, risk measures, volatility.

JEL classification: C10, C32, C53, C55, G17, G32

*The authors thank the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS), Brazil for their computational support. The second author acknowledges financial support from São Paulo Research Foundation (FAPESP), grants 2016/18599-4 and 2018/03012-3
1 Introduction

Value-at-Risk (VaR) and Expected Shortfall (ES) play an essential role in economics, finance, and insurance. They are lying at the core of the regulatory framework settled by the Solvency II regulation, the Swiss Solvency Test, and the Basel II and III accords, all intended to strengthen the financial stability of the banking system. Beyond this, VaR and ES constitute a basic tool for investors, hedge fund managers, traders, and all decision-makers involved in risk management, portfolio allocation, trading desk limits, investment strategies, and the development of trading algorithms. Their accurate estimation and forecasting, thus, is of primary importance.

Several procedures dealing with VaR and ES estimation in univariate and multivariate settings are available in the literature: see Braione and Scholtes (2016), Gao and Zhou (2016), Trucios et al. (2018), Patton et al. (2019), Taylor (2019), Francq and Zakoian (2020), Trucios et al. (2020b), ... to quote only a few recent references. In the multivariate framework, most procedures are based on the estimation of the conditional covariance matrix, which in a high-dimensional context is quite challenging due to the difficulty to guarantee positive definiteness and the fastly increasing number of parameters—the classical curse of dimensionality. As a consequence, traditional multivariate volatility estimation procedures, while remaining useful in small and moderate dimensions, turn out to be helpless or infeasible in high dimensions.

A high-dimensional context, unfortunately, is the rule rather than the exception in the area, where relevant information is scattered among large numbers of time series. Alternative procedures that are able to handle very large panels of financial data thus are badly needed. This need has aroused much interest, and several procedures have been proposed recently to estimate high-dimensional conditional covariance matrices; see Zumbach (2007), Fan et al. (2008), Alessi et al. (2009), Li et al. (2016), Chang et al. (2018), Engle et al. (2019), Trucios et al. (2019), Pakel et al. (2020), among others.

Recently, Trucios et al. (2020a) have proposed a high-dimensional conditional covariance matrix estimator based on the general dynamic factor model which does not restrict the factor space to be finite-dimensional. This procedure is based on the one-sided estimator of Forni et al. (2015, 2017), and has shown a good performance both in Monte Carlo experiments and empirical data, outperforming several among the aforementioned approaches.

On the other hand, two methodologies are commonly used by financial institutions to estimate the VaR in large panels, namely, the historical simulation and the filtered historical simulation methods (See, for instance; Pérignon and Smith, 2010; Aramonte et al. 2013). The former does not account for time-varying volatility and correlations while the latter does. The filtered historical simulation method actually applies the historical simulation method on devolatilized (or filtered) portfolio returns, where devolatilized returns are obtained via a univariate conditional
variance or a multivariate conditional covariance estimation procedure. See, for instance, Engle (2002), Giannopoulos and Tunaru (2005), Aramonte et al. (2013), and Gurrola-Perez and Murphy (2015).

We propose the use of the filtered historical simulation procedure with the high-dimensional conditional covariance matrix estimator of Trucós et al. (2020a) to estimate VaR and ES in large portfolios. Our choice for the covariance matrix estimator is motivated on its feasibility and good performance evidenced in the Monte Carlo experiments as well as the empirical application in Trucós et al. (2020a) and on its fast computation time compared to alternatives state-of-the-art procedures in a high-dimensional context. As for our choice of the filtered historical simulation methodology, it is based on practitioners’ familiarity. The methodology is applied on panels of 652 cross-section stock returns and 750 trading days in a rolling window scheme obtaining satisfactory results.

The rest of the paper is organized as follows. In Section 2, we introduce the general dynamic factor model and the conditional covariance matrix estimation procedure of Trucós et al. (2020a). In Section 3 we define the VaR and ES forecasts and describe the filtered historical simulation process. Section 4 introduces the various backtesting tools used to evaluate the accuracy of VaR and ES forecasts. In Section 5 we conduct an empirical out-of-sample analysis comparative study of our method and its competitors. Section 6 concludes.

2 The general dynamic factor model with conditional heteroscedastic factors

The general dynamic factor model (GDFM) was introduced by Forni et al. (2000) and encompasses most other high-dimensional factor models proposed in the econometric and time series literatures.

Let \( \{X_t := (X_{1t}, X_{2t}, \ldots, X_{dt}, \ldots)', t \in \mathbb{Z}\} \), be a double-indexed zero-mean second-order stationary stochastic process, where \( i \) is a cross-sectional index and \( t \) stand for time. The GDFM is based on the decomposition of \( X_{it} \) into a common component \( \chi_{it} \) and an idiosyncratic one \( \xi_{it} \) driven by common \( (u_t) \) and idiosyncratic \( (v_{it}) \) shocks, respectively.

Letting \( X_{nt} := \{X_{it} | i = 1, \ldots, n, t \in \mathbb{Z}\} \), \( \chi_{nt} := \{\chi_{it} | i = 1, \ldots, n, t \in \mathbb{Z}\} \), and \( \xi_{nt} := \{\xi_{it} | i = 1, \ldots, n, t \in \mathbb{Z}\} \), the GDFM decomposition in vector notation takes the form

\[
X_{nt} = \chi_{nt} + \xi_{nt} = B_n(L)u_t + D_n(L)v_{nt}, \quad n \in \mathbb{N}_0, \quad t \in \mathbb{Z} \tag{1}
\]

with

\[
B_n(L) := (b_1(L) \ldots b_n(L))', \quad D_n(L) := \text{diag}(d_1(L) \ldots d_n(L)),
\]

\[
u_t := (u_{1t} u_{2t} \ldots u_{qt})', \quad \text{and} \quad v_{nt} := (v_{1t} \ldots v_{nt})'.
\]
where \( \{u_t, v_{nt}\} \) is orthonormal white noise.

The models considered in [Bai and Ng (2002)] or [Stock and Watson (2002a,b)] are particular cases under which \( B_n(L)u_t \) reduces to \( B_nF_t \), where \( B_n \) is an \( n \times r \) matrix of loadings and the factors \( F_t = (F_{1t}, \ldots, F_{rt})' \) span an \( r \)-dimensional factor space. The loadings, in these models, are static; for convenience, call them static factor models.

Consistent estimation procedures for the GDFM have been proposed under various assumptions in [Forni et al. (2000, 2005)] and [Forni and Lippi (2011)]; the most general results are those of [Forni et al. (2015, 2017)]. Contrary to [Bai and Ng (2002)] or [Stock and Watson (2002a,b)], their estimation procedure, which we now briefly describe, accommodates infinite-dimensional factor spaces, yet only involves one-sided filters—an essential feature in a forecasting context where “future” observations are not available.

### 2.1 The GDFM estimation procedure

For an observed panel \( X_{nT} \) with cross-sectional dimension \( n \) and sample size \( T \), the GDFM estimation procedure of [Forni et al. (2017)] proceeds as follows; the assumptions made are the same as in [Trucios et al. (2020a)], where we refer to for details.

- **Step 1.** Determine the number \( q \) of common shocks via an information criterion, for instance, using [Hallin and Liska (2007)].
- **Step 2.** Randomly reorder the \( n \) observed series.
- **Step 3.** Estimate the spectral density matrix of \( X \) by
  \[
  \hat{\Sigma}_X^{nT}(-\theta) = \frac{1}{2\pi} \sum_{k=-M_T}^{M_T} e^{-ik\theta} K\left(\frac{k}{B_T}\right) \hat{\Gamma}_k^X \quad \theta \in [0, 2\pi]
  \]
  where \( K(\cdot) \) is a kernel function, \( M_T \) a truncation parameter, \( B_T \) a bandwidth, and \( \hat{\Gamma}_k^X \) the sample lag-\( k \) cross-covariance matrix.
- **Step 4.** Estimate the spectral density matrix of the common components by
  \[
  \hat{\Sigma}_{X}^{nT}(-\theta) := \hat{P}_X^{nT}(-\theta) \hat{\Lambda}_{X}^{nT}(-\theta) \hat{P}_X^{nT}(-\theta) \quad \theta \in [0, 2\pi]
  \]
  where \( \hat{\Lambda}_{X}^{nT}(\theta) \) is a \( q \times q \) diagonal matrix with diagonal elements the \( q \) largest eigenvalues of \( \hat{\Sigma}_X^{nT}(-\theta) \) and \( \hat{P}_X^{nT}(-\theta) \) (with complex conjugate \( \hat{P}_X^{nT}(-\theta) \)) is the \( n \times q \) matrix with the associated eigenvectors.
- **Step 5.** Let \( n^* := m(q+1) \) with \( m := \left\lceil \frac{n}{q+1} \right\rceil \) and denote by \( \hat{\Sigma}_{X}^{n^*T}(-\theta) \) the \( n^* \times n^* \) spectral density matrix corresponding to \( X_{n^*T} \).
Forni et al. (2015) show that there exists an $n^* \times n^*$ block-diagonal VAR filter $A_{n^*}(L)$ with diagonal blocks $A^k(L)$, $k = 1, \ldots, m$ such that

$$ Y_{n^*t} := A_{n^*}(L)X_{n^*t} = R_{n^*}u_t + A_{n^*}(L)\xi_{n^*t} $$

(2)

where the filtered process $Y_{n^*t}$ satisfies the assumptions of a static factor model with $r = q$ factors $F_t = u_t$, loading matrix $R_{n^*}$, and idiosyncratic component $A_{n^*}(L)\xi_{n^*t}$.

- **Step 6.** By inverse Fourier transform of $\hat{\Sigma}_{n^*T}(\theta)$, estimate the autocovariance matrices $\hat{\Gamma}^k$ of the $m$ sub-vectors

$$ \chi^k_t = (\chi(k-1)(q+1)+1,t \cdots \chi(k+1),t)^T, \quad k = 1, \ldots, m $$

of dimension $(q + 1)$. Based on these, compute, after order identification, the Yule-Walker estimators $\hat{A}^k(L)$ of the $m$ VAR filters $A^k(L)$ and stack them into a block-diagonal VAR matrix $\hat{A}_{n^*}(L)$. Compute $\hat{Y}_{n^*t} := \hat{A}_{n^*}(L)X_{n^*t}$.  

- **Step 7.** Based on the first $q$ standard principal components of $\hat{Y}_{n^*t}$, obtain\(^\dagger\) estimates $\hat{R}_{n^*}u_t$ of $R_{n^*}u_t$ and, via a Cholesky identification constraint\(^\ddagger\) the estimates $\hat{R}_{n^*}$ and $\hat{u}_t$ of $R_{n^*}$ and $u_t$; then, an estimate of the impulse-response function is $B_{n^*}(L) := [\hat{A}_{n^*}(L)]^{-1}\hat{R}_{n^*}$.  

- **Step 8.** Repeat Steps 2 through 7 (generating, as in Trucios et al. (2020a), 30 randomly chosen cross-sectional permutations): the final estimates (denoted as $\hat{R}_{n^*}$, $\hat{u}_t$, and $\hat{B}_{n^*}$) are obtained by averaging the estimates $\hat{R}_{n^*}$, $\hat{u}_t$, and $\hat{B}_{n^*}$ associated with each iteration. Let $\hat{x}_{nt} := \hat{B}_{n^*}(L)\hat{u}_t$ and $\xi_{nt} := X_{nt} - \hat{x}_{nt}$.  

The consistency of the method is established (see Forni et al. (2017) and Barigozzi and Hallin (2020)) under assumptions (stationarity, the existence of a spectral density matrix, etc.) which we do not reproduce here.

### 2.2 One-step ahead estimation of conditional covariance matrices

In order to exploit the dependence on the past of the conditional covariance matrix, additional assumptions on the volatilities of the common shocks and the idiosyncratic

\(^\dagger\)This consists in running, on the filtered $\hat{Y}_{n^*t}$, the estimation procedures proposed by Bai and Ng (2002) or Stock and Watson (2002a,b) for the common component of the static factor model (2); these procedures consist in projecting the observations onto the space spanned by the $q$ first eigenvectors of $\hat{Y}_{n^*t}$’s covariance matrix.

\(^\ddagger\)To be more precise, going back to model (2), consider the $q \times q$ matrix $B_q(0)$ with entries $b_{ij}(0)$, $i, j = 1, \ldots, q$, and the Cholesky factorization $HH'$ of $B_q(0)B_q(0)$: $H$ thus is the lower triangular matrix with positive diagonal satisfying $HH' = B_q(0)B_q(0)$. The identifiability constraints consist in imposing $u_i = HH'v^*(0)u_i$ and $b_{ij}(L) = b_{ij}(L)H_{ij}^{-1}(0)H$, $j = 1, \ldots, n$.  

5
components are to be made, which are similar to the assumptions in Alessi et al. (2009), Aramonte et al. (2013), and Trucios et al. (2020a). We assume that the heteroscedastic volatility dynamics of the common shocks are those of an MGARCH process stable by aggregation. The heteroscedastic volatility dynamics of each idiosyncratic component is modelled as a GARCH-type process, and the common shocks and idiosyncratic components are assumed to be conditionally uncorrelated.

Under these assumptions, Trucios et al. (2020a) show that the conditional co-variance matrix $V_{X_{nt}|t-1}$ of $X_{nt}$ (conditional on the sigma-field $\mathcal{F}_{t-1}$ generated by the observations $X_{n,t-1}, X_{n,t-2}, \ldots$) decomposes into

$$V_{X_{nt}|t-1} = R_n V^u_{t|t-1} R'_n + V^\xi_{nt|t-1}$$

where $V^u_{t|t-1}$ and $V^\xi_{nt|t-1}$ stand for the conditional covariance matrices of the common shocks and idiosyncratic components, respectively. Therefore, the estimation of (3) can be obtained as follows.

- **Step 9.** Estimate the one-step-ahead conditional covariance matrix of the whole panel $X_n$ as

$$\hat{V}_{X_{nt}|t-1} = \hat{R}_n \hat{V}^u_{t|t-1} \hat{R}'_n + \hat{V}^\xi_{nt|t-1},$$

where $\hat{V}^u_{t|t-1}$ and $\hat{V}^\xi_{nt|t-1}$ are the estimated one-step-ahead conditional covariance matrices of the estimated common shocks and idiosyncratic components, respectively.

Note that the estimator $\hat{V}_{X_{nt}|t-1}$ in (4) is feasible in high dimensions $n$ thanks to the fact that $\hat{V}^u_{t|t-1}$ is $q \times q$ with $q$ typically small, while $\hat{V}^\xi_{nt|t-1}$ is approximated by a diagonal matrix.

It is a well-documented fact, however, that estimation of MGARCH models such as the BEKK and VECH models are strongly affected by the initial values, suffer from convergence problems, and perform quite poorly (see, for instance; Lien et al., 2002; Manabu, 2015). Fortunately, the DCC estimators (Engle, 2002) achieve much better performance, even when the data-generating process is not a DCC (see, Chevallier, 2012; Laurent et al., 2012; de Almeida et al., 2018; Trucios et al., 2020a). The DCC estimator accordingly is used in all subsequent applications: call GDFM-CHF method this procedure.

3Stability by aggregation requires that if $y_t$ is driven by an MGARCH model, then $Py_t$ (where $P$ is a non-singular square matrix) is also driven by an MGARCH model of the same class. Examples of MGARCH models stable by aggregation are the BEKK and VECH models. See Chapter 10 of Francq and Zakoian (2019) for details.

4That diagonal matrix $\hat{V}^\xi_{t|t-1}$ thus neglects possible idiosyncratic cross-covariances. Idiosyncratic cross-covariances, however, are mild by definition (non-pervasive if not null) and do not affect the consistency of (4).
3 Value at Risk (VaR) and Expected Shortfall (ES) estimation

The one-step-ahead risk level $\alpha$ VaR of a series $\{r_t\}$ of returns at time $T+1$ is defined as

$$\text{VaR}^\alpha_{T+1} := \sup \{x \in \mathbb{R} : F(x|\mathcal{F}_T) \leq \alpha\},$$

where $x \mapsto F(x|\mathcal{F}_T)$ stands for the distribution function of $r_{T+1}$ conditional on the information $\mathcal{F}_T$ available up to time $T$: $\text{VaR}^\alpha_{T+1}$ thus is $r_{T+1}$’s $\alpha$-quantile conditional on $r_T, r_{T-1}, \ldots$. The one-step-ahead risk level $\alpha$ ES at time $T+1$ of $\{r_t\}$ similarly is defined as

$$\text{ES}^\alpha_{T+1} := \mathbb{E}[r_{T+1}|r_{T+1} < \text{VaR}^\alpha_{T+1}, \mathcal{F}_T].$$

In practice, of course, $\text{VaR}^\alpha_{T+1}$ and $\text{ES}^\alpha_{T+1}$ have to be estimated from a finite realization $r_1, \ldots, r_T$ of $\{r_t\}$ or, in case $r_t$ is the revenue of a portfolio (the notation $R_t$ will be used when this is to be emphasized), from a finite realization of the returns of the individual stocks composing the portfolio. Since these estimators of a quantity related to time $T + 1$ are to be computed at time $T$, we call them one-step-ahead forecasts.

Many procedures to estimate (forecast) the VaR and ES of a portfolio are available in the literature—see Nieto and Ruiz (2016) or Righi and Ceretta (2015) for recent reviews. The estimation of the conditional covariance matrix of individual stocks plays an important role in all these methods. When the number of stocks is high, however, this estimation is a major issue and, due to the curse of dimensionality, a number of methods cannot handle very large portfolios. To overcome this problem, we propose using the filtered historical simulation approach along with the estimator of the conditional covariance matrix of returns recently proposed by Trucós et al. (2020a).

3.1 Filtered Historical Simulation

Since its inception by Barone-Adesi et al. (1998) and Barone-Adesi et al. (1999), the so-called filtered historical simulation (FHS) method has been widely used to forecast VaRs (see, for instance, Engle (2002), Aramonte et al. (2013), Gurrola-Perez and Murphy (2015)). Its application in the ES estimation/forecast context was initiated by Giannopoulos and Tunaru (2005) and since then adopted by both academics and practitioners.

FHS modifies the classical historical simulation method to take into account the conditional heteroscedastic dynamics of financial returns by exploiting the power of conditional volatility models (Barone-Adesi et al. (1998), Giannopoulos and Tunaru (2005)). Unlike the widely-applied historical simulation method, which basically estimates the VaR and ES by using the $\alpha$-quantile of historical returns and the average of
returns falling below that VaR, FHS first scales the portfolio returns by their conditional volatilities and then re-scales back these devolatilized returns via the volatility forecast.

Let \( R_t := \omega' r_t \), where \( r_t \) stands for the \( N \)-dimensional vector of individual returns at time \( t \) in a portfolio with portfolio weights \( \omega = (\omega_1, \ldots, \omega_N) \). The general framework of the FHS method to forecast the one-step-ahead VaR and ES can be summarized in three steps (on top of the steps 1-9 previously described in Section 2).

- **Step 10.** First, using the historical returns, fit a volatility model and obtain the filtered (or devolatilized) portfolio returns \( \epsilon_t := R_t / \hat{\sigma}_t \), where \( \hat{\sigma}_t \) stands for the estimated volatility at time \( t = 1, \ldots, T \).

- **Step 11.** Second, generate \( B \) bootstrap samples \( \epsilon_1^*, \ldots, \epsilon_B^* \) from the devolatilized returns and then simulate \( B \) one-step-ahead portfolio returns by re-scaling back the bootstrapped devolatilized returns—namely, multiplying them by the volatility forecast \( \hat{\sigma}_{T+1} \), which yields \( B \) simulated one-step-ahead portfolio return forecasts \( R_{T+1}^{\epsilon_i} := \epsilon_i^* \hat{\sigma}_{T+1} \), for \( i = 1, \ldots, B \).

- **Step 12.** The one-step-ahead forecasts of VaR and ES are

\[
\hat{\text{VaR}}_{T+1}^\alpha := \sup \{ x : \hat{F}_{R_{T+1}^*}(x) \leq \alpha \} \quad (5)
\]

and

\[
\hat{\text{ES}}_{T+1}^\alpha = \sum_{i=1}^B \frac{R_{T+1}^{\epsilon_i} I[R_{T+1}^{\epsilon_i} < \hat{\text{VaR}}_{T+1}^\alpha]}{\sum_{i=1}^B I[R_{T+1}^{\epsilon_i} < \hat{\text{VaR}}_{T+1}^\alpha]}, \quad (6)
\]

respectively, where \( \hat{F}_{R_{T+1}^*}(x) \) is the empirical distribution of the simulated one-step-ahead portfolio returns \( R_{T+1}^{\epsilon_1}, \ldots, R_{T+1}^{\epsilon_B} \) and \( I[\cdot] \) denotes the indicator function.

Steps 10-12 above constitute the general FHS setting for one-step-ahead forecasting of VaR and ES. They can be combined with any estimators of the portfolio volatility, such as those proposed by Engle (2002), Giannopoulos and Tunaru (2005), Aramonte et al. (2013) and Gurrola-Perez and Murphy (2015). However, most covariance matrix estimators used in FHS are infeasible in high dimensions (Caporin and McAleer, 2014) or fail to exploit the information available on the dependence between individual returns.

Given the excellent performance of the high-dimensional conditional covariance matrix GDFM-based forecasts proposed in Trucós et al. (2020a), we propose to combine the flexibility of the FHS method with the power of the latter. Namely, the estimated volatilities we are using in Steps 10-11 are \( \hat{\sigma}_t := (\omega' \hat{H}_t \omega)^{1/2} \), where \( \hat{H}_t \) is the one-step ahead estimator of the conditional covariance matrix at time \( t \) proposed in Trucós et al. (2020a).
4 Backtesting

Backtesting is the widely-used set of statistical tools allowing for the assessment of confidence set and quantile-related estimation accuracy—in this case, the accuracy of VaR and ES forecasting. The empirical backtesting exercise we are conducting in Section 5 below is based on calibration tests and scoring functions, which we first briefly describe.

4.1 Calibration tests

Calibration tests are hypothesis testing methods aiming at the evaluation of VaR and ES forecasts accuracy. Different calibration tests are constructed taking into account different characteristics expected from the correct VaR and ES forecasts. In all cases, the null hypothesis can be interpreted as “the VaR/ES is correctly specified”, where the term “correct” is defined in different manners. If the null is not rejected the risk measure estimation procedure is considered adequate.

4.1.1 VaR Calibration tests

Call violation the occurrence of a portfolio return smaller than the corresponding estimated VaR.

More precisely, a sequence of violations is defined as

\[ I_t^\alpha := \mathbb{I}[r_t < \hat{\text{VaR}}^\alpha_t] \quad t = T + 1, \ldots, T + H, \quad (7) \]

where \( H \) stands for the length of some out-of-sample period. Let \( V_H := \sum_{t=T+1}^{T+H} I_t^\alpha \) denote the number of violations in the out-of-sample period: if the VaR forecasts \( \hat{\text{VaR}}_{T+1} \) were accurate, the observed proportion \( V_H/H \) of violations should be close to the nominal risk level \( \alpha \).

The unconditional coverage (UC) test of Kupiec (1995) is among the most popular methods in backtesting, and addresses the null hypothesis \( H_0^\alpha \) under which the violations \( (7) \) constitute an i.i.d. Bernoulli sequence with parameter \( \alpha \). The UC test statistic is usually defined as

\[ UC^\alpha = -2 \log \frac{L(\alpha)}{L(V_H/H)} \quad (8) \]

where \( L(p) = (1 - p)^{V_H} p^{V_H} \) is the likelihood function of a binomial distribution with parameter \( p \). The null distribution of that statistic is asymptotically \( \chi^2_1 \) as \( H \to \infty \) and chi-square quantiles are usually used as critical values to perform the test.

---

\[ ^5 \text{With a slight abuse of notation, } T \text{ is used here instead of } T_0 < T, \text{ say, to denote the size of the training sample. Similarly, } r_t \text{ stands for a series of returns—individual stock returns or portfolio returns, depending on the context.} \]
The UC test, however, is insensitive to alternatives of serial dependence among the violations (such as clustered violations, indicating positive serial dependence). In order to deal with such alternatives, Christoffersen (1998) developed the so-called conditional coverage (CC) test which is based on a first-order Markov chain model with transition matrix

\[ \Pi_1 := \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix}, \tag{9} \]

where \( \pi_{01} \) is the probability of a violation at time \( t + 1 \) conditional on no violation at time \( t \), \( \pi_{11} \) the probability of a violation at time \( t + 1 \) conditional on a violation taking place at time \( t \); under \( H_{\alpha}^0 \) (the violations are i.i.d. Bernoulli with parameter \( \alpha \)), one has \( \pi_{01} = \pi_{11} = \alpha \). Letting

\[ T_{ij} := \sum_{t=T+1}^{T+H-1} I[I_t^\alpha = i \text{ and } I_{t+1}^\alpha = j], \quad i, j = \{0, 1\}, \]

the CC test statistic is

\[ CC^\alpha := -2 \log \left[ \frac{L(\alpha)}{L_{\Pi_1}} \right] \tag{10} \]

where \( L(\alpha) \), as in (8), is the likelihood under \( H_{\alpha}^0 \),

\[ L_{\Pi_1} := (1 - \pi_{01})^{T_{01}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{11}} \pi_{11}^{T_{11}} \]

the likelihood under the alternative of a Markov chain with transition matrix \( \Pi_1 \), and

\[ \hat{\Pi}_1 := \left( \begin{array}{cc} 1 - T_{01}/V_{H-1} & T_{01}/V_{H-1} \\ 1 - T_{11}/(H - 1 - V_{H-1}) & T_{11}/(H - 1 - V_{H-1}) \end{array} \right) \]

the maximum likelihood estimator version of \( \Pi_1 \).

The dynamic quantile (DQ) test proposed by Engle and Manganelli (2004) is based on the concept of hits. A hit at time \( t \in \{T + 1, \ldots, T + H\} \), is defined as \( \text{Hit}_t^\alpha := I_t^\alpha - \alpha \). Under \( H_{\alpha}^0 \), the sequence of hits is a martingale difference sequence, hence has expected value zero and is uncorrelated with both its own lagged values and the estimated VaR. The DQ test is a Wald-type test of the hypothesis that all coefficients in the linear regression

\[ \text{Hit}_t^\alpha = \beta_0 + \sum_{i=1}^{L} \beta_i \text{Hit}_{t-i}^\alpha + \beta_{L+1} \text{VaR}_{t-1}^\alpha + \epsilon_t \tag{11} \]

are equal to zero. The DQ test statistic is

\[ DQ^\alpha := H^{-1} \hat{\beta}' X \hat{\beta} / \alpha (1 - \alpha) \tag{12} \]

where \( \hat{\beta} := (\hat{\beta}_0, \ldots, \hat{\beta}_{L+1})' \) is the vector of OLS estimators of \( \beta := (\beta_0, \ldots, \beta_{L+1})' \), \( X \) is the \( H \times (L + 2) \) matrix with columns \((1, \text{Hit}_{t-1}^\alpha, \ldots, \text{Hit}_{t-L}^\alpha, \text{VaR}_{t-1}^\alpha)'\).
for $t = T + 2, \ldots, T + H$; the number $L$ of lags is commonly set as $L = 4$. Critical values are obtained from the asymptotic (as $H \to \infty$) null distribution of $DQ^\alpha$, which is chi-square with $L + 2$ degrees of freedom.

Finally, based on the fact that the VaR actually is a quantile of the conditional distribution of returns, Gaglianone et al. (2011) proposed the VaR quantile (VQ) regression test in which the null is that the coefficients in the $\alpha$-quantile regression

$$r_t = \beta_0 + \beta_1 \overline{\text{VaR}}^\alpha_t + \epsilon_t,$$

are $\beta_0 = 0$ and $\beta_1 = 1$. The test statistic is

$$\text{VQ}^\alpha := \hat{\theta}' M^{-1} \hat{\theta}$$

where $\hat{\theta} := (\hat{\beta}_0, \hat{\beta}_1 - 1)'$ is the OLS estimator of $\theta := (\beta_0, \beta_1 - 1)'$ and $M$ is $\hat{\theta}$'s covariance matrix. Critical values are obtained from the asymptotic (as $H \to \infty$) null distribution of $\text{VQ}^\alpha$, which is chi-square with two degrees of freedom.

### 4.1.2 ES Calibration tests

Turning to expected shortfall, the most common calibration test is the exceedance residual (ER) test of McNeil and Frey (2000). It is based on the concept of standardized residual exceedances over the VaR, namely,

$$e_t := \frac{r_t - \overline{\text{ES}}^\alpha_t}{\sigma_t} \mathbb{I}[r_t < \text{VaR}^\alpha_t], \text{ for } t = T + 1, \ldots, T + H.$$  

(15)

The null hypothesis $K^\alpha_0$ to be tested is $E(e_t| r_t < \text{VaR}^\alpha_t) = 0$, with one-sided alternatives of the form $E(e_t| r_t < \text{VaR}^\alpha_t) < 0$. Rejecting $K^\alpha_0$ thus implies that the risk is underestimated (a dangerous situation). The sample counterpart $\hat{e}_t$ of $e_t$ is obtained by replacing $\overline{\text{ES}}^\alpha_t$, $\text{VaR}^\alpha_t$, and $\sigma_t$ by their estimated values $\hat{\overline{\text{ES}}}^\alpha_t$, $\hat{\text{VaR}}^\alpha_t$, and $\hat{\sigma}_t$, yielding the test statistic

$$\text{ER}^\alpha := \sqrt{V_H m_{\hat{e}}/s_{\hat{e}}},$$

(16)

where $V_H$ is the number of VaR violations as defined in Section 4.1.1 $m_{\hat{e}}$ and $s_{\hat{e}}$ the sample mean and sample standard deviation of $\{\hat{e}_t| I^\alpha_t = 1\}$—the subsample of $\hat{e}_t$ values at which violations occur. The null distribution of $\text{ER}^\alpha$ is obtained via non-parametric bootstrap.

Another popular calibration test is the conditional calibration test (CoC) proposed by Nolde et al. (2017). The idea consists in checking whether the forecasts $\hat{\text{VaR}}^\alpha_t$ and $\hat{\overline{\text{ES}}}^\alpha_t$ are “optimal.” Optimality here means satisfying

$$E(W(\hat{\text{VaR}}^\alpha_t, \hat{\overline{\text{ES}}}^\alpha_t, r_t)| F_{t-1}) = 0,$$

(17)
for some adequate (vector-valued) function $W$ (see Nolde et al., 2017). Here, we use the $W$ function

$$W(R_1, R_2, x) := \left( \alpha - I(x < R_1) \right) \left( R_2 - R_1 + I(x < R_1)(R_1 - x)/\alpha \right).$$

(17)

The CoC test statistic, of the Wald type, is

$$\text{CoC}^\alpha := H \left( \frac{1}{H} \sum_{t=T+1}^{T+H} W(\hat{\text{VaR}}^\alpha_t, \hat{\text{ES}}^\alpha_t, r_t) \right)' \hat{\Omega}_H^{-1} \left( \frac{1}{H} \sum_{t=T+1}^{T+H} W(\hat{\text{VaR}}^\alpha_t, \hat{\text{ES}}^\alpha_t, r_t) \right),$$

(18)

where

$$\hat{\Omega}_H := \frac{1}{H} \sum_{t=T+1}^{T+H} \left( W(\hat{\text{VaR}}^\alpha_t, \hat{\text{ES}}^\alpha_t, r_t) \right) \left( W(\hat{\text{VaR}}^\alpha_t, \hat{\text{ES}}^\alpha_t, r_t) \right)' ;$$

its (asymptotic, as $H \to \infty$) null distribution is chi-square with two degrees of freedom.

More recently, Bayer and Dimitriadis (2020) proposed yet another approach, the so-called Expected Shortfall Regression (ESR) method. That method considers (Dimitriadis et al., 2019; Patton et al., 2019) a double regression model of the form (similar to (13))

$$r_t = V_t' \beta + u^q_t \quad \text{and} \quad r_t = W_t' \gamma + u^e_t, \quad t = T + 1, \ldots, T + H$$

(19)

where the conditional $\alpha$-quantile of $u^q_t$ given $u^q_{t-1}, u^q_{t-2}, \ldots$ (viz., the level-$\alpha$ VaR$^\alpha$ associated with $u^q_t$) and the level-$\alpha$ expected shortfall ES$^\alpha$ associated with $u^e_t$ are zero; $V_t$ and $W_t$ are covariate vectors.

The Wald-type test statistic for a null hypothesis of the form $H_0 : \gamma = \gamma_0$ is

$$H(\hat{\gamma} - \gamma_0)\hat{\Omega}_\hat{\gamma}^{-1}(\hat{\gamma} - \gamma_0)' ,$$

(20)

where $\hat{\gamma}$ is a root-$H$ consistent estimator of $\gamma$ and $\hat{\Omega}_\hat{\gamma}$ a consistent estimator of $\hat{\gamma}$’s covariance matrix; its null distribution is asymptotically (as $H \to \infty$) chi-square with $q$ degrees of freedom ($q$ the dimension of $\gamma$, which is either one or two below).

The authors proposed three different versions of the ESR test. Denoting by $\hat{\text{VaR}}^\alpha_t$ and $\hat{\text{ES}}^\alpha_t$ the VaRs and ESs estimated in the previous sections, the first version, called the strict ESR test, is obtained for $V_t = W_t = (1, \hat{\text{ES}}^\alpha_t)'$. The second version, called auxiliary ESR test, is obtained with $V_t = (1, \hat{\text{VaR}}^\alpha_t)'$ and $W_t = (1, \hat{\text{ES}}^\alpha_t)'$. In both cases, $\gamma_0 = (0, 1)'$ is used. These two versions are intrinsically two-sided, hence provide no information about under- or over-estimation of VaR and ES. A third

---

6Note that the sigma-fields generated by $u^q_{t-1}, u^q_{t-2}, \ldots$ and $u^e_{t-1}, u^e_{t-2}, \ldots$ actually coincide.
version of ESR, called *intercept* ESR test, has been proposed to take care of that drawback. The regression equations (19) are replaced with

\[ r_t - \hat{E}S_t^\alpha = \beta_0 + \beta_1 \hat{ES}_t^\alpha + u_t^r \quad \text{and} \quad r_t - \hat{E}S_t^\alpha = \gamma_1 + u_t^r, \quad t = T + 1, \ldots, T + H \]

and the problem consists in testing \( H_0 : \gamma_1 \geq 0 \) against \( H_1 : \gamma_1 < 0 \). The test statistic is \( H^{1/2}/\hat{\Omega} \hat{\gamma}_1 \) which under the null is asymptotically standard normal.

Denote by ESR1, ESR2, and ESR3, respectively, these three versions of the ESR method.

### 4.2 Scoring functions

It often happens that several forecast procedures successfully pass (no rejection) the VaR and ES calibration tests. A criterion then is needed to choose the “best one(s)”. This is the purpose of scoring functions (also known as loss functions): the smaller the scoring function, the better the forecasting procedure.

A scoring function \( S \) is called **strictly consistent** for a risk measure \( \rho(X) \) if

\[ \mathbb{E}(S(\rho(X), X)) < \mathbb{E}(S(r(X), X)) \quad \text{for all} \quad r(X) \neq \rho(X); \]

the risk measure \( \rho(X) \) is called **elicitable** if it admits a strictly consistent scoring function.

The VaR is an elicitable risk measure and \( \text{VaR}^\alpha_t \) admits the strictly consistent scoring function

\[
S(\text{VaR}^\alpha_t, r_t) = (\alpha - \mathbb{I}[r_t \leq \text{VaR}^\alpha_t])(G(r_t) - G(\text{VaR}^\alpha_t)), \quad (21)
\]

where \( \alpha \) is the risk level, \( \mathbb{I}[\cdot] \) is the indicator function, and \( G(\cdot) \) is a strictly increasing function. A commonly used scoring function for the VaR is the quantile loss function (González-Rivera et al. 2004), which is obtained when \( G(\cdot) \) is the identity function.

Contrary to the VaR, the ES is not elicitable but, as showed by Fissler et al. (2016), the pair (VaR, ES) is. Consistent scoring functions for \( (\text{VaR}^\alpha_t, \text{ES}^\alpha_t) \) take the form

\[
S((\text{VaR}^\alpha_t, \text{ES}^\alpha_t), r_t) = (\mathbb{I}[r_t \leq \text{VaR}^\alpha_t] - \alpha)G_1(\text{VaR}^\alpha_t) - \mathbb{I}[r_t \leq \text{VaR}^\alpha_t] G_1(r_t) + G_2(\text{ES}^\alpha_t)(\text{ES}^\alpha_t - \text{VaR}^\alpha_t + \mathbb{I}[r_t \leq \text{VaR}^\alpha_t])(\text{VaR}^\alpha_t - r_t)/\alpha + G_3(\text{ES}^\alpha_t) + G_4(r_t), \quad (22)
\]

provided that the functions \( G_1-G_4 \) satisfy some conditions for which we refer to Fissler et al. (2016), Fissler and Ziegel (2016) and Taylor (2020).

In Section 5, we are using three distinct scoring functions for a joint evaluation of VaR and ES forecasts. The first one, denoted as FZG, was proposed by Fissler et al. (2016) and is obtained for

\[ G_1(x) = x, \quad G_2(x) = e^x/(1 + e^x), \quad G_3(x) = \log(1 + e^x), \quad \text{and} \quad G_4(x) = \log(2). \]
The second one, denoted as NZ and proposed by Nolde et al. (2017), with
\[ G_1(x) = 0, \quad G_2(x) = \sqrt{-x}/2, \quad G_3(x) = -\sqrt{-x}, \quad \text{and} \quad G_4(x) = 0. \]

The last one, denoted as AL was proposed by Taylor (2019) and is obtained for
\[ G_1(x) = 0, \quad G_2(x) = -1/x, \quad G_3(x) = -\log(-x), \quad \text{and} \quad G_4(x) = 1 - \log(1 - \alpha). \]

As previously mentioned, a pair (VaR, ES) failing to reject the null hypotheses in calibration tests and achieving the smallest scoring function value is preferable.

5 Empirical Application

5.1 Data

Our dataset consists of daily closing prices of stocks used in the composition of the Standard & Poor’s 500 Index (S&P 500), the National Association of Securities Dealers Automated Quotations 100 Index (NASDAQ-100), and the NYSE Amex Composite Index (AMEX) from January 3, 2012 to July 1, 2020 (\( T = 2136 \) observations). Only data with no missing values in the entire sample period was considered, yielding a panel of \( N = 655 \) stocks.

Daily returns (in percentage) are obtained as \( r_{i,t} := 100 \times (P_{i,t}/P_{i,t-1} - 1) \) where \( P_{i,t} \) denotes the closing price of the \( i \)th stock at time \( t \). Following Engle et al. (2019), if pairs of stock returns are highly correlated (sample correlation larger than 0.95), the one with lower volume is removed from the panel. Using this criterion, the stocks DISCK, GOOGL and LBTYA were removed from the panel.

Figure 1 plots the sample correlation among the 652 stock returns over the full sample period, the highest correlation is 0.91 between CMS and XEL stocks while the lowest one is -0.19 between the stocks APT and CCL.

5.2 Out-of-sample analysis

For the sake of simplicity, we throughout consider an equal-weight portfolio, which means that the portfolio returns are \( R_t = \sum_{i=1}^{652} \omega_i r_{i,t} = (1/652) \sum_{i=1}^{652} r_{i,t} \). We consider the one-step-ahead forecasts of the VaR and ES of this portfolio at risk levels \( \alpha = 1\%, \ 2.5\%, \ \text{and} \ 5\% \) using the methodology described in Section 3.1 over a rolling window of 750 days (concentration ratio 652/750 = 0.87): observations 1–750 thus were used to forecast VaR and ES, as explained in Section 3.1 at time 751, then observations 2–751 to forecast VaR and ES at time 752, . . . , observations 1386–2135 to forecast VaR and ES at time 2136. This produces a total of 1386 forecasts, from December 29, 2014 through July 1, 2020.

For the sake of comparison, we are basing our forecasts on various conditional covariance estimators which were selected for their feasibility in a high-dimensional
Figure 1: Sample unconditional contemporaneous correlations between stock returns.

classical context (the dimension of the observations here is 652). Specifically, we include the Risk Metrics 2006 methodology (Zumbach 2007) denoted as RM2006, the dynamic conditional correlation with composite likelihood (Pakel et al., 2020) denoted as DCCc, and the conditional covariance estimation using the dynamic factor model approach with finite-dimensional space (Alessi et al., 2009; Aramonte et al., 2013) denoted as ABC; GDFM stands for the GDFM-CHF method we are proposing. Whenever univariate volatility models are needed (typically, for the prediction of idiosyncratic volatilities and the marginals of the MGARCH model in the common shocks), the GJR-GARCH model of Glosten et al. (1993) with Student-t distribution was used.

Table 1 reports the results of the out-of-sample VaR and ES backtesting exercise: the p-values are reported for the various forecasting methods (RM2006, DCCc, ABC, and GDFM) and calibration tests (UC, CC, DQ, and VQ; ER, CoC, ESR$_1$-ESR$_3$) described in Section 4.1. Shadowed cells indicate p-values larger than 0.05 (non-rejection at nominal level 5%); boldface scoring values indicate the lowest scoring
function in each column (corresponding to the “best” method). The model confidence set approach of Hansen et al. (2011) was also performed (at significance level 10%); the results, however, are non-informative (in almost all of cases, all four procedures are in the model confidence set) and are not reported here.

All procedures report percentages of violations that are close to the nominal risk levels. Regarding VaR forecasts, the GDFM-CHF method is doing better than its competitors: at significance level 5%, out of four distinct calibration tests, not a single rejection, while all other procedures get rejected at least once, indicating that, irrespective of the calibration method adopted, GDFM-CHF VaR forecasts are more accurate than RM2006, DCCc, and ABC. Note also that the QL scoring function, which only takes into account VaR performance, reaches its lowest values when the GDFM-CHF method is used. For ES forecasts, only the RM2006 and GDFM-CHF methods are achieving non-rejection in all ES calibration tests and for all risk levels considered, confirming the excellent performance of the GDFM approach.

As for scoring function evaluations, irrespective of the choice of the scoring function, the lowest average values are achieved by the GDFM VaR and ES estimates, meaning again that the GDFM procedure is performing best. Overall, our results support evidence in favour of the GDFM-CHF superiority to forecast the one-step-ahead VaR and ES.

<table>
<thead>
<tr>
<th>%</th>
<th>Viol</th>
<th>UC</th>
<th>CC</th>
<th>DQ</th>
<th>Calibration tests</th>
<th>Avg. Scoring functions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VQ</td>
<td>ER</td>
<td>CoC</td>
<td>ESR</td>
<td>ESR</td>
<td>ESR</td>
</tr>
<tr>
<td>1%</td>
<td>RM2006</td>
<td>1.40</td>
<td>0.189</td>
<td>0.000</td>
<td>0.000</td>
<td>0.558</td>
</tr>
<tr>
<td></td>
<td>DCCc</td>
<td>1.80</td>
<td>0.007</td>
<td>0.001</td>
<td>0.000</td>
<td>0.291</td>
</tr>
<tr>
<td></td>
<td>ABC</td>
<td>1.70</td>
<td>0.013</td>
<td>0.001</td>
<td>0.000</td>
<td>0.233</td>
</tr>
<tr>
<td></td>
<td>GDFM</td>
<td>1.20</td>
<td>0.413</td>
<td>0.317</td>
<td>0.100</td>
<td>0.299</td>
</tr>
<tr>
<td>2.5%</td>
<td>RM2006</td>
<td>2.70</td>
<td>0.570</td>
<td>0.059</td>
<td>0.002</td>
<td>0.883</td>
</tr>
<tr>
<td></td>
<td>DCCc</td>
<td>3.00</td>
<td>0.288</td>
<td>0.014</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>ABC</td>
<td>3.10</td>
<td>0.166</td>
<td>0.014</td>
<td>0.000</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>GDFM</td>
<td>3.00</td>
<td>0.288</td>
<td>0.453</td>
<td>0.065</td>
<td>0.767</td>
</tr>
<tr>
<td>5%</td>
<td>RM2006</td>
<td>5.60</td>
<td>0.293</td>
<td>0.003</td>
<td>0.000</td>
<td>0.665</td>
</tr>
<tr>
<td></td>
<td>DCCc</td>
<td>6.10</td>
<td>0.061</td>
<td>0.000</td>
<td>0.000</td>
<td>0.326</td>
</tr>
<tr>
<td></td>
<td>ABC</td>
<td>5.60</td>
<td>0.351</td>
<td>0.005</td>
<td>0.001</td>
<td>0.435</td>
</tr>
<tr>
<td></td>
<td>GDFM</td>
<td>5.80</td>
<td>0.160</td>
<td>0.064</td>
<td>0.086</td>
<td>0.144</td>
</tr>
</tbody>
</table>

Table 1: One-step-ahead VaR and ES backtesting for risk levels $\alpha = 1\%$ (top panel), 2.5% (middle panel), and 5% (bottom panel), based on various calibration tests and scoring functions. Out-of-sample period from September 18, 2019 through July 1, 2020 (1386 observations). RM2006 is the Risk Metrics 2006 methodology (Zumbach, 2007); DCCc is the Dynamic Conditional Correlation with composite likelihood (Pakel et al., 2020); ABC is the conditional covariance estimation based on a Dynamic Factor Model with finite-dimensional factor space (Alessi et al., 2009; Arancione et al., 2013); GDFM is the method we are proposing. Shadowed cells indicate calibration tests with $p$-values larger than 0.05; boldface averaged scoring function values indicate the lowest values in the column (that is, the “winning method”).

Figure 2 plots the out-of-sample equal-weight portfolio returns along with their
one-step-ahead VaR forecasts at risk levels $\alpha = 1\%$ (top panel), $2.5\%$ (middle panel), and $5\%$ (bottom panel), for the four methods (RM2006, DCCc, ABC, and GDFM). The vertical dotted line stands for March 12, 2020, one day after the COVID-19 virus has been declared a pandemic by the World Health Organisation. Note that, after March 12, 2020, the VaR estimates based on the RM2006 and DCCc methods still are strongly affected by the large March 12 shock. As a consequence, the VaR forecasts are larger than should be for quite some time. For a better visualization, Figure 3 in Appendix reports the same information over a smaller out-of-sample period ranging from January 2, 2018 to June 28, 2019.

The early February 2018 VaR spike observed in the DCCc column takes place when the stock market smacked into spiking bond rates that were pricing in the threat of inflation\footnote{https://money.cnn.com/2018/02/28/investing/stock-market-february-dow-jones/index.html}. In that period the VIX Volatility Index doubled, reflecting the investors’ sentiment. As expected, volatility forecasts were also affected. Although the increase in the estimated volatility is observed in all procedures, the DCCc forecasts are by far the most seriously affected. On the other hand, the 2019 VaR spikes in the ABC column are due to the presence of extreme observation in the in-sample period, which cause bad estimation of the common shocks. Although the influence of outliers in the GDFM estimation process when the space spanned by the common factor is finite have not been analyzed yet, our own analysis in such case (not showed here) reveals that the common shocks present large peaks and spikes at some points of the in-sample period and very small volatility in the remaining periods. This common shock behaviour is very similar to that observed in the supplementary material of [Trucios et al. 2020] when dealing with outliers in the GDFM with infinite-dimensional factor space. The common shock behaviour in the ABC procedure, thus, can be explained by the presence of outliers in the in-sample period.

The consequences of neglecting the possibility of additive outliers in the volatility estimation process in most of the recently proposed high-dimensional conditional covariance matrix estimators have not been analyzed yet, but it is widely known that additive outliers badly affect volatility estimation in other contexts (univariate and moderate-dimensional multivariate). Therefore, the fact that the presence of extreme observations can affect the performance of high-dimensional conditional covariance matrix estimators is not surprising at all. We refer to [Boudt et al. 2013], [Grané and Veiga 2014], [Hotta and Trucio 2018], [Trucios et al. 2019] and the references therein for a discussion of the influence of outliers on other volatility models. Note, however, that while extreme observations have an impact on all methods, the less seriously affected ones are the RM2006 and the GDFM-CHF procedures.

The lack of performance of the alternatives procedures is not only due to the
Figure 2: Out-of-sample portfolio returns (black solid line) and the one-step-ahead forecasts VaR (red dashed line) at risk levels $\alpha = 1\%$ (top panel), 2.5\% (middle panel), and 5\% (bottom panel) based on the RM2006, DCCc, ABC, and GDFM-CHF procedures, respectively.
COVID-19 period (starting on March 12, 2020). Table 2 reports the same backtesting exercise displayed in Table 1 but using out-of-sample data until March 11, 2020 (a day before the largest spike observed in the out-of-sample period). The results in all scenarios are qualitatively the same as the obtained in Table 1, evidencing the superiority of the GDFM-CHF against the alternative approaches.

<table>
<thead>
<tr>
<th>Calibration tests</th>
<th>Avg. Scoring functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>RM2006</td>
<td>DCCc</td>
</tr>
<tr>
<td>Viol. UC CC DQ VQ ER CoC ESR ER2 ESR3</td>
<td>QL FZG NZ AL</td>
</tr>
<tr>
<td>1%</td>
<td>1.50</td>
</tr>
<tr>
<td>2.5%</td>
<td>2.70</td>
</tr>
<tr>
<td>5%</td>
<td>5.70</td>
</tr>
</tbody>
</table>

Table 2: One-step-ahead VaR and ES backtesting for risk levels $\alpha = 1\%$ (top panel), 2.5% (middle panel), and 5% (bottom panel), based on various calibration tests and scoring functions. Out-of-sample period from January 2, 2018 through June 28, 2019. RM2006 is the Risk Metrics 2006 methodology [Zumbach, 2007]; DCCc is the Dynamic Conditional Correlation with composite likelihood [Pakel et al., 2020]; ABC is the conditional covariance estimation based on a Dynamic Factor Model with finite-dimensional factor space [Alessi et al., 2009; Aramonte et al., 2013]; GDFM is the method we are proposing. Shadowed cells indicate calibration tests with $p$-values larger than 0.05; boldface averaged scoring function values indicate the lowest values in the column (that is, the “winning method”).

Due to the short COVID-19 period in our sample (77 trading days), results solely for this period are not performed since the power of the statistical tests in this small sample size should be strongly affected and could yield misleading conclusions.

6 Conclusions

In this paper, we deal with one-step-ahead risk measures (VaR and ES) forecasts in high-dimensional portfolios. We suggest to use a filtered historical simulation method with the conditional covariance estimator of Trucios et al. (2020a), itself based on the general dynamic factor model with infinite-dimensional factor space of Forni et al. (2013, 2017).

The procedure is empirically evaluated on a large portfolio of 652 stocks returns and 750 trading days using a comprehensive backtesting exercise based on calibration tests and scoring functions. Our results indicate that, both for VaR and ES, our procedure is outperforming its competitors—including the ABC method based on a
dynamic factor model with finite-dimensional factor space (a static factor model). Some high-dimensional conditional covariance matrix estimators seem to be more affected by extreme observation and outliers than some others. A detailed analysis of this fact as well as robust versions of those estimators are the subject of our current research.

References


23
Figure 3: Out-of-sample portfolio returns (black solid line) and the one-step-ahead VaR forecasts (red dashed line) at risk levels $\alpha = 1\%$ (top panel), 2.5\% (middle panel), and 5\% (bottom panel) based on the RM2006, DCCc, ABC, and GDFM-CHF procedures, respectively, from January 2, 2018 through June 28, 2019.