Econometric Theory, **41**, 2025, 341–390.
doi:10.1017/S0266466623000233

THE ESTIMATION RISK IN EXTREME SYSTEMIC RISK FORECASTS

YANNICK HOGA
University of Duisburg-Essen

Systemic risk measures have been shown to be predictive of financial crises and declines in real activity. Thus, forecasting them is of major importance in finance and economics. In this paper, we propose a new forecasting method for systemic risk as measured by the marginal expected shortfall (MES). It is based on first de-volatilizing the observations and, then, calculating systemic risk for the residuals using an estimator based on extreme value theory. We show the validity of the method by establishing the asymptotic normality of the MES forecasts. The good finite-sample coverage of the implied MES forecast intervals is confirmed in simulations. An empirical application to major U.S. banks illustrates the significant time variation in the precision of MES forecasts, and explores the implications of this fact from a regulatory perspective.

1. MOTIVATION

The financial crisis of 2007–2009 has sparked regulatory and academic interest in assessing systemic risk. For instance, in April 2009 G20 leaders asked national regulators to develop guidelines for the assessment of the systemic importance of financial institutions, which were provided in a joint report by the International Monetary Fund (IMF), the Bank for International Settlements (BIS), and the Financial Stability Board (FSB) (IMF/BIS/FSB, 2009). By now, these early developments have manifested themselves in official regulations. For example, the class of global systemically important banks (G-SIBs) is divided into buckets from 1 to 5, where banks in bucket 5 have to hold the highest additional capital buffers (FSB, 2021). The Basel framework, that sets out the methodology determining G-SIB membership, closely relies on systemic risk assessments (Basel Committee on Banking Supervision, 2019, SCO40).

Next to its regulatory importance, forecasting systemic risk is important in various other contexts. First, one hallmark of financial crises is that asset prices start to co-move. To measure the extent to which prices move in lockstep, several systemic risk measures may be used (Acharya et al., 2017; Adrian and Brunnermeier, 2016; Billio et al., 2012). Since the tendency of asset prices to co-move

The author would like to thank the Co-Editor Eric Renault and the three anonymous referees for their insightful comments that significantly improved the quality of the paper. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through project 460479886. Address correspondence to Yannick Hoga, Faculty of Economics and Business Administration, University of Duisburg-Essen, Essen, Germany; e-mail: yannick.hoga@vwl.uni-due.de

implies that diversification benefits are seriously reduced, it becomes important to predict systemic risk measures as indicators of diversification meltdown. Second, investigating the link between the real economy and the financial sector, Allen, Bali, and Tang (2012), Giglio, Kelly, and Pruitt (2016), and Brownlees and Engle (2017) find an increase in systemic risk to be predictive of future declines in real activity. These examples underscore the importance of accurately forecasting systemic risk.

Predictions of systemic risk may be produced from a variety of models, such as multivariate GARCH models or quantile regression models (Adrian and Brunnermeier, 2016; Girardi and Tolga Ergün, 2013). However, very little is known about the asymptotic properties (e.g., consistency or asymptotic normality) of systemic risk forecasts issued from these or other models. This contrasts with the large literature developing asymptotic theory for forecasts of *univariate* risk measures, such as the Value-at-Risk (VaR) or the expected shortfall (ES) (Chan et al., 2007; Francq and Zakoïan, 2015; Gao and Song, 2008; Hoga, 2019). Therefore, it is the main aim of this paper to fill this gap for systemic risk forecasts. Specifically, we establish conditions under which systemic risk forecasts issued from a general class of multivariate GARCH-type models are (consistent and) asymptotically normal.

Of course, consistency is vital for the point forecasts to reflect actual levels of systemic risk. However, point forecasts of systemic risk are only of limited value, since they lack a measure of uncertainty. To illustrate the importance of confidence intervals around point risk forecasts, Christoffersen and Gonçalves (2005) give the example of a portfolio manager allowed to take on portfolios with a VaR of at most 15% of current capital. A VaR point estimate of 13% would not indicate any need for rebalancing, yet a (say) 90%-confidence interval of 10–16% would induce the prudent portfolio manager to do so. Clearly, a similar case can be made for the importance of confidence intervals for systemic risk forecasts, which our asymptotic normality result allows us to construct.

As our systemic risk measure, we use the marginal expected shortfall (MES) of Acharya et al. (2017). We do so because of the ability of MES to identify key contributors to systemic risk during financial crises and due to its predictive content for downturns in real economic activity (Acharya et al., 2017; Giglio et al., 2016). Furthermore, MES has an additivity property that is crucial for attributing systemic risk (Chen, Iyengar, and Moallemi, 2013). This may be useful for individual banks as well as the financial system as a whole. For the purposes of risk management or asset allocation, an individual bank may want to break down firm-wide losses into contributions from single units or trading desks. From the wider perspective of the financial system, the additivity property allows to decompose the system-wide ES into the sum of the MESs of all banks in the system. We also refer to the empirical application for an illustration of this property.

For MES to truly capture *systemic* risk, it needs to be forecasted far out in the tail. For instance, Acharya et al. (2017, p. 13) state that "[w]e can think of systemic events [...] as extreme tail events that happen once or twice a decade."

Consequently, only few meaningful observations are available for forecasting MES. To deal with this, our MES estimator is motivated by extreme value theory (EVT). By imposing weak assumptions on the joint tail, EVT-based methods alleviate the problem of data scarceness outside the center of the distribution. Indeed, numerous studies show that EVT-based estimators improve the forecast quality of univariate risk measures, such as the VaR or the ES (Bali, 2007; Bao, Lee, and Saltoğlu, 2006; Hoga, 2022; Kuester, Mittnik, and Paolella, 2006; McNeil and Frey, 2000). Hence, these methods have caught on in empirical work as well (Gupta and Liang, 2005). We stress that the case for using EVT-based estimators for systemic risk measures, where *joint* extremes are of interest, seems to be even stronger than in the univariate case because data are even scarcer in the joint tail.

A key ingredient of our MES forecast is the MES estimator of Cai et al. (2015), and in deriving asymptotic properties of MES forecasts, we build on their work. However, Cai et al. (2015) deal with unconditional MES estimation for independent and identically distributed (i.i.d.) random variables and, thus, their framework is inherently static. In contrast, we consider (conditional) MES forecasting in dynamic models. In particular, this requires taking volatility dynamics into account in the forecasts. From an econometric perspective, conditional MES has the advantage of incorporating current market conditions. Therefore, changes in market conditions are reflected in conditional MES, but not in its static version. This is akin to the standard deviation as a measure of risk. The conditional standard deviation (a.k.a. volatility) is a good measure of current risk, whereas the unconditional standard deviation only provides an average measure of risk.

We also extend our results (and the results of Cai et al., 2015) to a higher-dimensional setting, where we consider MES forecasts for multiple variables jointly. This is important because one of the main purposes of systemic risk measures is to explore linkages in complex systems. We prove the joint asymptotic normality of MES forecasts in the higher-dimensional case. To enable inference, we propose an estimator of the asymptotic variance–covariance matrix and show its consistency. Then, we demonstrate how our results can be used to test for equal systemic risk contributions of the different units in the system. This test is later on used in the empirical application. From a technical perspective, our higher-dimensional results draw on Hoga (2018), who explores tail index estimation for multivariate time series.

We confirm the good finite-sample coverage of our asymptotic confidence intervals for MES in simulations. We do so for the constant conditional correlation (CCC) GARCH model of Bollerslev (1990). Our main findings are that coverage improves the more extreme the risk level of the MES forecasts. Also, the better the (marginal and joint) tail can be approximated using extreme value methods, the more precise the forecasts tend to be in terms of root mean square error (RMSE) and the lengths of the forecast intervals.

Our empirical application considers the eight G-SIBs from the US. As expected, there is significant time variation in the levels of systemic risk as measured by the MES. Significant peaks in systemic risk can be observed during the financial crisis of 2007–2009, the European sovereign debt crisis in 2011 and the Corona stock market crash of March 2020. Computing our MES forecast intervals over time shows that it is particularly during times of crises (when accurate systemic risk assessments are needed most) that forecasts tend to be least precise (as measured by the lengths of the forecast intervals). We also apply our test for equal systemic risk contributions of each of the eight banks. Not surprisingly, the null of equality can be rejected for every single time point in our sample. This is consistent with the fact that the eight banks are assigned to different buckets in the G-SIB classification.

The remainder of the paper is structured as follows: Section 2 introduces the multivariate volatility model and our MES forecasts together with all required regularity conditions. Section 3 derives limit theory for MES forecasts and Section 4 extends this to multiple MES forecasts. Coverage of the confidence intervals for MES is assessed in the simulations in Section 5. The empirical application in Section 6 investigates MES forecasts for the eight U.S. G-SIBs. The final Section 7 concludes. All proofs are relegated to the Appendix.

2. PRELIMINARIES

We adopt the following notational conventions. Throughout, we use bold letters to denote vectors and matrices. In particular, $\boldsymbol{I}$ is the (2×2)-identity matrix. For any matrix $\boldsymbol{A}=(a_{ij})$, we will use the norm defined by $\|\boldsymbol{A}\|=\sum_{i,j}|a_{ij}|$. The transpose of a matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}'$ and its vectorization by $\mathrm{vec}(\boldsymbol{A})$, where the columns of $\boldsymbol{A}$ are stacked on top of each other. The diagonal matrix containing the elements of the vector $\boldsymbol{v}$ on the main diagonal is $\mathrm{diag}(\boldsymbol{v})$. We let $\overset{d}{=}$ stand for equality in distribution. For some random variable X, put $X^+=\max\{X,0\}$ and $X^-=X-X^+$. For scalar sequences a_n and b_n, we write $a_n\asymp b_n$ if both $a_n=O(b_n)$ and $b_n=O(a_n)$ hold, as $n\to\infty$, and we write $a_n\sim b_n$ if $a_n/b_n\to 1$.

2.1. Defining MES

Consider a sample $\{(X_t,Y_t)'\}_{t=-\ell_n+1,\dots,n}$ from the random variables of interest. In a forecasting situation, interest focuses on predicting systemic risk based on the current state of the market, which is captured by the information set $\mathcal{F}_n=\sigma\big((X_n,Y_n)',(X_{n-1},Y_{n-1})',\dots\big)$. Define $F_n(x,y)=\mathrm{P}\{X_{n+1}\le x, Y_{n+1}\le y\mid\mathcal{F}_n\}$ to be the conditional joint distribution function (d.f.) with marginals $F_{n,0}(x)=F_n(x,\infty)$ and $F_{n,1}(y)=F_n(\infty,y)$. Then, the (conditional) MES is defined as

$$\theta_{n,p}=\mathrm{E}\big[Y_{n+1}\mid X_{n+1}>\mathrm{VaR}_n(p),\mathcal{F}_n\big],$$

where $\mathrm{VaR}_n(p)=F_{n,0}^{\leftarrow}(1-p)$, with "$\leftarrow$" indicating the left-continuous inverse, is the VaR of the reference position for some "small" $p\in(0,1)$. Thus, under current market conditions, MES measures next period's average loss Y_{n+1} given that X_{n+1} is in distress.

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

An MES forecast (based on $\mathcal{F}_n$) may be of interest in a number of situations, for instance, when the Y_t are losses of one's own portfolio, and the X_t denote losses of some reference index, such as the S&P 500. The Y_t may also denote the losses of a single trading desk, and the X_t firm-wide losses. Alternatively, the Y_t may be the losses of a financial institution with X_t standing for system-wide losses. In each case, it may be of interest to understand how the two risk factors are connected, e.g., for the purposes of stress testing or to assess portfolio sensitivities. In these situations, the MES can serve as a (real-valued) summary of the degree of connectedness.

2.2. The Data Generating Process

Throughout, we suppose that the losses are generated from the model

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \boldsymbol{\Sigma}_t(\boldsymbol{\theta}^\circ)\boldsymbol{\varepsilon}_t, \tag{1}$$

where the true parameter vector $\boldsymbol{\theta}^\circ$ is an element of some parameter space $\boldsymbol{\Theta}$ and the diagonal matrix $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}^\circ) = \boldsymbol{\Sigma}_t = \operatorname{diag}(\boldsymbol{\sigma}_t)$ with $\boldsymbol{\sigma}_t = (\sigma_{t,X}, \sigma_{t,Y})'$ is $\mathcal{F}_{t-1}$-measurable. Moreover, the $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,X}, \varepsilon_{t,Y})'$ are independent of $\mathcal{F}_{t-1}$ and i.i.d. with mean zero, unit variance and correlation matrix $\boldsymbol{R}$. Thus, $\boldsymbol{\Sigma}_t$ contains the individual volatilities on the main diagonal, since

$$\operatorname{Var}\big((X_t, Y_t)' \mid \mathcal{F}_{t-1}\big) = \boldsymbol{\Sigma}_t \operatorname{Var}(\boldsymbol{\varepsilon}_t)\boldsymbol{\Sigma}_t' = \boldsymbol{\Sigma}_t \boldsymbol{R} \boldsymbol{\Sigma}_t' = \begin{pmatrix} \sigma_{t,X}^2 & \rho_{X,Y}\sigma_{t,X}\sigma_{t,Y} \\ \rho_{X,Y}\sigma_{t,X}\sigma_{t,Y} & \sigma_{t,Y}^2 \end{pmatrix},$$

where $\rho_{X,Y} = \operatorname{corr}(\varepsilon_{t,X}, \varepsilon_{t,Y})$ is the CCC of $(X_t, Y_t)' \mid \mathcal{F}_{t-1}$. We already mention here that estimating the correlation $\rho_{X,Y}$ of $\boldsymbol{\varepsilon}_t$ is not required for conditional MES forecasting. Rather, it is the unconditional MES of $\boldsymbol{\varepsilon}_t$ that will be required (see (2)).

The best known among the class of models in (1) is the CCC–GARCH model of Bollerslev (1990). But our framework also covers models incorporating volatility spillover, such as the extended (E)CCC–GARCH of Jeantheau (1998).

Remark 1. We work with CCC–GARCH-type models here. However, DCC–GARCH models, due to Engle (2002), have attained benchmark status among multivariate GARCH models because of their forecasting accuracy (Laurent, Rombouts, and Violante, 2012). We focus here on the former class of models for several reasons. First, the former models continue to be studied extensively in the literature (Conrad and Karanasos, 2010; He and Teräsvirta, 2004; Jeantheau, 1998; Nakatani and Teräsvirta, 2009). Second, as Francq and Zakoïan (2016, p. 620) point out, a full estimation theory for DCC–GARCH models is not available. Since MES forecasts necessarily require a parameter estimator with known asymptotic properties, developing limit theory for MES forecasts based on DCC–GARCH models is beyond the scope of the present paper. Third, our EVT-based estimation method exploits dependence between $\varepsilon_{t,X}$ and $\varepsilon_{t,Y}$. However, in DCC–GARCH models, the innovations are decorrelated (to ensure identification), in which case our MES estimator cannot be expected to work well. Fourth, higher-order measures of risk, such as MES, are not properly identified in DCC–GARCH models (Hafner, Herwartz, and Maxand, 2022). To see this, recall that in a DCC–GARCH framework, $(X_t, Y_t)' = \boldsymbol{\Sigma}_t \boldsymbol{\varepsilon}_t$ for i.i.d. $\boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \boldsymbol{I})$ and not necessarily diagonal $\boldsymbol{\Sigma}_t$. However, only $\mathrm{Var}\big((X_t, Y_t)' \mid \mathcal{F}_{t-1}\big) = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t' =: \boldsymbol{H}_t$ is modeled, while the model stays silent on the choice of (the non-unique) $\boldsymbol{\Sigma}_t$. However, different $\boldsymbol{\Sigma}_t$ imply different conditional distributions $(X_t, Y_t)' \mid \mathcal{F}_{t-1}$ and, hence, different values for MES. For instance, $\boldsymbol{\Sigma}_t$ may denote the symmetric square root implied by the eigenvalue decomposition ($\boldsymbol{\Sigma}_t^s$), or it may be the lower triangular matrix of the Cholesky decomposition ($\boldsymbol{\Sigma}_t^l$). Both decompositions are equivalent in the sense that both imply the same dynamics in second-order moments (as given in $\boldsymbol{H}_t$). Yet, when it comes to higher-order measures of risk (such as MES), the two decompositions imply different values for MES, as the next example illustrates.

Example 1. Suppose that $\varepsilon_{t,X}$ and $\varepsilon_{t,Y}$ have a (standardized) Student's t_5-distribution, independently of each other. Assume for simplicity that

$$\boldsymbol{H}_{n+1} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}, \quad \text{implying} \quad \boldsymbol{\Sigma}_{n+1}^s = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{n+1}^l = \begin{pmatrix} 2.23\ldots & 0 \\ 1.78\ldots & 1.34\ldots \end{pmatrix}.$$

Then, when the underlying structural model is $\boldsymbol{\Sigma}_t^s \boldsymbol{\varepsilon}_t$ (resp. $\boldsymbol{\Sigma}_t^l \boldsymbol{\varepsilon}_t$), we have that $\theta_{n,p} = 3.82\ldots$ (resp. $\theta_{n,p} = 4.00\ldots$). Thus, while second-order (cross-)moments of X_t and Y_t are identical for both $\boldsymbol{\Sigma}_t^s \boldsymbol{\varepsilon}_t$ and $\boldsymbol{\Sigma}_t^l \boldsymbol{\varepsilon}_t$, MES depends on the assumed structural model. The underlying reason for this is that the conditional distribution $\mathrm{P}\{X_{n+1} \leq \cdot, Y_{n+1} \leq \cdot \mid \mathcal{F}_n\}$ depends on the decomposition of the variance–covariance matrix.

From a structural perspective, the triangular structure of $\boldsymbol{\Sigma}_t^l$ implies that $\varepsilon_{t,X}$ is the idiosyncratic shock pertaining to (say) the market return X_t, which also affects the (say) portfolio return Y_t. However, the market is not affected by $\varepsilon_{t,Y}$. This contrasts with, e.g., a symmetric assumption on $\boldsymbol{\Sigma}_t^s$, where a unit shock in $\varepsilon_{t,X}$ has the same effect on Y_t as a unit shock in $\varepsilon_{t,Y}$ on X_t.

2.3. Model Assumptions

Denote the d.f. of the innovations in (1) by $F(x,y) = \mathrm{P}\{\varepsilon_{t,X} \leq x,\ \varepsilon_{t,Y} \leq y\}$ (which we assume to be continuous) and the marginal d.f.s by $F_0(x) = F(x, \infty)$ and $F_1(y) = F(\infty, y)$. For model (1), we have that $X_{n+1} = \sigma_{n+1,X}\varepsilon_{n+1,X}$, such that $\mathrm{VaR}_n(p) = \sigma_{n+1,X} F_0^{\leftarrow}(1-p)$. Therefore, it is easy to check that the MES becomes

$$\theta_{n,p} = \sigma_{n+1,Y} \mathrm{E}\big[\varepsilon_{t,Y} \mid \varepsilon_{t,X} > F_0^{\leftarrow}(1-p)\big]. \tag{2}$$

Hence, a forecast of $\theta_{n,p}$ consists of two parts. First, volatility $\sigma_{n+1,Y}$ must be forecasted and, second, we must estimate $\theta_p := \mathrm{E}[\varepsilon_{t,Y} \mid \varepsilon_{t,X} > F_0^{\leftarrow}(1-p)]$, i.e., the *unconditional* MES of $(\varepsilon_{t,X}, \varepsilon_{t,Y})'$. Thus, in the following, we have to impose some regularity conditions on the parameter estimator and the volatility model (to forecast volatility via some $\widehat{\sigma}_{n+1,Y}(\widehat{\boldsymbol{\theta}})$) and on the joint tail of the $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,X}, \varepsilon_{t,Y})'$ (to estimate θ_p).

We begin by imposing assumptions on the estimator and the volatility model. Regarding the former, we work with a generic parameter estimator $\widehat{\boldsymbol{\theta}}$ that satisfies the following assumption.

Assumption 1. The parameter estimator $\widehat{\boldsymbol{\theta}}$ satisfies $n^{\xi}\big(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^{\circ}\big)=O_{\mathrm{P}}(1)$, as $n\to\infty$, for some $\xi>0$.

The standard case of $\sqrt{n}$-consistent estimators is covered by $\xi=1/2$. For some examples of such estimators in multivariate volatility models, we refer to Francq and Zakoïan (2010). When errors are heavy-tailed, Assumption 1 may only hold for $\xi<1/2$. For example, in standard univariate GARCH models, the quasi-maximum likelihood estimator (QMLE) satisfies Assumption 1 with $\xi=1/2$ ($\xi<1/2$), when the innovations have finite (infinite) fourth moments (Hall and Yao, 2003). Thus, the generality afforded by Assumption 1 is not vacuous.

Next, we introduce some assumptions on the volatility model, i.e., $\boldsymbol{\Sigma}_t$. The $\boldsymbol{\Sigma}_t$'s in sufficiently general volatility models often depend on the infinite past $(X_{t-1},Y_{t-1})'$, $(X_{t-2},Y_{t-2})'$, ... Therefore, to approximate the $\boldsymbol{\Sigma}_t$'s, we use fixed artificial initial values $(\widehat{X}_{-\ell_n},\widehat{Y}_{-\ell_n})'$, $(\widehat{X}_{-\ell_n-1},\widehat{Y}_{-\ell_n-1})',\ldots$ in

$$\widehat{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})=\boldsymbol{\Sigma}_t\big((X_{t-1},Y_{t-1})',\ldots,(X_{-\ell_n+1},Y_{-\ell_n+1})',(\widehat{X}_{-\ell_n},\widehat{Y}_{-\ell_n})',(\widehat{X}_{-\ell_n-1},\widehat{Y}_{-\ell_n-1})',\ldots;\boldsymbol{\theta}\big),$$

where $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. This suggests to approximate $\boldsymbol{\Sigma}_t$ by $\widehat{\boldsymbol{\Sigma}}_t:=\widehat{\boldsymbol{\Sigma}}_t(\widehat{\boldsymbol{\theta}})=\operatorname{diag}\big(\widehat{\boldsymbol{\sigma}}_t(\widehat{\boldsymbol{\theta}})\big)$, where $\widehat{\boldsymbol{\sigma}}_t(\widehat{\boldsymbol{\theta}})=\big(\widehat{\sigma}_{t,X}(\widehat{\boldsymbol{\theta}}),\widehat{\sigma}_{t,Y}(\widehat{\boldsymbol{\theta}})\big)'$.

We impose the following assumptions on the initialization and on the volatility model.

Assumption 2. For any $M>0$, there exists a neighborhood $\mathcal{N}(\boldsymbol{\theta}^{\circ})$ of $\boldsymbol{\theta}^{\circ}$, and $p_*>0$, $q_*>0$ with $p_*^{-1}+q_*^{-1}=1$, such that for all $i=1,\ldots,\dim(\boldsymbol{\Theta})$,

$$\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^{\circ})}\left\|\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\frac{\partial\boldsymbol{\Sigma}_t(\boldsymbol{\theta})}{\partial\theta_i}\right\|^{Mp_*}\right]<\infty,$$
$$\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^{\circ})}\left\|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}^{\circ})\right\|^{Mq_*}\right]<\infty.$$

Assumption 3. There exist some constants $C>0$ and $\rho\in(0,1)$, and some random variable $C_0>0$, such that for all $t\in\mathbb{N}$ it holds almost surely (a.s.) that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right\|\leq C,$$
$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\|\boldsymbol{\Sigma}_t(\boldsymbol{\theta})-\widehat{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})\right\|\leq C_0\rho^{t+\ell_n-1}.$$

Assumption 2 for $M=4$ is almost identical to assumption A8 in Francq, Jiménez-Gamero, and Meintanis (2017). Together with Assumption 1, it can be used to show that parameter estimation effects in the MES forecasts vanish asymptotically. At first sight, the Hölder conjugate exponents p_* and q_* in Assumption 2 do not appear useful, since M is allowed to be arbitrarily large. Yet, when (e.g.)

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

$q_* = C/M$ for some finite constant C, one may accommodate an infinite $(C+\varepsilon)$th moment of $\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)} \left\| \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}^\circ)\right\|$. This, however, comes at the price of requiring moments of arbitrary order to exist for $\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)} \left\| \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\partial\boldsymbol{\Sigma}_t(\boldsymbol{\theta})/\partial\theta_i\right\|$.

Assumption 3 is almost identical to assumptions A1 and A2 in Francq et al. (2017), and ensures that initialization effects vanish asymptotically at a suitable rate. Initialization effects may occur because $\boldsymbol{\Sigma}_t$ often depends on the infinite past in $\mathcal{F}_{t-1}$, yet for estimation via $\widehat{\boldsymbol{\Sigma}}_t$ only the truncated information set $\widehat{\mathcal{F}}_{t-1} = \sigma\big((X_{t-1}, Y_{t-1})', \ldots, (X_{-\ell_n+1}, Y_{-\ell_n+1})'\big)$ is available.

Note that in Assumptions 2 and 3, we have tacitly assumed that $\boldsymbol{\Sigma}_t(\boldsymbol{\theta})$ is differentiable (in a neighborhood of the true parameter) and invertible, where the latter is equivalent to $\sigma_{t,X}(\boldsymbol{\theta}) > 0$ and $\sigma_{t,Y}(\boldsymbol{\theta}) > 0$, since $\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) = \operatorname{diag}\big(\boldsymbol{\sigma}_t(\boldsymbol{\theta})\big)$. Of course, assuming positive volatilities is rather innocuous.

Example 2. This example gives an instance of a model that satisfies Assumptions 2 and 3. Consider the ECCC–GARCH model of Jeantheau (1998), which models the squared volatilities $\boldsymbol{\sigma}_t^2 = (\sigma_{t,X}^2, \sigma_{t,Y}^2)'$ as

$$\boldsymbol{\sigma}_t^2 = \boldsymbol{\sigma}_t^2(\boldsymbol{\theta}) = \boldsymbol{\omega} + \sum_{j=1}^{\overline{p}} \boldsymbol{B}_j \begin{pmatrix} X_{t-j}^2 \\ Y_{t-j}^2 \end{pmatrix} + \sum_{j=1}^{\overline{q}} \boldsymbol{\Gamma}_j \boldsymbol{\sigma}_{t-j}^2(\boldsymbol{\theta}), \qquad t\in\mathbb{Z}, \tag{3}$$

where $\boldsymbol{\theta} = \big(\boldsymbol{\omega}', \operatorname{vec}'(\boldsymbol{B}_1), \ldots, \operatorname{vec}'(\boldsymbol{B}_{\overline{p}}), \operatorname{vec}'(\boldsymbol{\Gamma}_1), \ldots, \operatorname{vec}'(\boldsymbol{\Gamma}_{\overline{q}})\big)'$. The classical CCC–GARCH model of Bollerslev (1990) only allows for diagonal $\boldsymbol{B}_j$'s and $\boldsymbol{\Gamma}_j$'s, while non-diagonal matrices (and, hence, volatility spillovers) are accommodated only by ECCC–GARCH models. Under the conditions of their Theorem 5.1, Francq et al. (2017) show that a solution to the stochastic recurrence equations (1) and (3) exists, and Assumptions 2 and 3 are satisfied for

$$\widehat{\boldsymbol{\sigma}}_t^2(\boldsymbol{\theta}) = \boldsymbol{\omega} + \sum_{j=1}^{\overline{p}} \boldsymbol{B}_j \begin{pmatrix} X_t^2 \\ Y_t^2 \end{pmatrix} + \sum_{j=1}^{\overline{q}} \boldsymbol{\Gamma}_j \widehat{\boldsymbol{\sigma}}_{t-j}^2(\boldsymbol{\theta}), \qquad t\geq 1,$$

with fixed initial values $(X_{-\ell_n}, Y_{-\ell_n})' = (\widehat{X}_{-\ell_n}, \widehat{Y}_{-\ell_n})'$, $(X_{-\ell_n-1}, Y_{-\ell_n-1})' = (\widehat{X}_{-\ell_n-1}, \widehat{Y}_{-\ell_n-1})', \ldots$ and fixed initial $\widehat{\boldsymbol{\sigma}}_{-\ell_n}^2(\boldsymbol{\theta}), \widehat{\boldsymbol{\sigma}}_{-\ell_n-1}^2(\boldsymbol{\theta}), \ldots$.

As pointed out above, we also need to impose some regularity conditions on the tail of the $\boldsymbol{\varepsilon}_t$ (to estimate θ_p). Specifically, we assume that the following limit exists for all $(x,y)' \in [0,\infty]^2 \setminus \{(\infty,\infty)\}$:

$$\lim_{s\to\infty} s\,\mathrm{P}\big\{1-F_0(\varepsilon_{t,X}) \leq x/s,\ 1-F_1(\varepsilon_{t,Y}) \leq y/s\big\} =: R(x,y). \tag{4}$$

Schmidt and Stadtmüller (2006) call $R(\cdot,\cdot)$ the (upper) *tail copula*. Much like a copula, $R(x,y)$ only depends on the (extremal) dependence structure of $\boldsymbol{\varepsilon}_t$, and not on the marginal distributions. Regarding the marginals, we assume heavy right tails in the sense that there exist $\gamma_i > 0$, such that

$$\lim_{s\to\infty} U_i(sx)/U_i(s) = x^{\gamma_i} \qquad \text{for all } x>0,\ i=0,1, \tag{5}$$

where $U_i = (1/[1-F_i])^{\leftarrow}$ (de Haan and Ferreira, 2006). This condition means that far out in the tail, the distribution can roughly be modeled as a Pareto distribution (for which (5) holds even without the limit). Ever since the work of Bollerslev (1987), heavy-tailed innovations are standard ingredients of volatility models. For instance, the popular Student's t_ν-distribution satisfies (5) with $\gamma_i = 1/\nu$. In EVT, γ_i is known as the *extreme value index*.

2.4. MES Estimator

To estimate θ_p, we build on Cai et al. (2015). They use an extrapolation argument that is often applied in EVT, such as in estimating high quantiles (Weissman, 1978). The general idea is to first estimate the quantity of interest at a less extreme level (say, $\theta_{k/n}$ for $k/n \gg p$) and then, in a second step, to extrapolate to the desired level (say, θ_p) by exploiting the tail shape. These arguments rely on $p = p_n$ tending to zero, as $n \to \infty$.

Specifically, under (4) and (5) with $\gamma_1 \in (0,1)$, Cai et al. (2015) show that

$$\lim_{p\downarrow 0} \frac{\theta_p}{U_1(1/p)} = \int_0^\infty R(1, y^{-1/\gamma_1})\mathrm{d}y. \tag{6}$$

Now, let $k = k_n$ be an *intermediate sequence* of integers, such that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$. Then, for $n \to \infty$,

$$\theta_p \overset{(6)}{\sim} \frac{U_1(1/p)}{U_1(n/k)} \theta_{k/n} \overset{(5)}{\sim} \left(\frac{k}{np}\right)^{\gamma_1} \theta_{k/n}. \tag{7}$$

This relation suggests the following two-step procedure to estimate θ_p. First, estimate the less extreme ("within-sample") $\theta_{k/n}$ and, then, use the tail shape (characterized here by γ_1) to extrapolate to the desired (possibly "beyond the sample") θ_p.

One key difference to Cai et al. (2015) is that the $\boldsymbol{\varepsilon}_t$ are not available for estimation, but need to be approximated by the standardized residuals

$$\widehat{\boldsymbol{\varepsilon}}_t = \widehat{\boldsymbol{\varepsilon}}_t(\widehat{\boldsymbol{\theta}}) = \widehat{\boldsymbol{\Sigma}}_t^{-1}(X_t, Y_t)'.$$

To estimate $\theta_{k/n}$, we then use the non-parametric estimator

$$\widehat{\theta}_{k/n} = \frac{1}{k}\sum_{t=1}^{n} \widehat{\varepsilon}_{t,Y}^{+} I_{\{\widehat{\varepsilon}_{t,X} > \widehat{\varepsilon}_{(k+1),X}\}},$$

where $\widehat{\varepsilon}_{(1),Z} \geq \cdots \geq \widehat{\varepsilon}_{(n),Z}$ denote the order statistics of $\widehat{\varepsilon}_{1,Z}, \ldots, \widehat{\varepsilon}_{n,Z}$ ($Z \in \{X, Y\}$), such that $\widehat{\varepsilon}_{(k+1),X}$ estimates $F_0^{\leftarrow}(1-k/n)$ in $\theta_{k/n} = \mathrm{E}[\varepsilon_{t,Y} \mid \varepsilon_{t,X} > F_0^{\leftarrow}(1-k/n)]$, and $I_{\{\cdot\}}$ denotes the indicator function.

Remark 2. The estimator $\widehat{\theta}_{k/n}$ uses $\widehat{\varepsilon}_{t,Y}^{+}$ instead of $\widehat{\varepsilon}_{t,Y}$. Thus, it is in fact an estimator of $\theta_{k/n}^{+} = \mathrm{E}[\varepsilon_{t,Y}^{+} \mid \varepsilon_{t,X} > F_0^{\leftarrow}(1-k/n)]$. This is because the proofs closely

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

exploit the relation that $\mathrm{E}[Z] = \int_0^\infty \mathrm{P}\{Z > z\}\,\mathrm{d}z$ for $Z \geq 0$, such that

$$\theta_{k/n}^{+} = \frac{n}{k}\int_0^\infty \mathrm{P}\{\varepsilon_{t,X} > U_0(n/k),\ \varepsilon_{t,Y}^{+} > y\}\,\mathrm{d}y. \tag{8}$$

However, since Cai et al. (2015, Proof of Theorem 2) show that $\theta_p/\theta_p^{+} = 1 + o\big(1/\sqrt{k}\big)$, this does not impair the asymptotic validity of the (to be introduced) estimator $\widehat{\theta}_p$ based on $\widehat{\theta}_{k/n}$.

To estimate γ_1, we use the Hill (1975) estimator

$$\widehat{\gamma}_1 = \frac{1}{k_1}\sum_{t=1}^{k_1} \log\left(\widehat{\varepsilon}_{(t),Y}/\widehat{\varepsilon}_{(k_1+1),Y}\right),$$

where k_1 is another intermediate sequence of integers. Plugging the estimators $\widehat{\theta}_{k/n}$ and $\widehat{\gamma}_1$ into (7), we obtain the desired estimator

$$\widehat{\theta}_p = \left(\frac{k}{np}\right)^{\widehat{\gamma}_1}\widehat{\theta}_{k/n}.$$

To establish the asymptotic normality of $\widehat{\theta}_p$, we impose essentially the same regularity conditions as Cai et al. (2015). First, we specify the speed of convergence in (4) via Assumption 4 and that in (5) via Assumption 5. Here and elsewhere, $x \wedge y = \min\{x, y\}$.

Assumption 4. There exist $\beta > \gamma_1$ and $\tau < 0$ such that, as $s \to \infty$,

$$\sup_{\substack{x\in[1/2,2]\\ y\in(0,\infty)}} \frac{\left|s\mathrm{P}\{1-F_0(\varepsilon_{t,X}) \leq x/s,\ 1-F_1(\varepsilon_{t,Y}) \leq y/s\} - R(x,y)\right|}{y^{\beta}\wedge 1} = O(s^{\tau}).$$

Assumption 5. For $i = 0,1$ there exist $\rho_i < 0$ and an eventually positive or negative function $A_i(\cdot)$ such that, as $s \to \infty$, $A_i(sx)/A_i(s) \to x^{\rho_i}$ for all $x > 0$ and, for any $x_0 > 0$,

$$\sup_{x\geq x_0}\left|x^{-\gamma_i}\frac{U_i(sx)}{U_i(s)} - 1\right| = O\{A_i(s)\}.$$

Assumptions 4 and 5 provide second-order refinements of the convergences in (4) and (5). They ensure that bias terms arising from extrapolation vanish asymptotically. Note that the more negative τ (ρ_i) in Assumption 4 (Assumption 5), the better the approximation.

Remark 3. (i) Replacing s with n/k in the probability in the numerator, Assumption 4 requires $(n/k)\mathrm{P}\big\{\varepsilon_{t,X} > U_0(n/[kx]),\ \varepsilon_{t,Y} > U_1(n/[ky])\big\}$ to converge uniformly to its limit. In view of (8), it is therefore sufficient to impose uniformity only in a neighborhood of 1 for x (which, following Cai et al., 2015, we take to be $[1/2, 2]$ here). However, uniformity in y over $(0,\infty)$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

is required, as the integration in (8) extends over all positive y-values. For a specific dependence structure of $(\varepsilon_{t,X}, \varepsilon_{t,Y})'$, Assumption 4 may be checked by drawing on the results in Fougères et al. (2015, Sect. 4) (see also Remark 3 in that paper). Some specific distributions for which Assumption 4 holds are also given by Cai et al. (2015, Sect. 3).

(ii) As pointed out by Cai et al. (2015, Rem. 2), Assumption 4 excludes the case of asymptotically independent $(\varepsilon_{t,X}, \varepsilon_{t,Y})'$, where $R \equiv 0$ (to see this, let $y = s$). This rules out Gaussian copulas, but covers t-copulas (Heffernan, 2000), which seem to be empirically more relevant for financial data (Breymann, Dias, and Embrechts, 2003). Building on Cai and Musta (2020), it may be possible to allow asymptotically independent innovations in MES estimation. This is, however, beyond the scope of the present paper.

Remark 4. (i) Cai et al. (2015) also impose Assumption 5 on the tail of $\varepsilon_{t,Y}$. Together with $\sqrt{k_1}A_1(n/k_1) \to 0$ (see Assumption 6), it ensures that $\sqrt{k_1}(\widehat{\gamma}_1 - \gamma_1)$ does not have any asymptotic bias terms (see Lemma 2 or also de Haan and Ferreira, 2006, Example 5.1.5).

(ii) Note that Cai et al. (2015) do not have to impose Assumption 5 on the tail of $\varepsilon_{t,X}$, because the MES does not depend on its distribution. However, in estimating MES based on the estimated residuals $(\widehat{\varepsilon}_{t,X}, \widehat{\varepsilon}_{t,Y})'$, we have to justify the replacement of the unobservable $\varepsilon_{t,X}$ by the feasible $\widehat{\varepsilon}_{t,X}$ even in the tails. To that end, we require a sufficiently well-behaved tail also of the $\varepsilon_{t,X}$.

We stress that imposing Assumption 5 for $\varepsilon_{t,X}$ (i.e., for $i = 0$) is mainly a convenience. It can be replaced by any other condition ensuring the conclusion of Lemma 4 holds, as a careful reading of the proofs reveals. For instance, it can easily be shown that Lemma 4 remains valid, e.g., for light-tailed (standardized) exponentially distributed $\varepsilon_{t,X}$. However, ever since the work of Bollerslev (1987), heavy-tailed errors satisfying Assumption 5 (such as t_ν-distributed errors with $\gamma_i = 1/\nu$ and $\rho_i = -2$) are regarded as more suitable in volatility modeling. We refer to Hua and Joe (2011, Examples 1–3) for further heavy-tailed distributions with corresponding values for γ_i and ρ_i.

We mention that the problem of estimating a MES based on approximated conditioning variables (here, the $\widehat{\varepsilon}_{t,X}$) also appears in the work of Di Bernardino and Prieur (2018), who deal with unconditional MES estimation for i.i.d. random variables. They have to ensure that replacing their (latent) conditioning variable Z_j with some feasible $\widetilde{Z}_j$ has no asymptotic impact. Instead of imposing a distributional assumption (as we do) on the conditioning variables, Di Bernardino and Prieur (2018) assume that the $\widetilde{Z}_j$'s are estimated from an initial pre-sample of length n_2 to rule out any asymptotic effects. Their Assumption 1(a.3) (with $p_0 = q_0 = 1/2$ in their notation) then requires $n_1 = o(n_2^{(1-\varepsilon)/2})$ for the actual estimation sample size n_1. This allows them to justify the replacement of the latent Z_j with the $\widetilde{Z}_j$ without imposing distributional assumptions, as we do here. However, adopting a similar approach in our present time series context would be highly unnatural.

We need two additional assumptions.

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

Assumption 6. As $n \to \infty$,

$$\sqrt{k}A_0(n/k) \to 0, \quad \sqrt{k_1}A_1(n/k_1) \to 0,$$
$$k = O(n^{\alpha}) \quad \text{for some } \alpha < -2\tau/(-2\tau+1) \wedge 2\gamma_1\rho_1/(2\gamma_1\rho_1+\rho_1-1),$$
$$k = o(p^{2\tau(1-\gamma_1)}),$$
$$\min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\} = O(n^{\widetilde{\alpha}}) \quad \text{and} \quad \sqrt{k_1} = O(n^{\widetilde{\alpha}}) \quad \text{for some } \widetilde{\alpha} < \xi,$$

with $A_0(\cdot)$ and $A_1(\cdot)$ from Assumption 5, $\xi > 0$ from Assumption 1, and $d_n := k/(np)$.

Assumption 7. $\mathrm{E}|\varepsilon_{t,Y}^{-}|^{1/\gamma_1} < \infty$.

The purpose of Assumption 6 is to restrict the speed of divergence of k and k_1. While this is obvious for most items, it is less clear for the first conditions involving $A_0(\cdot)$ and $A_1(\cdot)$. However, de Haan and Ferreira (2006, p. 77) show that these conditions imply that $k = o(n^{-2\rho_0/(1-2\rho_0)})$ and $k_1 = o(n^{-2\rho_1/(1-2\rho_1)})$, respectively. While large values of the intermediate sequences k and k_1 imply a small asymptotic variance, a bias is incurred by using possibly "non-tail" observations in the estimates. Therefore, a bound on the growth of k and k_1 is required for asymptotically unbiased estimates. The requirement that $k = o(p^{2\tau(1-\gamma_1)})$ together with Assumption 7 is only used to show that $\theta_p/\theta_p^{+} = 1 + o(1/\sqrt{k})$ (cf. Remark 2). This relation ensures that $\widehat{\theta}_p$, which actually estimates θ_p^{+}, also estimates θ_p. The final condition in Assumption 6 ensures that the parameter estimator (which is n^{ξ}-consistent) converges sufficiently fast relative to our MES estimator (which is $\min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}$-consistent by Proposition 3). In the standard case where $\xi = 1/2$, this condition is redundant, because $\min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\} \le \sqrt{k} = o(n^{1/2})$ and $\sqrt{k_1} = o(n^{1/2})$ as k and k_1 are intermediate sequences and hence $o(n)$.

3. ASYMPTOTIC NORMALITY OF MES FORECASTS

With the MES estimator of the previous subsection, our MES forecast becomes

$$\widehat{\theta}_{n,p} = \widehat{\sigma}_{n+1,Y}\widehat{\theta}_p, \tag{9}$$

where $\widehat{\sigma}_{n+1,Y} = \widehat{\sigma}_{n+1,Y}(\widehat{\boldsymbol{\theta}})$. We can now state our first main theoretical result.

THEOREM 1. *Let $(X_t, Y_t)'$ be a strictly stationary solution to (1) that is measurable with respect to the sigma-field generated by $\{\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1}, \ldots\}$. Suppose Assumptions 1–7 hold, and $\gamma_1 \in (0, 1/2)$. Suppose further that $d_n \ge 1$, $r := \lim_{n\to\infty}\sqrt{k}\log(d_n)/\sqrt{k_1} \in [0,\infty]$, $q := \lim_{n\to\infty} k_1/k \in (0,\infty)$ and $\lim_{n\to\infty} \log(d_n)/\sqrt{k_1} = 0$. Moreover, suppose that the truncation sequence ℓ_n satisfies $\ell_n/\log n \to \infty$. Then, as $n \to \infty$,*

$$\min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\} \log\left(\frac{\widehat{\theta}_{n,p}}{\theta_{n,p}}\right) \xrightarrow{d} \begin{cases} \Theta + r\Gamma, & \text{if } r \le 1, \\ (1/r)\Theta + \Gamma, & \text{if } r > 1, \end{cases}$$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

where Θ and Γ are zero-mean Gaussian random variables with

$$\begin{aligned}
\mathrm{Var}(\Theta) &= \gamma_1^2 - 1 - b^2\int_0^\infty R(1,s)\,\mathrm{d}s^{-2\gamma_1}, \qquad b = 1\Big/\int_0^\infty R(1,s)\,\mathrm{d}s^{-\gamma_1},\\
\mathrm{Var}(\Gamma) &= \gamma_1^2,\\
\mathrm{Cov}(\Gamma,\Theta) &= \frac{\gamma_1}{\sqrt{q}}\Big(1-\gamma_1+\frac{b}{q^{\gamma_1}}\Big)R(1,q)\\
&\quad -\frac{\gamma_1}{\sqrt{q}}\int_0^q\Big[(1-\gamma_1)+bs^{-\gamma_1}\big\{1-\gamma_1-\gamma_1\log(s/q)\big\}\Big]R(1,s)s^{-1}\,\mathrm{d}s.
\end{aligned}$$

Proof. See Appendix A. □

The assumptions of Theorem 1 can be roughly divided into two parts. Assumptions 1–3, which are similar to conditions maintained by Francq et al. (2017), ensure that the innovations of the volatility model can be recovered from the observations with sufficient precision (see Proposition 2). Assumptions 4–7, which closely resemble the conditions in Cai et al. (2015), imply the asymptotic normality of the MES estimator for the innovations. Then, Assumptions 1–7 jointly ensure that the MES estimator $\widehat{\theta}_p$ based on the filtered residuals is also asymptotically normal (see Proposition 3). Here, the requirement that the truncation sequence ℓ_n be sufficiently long (i.e., $\ell_n/\log n \to \infty$) together with Assumption 3 ensures that initialization effects are negligible in the limit.

The case most often considered in EVT is that where $d_n \to \infty$ (see, e.g., de Haan and Ferreira, 2006, Thm. 4.3.1, for high quantile estimation). When additionally $k \asymp k_1$, we have that $r = \infty$, implying that Γ is the asymptotic limit. The proof of Theorem 1 shows that $\widehat{\gamma}_1$ consistently estimates γ_1 (see Lemma 2 in Appendix C). Hence, a feasible asymptotic $(1-\iota)$-confidence interval for $\theta_{n,p}$ in case $r = \infty$ is given by

$$\Big[\widehat{\theta}_{n,p}\exp\big\{\mp\Phi^{-1}(1-\iota/2)\widehat{\gamma}_1\log(d_n)/\sqrt{k_1}\big\}\Big], \tag{10}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse of the standard normal d.f.

This is not a "classical" confidence interval because $\theta_{n,p}$ is not a fixed parameter but a random quantity. However, it can be interpreted as such, since (by Theorem 1 and Lemma 2)

$$\begin{aligned}
\mathrm{P}\Big\{&\widehat{\theta}_{n,p}\exp\big\{-\Phi^{-1}(1-\iota/2)\widehat{\gamma}_1\log(d_n)/\sqrt{k_1}\big\} \le \theta_{n,p}\\
&\le \widehat{\theta}_{n,p}\exp\big\{\Phi^{-1}(1-\iota/2)\widehat{\gamma}_1\log(d_n)/\sqrt{k_1}\big\}\Big\} \underset{(n\to\infty)}{\longrightarrow} 1-\iota.
\end{aligned}$$

Such a straightforward interpretation of confidence intervals for random quantities is often not possible (Beutner, Heinemann, and Smeekes, 2021). Yet, it is possible here, because in Theorem 1 the asymptotic estimation uncertainty comes solely from the non-random θ_p component in

$$\log(\widehat{\theta}_{n,p}/\theta_{n,p}) = \log(\widehat{\sigma}_{n+1,Y}/\sigma_{n+1,Y}) + \log(\widehat{\theta}_p/\theta_p).$$

Specifically, the proof of Theorem 1 shows that parameter estimation effects vanish because volatility in $\widehat{\theta}_{n,p}$ can be estimated n^{ξ}-consistently, yet the unconditional MES estimate has a slower rate of convergence, thus dominating asymptotically. In the context of EVT-based VaR and ES forecasting, this was noted before by, e.g., Chan et al. (2007), Martins-Filho, Yao, and Torero (2018), and Hoga (2019).

Inference when $r < \infty$ remains an unsolved issue, even for unconditional MES estimation; see Cai et al. (2015) and Di Bernardino and Prieur (2018). However, the case $k \asymp k_1$ and $d_n \to \infty$ seems to be the case of most practical interest, because $d_n \to \infty$ corresponds to situations of strong extrapolation, where $k/n \gg p$. Furthermore, we demonstrate the good finite-sample coverage of (10) in simulations in Section 5. Nonetheless, it may be possible to explicitly deal with the case $r < \infty$. This may be possible by using self-normalization as in Hoga (2019) or by employing suitable bootstrap methods along the lines of Li, Peng, and Song (2023). We leave these challenging extensions for future research.

4. HIGHER-DIMENSIONAL EXTENSIONS

One desirable property of MES as a systemic risk measure is its additivity property. To illustrate, suppose that $Y_{t,1},\ldots,Y_{t,D}$ denote the losses of all trading desks of a business unit. The weighted losses of the business unit then sum to $X_t = \sum_{d=1}^{D} w_{t-1,d}Y_{t,d}$, where the weights $w_{t-1,d}$ are determined by how much capital is allocated to each trading desk in advance and, thus, are known at time $t-1$. The total riskiness of the business unit, as measured by the ES, can then be decomposed as $\mathrm{E}[X_t \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}] = \sum_{d=1}^{D} w_{t-1,d}\mathrm{E}[Y_{t,d} \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}]$. In allocating capital among the trading desks, one may want to ensure an equal risk contribution of each trading desk, such that $w_{t-1,1}\mathrm{E}[Y_{t,1} \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}] = \cdots = w_{t-1,D}\mathrm{E}[Y_{t,D} \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}]$. Therefore, it becomes important to develop tools for the joint inference on different MES forecasts.

Clearly, drawing inferences on many MES forecasts jointly is also important in other contexts. For instance, suppose the $Y_{t,1},\ldots,Y_{t,D}$ denote individual losses of all banks in the financial system. Then, the regulator seeks to control the system's total risk as measured by the ES $\mathrm{E}[X_t \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}]$ (see Qin and Zhou, 2021). Since $\mathrm{E}[X_t \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}] = \sum_{d=1}^{D} w_{t-1,d}\mathrm{E}[Y_{t,d} \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}] =: \sum_{d=1}^{D} w_{t-1,d}\theta_{t-1,p,d}$, it becomes clear that regulators should take into account the estimation risk of the individual MES forecasts.

Remark 5. To the extent that $\mathrm{E}[Y_{t,d} \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}]$ measures the contribution of institution d to the total risk in the financial system, it may be viewed as a measure of the systemic riskiness of that institution (Acharya et al., 2017, p. 7). However, the measure with reversed conditioning, i.e., $\mathrm{E}[X_t \mid Y_{t,d} >$

$\mathrm{VaR}_{t,d}(p),\ \mathcal{F}_{t-1}]$, also describes the systemic riskiness of bank d. The higher this particular MES, the greater the impact of distress of institution d on the financial system.[1] Yet, using $\mathrm{E}[Y_{t,d} \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}]$ has the advantage of the above additivity property, which is a crucial property for systemic risk measures as it allows for risk attribution (Chen et al., 2013). Acharya et al. (2017, p. 7), who also use the definition $\mathrm{E}[Y_{t,d} \mid X_t > \mathrm{VaR}_t(p),\ \mathcal{F}_{t-1}]$, additionally mention risk management, transfer pricing and strategic capital allocation as tasks where the additivity is important. Therefore, we work with Acharya et al.'s (2017) original definition in the following.

To enable joint hypothesis testing, we consider a high-dimensional extension of model (1), viz.,

$$(X_t, Y_{t,1}, \ldots, Y_{t,D})' = \boldsymbol{\Sigma}_t \boldsymbol{\varepsilon}_t. \tag{11}$$

We take the diagonal matrix $\boldsymbol{\Sigma}_t = \mathrm{diag}(\boldsymbol{\sigma}_t)$ with $\boldsymbol{\sigma}_t = (\sigma_{t,X}, \sigma_{t,Y_1}, \ldots, \sigma_{t,Y_D})'$ to be measurable with respect to $\mathcal{F}_{t-1} = \sigma\big((X_{t-1}, Y_{t-1,1}, \ldots, Y_{t-1,D})', (X_{t-2}, Y_{t-2,1}, \ldots, Y_{t-2,D})', \ldots\big)$, and the $\boldsymbol{\varepsilon}_t = (\varepsilon_{t,X}, \varepsilon_{t,Y_1}, \ldots, \varepsilon_{t,Y_D})'$ to be independent of $\mathcal{F}_{t-1}$ and i.i.d. with mean zero, unit variance and correlation matrix $\boldsymbol{R}$. We again assume the innovations $\boldsymbol{\varepsilon}_t$ to have a continuous d.f.

To forecast the individual MESs, we use the same estimator as before, i.e., $\widehat{\theta}_{n,p,d} = \widehat{\sigma}_{n+1,Y_d}\widehat{\theta}_{p,d}$ with

$$\widehat{\theta}_{p,d} = \left(\frac{k}{np}\right)^{\widehat{\gamma}_d} \widehat{\theta}_{k/n,d},$$

where $\widehat{\gamma}_d = \frac{1}{k_d}\sum_{t=1}^{k_d} \log\big(\widehat{\varepsilon}_{(t),Y_d}/\widehat{\varepsilon}_{(k_d+1),Y_d}\big)$ and $\widehat{\theta}_{k/n,d} = \frac{1}{k}\sum_{t=1}^{n} \widehat{\varepsilon}^{+}_{t,Y_d} I_{\{\widehat{\varepsilon}_{t,X} > \widehat{\varepsilon}_{(k+1),X}\}}$ are defined in the expected way.

To derive the asymptotic limit of $\big(\widehat{\theta}_{n,p,1}, \ldots, \widehat{\theta}_{n,p,D}\big)'$, Assumptions 1–3 do not have to be changed. The remaining Assumptions 4–7 have to be generalized slightly as follows. To that end, we extend the notation in the obvious way. For example, we denote the d.f. of $Y_{t,d}$ by $F_d(\cdot)$, and set $U_d = (1/[1-F_d])^{\leftarrow}$. The extreme value index of the $Y_{t,d}$ is denoted by γ_d.

Assumption 4*. For all $d = 1, \ldots, D$, there exist $\beta_d > \gamma_d$, $\tau_d < 0$ and $R_d(\cdot,\cdot)$ such that, as $s \to \infty$,

$$\sup_{\substack{x\in[1/2,2] \\ y\in(0,\infty)}} \frac{\big|s\mathrm{P}\{1-F_0(\varepsilon_{t,X}) \le x/s,\ 1-F_d(\varepsilon_{t,Y_d}) \le y/s\} - R_d(x,y)\big|}{y^{\beta_d} \wedge 1} = O(s^{\tau_d}).$$

[1] Of course, the MES $\mathrm{E}[X_t \mid Y_{t,d} > \mathrm{VaR}_{t,d}(p),\ \mathcal{F}_{t-1}]$ can also be forecasted using the theory of Section 3 by reversing the roles of X and Y.

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

Moreover, for all $i,j = 1,\dots,D$, there exists a function $R_{i,j}(\cdot,\cdot)$, such that for all $x,y \in [0,\infty]^2 \setminus \{(\infty,\infty)\}$,

$$\lim_{s\to\infty} s\,\mathrm{P}\{1-F_i(\varepsilon_{t,Y_i}) \le x/s,\ 1-F_j(\varepsilon_{t,Y_j}) \le y/s\} = R_{i,j}(x,y). \tag{12}$$

Assumption 5*. For all $d = 0,\dots,D$, there exist $\rho_d < 0$ and an eventually positive or negative function $A_d(\cdot)$ such that, as $s \to \infty$, $A_d(sx)/A_d(s) \to x^{\rho_d}$ for all $x > 0$ and, for any $x_0 > 0$,

$$\sup_{x\ge x_0}\left|x^{-\gamma_d}\frac{U_d(sx)}{U_d(s)}-1\right| = O\{A_d(s)\}.$$

Assumption 6*. As $n \to \infty$, for each $d = 1,\dots,D$,

$$\begin{aligned}
&\sqrt{k}A_0(n/k) \to 0, \quad \sqrt{k_d}A_1(n/k_d) \to 0,\\
&k = O(n^{\alpha}) \quad \text{for some } \alpha < -2\tau_d/(-2\tau_d+1) \wedge 2\gamma_d\rho_d/(2\gamma_d\rho_d+\rho_d-1),\\
&k = o(p^{2\tau_d(1-\gamma_d)}),\\
&\min\left\{\sqrt{k}, \sqrt{k_d}/\log(d_n)\right\} = O(n^{\widetilde{\alpha}}) \quad \text{and} \quad \sqrt{k_d} = O(n^{\widetilde{\alpha}}) \quad \text{for some } \widetilde{\alpha} < \xi,
\end{aligned}$$

with $A_0(\cdot)$ and $A_d(\cdot)$ from Assumption 5*, and $\xi > 0$ from Assumption 1.

Assumption 7*. $\mathrm{E}\,|\varepsilon_{t,Y_d}^{-}|^{1/\gamma_d} < \infty$ for all $d = 1,\dots,D$.

The only nontrivial extension of Assumptions 4*–7* relative to Assumptions 4–7 is condition (12), which is closely related to (4). It is needed to derive the joint convergence of $(\widehat{\gamma}_1,\dots,\widehat{\gamma}_D)'$, and the limit function $R_{i,j}(\cdot,\cdot)$ features prominently in its asymptotic limit and that of Theorem 2. For simplicity, Theorem 2 only considers the case where $r_d := \lim_{n\to\infty}\sqrt{k}\log(d_n)/\sqrt{k_d} = \infty$ for all $d = 1,\dots,D$, which implies individual confidence intervals of the form in (10).

THEOREM 2. *Let $(X_t, Y_{t,1},\dots,Y_{t,D})'$ be a strictly stationary solution to (11) that is measurable with respect to the sigma-field generated by $\{\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-1},\dots\}$. Suppose Assumptions 1–3 and Assumptions 4*–7* hold, and $\gamma_d \in (0,1/2)$ for all $d = 1,\dots,D$. Suppose further that $d_n \ge 1$, $r_d = \lim_{n\to\infty}\sqrt{k}\log(d_n)/\sqrt{k_d} = \infty$, $q_d := \lim_{n\to\infty} k_d/k \in (0,\infty)$ and $\lim_{n\to\infty}\log(d_n)/\sqrt{k_d} = 0$ for all $d = 1,\dots,D$. Moreover, suppose that the truncation sequence ℓ_n satisfies $\ell_n/\log n \to \infty$. Then, as $n \to \infty$,*

$$\left(\frac{\sqrt{k_1}}{\log(d_n)}\log\left(\frac{\widehat{\theta}_{n,p,1}}{\theta_{n,p,1}}\right),\dots,\frac{\sqrt{k_D}}{\log(d_n)}\log\left(\frac{\widehat{\theta}_{n,p,D}}{\theta_{n,p,D}}\right)\right)' \xrightarrow{d} (\Gamma_1,\dots,\Gamma_D)',$$

where $(\Gamma_1,\dots,\Gamma_D)'$ is a zero-mean Gaussian random vector with variance–covariance matrix $\boldsymbol{\Sigma} = (\sigma_{i,j})_{i,j=1,\dots,D}$,

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

$$\sigma_{i,j} := \mathrm{Cov}(\Gamma_i,\Gamma_j) = \frac{\gamma_i\gamma_j}{\sqrt{q_iq_j}}\frac{R_{i,j}(q_i,q_j)+R_{i,j}(q_j,q_i)}{2}.$$

Proof. See Appendix E. □

For joint inference on the MES forecasts, we have to estimate $\boldsymbol{\Sigma}$, i.e., the γ_d's, q_d's, $R_{i,j}(q_i,q_j)$'s, and $R_{i,j}(q_j,q_i)$'s. The extreme value indices γ_d can be consistently estimated via $\widehat{\gamma}_d$ (see Lemma 2) and the q_d's can easily be "estimated" via k_d/k. We propose to estimate the remaining quantities $R_{i,j}(q_i,q_j)$ and $R_{i,j}(q_j,q_i)$ via

$$\widehat{R}_{i,j}(q_i,q_j) = \frac{1}{k}\sum_{t=1}^{n} I_{\left\{\widehat{\varepsilon}_{t,Y_i}>\widehat{\varepsilon}_{(k_i+1),Y_i},\ \widehat{\varepsilon}_{t,Y_j}>\widehat{\varepsilon}_{(k_j+1),Y_j}\right\}},$$

$$\widehat{R}_{i,j}(q_j,q_i) = \frac{1}{k}\sum_{t=1}^{n} I_{\left\{\widehat{\varepsilon}_{t,Y_i}>\widehat{\varepsilon}_{(k_j+1),Y_i},\ \widehat{\varepsilon}_{t,Y_j}>\widehat{\varepsilon}_{(k_i+1),Y_j}\right\}}.$$

The next proposition shows that these estimates are consistent:

PROPOSITION 1. *Under the conditions of Theorem 2, it holds, for all* $i,j = 1,\ldots,D$, *that* $\widehat{R}_{i,j}(q_i,q_j) \xrightarrow{\mathrm{P}} R_{i,j}(q_i,q_j)$ *and* $\widehat{R}_{i,j}(q_j,q_i) \xrightarrow{\mathrm{P}} R_{i,j}(q_j,q_i)$, *as* $n\to\infty$.

Proof. See Appendix F. □

In sum, the asymptotic variance–covariance matrix $\boldsymbol{\Sigma}$ from Theorem 2 may be estimated consistently via $\widehat{\boldsymbol{\Sigma}} = (\widehat{\sigma}_{i,j})_{i,j=1,\ldots,D}$ with typical element

$$\widehat{\sigma}_{i,j} = k\frac{\widehat{\gamma}_i\widehat{\gamma}_j}{\sqrt{k_ik_j}}\frac{\widehat{R}_{i,j}(q_i,q_j)+\widehat{R}_{i,j}(q_j,q_i)}{2}.$$

Note that for $i=j=d$, we get (as expected from Theorem 1) that $\widehat{\sigma}_{d,d} = \widehat{\gamma}_d^2$.

We mention that Di Bernardino and Prieur (2018) also consider MES estimation (albeit in an i.i.d. static setting) in a higher-dimensional framework. However, they focus on limit theory for individual MES estimates. This contrasts with our joint convergence result in Theorem 2, with appertaining inference tools provided by Proposition 1 and Lemma 2.

As outlined above, in applications, one may want to test the equality of (value-weighted) risk contributions by several trading desks. Similarly, one may want to test the equality of the risk contributions by banks in a financial system. This latter application is further explored in Section 6. In each case, the null is that

$$H_0\ :\ w_{n,1}\theta_{n,p,1} = \cdots = w_{n,D}\theta_{n,p,D}$$

for positive weights $w_{n,D}>0$ summing to one.

We test this null by comparing each of $\log(w_{n,1}\widehat{\theta}_{n,p,1}),\ldots,\log(w_{n,D-1}\widehat{\theta}_{n,p,D-1})$ with the "average" forecast $(1/D)\sum_{d=1}^{D}\log(w_{n,d}\widehat{\theta}_{n,p,d})$. Then, any "large"

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

difference is evidence against the null. With the $(D-1)\times D$-transformation matrix

$$\boldsymbol{T}=\begin{pmatrix}1 & & 0 & 0\\ & \ddots & & \vdots\\ 0 & & 1 & 0\end{pmatrix}-(1/D)\begin{pmatrix}1 & \dots & 1\\ \vdots & \ddots & \vdots\\ 1 & \dots & 1\end{pmatrix}$$

this suggests using

$$\boldsymbol{T}\begin{pmatrix}\log(w_{n,1}\widehat{\theta}_{n,p,1})\\ \vdots\\ \log(w_{n,D}\widehat{\theta}_{n,p,D})\end{pmatrix}=\begin{pmatrix}\log(w_{n,1}\widehat{\theta}_{n,p,1})-(1/D)\sum_{d=1}^{D}\log(w_{n,d}\widehat{\theta}_{n,p,d})\\ \vdots\\ \log(w_{n,D-1}\widehat{\theta}_{n,p,D-1})-(1/D)\sum_{d=1}^{D}\log(w_{n,d}\widehat{\theta}_{n,p,d})\end{pmatrix}$$

in a Wald test. To ease the exposition, we assume that $k=k_1=\dots=k_D$, such that the test statistic becomes

$$\mathcal{T}_n=\frac{k}{\log(d_n)^2}\left(\log(w_{n,1}\widehat{\theta}_{n,p,1}),\dots,\log(w_{n,D}\widehat{\theta}_{n,p,D})\right)\boldsymbol{T}'(\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T}')^{-1}$$
$$\times\boldsymbol{T}\left(\log(w_{n,1}\widehat{\theta}_{n,p,1}),\dots,\log(w_{n,D}\widehat{\theta}_{n,p,D})\right)',$$

where the inverse of $\boldsymbol{T}\widehat{\boldsymbol{\Sigma}}\boldsymbol{T}'$ is assumed to exist.

COROLLARY 1. *Suppose that the conditions of Theorem 2 hold and that $k=k_1=\dots=k_D$. Furthermore, assume that $\boldsymbol{T}\boldsymbol{\Sigma}\boldsymbol{T}'$ is positive definite. Then, it holds under H_0 that $\mathcal{T}_n\xrightarrow{d}\chi^2_{D-1}$, as $n\to\infty$.*

Proof. Note that under H_0,

$$\boldsymbol{T}\begin{pmatrix}\log(w_{n,1}\widehat{\theta}_{n,p,1})\\ \vdots\\ \log(w_{n,D}\widehat{\theta}_{n,p,D})\end{pmatrix}=\begin{pmatrix}\log(\widehat{\theta}_{n,p,1}/\theta_{n,p,1})-(1/D)\sum_{d=1}^{D}\log(\widehat{\theta}_{n,p,d}/\theta_{n,p,d})\\ \vdots\\ \log(\widehat{\theta}_{n,p,D-1}/\theta_{n,p,D-1})-(1/D)\sum_{d=1}^{D}\log(\widehat{\theta}_{n,p,d}/\theta_{n,p,d})\end{pmatrix}$$
$$=\boldsymbol{T}\begin{pmatrix}\log(\widehat{\theta}_{n,p,1}/\theta_{n,p,1})\\ \vdots\\ \log(\widehat{\theta}_{n,p,D}/\theta_{n,p,D})\end{pmatrix}.$$

Then, combine the continuous mapping theorem with Theorem 2, Proposition 1 and the consistency of $\widehat{\gamma}_d$ for $d=1,\dots,D$ (from Lemma 2). □

5. SIMULATIONS

Here, we investigate the coverage of the confidence interval in (10) for $\iota=0.05$. We do so for $n\in\{500,\ 1{,}000\}$ and $p\in\{1\%,\ 0.5\%,\ 0.1\%,\ 0.05\%,\ 0.01\%,\ 0.005\%,\ 0.001\%\}$. Throughout, we use 1,000 replications and we clip off the first $\ell_n=10$ residuals to reduce the impact of initialization effects on our MES estimator. Following Qin and Zhou (2021, Footnote 6), we use identical $k=k_1$ in the simulations and the empirical application. Specifically, we set $k=k_1=\lfloor 0.1\cdot$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

$\log(n)^4\rfloor$ to satisfy Assumption 6. Since the probabilities p are quite small relative to the sample sizes n, our confidence intervals (valid when $d_n = k/(np)$ tends to infinity) should provide reasonable coverage.

We simulate from the simple CCC–GARCH model $(X_t, Y_t)' = \mathrm{diag}(\boldsymbol{\sigma}_t)\boldsymbol{\varepsilon}_t$, where $\boldsymbol{\sigma}_t = (\sigma_{t,X}, \sigma_{t,Y})'$ with

$$\begin{aligned}\sigma_{t,X}^2 &= 0.001 + 0.2 \cdot X_t^2 + 0.75 \cdot \sigma_{t-1,X}^2,\\ \sigma_{t,Y}^2 &= 0.001 + 0.1 \cdot Y_t^2 + 0.85 \cdot \sigma_{t-1,Y}^2.\end{aligned} \tag{13}$$

The parameter values of the volatility equations are chosen to resemble typical estimates obtained for financial data. The innovations $\boldsymbol{\varepsilon}_t$ are i.i.d. draws from a t-copula with ν degrees of freedom and correlation coefficient $\rho_{X,Y}$. The marginals of $\boldsymbol{\varepsilon}_t$ are from a (standardized and symmetrized) Burr(a,b)-distribution with d.f. $F(x) = 1-(1+x^b)^{-a}$, $x > 0$, $a, b > 0$. Hence, $\varepsilon_{t,Z} \overset{d}{=} R_t B_t / \sqrt{\mathrm{E}[B_t^2]}$ for $Z \in \{X, Y\}$, where R_t are Rademacher random variables (i.e., equal to ± 1 with probability 1/2), independent of the $B_t \sim \mathrm{Burr}(a,b)$.

We choose the marginal Burr distribution because it allows to vary the quality of the approximation in Assumption 5 via the parameters a and b without changing the extreme value index γ. Example 2 in Hua and Joe (2011) shows that the extreme value index of a Burr(a,b)-distribution is given by $1/(ab)$ and its second-order parameter by $-b$. Thus, we have that $\gamma_0 = \gamma_1 = 1/(ab)$ and $\rho_0 = \rho_1 = -b$ in the notation of Assumption 5. This implies that the larger b, the faster the convergence of $U_i(sx)/U_i(s)$ to the Pareto-type limit x^{γ_i} takes place in Assumption 5. By suitable choices of a and b, this allows us to assess the implications of a better tail approximation on the precision of our MES estimator, while keeping the tail index constant. Specifically, we choose $(a,b) \in \{(0.25, 20),\ (0.2, 25)\}$ to always obtain $\gamma_0 = \gamma_1 = 1/5$. However, the Pareto approximation is better for $(a,b) = (0.2, 25)$ because b is larger.

The choice of the t-copula for $\boldsymbol{\varepsilon}_t$ implies that Assumption 4 is satisfied with $\tau = -2/\nu$, $\beta = 1 + 2/\nu$ and $R(x,y) = x\overline{F}_{\nu+1}\Big(\frac{(x/y)^{1/\nu} - \rho_{X,Y}}{\sqrt{1-\rho_{X,Y}^2}}\sqrt{\nu+1}\Big) + y\overline{F}_{\nu+1}\Big(\frac{(y/x)^{1/\nu} - \rho_{X,Y}}{\sqrt{1-\rho_{X,Y}^2}}\sqrt{\nu+1}\Big)$, where $\overline{F}_{\nu+1}(\cdot) = 1 - F_{\nu+1}(\cdot)$ denotes the survivor function of the $t_{\nu+1}$-distribution (see Fougères et al., 2015, Rem. 3 and Sect. 4.1). A more negative τ (smaller degrees of freedom ν of the t-copula) implies a better approximation in Assumption 4. Our copula construction allows us to vary the dependence structure (in particular, the quality of the approximation in Assumption 4 as a function of ν) without changing the marginals of the innovations. Specifically, we choose $\nu \in \{3, 5\}$, and set the correlation coefficient of the copula to $\rho_{X,Y} = 0.95$.

To estimate the GARCH parameters in (13), we use the QMLE. Since we choose a and b to give $\gamma_0 = \gamma_1 = 1/(ab) = 1/5$, the innovations have finite fourth moments, implying $\sqrt{n}$-consistency of the QMLE (Francq and Zakoïan, 2010).

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

TABLE 1. Bias, RMSEs, average interval lengths of (10) and coverage (in %) for $n = 500$ and $1-\iota = 95\%$

n	ν	(a,b)	p	Bias	RMSE	RMSE_{ML}	RMSE_{NP}	Length	Coverage
500	3	(0.25, 20)	1%	0.3	2.6	2.1	4.8	6.6	83.9
			0.5%	0.6	3.4	2.7	9.6	9.2	86.9
			0.1%	1.3	6.0	4.9	46.1	18.0	90.5
			0.05%	1.6	7.6	6.3	52.9	23.4	91.6
			0.01%	3.2	12.9	10.8	72.7	41.3	93.1
			0.005%	4.3	16.1	13.7	83.8	52.0	93.4
			0.001%	11.0	27.7	23.1	114.0	86.9	93.9
	3	(0.2, 25)	1%	0.1	2.5	2.0	4.6	6.4	83.2
			0.5%	0.3	3.3	2.6	9.3	8.9	85.2
			0.1%	0.8	5.8	4.5	46.1	17.4	90.2
			0.05%	1.2	7.3	5.6	53.1	22.6	01.4
			0.01%	2.6	12.3	9.3	73.7	39.6	93.1
			0.005%	3.9	15.4	11.4	84.9	49.8	93.6
			0.001%	6.0	24.8	18.6	117.4	82.9	94.7
	5	(0.25, 20)	1%	0.7	2.7	2.0	4.6	6.7	82.7
			0.5%	1.1	3.6	2.7	9.2	9.3	85.7
			0.1%	2.6	6.6	4.7	45.9	18.2	88.9
			0.05%	3.5	8.4	5.9	52.6	23.7	89.8
			0.01%	6.8	14.5	9.9	72.6	41.8	90.4
			0.005%	8.1	17.8	12.3	83.5	52.6	91.6
			0.001%	10.2	27.6	20.5	117.6	88.0	94.1

Note: Bias, RMSEs, and interval lengths are all multiplied by 100 for better readability.

Thus, Assumption 1 is met for $\xi = 1/2$. By Example 2, Assumptions 2 and 3 are also satisfied.

We draw the following conclusions from the simulation results in Tables 1 and 2:

1. The more extreme p, the better the coverage of the 95%-confidence intervals (10). This may be explained as follows. For more extreme p, the pre-asymptotic value of r in Theorem 1 (i.e., $\sqrt{k}\log(d_n)/\sqrt{k_1}$) is closer to infinity, suggesting that the limit in Theorem 1 can be better approximated by Γ, i.e., the limit distribution exploited in the construction (10) for $r = \infty$. As pointed out in Section 1, it is precisely the small p's for which systemic risk measures are of most interest. For these small p's, our confidence intervals are reasonably accurate, particularly, when compared with other EVT-based confidence intervals (see, e.g., Chan et al., 2007, Fig. 1). To improve coverage for less extreme p, one may entertain self-normalized confidence intervals as in Hoga (2019) or one may use bootstrap approximations, as explored by Li et al. (2023) in the context

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

TABLE 2. Bias, RMSEs, average interval lengths of (10) and coverage (in %) for $n = 1{,}000$ and $1-\iota = 95\%$

n	ν	(a,b)	p	Bias	RMSE	RMSE_{ML}	RMSE_{NP}	Length	Coverage
1,000	3	(0.25, 20)	1%	0.2	1.9	1.4	3.3	4.8	83.1
			0.5%	0.4	2.5	1.8	5.3	6.7	86.3
			0.1%	1.0	4.4	3.2	14.5	13.4	90.4
			0.05%	1.3	5.6	4.1	53.2	17.4	90.8
			0.01%	2.2	9.4	6.8	73.5	30.7	92.5
			0.005%	2.4	11.6	8.5	83.3	38.6	93.1
			0.001%	2.3	18.5	14.2	114.2	64.1	93.5
	3	(0.2, 25)	1%	0.1	1.9	1.5	3.3	4.7	82.1
			0.5%	0.2	2.4	2.0	5.4	6.7	85.8
			0.1%	0.5	4.3	3.4	15.5	13.2	89.6
			0.05%	0.8	5.5	4.3	52.6	17.2	91.1
			0.01%	1.9	9.3	7.0	72.2	30.2	92.5
			0.005%	3.0	11.7	8.7	82.7	37.9	92.9
			0.001%	7.9	19.9	14.2	114.5	62.9	92.9
	5	(0.25, 20)	1%	0.6	2.0	1.4	3.2	4.8	82.0
			0.5%	0.8	2.7	1.9	5.2	6.8	85.2
			0.1%	1.6	4.8	3.3	14.3	13.5	89.6
			0.05%	2.2	6.1	4.2	52.3	17.5	90.6
			0.01%	3.8	10.3	7.0	71.7	30.9	92.3
			0.005%	4.8	12.8	8.7	82.1	38.8	92.7
			0.001%	8.7	21.2	15.7	107.9	64.6	93.3

Note: Bias, RMSEs, and interval lengths are all multiplied by 100 for better readability.

of extreme VaR and ES forecasting. In principle, coverage for larger p could also be improved by using forecast intervals based on the limiting distribution $(1/r)\Theta + \Gamma$. However, to the best of our knowledge, no consistent estimator has been proposed for its asymptotic variance, which seems difficult to estimate.

2. As expected, the more extreme p, the larger the bias and the RMSE of the MES forecasts. Also, the larger b and the smaller ν, the lower the bias and RMSE tend to be for fixed risk level p. This may be explained by the fact that, for larger b, the Pareto approximation is more accurate for the marginals. Similarly, for smaller ν, the approximation in Assumption 4 (that is used in extrapolating) is more precise, leading to lower bias and RMSE. Of course, bias and particularly the RMSE are also reduced when the sample size n increases.
3. Similarly as the RMSE, the lengths of the confidence intervals also decrease the larger n and the better the approximations in Assumptions 4 and 5 (i.e., the smaller ν and the larger b). This is also as expected since confidence intervals provide a measure of the statistical uncertainty around the point MES forecasts.

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

4. For purposes of comparison, we have also included the RMSEs of two additional MES forecasts with quite different robustness–efficiency trade-off: one based on a parametric maximum likelihood (ML) estimator and another based on a non-parametric (NP) estimator.[2] Our semi-parametric estimator $\widehat{\theta}_p$ provides a balance between the advantages of these two estimators by producing reasonably robust estimates with good efficiency. The results are as expected. Because the assumptions underlying the use of the ML estimator and our $\widehat{\theta}_p$ are met, they are more efficient than the NP estimates in terms of RMSE, with the ML estimator being the favorite. But of course the ML estimator may be dangerous to use under misspecification. As Drees (2008, Sect. 2.2) warns, the consequences of misspecification of model-based estimates are even magnified in the tail and, therefore, such estimates are not recommended in applications of EVT. Indeed, unreported simulations with the ML estimator based on assuming (misspecified) marginal t-densities show an inferior performance compared to $\widehat{\theta}_p$ with RMSEs 2–4 times as large.

Overall, our MES forecasts are reasonably accurate, with the forecast intervals providing good measures of their uncertainty (particularly, for extreme risk levels p).

6. EMPIRICAL APPLICATION

Here, we consider the eight U.S. G-SIBs as determined by the FSB (2021).[3] We use daily data from 2001 to 2021, because data for JP Morgan Chase (one of the G-SIBs) is only available after the merger of JP Morgan and Chase Manhattan in 2000. This gives us $N = 5{,}283$ observations. All data are downloaded from Datastream.

Since we consider eight banks, we have $D = 8$ in the notation of Section 4. The $Y_{t,d}$ $(d = 1, \ldots, D)$ then denote the log-losses of the individual institutions' shares (calculated based on adjusted closing prices). The value-weighted losses of the whole financial system are $X_t = \sum_{t=1}^{D} w_{t-1,d} Y_{t,d}$. Specifically, the weights sum to one (i.e., $\sum_{d=1}^{D} w_{t-1,d} = 1$) and are calculated as $w_{t-1,d} = M_{t-1,d}/(\sum_{d=1}^{D} M_{t-1,d})$ with $M_{t-1,d}$ the market capitalization of institution d at time $t-1$ (see Qin and Zhou, 2021, for a similar approach). To forecast MES, we use a CCC–GARCH model of the form (11) with standard GARCH(1,1) volatility models for the marginals. Note that a model only for $(Y_{t,1}, \ldots, Y_{t,D})'$ is not sufficient for this purpose, as X_t is not only composed of the $Y_{t,d}$'s but also of the time-varying weights $w_{t-1,d}$. We use rolling-window MES forecasts, where a window of length

[2]The ML estimator separately estimates the copula parameters $(\nu, \rho_{X,Y})$ and the marginal parameters (a, b) via ML. It then computes the MES θ_p implied by the estimated values. In contrast, the NP estimate is simply $\widehat{\theta}_p^{\mathrm{NP}} = \frac{1}{np} \sum_{t=1}^{n} \widehat{\varepsilon}_{t,Y} I_{\{\widehat{\varepsilon}_{t,X} > \widehat{\varepsilon}_{(\lfloor np \rfloor+1),X}\}}$. In both cases, the MES forecasts are obtained by pre-multiplying the MES estimate with $\widehat{\sigma}_{n+1}^{Y}$ (cf. (9)).

[3]Specifically, the eight G-SIBs are Bank of America, Bank of New York Mellon, Citigroup, Goldman Sachs, JP Morgan Chase, Morgan Stanley, State Street, and Wells Fargo.

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

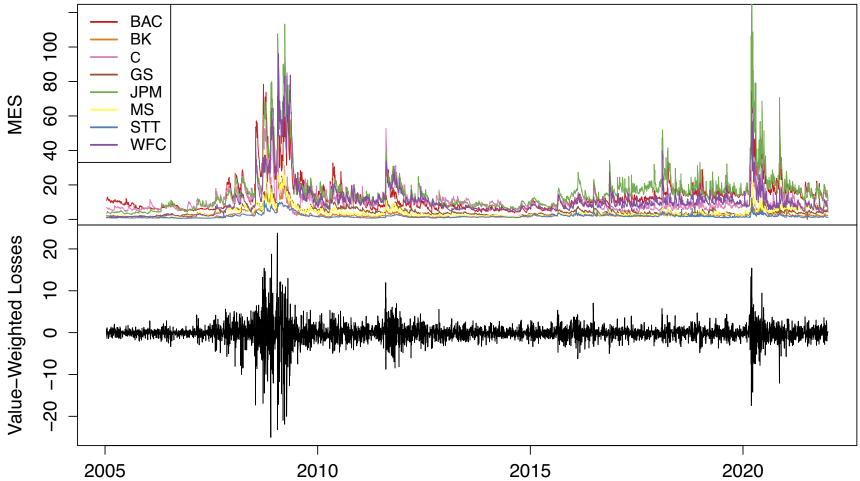

FIGURE 1. Top panel: Value-weighted MES forecasts for eight U.S. G-SIBs are Bank of America (BAC), Bank of New York Mellon (BK), Citigroup (C), Goldman Sachs (GS), JP Morgan Chase (JPM), Morgan Stanley (MS), State Street (STT), and Wells Fargo (WFC). Bottom panel: Time series of value-weighted losses X_t.

$(n+\ell_n)$ is rolled through the $N=5{,}283$ observations to yield $N-(n+\ell_n)$ one-step ahead forecasts. We pick $n=1{,}000$ and $\ell_n=10$, corresponding to roughly four years of daily returns in each moving window. To reflect the scarcity of systemic events, we choose $p=1/n=0.001$. Results for different values of p are available upon request. As in the simulations, we set $k=k_1=\lfloor 0.1\cdot\log(1{,}000)^4\rfloor$.

From a regulatory perspective, it is important (at each point t in time) to limit the risk of the overall banking system as measured by the ES, viz.,

$$\mathrm{E}[X_t \mid X_t > \mathrm{VaR}_t(p), \mathcal{F}_{t-1}] = \sum_{d=1}^{D} w_{t-1,d}\,\mathrm{E}[Y_{t,d} \mid X_t > \mathrm{VaR}_t(p), \mathcal{F}_{t-1}] = \sum_{d=1}^{D} w_{t-1,d}\theta_{t-1,p,d}.$$

The bottom panel of Figure 1 shows a plot of the value-weighted losses X_t. During periods of high volatility, aggregate risk of the system is high. Such periods are noticeable during the financial crisis of 2007–2009, the European sovereign debt crisis in the early 2010s, and the Corona stock market crash of March 2020.

The top panel of Figure 1 plots the systemic risk contributions of each bank as measured by the value-weighted MES $w_{t-1,d}\theta_{t-1,p,d}$. Clearly, the contributions vary through time. Notice also that the MES forecasts during 2021 largely reflect the systemic risk contributions of the different institutions as determined by the FSB (2021). For instance, JP Morgan Chase is listed as the systemically most risky bank by the FSB (2021), which is mirrored by the high MES forecasts in the top panel of Figure 1. At the other end of the spectrum, State Street, Bank of New York

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

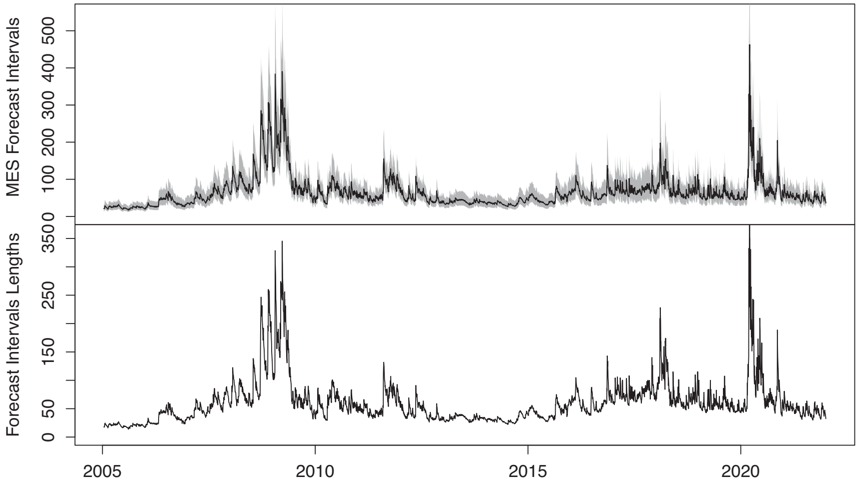

FIGURE 2. Top panel: MES forecasts (solid black line) together with 95% confidence intervals (shaded grey area) for JP Morgan Chase. Bottom panel: Lengths of 95% confidence intervals.

Mellon, and Morgan Stanley are deemed least risky by the FSB (2021), similarly as their low MES forecasts suggest.

The placement of the institutions in different buckets by the FSB (2021) (with bucket 5 containing the most systemically risky institutions and bucket 1 the least systemically risky ones) suggests that the systemic risk contributions of the different banks may also be distinguishable in statistical terms. Thus, for each point in time t, for which MES forecasts are issued, we test equality of the value-weighted MES forecasts, i.e., $w_{t-1,1}\theta_{t-1,p,1} = \cdots = w_{t-1,D}\theta_{t-1,p,D}$. To do so, we use the test statistic $\mathcal{T}_n$ from Corollary 1. For each point in time, the p-value is virtually zero, indicating (as expected) large heterogeneity of systemic risk contributions among banks.

To limit the overall riskiness of the system, it seems prudent for the regulator to not only rely on the point MES forecasts, but also to take into account the estimation risk associated with those forecasts. To illustrate the significant impact of estimation risk, Figure 2 displays the MES forecasts together with the 95%-confidence intervals given in (10). It does so exemplarily for JP Morgan Chase, which is ranked as the world's most systemically risky bank by the FSB (2021). The financial crisis of 2007–2009 and the Corona stock market crash of March 2020 are associated with the largest spikes in systemic riskiness. It is exactly during these times, where precise risk assessments are needed most, that our intervals suggest that systemic risk is forecasted with the greatest uncertainty. This can be seen most clearly from the bottom panel of Figure 2, where lengths of the 95%-forecast intervals are shown. During crises (e.g., around 2008) the estimation risk in the forecasts is highest. In contrast, during calmer times, the estimation risk is comparatively smaller. Formula (10) suggests that the main driver of these differences in forecast quality is the volatility of the JP Morgan Chase shares.

7. CONCLUSION

For univariate risk measures, such as VaR and ES, there is a voluminous literature on asymptotic properties of forecasts. In contrast, asymptotic properties of systemic risk forecasts are largely unexplored. This paper fills this gap by deriving limit theory for EVT-based MES forecasts. In doing so, we extend the unconditional MES estimator of Cai et al. (2015) along two dimensions. First, we prove its validity when applied to residuals of multivariate volatility models, thus allowing it to be used for conditional MES forecasting and confidence interval construction. In simulations, we illustrate the good finite-sample coverage of the forecast intervals, which provide valuable information beyond the mere point forecast of MES. Second, we derive limit theory also in higher-dimensional systems, therefore enabling joint inference on multiple MES forecasts. This may be beneficial as illustrated in the empirical application to the losses of the eight U.S. G-SIBs.

The following avenues may be worth exploring in future research. First, one may develop bootstrap-based confidence intervals for MES to improve coverage, particularly for not so extreme p. Second, it may be interesting to explore the properties of our forecasts for data-adaptive choices of k and k_1, which may improve finite-sample properties. The results of Drees et al. (2020) in the context of extreme value index estimation suggest that this may influence the asymptotic behavior. Third, we have exclusively focused on MES as a measure of systemic risk in this paper. However, there are many other popular systemic risk measures in the literature, such as the CoVaR and CoES of Adrian and Brunnermeier (2016). Thus, future research could develop limit theory for (EVT-based) forecasts of these measures as well.

APPENDICES

If not specified otherwise, all limits and all $o_{(\mathrm{P})}$- and $O_{(\mathrm{P})}$-symbols are to be interpreted with respect to $n \to \infty$. We denote by $K > 0$ a large positive constant that may change from line to line, and by $\boldsymbol{I}$ the identity matrix of appropriate dimension.

A. Proof of Theorem 1

The proof of Theorem 1 requires Propositions 2 and 3.

PROPOSITION 2. *Suppose Assumptions 1–3 hold, and the truncation sequence ℓ_n satisfies $\ell_n/\log n \to \infty$. Then, we have for any $\iota > 0$ that, as $n \to \infty$,*

$$\widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{\varepsilon}_t\big\{1 + o_{\mathrm{P}}(n^{\iota-\xi})\big\},$$

$$\widehat{\boldsymbol{\sigma}}_t = \boldsymbol{\sigma}_t\big\{1 + o_{\mathrm{P}}(n^{\iota-\xi})\big\},$$

uniformly in $t = 1, \ldots, n+1$.

Proof. See Appendix B. □

PROPOSITION 3. *Under the conditions of Theorem 1, we have that, as $n \to \infty$,*

$$\min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}\log\left(\frac{\widehat{\theta}_p}{\theta_p}\right) \xrightarrow{d} \begin{cases} \Theta + r\Gamma, & \text{if } r \le 1, \\ (1/r)\Theta + \Gamma, & \text{if } r > 1, \end{cases}$$

with r, Γ, and Θ defined as in Theorem 1.

Proof. See Appendix C. □

Proof of Theorem 1. Use (2) and (9) to write

$$\begin{aligned} &\min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}\log\left(\frac{\widehat{\theta}_{n,p}}{\theta_{n,p}}\right) \\ &\quad = \min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}\log\left(\frac{\widehat{\sigma}_{n+1,Y}}{\sigma_{n+1,Y}}\cdot\frac{\widehat{\theta}_p}{\theta_p}\right) \\ &\quad = \min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}\log\left(\frac{\widehat{\sigma}_{n+1,Y}}{\sigma_{n+1,Y}}\right) + \min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}\log\left(\frac{\widehat{\theta}_p}{\theta_p}\right). \end{aligned} \tag{A.1}$$

From Proposition 2 and Assumption 6, it follows that for $\iota < \xi - \widetilde{\alpha}$,

$$\begin{aligned} \min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}\log\left(\frac{\widehat{\sigma}_{n+1,Y}}{\sigma_{n+1,Y}}\right) &= \min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}\log\left(1 + o_{\mathrm{P}}(n^{\iota-\xi})\right) \\ &= \min\left\{\sqrt{k}, \sqrt{k_1}/\log(d_n)\right\}o_{\mathrm{P}}(n^{\iota-\xi}) \\ &= O(n^{\widetilde{\alpha}})o_{\mathrm{P}}(n^{\iota-\xi}) \\ &= o_{\mathrm{P}}(n^{\widetilde{\alpha}+\iota-\xi}) \\ &= o_{\mathrm{P}}(1), \end{aligned} \tag{A.2}$$

because $\log x \sim x - 1$, $x \to 1$. For the second right-hand side term in (A.1) Proposition 3 applies. Overall, the conclusion follows. □

B. Proof of Proposition 2

Proof of Proposition 2. Fix $\iota > 0$. Choose $M > 0$ from Assumption 2 sufficiently large, such that $1/M < \iota$. Consider n sufficiently large, such that Assumption 2 holds for the neighborhood $\mathcal{N}(\boldsymbol{\theta}^\circ) := \left\{\boldsymbol{\theta} : n^{\xi}\|\boldsymbol{\theta} - \boldsymbol{\theta}^\circ\| \le K_0\right\}$ for some $K_0 > 0$. For $\boldsymbol{\theta} \in \mathcal{N}(\boldsymbol{\theta}^\circ)$, write

$$\begin{aligned} \widehat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta}) &= \widehat{\boldsymbol{\Sigma}}_t^{-1}(\boldsymbol{\theta})(X_t, Y_t)' = \widehat{\boldsymbol{\Sigma}}_t^{-1}(\boldsymbol{\theta})\boldsymbol{\Sigma}_t(\boldsymbol{\theta}^\circ)\boldsymbol{\varepsilon}_t \\ &= \left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\widehat{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})\right]^{-1}\left[\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}^\circ)\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]^{-1}\boldsymbol{\varepsilon}_t \\ &= \left\{\boldsymbol{I} + \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\left[\widehat{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right]\right\}^{-1}\left\{\boldsymbol{I} + \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}^\circ)\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta}) - \boldsymbol{\Sigma}_t(\boldsymbol{\theta}^\circ)\right]\right\}^{-1}\boldsymbol{\varepsilon}_t. \end{aligned} \tag{B.1}$$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

Use Assumption 3 (where C_0 is random) to conclude that

$$\max_{t=1,\dots,n+1}\left\|\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\big[\widehat{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})-\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\big]\right\| \leq \max_{t=1,\dots,n+1}\left\|\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta})\right\| \max_{t=1,\dots,n+1}\left\|\widehat{\boldsymbol{\Sigma}}_t(\boldsymbol{\theta})-\boldsymbol{\Sigma}_t(\boldsymbol{\theta})\right\|$$
$$\leq CC_0\rho^{\ell_n}=o_{\mathrm{P}}(n^{-K}) \tag{B.2}$$

for any $K>0$, because

$$\begin{aligned}
\rho^{\ell_n} &= \exp\left\{\log\left(\rho^{(\ell_n/\log n)\log n}\right)\right\} \\
&= \exp\left\{\log(n)\log\left(\rho^{\ell_n/\log n}\right)\right\} = \left(\exp\{\log n\}\right)^{\log\left(\rho^{\ell_n/\log n}\right)} \\
&= n^{\log\left(\rho^{\ell_n/\log n}\right)} = o\big(n^{-K}\big),
\end{aligned}$$

since $\ell_n/\log n \to \infty$ by assumption, such that $\rho^{\ell_n/\log n}\to 0$ and, hence, $\log\left(\rho^{\ell_n/\log n}\right)\to-\infty$.

For the second right-hand side term in (B.1), we obtain that

$$\max_{t=1,\dots,n+1}\left\|\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}^\circ)\big[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})-\boldsymbol{\Sigma}_t(\boldsymbol{\theta}^\circ)\big]\right\| \leq \max_{t=1,\dots,n+1}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})-\sigma_{t,X}(\boldsymbol{\theta}^\circ)}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|$$
$$+\max_{t=1,\dots,n+1}\left|\frac{\sigma_{t,Y}(\boldsymbol{\theta})-\sigma_{t,Y}(\boldsymbol{\theta}^\circ)}{\sigma_{t,Y}(\boldsymbol{\theta}^\circ)}\right|. \tag{B.3}$$

Consider the first term on the right-hand side of (B.3). Define $\dot{\sigma}_{t,X}(\boldsymbol{\theta})=\partial\sigma_{t,X}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Use the mean value theorem to deduce that for some $\boldsymbol{\theta}^*$ on the line connecting $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^\circ$,

$$\begin{aligned}
\max_{t=1,\dots,n+1}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})-\sigma_{t,X}(\boldsymbol{\theta}^\circ)}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right| &= \max_{t=1,\dots,n+1}\left|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right| \\
&= \max_{t=1,\dots,n+1}\left|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta}^*)}{\sigma_{t,X}(\boldsymbol{\theta}^*)}\cdot\frac{\sigma_{t,X}(\boldsymbol{\theta}^*)}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\cdot(\boldsymbol{\theta}-\boldsymbol{\theta}^\circ)\right| \\
&\leq \max_{t=1,\dots,n+1}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left\|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta})}\right\|\cdot\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|\right\}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^\circ\right\|.
\end{aligned} \tag{B.4}$$

We obtain that

$$\begin{aligned}
&\mathrm{P}\left\{\max_{t=1,\dots,n+1}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left\|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta})}\right\|\cdot\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|\right\}>\varepsilon n^\iota\right\} \\
&\leq \sum_{t=1}^{n+1}\mathrm{P}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left\|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta})}\right\|\cdot\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|>\varepsilon n^\iota\right\} \\
&\leq \sum_{t=1}^{n+1}\frac{1}{\varepsilon^M n^{M\iota}}\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left\|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta})}\right\|^M\cdot\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|^M\right] \\
&\leq \sum_{t=1}^{n+1}\frac{1}{\varepsilon^M n^{M\iota}}\left\{\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left\|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta})}\right\|^{Mp_*}\right]\right\}^{1/p_*}\cdot\left\{\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|^{Mq_*}\right]\right\}^{1/q_*} \\
&\leq Kn^{1-M\iota},
\end{aligned}$$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

where we used subadditivity in the first step, Markov's inequality in the second step, Hölder's inequality in the third step, and Assumption 2 for the final inequality. Therefore, since $1/M < \iota$,

$$\max_{t=1,\dots,n+1}\left\{\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left\|\frac{\dot{\sigma}_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta})}\right\|\cdot\sup_{\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|\right\} = o_{\mathrm{P}}(n^{\iota}),$$

whence from (B.4) and Assumption 1,

$$\max_{t=1,\dots,n+1}\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})-\sigma_{t,X}(\boldsymbol{\theta}^\circ)}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right| = o_{\mathrm{P}}(n^{\iota})n^{-\xi}n^{\xi}\left\|\boldsymbol{\theta}-\boldsymbol{\theta}^\circ\right\| = o_{\mathrm{P}}(n^{\iota-\xi})K_0 = o_{\mathrm{P}}(n^{\iota-\xi}). \tag{B.5}$$

Using identical arguments, we may also show that

$$\max_{t=1,\dots,n+1}\left|\frac{\sigma_{t,Y}(\boldsymbol{\theta})-\sigma_{t,Y}(\boldsymbol{\theta}^\circ)}{\sigma_{t,Y}(\boldsymbol{\theta}^\circ)}\right| = o_{\mathrm{P}}(n^{\iota-\xi}).$$

Thus, from (B.3),

$$\max_{t=1,\dots,n+1}\left\|\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\theta}^\circ)\left[\boldsymbol{\Sigma}_t(\boldsymbol{\theta})-\boldsymbol{\Sigma}_t(\boldsymbol{\theta}^\circ)\right]\right\| = o_{\mathrm{P}}(n^{\iota-\xi}). \tag{B.6}$$

Plugging (B.2) and (B.6) into (B.1), we get that for all $\boldsymbol{\theta}\in\mathcal{N}(\boldsymbol{\theta}^\circ)$,

$$\widehat{\boldsymbol{\varepsilon}}_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t\left\{1+o_{\mathrm{P}}(n^{\iota-\xi})\right\},$$

where we recall that the matrices $\boldsymbol{\Sigma}(\cdot)$ and $\widehat{\boldsymbol{\Sigma}}(\cdot)$ in (B.1) are diagonal. Since, by Assumption 1, $\widehat{\boldsymbol{\theta}}$ is an element of $\mathcal{N}(\boldsymbol{\theta}^\circ)$ with probability approaching 1, as $n\to\infty$ followed by $K_0\to\infty$, the conclusion for the residuals follows.

The uniform approximability of the volatilities follows from

$$\begin{aligned}
\left|\frac{\widehat{\sigma}_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}-1\right| &= \left|\frac{\widehat{\sigma}_{t,X}(\boldsymbol{\theta})-\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}+\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}-1\right| \\
&\le \left|\frac{\widehat{\sigma}_{t,X}(\boldsymbol{\theta})-\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}\right|+\left|\frac{\sigma_{t,X}(\boldsymbol{\theta})}{\sigma_{t,X}(\boldsymbol{\theta}^\circ)}-1\right| \\
&= o_{\mathrm{P}}(n^{-K})+o_{\mathrm{P}}(n^{\iota-\xi}) = o_{\mathrm{P}}(n^{\iota-\xi})
\end{aligned}$$

(from (B.2) and (B.5)) together with the n^{ξ}-consistency of $\widehat{\boldsymbol{\theta}}$ (from Assumption 1). A similar result obtains for $\widehat{\sigma}_{t,Y}$. □

C. Proof of Proposition 3

Let $F_+(\cdot)$ denote the d.f. of $\varepsilon_{t,Y}^{+}$, and set $U_+ = [1/(1-F_+)]^{\leftarrow}$. Then, the proof of Theorem 2 in Cai et al. (2015) shows that Assumptions 4 and 5 (phrased in terms of $(\varepsilon_{t,X},\varepsilon_{t,Y})'$) continue to hold for $(\varepsilon_{t,X},\varepsilon_{t,Y}^{+})'$ with the same constants and functions $A_i(\cdot)$. This will be exploited in some of the following proofs without further mention.

For $(x,y)' \in [0,\infty]^2 \setminus \{(\infty,\infty)'\}$, we define

$$R_n(x,y) = \frac{n}{k} \mathrm{P}\left\{F_0(\varepsilon_{t,X}) > 1 - kx/n,\ F_+(\varepsilon_{t,Y}^+) > 1 - ky/n\right\},$$

$$T_n(x,y) = \frac{1}{k} \sum_{t=1}^{n} I_{\left\{F_0(\varepsilon_{t,X}) > 1-kx/n,\ F_+(\varepsilon_{t,Y}^+) > 1-ky/n\right\}},$$

$$\widehat{T}_n(x,y) = \frac{1}{k} \sum_{t=1}^{n} I_{\left\{F_0(\widehat{\varepsilon}_{t,X}) > 1-kx/n,\ F_+(\widehat{\varepsilon}_{t,Y}^+) > 1-ky/n\right\}}.$$

The limit distribution of $\widehat{\theta}_p$ is characterized by the zero-mean Gaussian process

$$\left\{W_R(x,y)\right\}_{(x,y)' \in [0,\infty]^2 \setminus \{(\infty,\infty)'\}}$$

with covariance structure given by

$$\mathrm{E}\left[W_R(x_1,y_1) W_R(x_2,y_2)\right] = R(x_1 \wedge x_2, y_1 \wedge y_2).$$

Then,

$$\Theta = (\gamma_1 - 1) W_R(1,\infty) + \left\{\int_0^\infty R(1,s)\,\mathrm{d}s^{-\gamma_1}\right\}^{-1} \int_0^\infty W_R(1,s)\,\mathrm{d}s^{-\gamma_1},$$

$$\Gamma = \frac{\gamma_1}{\sqrt{q}} \left\{\int_0^q s^{-1} W_R(\infty,s)\,\mathrm{d}s - W_R(\infty,q)\right\}$$

are the zero-mean Gaussian random variables from Theorem 1.

The proof of Proposition 3 requires Lemmas 1–3. These lemmas build on Proposition 3.1 in Einmahl, de Haan, and Li (2006). Invoking a Skorohod construction, the limit processes involved in that proposition may be assumed to be defined on the same probability space. This leads to an easier presentation of some of the subsequent results. We state the version of Proposition 3.1 as given in Cai et al. (2015, Lemma 1).

LEMMA 1. *Suppose that (4) holds. Then, for any $\eta \in [0, 1/2)$ and $T > 0$, it holds that, as $n \to \infty$,*

$$\begin{aligned}
&\sup_{x,y \in (0,T]} y^{-\eta} \left| \sqrt{k}\left\{T_n(x,y) - R_n(x,y)\right\} - W_R(x,y) \right| \xrightarrow{a.s.} 0,\\
&\sup_{x \in (0,T]} x^{-\eta} \left| \sqrt{k}\left\{T_n(x,\infty) - x\right\} - W_R(x,\infty) \right| \xrightarrow{a.s.} 0,\\
&\sup_{y \in (0,T]} y^{-\eta} \left| \sqrt{k}\left\{T_n(\infty,y) - y\right\} - W_R(\infty,y) \right| \xrightarrow{a.s.} 0.
\end{aligned} \tag{C.1}$$

LEMMA 2. *Under the conditions of Theorem 1, we have that, as $n \to \infty$,*

$$\sqrt{k_1}\left(\widehat{\gamma}_1 - \gamma_1\right) \xrightarrow{\mathrm{P}} \Gamma.$$

Proof. See Appendix D. □

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

For the next lemma, we introduce the following (with the exception of $\widehat{\theta}_{k/n}$) infeasible estimators of $\theta^{+}_{k/n}$:

$$\widetilde{\theta}_{k/n} = \frac{1}{k}\sum_{t=1}^{n} \varepsilon^{+}_{t,Y} I_{\{\varepsilon_{t,X} > \varepsilon_{(k+1),X}\}},$$

$$\widetilde{\theta}^{*}_{k/n} = \frac{1}{k}\sum_{t=1}^{n} \varepsilon^{+}_{t,Y} I_{\{\varepsilon_{t,X} > U_0(n/k)\}},$$

$$\widehat{\theta}_{k/n} = \frac{1}{k}\sum_{t=1}^{n} \widehat{\varepsilon}^{+}_{t,Y} I_{\{\widehat{\varepsilon}_{t,X} > \widehat{\varepsilon}_{(k+1),X}\}},$$

$$\widehat{\theta}^{*}_{k/n} = \frac{1}{k}\sum_{t=1}^{n} \widehat{\varepsilon}^{+}_{t,Y} I_{\{\widehat{\varepsilon}_{t,X} > U_0(n/k)\}}.$$

Moreover, define

$$e_n = (n/k)\left\{1 - F_0(\varepsilon_{(k+1),X})\right\}, \tag{C.2}$$

$$\widehat{e}_n = (n/k)\left\{1 - F_0(\widehat{\varepsilon}_{(k+1),X})\right\}, \tag{C.3}$$

such that $\widehat{e}_n \widehat{\theta}^{*}_{k\widehat{e}_n/n} = \widehat{\theta}_{k/n}$ and $e_n \widetilde{\theta}^{*}_{ke_n/n} = \widetilde{\theta}_{k/n}$.

LEMMA 3. *Under the conditions of Theorem 1, we have that, as $n \to \infty$,*

$$\frac{\sqrt{k}}{U_1(n/k)}\left(\widehat{\theta}_{k/n} - \widetilde{\theta}_{k/n}\right) = o_{\mathrm{P}}(1).$$

Proof. See Appendix D. □

Now, we can prove Proposition 3.

Proof of Proposition 3. Write

$$\frac{\widehat{\theta}_p}{\theta_p} = \frac{d_n^{\widehat{\gamma}_1}}{d_n^{\gamma_1}} \cdot \frac{\widehat{\theta}_{k/n}}{\theta^{+}_{k/n}} \cdot \frac{d_n^{\gamma_1}\theta^{+}_{k/n}}{\theta^{+}_{p}} \cdot \frac{\theta^{+}_{p}}{\theta_p} =: B_1 \cdot B_2 \cdot B_3 \cdot B_4.$$

First consider B_1. By the mean value theorem and $(\partial/\partial x) d^x = d^x \log d$, there exists $\vartheta \in (0,1)$, such that

$$\begin{aligned}
\frac{\sqrt{k_1}}{\log d_n}(B_1 - 1) &= \frac{\sqrt{k_1}}{\log d_n}\left(d_n^{\widehat{\gamma}_1 - \gamma_1} - d_n^{0}\right) \\
&= \frac{\sqrt{k_1}}{\log d_n} d_n^{\vartheta(\widehat{\gamma}_1 - \gamma_1)} \log(d_n)(\widehat{\gamma}_1 - \gamma_1) \\
&= \sqrt{k_1}(\widehat{\gamma}_1 - \gamma_1)\exp\left\{\log\left(d_n^{\vartheta(\widehat{\gamma}_1 - \gamma_1)}\right)\right\} \\
&= \sqrt{k_1}(\widehat{\gamma}_1 - \gamma_1)\exp\left\{\vartheta(\widehat{\gamma}_1 - \gamma_1)\log d_n\right\} \\
&= \sqrt{k_1}(\widehat{\gamma}_1 - \gamma_1)\left\{1 + o_{\mathrm{P}}(1)\right\} \\
&= \Gamma + o_{\mathrm{P}}(1),
\end{aligned} \tag{C.4}$$

where the second-to-last line follows from $\vartheta(\widehat{\gamma}_1-\gamma_1)\log d_n = O_{\mathrm{P}}(\log d_n/\sqrt{k_1}) = o_{\mathrm{P}}(1)$ (from Lemma 2 and our assumption that $\log d_n/\sqrt{k_1}=o(1)$), and the final line follows from Lemma 2.

Combine our Lemma 3 with Proposition 3 in Cai et al. (2015) to get that

$$\sqrt{k}(B_2-1)=\sqrt{k}\left(\frac{\widehat{\theta}_{k/n}}{\theta^{+}_{k/n}}-1\right)\xrightarrow{\mathrm{P}}\Theta.$$

Equation (32) of Cai et al. (2015) yields that

$$B_3=1+o\big(1/\sqrt{k}\big).$$

Finally, the proof of Theorem 2 in Cai et al. (2015) shows that

$$B_4=1+o\big(1/\sqrt{k}\big).$$

The rest of the proof follows as that of Theorem 1 in Cai et al. (2015). We give it here for the sake of completeness. Combining the above displays leads to

$$\begin{aligned}
&\frac{\widehat{\theta}_p}{\theta_p}-1=B_1\cdot B_2\cdot B_3\cdot B_4-1\\
&=\left[1+\frac{\log d_n}{\sqrt{k_1}}\Gamma+o_{\mathrm{P}}\left(\frac{\log d_n}{\sqrt{k_1}}\right)\right]\left[1+\frac{\Theta}{\sqrt{k}}+o_{\mathrm{P}}\left(\frac{1}{\sqrt{k}}\right)\right]\left[1+o\left(\frac{1}{\sqrt{k}}\right)\right]\left[1+o\left(\frac{1}{\sqrt{k}}\right)\right]-1\\
&=\frac{\log d_n}{\sqrt{k_1}}\Gamma+\frac{\Theta}{\sqrt{k}}+o_{\mathrm{P}}\left(\frac{\log d_n}{\sqrt{k_1}}\right)+o_{\mathrm{P}}\left(\frac{1}{\sqrt{k}}\right).
\end{aligned}\tag{C.5}$$

Since $\log x\sim x-1$ as $x\to 1$, the claimed convergence follows. The variances and covariances of Γ and Θ can be computed by using their definition and exploiting the covariance structure of $W_R(\cdot,\cdot)$ together with Fubini's theorem. □

D. Proofs of Lemmas 2 and 3

Define

$$s_n(y)=s_{n,k_1}(y)=\frac{n}{k_1}\left[1-F_+\big(y^{-\gamma_1}U_+(n/k_1)\big)\right],\qquad y>0.$$

Proof of Lemma 2. As a first step, we show that for any $T>0$,

$$\sup_{y\in(0,T]}\left|\frac{s_n(y)}{y}-1\right|=o\big(1/\sqrt{k_1}\big).\tag{D.1}$$

Since $U_1(s)=U_+(s)$, for $s>1/\{1-F_1(0)\}$, we have from Assumption 5 that for any $y_0>0$,

$$\sup_{y\geq y_0}\left|y^{-\gamma_1}\frac{U_+(sy)}{U_+(s)}-1\right|=O\{A_1(s)\},\qquad s\to\infty.$$

Insert $s=n/k_1$ and $y=1/s_n(y)$ in that relation to obtain that

$$\sup_{y\in(0,T]}\left|\left(\frac{s_n(y)}{y}\right)^{\gamma_1}-1\right|=O\{A_1(n/k_1)\},\qquad n\to\infty.$$

From $A_1(n/k_1)=o\big(1/\sqrt{k_1}\big)$ (by Assumption 6) and from a Taylor expansion, (D.1) follows.

Lemma 1 implies, for any $\eta\in[0,1/2)$ and any $T>0$, that

$$\sup_{y\in(0,T]} y^{-\eta}\left|\sqrt{k}\big\{T_n(\infty,y)-y\big\}-W_R(\infty,y)\right|\xrightarrow{a.s.}0. \tag{D.2}$$

Since $k_1/k\to q\in(0,\infty)$, we may replace y by $y(k_1/k)$ in that relation to obtain

$$\sup_{y\in(0,T]}\left(y\frac{k_1}{k}\right)^{-\eta}\left|\sqrt{k}\left\{\frac{1}{k}\sum_{t=1}^{n}I_{\{F_+(\varepsilon_{t,Y}^+)>1-k_1y/n\}}-y\frac{k_1}{k}\right\}-W_R\Big(\infty,y\frac{k_1}{k}\Big)\right|\xrightarrow{a.s.}0$$

or, multiplying through with $\sqrt{k/k_1}$,

$$\sup_{y\in(0,T]} y^{-\eta}\left(\frac{k_1}{k}\right)^{-\eta}\left|\sqrt{k_1}\left\{\frac{1}{k_1}\sum_{t=1}^{n}I_{\{F_+(\varepsilon_{t,Y}^+)>1-k_1y/n\}}-y\right\}-\sqrt{\frac{k_1}{k}}W_R\Big(\infty,y\frac{k_1}{k}\Big)\right|\xrightarrow{a.s.}0.$$

Again, since $k_1/k\to q\in(0,\infty)$, we may drop the pre-factor $(k_1/k)^{-\eta}$ to get that

$$\sup_{y\in(0,T]} y^{-\eta}\left|\sqrt{k_1}\left\{\frac{1}{k_1}\sum_{t=1}^{n}I_{\{F_+(\varepsilon_{t,Y}^+)>1-k_1y/n\}}-y\right\}-\sqrt{\frac{k_1}{k}}W_R\Big(\infty,y\frac{k_1}{k}\Big)\right|\xrightarrow{a.s.}0.$$

Due to the uniform continuity of the weighted Wiener process,

$$\sup_{y\in(0,T]} y^{-\eta}\left|\sqrt{\frac{k}{k_1}}W_R\Big(\infty,y\frac{k_1}{k}\Big)-\frac{1}{\sqrt{q}}W_R(\infty,yq)\right|\xrightarrow{a.s.}0.$$

The last two displays imply that

$$\sup_{y\in(0,T]} y^{-\eta}\left|\sqrt{k_1}\left\{\frac{1}{k_1}\sum_{t=1}^{n}I_{\{F_+(\varepsilon_{t,Y}^+)>1-k_1y/n\}}-y\right\}-q^{-1/2}W_R(\infty,qy)\right|\xrightarrow{a.s.}0. \tag{D.3}$$

With this and $s_n(y)\to y$, as $n\to\infty$, uniformly in $y\in(0,T]$ from (D.1), it follows, for any $0<T_1<T$, that

$$\sup_{y\in(0,T_1]} s_n^{-\eta}(y)\left|\sqrt{k_1}\left\{\frac{1}{k_1}\sum_{t=1}^{n}I_{\{\varepsilon_{t,Y}^+>y^{-\gamma_1}U_+(n/k_1)\}}-s_n(y)\right\}-q^{-1/2}W_R\{\infty,qs_n(y)\}\right|\xrightarrow{a.s.}0. \tag{D.4}$$

Our next goal is to show that $s_n(y)$ can be replaced by y at all three appearances in (D.4). From (D.1), it follows that, without changing the limit, $s_n(y)$ may be replaced by y in (D.4) at the first two appearances. Finally, the uniform continuity of the weighted Wiener process ensures that, as $n\to\infty$,

$$\sup_{y\in(0,T_1]} y^{-\eta}q^{-1/2}\left|W_R\{\infty,qs_n(y)\}-W_R(\infty,qy)\right|\xrightarrow{a.s.}0.$$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

Thus, since T and $T_1 < T$ can be chosen arbitrarily large, we get for any $y_0 > 0$ that

$$\sup_{y \geq y_0} y^{\eta/\gamma_1} \left| \sqrt{k_1} \left\{ \frac{1}{k_1} \sum_{t=1}^{n} I_{\left\{\varepsilon_{t,Y}^{+} > yU_{+}(n/k_1)\right\}} - y^{-1/\gamma_1} \right\} - q^{-1/2} W_R\left(\infty, qy^{-1/\gamma_1}\right) \right| \xrightarrow{a.s.} 0.$$

Because $U_{+}(n/k_1) = U_1(n/k_1) > 0$ for sufficiently large n, we get from this that

$$\sup_{y \geq y_0} y^{\eta/\gamma_1} \left| \sqrt{k_1} \left\{ T_n(y) - y^{-1/\gamma_1} \right\} - q^{-1/2} W_R\left(\infty, qy^{-1/\gamma_1}\right) \right| \xrightarrow{a.s.} 0, \tag{D.5}$$

where $T_n(y) = \frac{1}{k_1} \sum_{t=1}^{n} I_{\left\{\varepsilon_{t,Y} > yU_1(n/k_1)\right\}}$.

Our next goal is to show that $T_n(y)$ in (D.5) can be replaced by $\widehat{T}_n(y) = \frac{1}{k_1} \sum_{t=1}^{n} I_{\left\{\widehat{\varepsilon}_{t,Y} > yU_1(n/k_1)\right\}}$. To task this, let $\delta_n = n^{\iota - \xi}$ with $\iota > 0$ chosen sufficiently small to ensure that $\sqrt{k_1}\delta_n = o(1)$ (which is possible due to the Assumption 6 requirement that $\sqrt{k_1} = O(n^{\widetilde{\alpha}})$). By Proposition 2, $\widehat{\varepsilon}_{t,Y} = \varepsilon_{t,Y}\left\{1 + o_{\mathrm{P}}\left(n^{\iota - \xi}\right)\right\}$ uniformly in $t = 1, \ldots, n$. Thus, since $U_1(n/k_1) > 0$ for sufficiently large n (by Assumption 5),

$$I_{\left\{\varepsilon_{t,Y} > (1+\delta_n)yU_1(n/k_1)\right\}} \leq I_{\left\{\widehat{\varepsilon}_{t,Y} > yU_1(n/k_1)\right\}} \leq I_{\left\{\varepsilon_{t,Y} > (1-\delta_n)yU_1(n/k_1)\right\}}$$

holds for all $t = 1, \ldots, n$ with probability approaching 1 (w.p.a. 1), as $n \to \infty$. Hence, for any $\varepsilon > 0$, we can ensure that $\mathrm{P}\{\mathcal{W}_1\} > 1 - \varepsilon/3$ for sufficiently large n, where

$$\mathcal{W}_1 := \left\{ T_n\left\{(1+\delta_n)y\right\} - T_n(y) \leq \widehat{T}_n(y) - T_n(y) \leq T_n\left\{(1-\delta_n)y\right\} - T_n(y) \right\}.$$

We now show that

$$\sup_{y \geq y_0} y^{\eta/\gamma_1} \left| \sqrt{k_1} \left[T_n\left\{(1 \pm \delta_n)y\right\} - T_n(y) \right] \right| \xrightarrow{a.s.} 0. \tag{D.6}$$

We only show (D.6) for $T_n\left\{(1+\delta_n)y\right\}$, as the proof for $T_n\left\{(1-\delta_n)y\right\}$ is similar. Bound

$$\begin{aligned}
&\sup_{y \geq y_0} y^{\eta/\gamma_1} \left| \sqrt{k_1} \left[T_n\left\{(1+\delta_n)y\right\} - T_n(y) \right] \right| \\
&\quad \leq \sup_{y \geq y_0} y^{\eta/\gamma_1} \Big| \sqrt{k_1} \left[T_n\left\{(1+\delta_n)y\right\} - (1+\delta_n)^{-1/\gamma_1} y^{-1/\gamma_1} \right] \\
&\qquad\qquad\qquad\qquad - q^{-1/2} W_R\left(\infty, q(1+\delta_n)^{-1/\gamma_1} y^{-1/\gamma_1}\right) \Big| \\
&\qquad + \sup_{y \geq y_0} y^{\eta/\gamma_1} \left| \sqrt{k_1} \left[T_n(y) - y^{-1/\gamma_1} \right] - q^{-1/2} W_R\left(\infty, qy^{-1/\gamma_1}\right) \right| \\
&\qquad + \sup_{y \geq y_0} y^{\eta/\gamma_1} q^{-1/2} \left| W_R\left(\infty, q(1+\delta_n)^{-1/\gamma_1} y^{-1/\gamma_1}\right) - W_R\left(\infty, qy^{-1/\gamma_1}\right) \right| \\
&\qquad + \sqrt{k_1} \sup_{y \geq y_0} y^{\eta/\gamma_1} \gamma_1^{-1} \left\{ \delta_n + o(\delta_n) \right\} y^{-1/\gamma_1} \\
&\quad = o(1),
\end{aligned}$$

where we used (D.5), the uniform continuity of the weighted Wiener process, and the fact that $\sqrt{k_1}\delta_n = o(1)$ by our choice of δ_n. This proves (D.6).

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

Now, we may use (D.6) to conclude that for any $\delta>0$ it holds for sufficiently large n that

$$\begin{aligned}
&\mathrm{P}\Big\{\sup_{y\geq y_0} y^{\eta/\gamma_1}\Big|\sqrt{k_1}\big[\widehat{T}_n(y)-T_n(y)\big]\Big|\geq\delta\Big\}\\
&\leq \mathrm{P}\Big\{\sup_{y\geq y_0} y^{\eta/\gamma_1}\Big|\sqrt{k_1}\big[\widehat{T}_n(y)-T_n(y)\big]\Big|\geq\delta,\ \mathcal{W}_1\Big\}+\mathrm{P}\big\{\mathcal{W}_1^C\big\}\\
&\leq \mathrm{P}\Big\{\sup_{y\geq y_0} y^{\eta/\gamma_1}\Big|\sqrt{k_1}\big[\widehat{T}_n\{(1+\delta_n)y\}-T_n(y)\big]\Big|\geq\delta,\ \mathcal{W}_1\Big\}\\
&\qquad+\mathrm{P}\Big\{\sup_{y\geq y_0} y^{\eta/\gamma_1}\Big|\sqrt{k_1}\big[\widehat{T}_n\{(1-\delta_n)y\}-T_n(y)\big]\Big|\geq\delta,\ \mathcal{W}_1\Big\}+\mathrm{P}\big\{\mathcal{W}_1^C\big\}\\
&\leq\frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\varepsilon.
\end{aligned}$$

Combine this with (D.5) to obtain that

$$\sup_{y\geq y_0} y^{\eta/\gamma_1}\Big|\sqrt{k_1}\big[\widehat{T}_n(y)-y^{-1/\gamma_1}\big]-q^{-1/2}W_R\big(\infty,qy^{-1/\gamma_1}\big)\Big|=o_{\mathrm{P}}(1).$$

From this convergence, it follows as in the proof of Corollary 1 in Hoga (2017) that

$$\begin{aligned}
\sup_{y\geq y_0} y^{\eta/\gamma_1}\Big|\sqrt{k_1}\Big[\frac{1}{k_1}\sum_{t=1}^{n}I_{\{\widehat{\varepsilon}_{t,Y}>y\widehat{\varepsilon}_{(k_1+1),Y}\}}-y^{-1/\gamma_1}\Big]&\\
-q^{-1/2}\Big[W_R\big(\infty,qy^{-1/\gamma_1}\big)-y^{-1/\gamma_1}W_R(\infty,q)\Big]\Big|&=o_{\mathrm{P}}(1).
\end{aligned}$$

As in Hoga (2017, Exam. 4), we then obtain that

$$\begin{aligned}
\sqrt{k_1}(\widehat{\gamma}_1-\gamma_1)&=\sqrt{k_1}\int_1^\infty\Big[\frac{1}{k_1}\sum_{t=1}^{n}I_{\{\widehat{\varepsilon}_{t,Y}>y\widehat{\varepsilon}_{(k_1+1),Y}\}}-y^{-1/\gamma_1}\Big]\frac{\mathrm{d}y}{y}\\
&\underset{(n\to\infty)}{\overset{\mathrm{P}}{\longrightarrow}} q^{-1/2}\int_1^\infty W_R(\infty,qy^{-1/\gamma_1})\frac{\mathrm{d}y}{y}-q^{-1/2}\int_1^\infty y^{-1/\gamma_1-1}\,\mathrm{d}y\cdot W_R(\infty,q)\\
&=q^{-1/2}\gamma_1\int_0^q s^{-1}W_R(\infty,s)\,\mathrm{d}s-q^{-1/2}\gamma_1W_R(\infty,q)\\
&=\Gamma,
\end{aligned}$$

where we used the substitution $s=qy^{-1/\gamma_1}$ in the third step. This ends the proof. □

The proof of Lemma 3 builds on the preliminary Lemmas 4–6. These require the following additional notation:

$$\begin{aligned}
s_n^{\pm}(y)&=s_{n,k}\big(y\{1\pm\delta_n\}^{-1/\gamma_1}\big),\\
r_n^{\pm}(x)&=\frac{n}{k}\Big[1-F_0\big\{(1\pm\delta_n)U_0(n/[kx])\big\}\Big],
\end{aligned}\tag{D.7}$$

where $\delta_n=n^{\iota-\xi}$ as in the above proof of Lemma 2. Also let $s_n(y)=s_{n,k}(y)$ for brevity from now on.

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

LEMMA 4. *Under Assumptions 5 and 6, it holds that, as $n \to \infty$,*

$$\sup_{x\in[1/2,2]} \left|r_n^{\pm}(x) - x\right| = o\big(1/\sqrt{k}\big).$$

Proof. By the regular variation condition (5) and de Haan and Ferreira (2006, Thm. 1.2.1),

$$\frac{r_n^{\pm}(x)}{x} = \frac{\frac{n}{k}\big[1 - F_0\{(1\pm\delta_n)U_0(n/[kx])\}\big]}{\frac{n}{k}\big[1 - F_0\{U_0(n/[kx])\}\big]} \underset{(n\to\infty)}{\longrightarrow} (1\pm\delta_n)^{-1/\gamma_0}. \tag{D.8}$$

Moreover, Assumption 5 implies that there exists $t_0 > 0$ such that for $t > t_0$ and $u > 1/2$,

$$\left|\frac{u^{-\gamma_0}U_0(tu)/U_0(t) - 1}{A_0(t)}\right| < K.$$

Inserting $t = \frac{n}{kx}$ and $u = x/r_n^{\pm}(x)$ in that relation, it follows that

$$\left|\frac{\left(\frac{r_n^{\pm}(x)}{x}\right)^{\gamma_0}(1\pm\delta_n) - 1}{A_0\left(\frac{n}{kx}\right)}\right| < K.$$

Multiplying through with $|A_0\left(\frac{n}{kx}\right)|$ gives

$$\left|\left(\frac{r_n^{\pm}(x)}{x}\right)^{\gamma_0} - 1 \pm \delta_n\left(\frac{r_n^{\pm}(x)}{x}\right)^{\gamma_0}\right| < K\left|A_0\left(\frac{n}{kx}\right)\right|.$$

By (D.8), this implies

$$\left|\left(\frac{r_n^{\pm}(x)}{x}\right)^{\gamma_0} - 1\right| < K\left[\left|A_0\left(\frac{n}{kx}\right)\right| + \delta_n\right].$$

Use the Taylor expansion

$$\left(\frac{r_n^{\pm}(x)}{x}\right)^{\gamma_0} - 1 = \gamma_0\left[\frac{r_n^{\pm}(x)}{x} - 1\right] + o\left(\frac{r_n^{\pm}(x)}{x} - 1\right)$$

to deduce that

$$\left|\frac{r_n^{\pm}(x)}{x} - 1\right| < K\left[\left|A_0\left(\frac{n}{kx}\right)\right| + \delta_n\right].$$

By Assumption 6 and the Potter bounds (de Haan and Ferreira, 2006, Prop. B.1.9(5)), it holds for any $\delta > 0$ and n sufficiently large,

$$\begin{aligned}\sqrt{k}\left|A_0\left(\frac{n}{kx}\right)\right| &= \sqrt{k}|A_0(n/k)|\frac{A_0\big(n/[kx]\big)}{A_0(n/k)}\\ &= o(1)x^{-\rho_0}\max\{x^{\delta},\ x^{-\delta}\} = o(1)\end{aligned}$$

uniformly in $x \in [1/2,2]$. Since also $\sqrt{k}\delta_n = o(1)$ by Assumption 6, the conclusion follows. □

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

LEMMA 5. *Under the conditions of Theorem 1, we have that, as $n \to \infty$,*

$$\sup_{x\in[1/2,2]}\left|\frac{\sqrt{k}}{U_1(n/k)}\left[x\widehat{\theta}^*_{kx/n} - x\widetilde{\theta}^*_{kx/n}\right]\right| = o_{\mathrm{P}}(1). \tag{D.9}$$

Proof. Using the substitution $s = U_1(n/k)y^{-\gamma_1}$, we get that

$$\begin{aligned}
&-U_1(n/k)\int_0^\infty \widehat{T}_n\{x, s_n(y)\}\,\mathrm{d}y^{-\gamma_1}\\
&\quad= -U_1(n/k)\int_0^\infty \frac{1}{k}\sum_{t=1}^n I_{\left\{F_0(\widehat{\varepsilon}_{t,X})>1-kx/n,\ F_+(\widehat{\varepsilon}^+_{t,Y})>F_+(y^{-\gamma_1}U_+(n/k))\right\}}\,\mathrm{d}y^{-\gamma_1}\\
&\quad= -U_1(n/k)\frac{1}{k}\sum_{t=1}^n\int_0^\infty I_{\left\{\widehat{\varepsilon}_{t,X}>U_0(n/[kx]),\ \widehat{\varepsilon}^+_{t,Y}>y^{-\gamma_1}U_+(n/k)\right\}}\,\mathrm{d}y^{-\gamma_1}\\
&\quad= \frac{1}{k}\sum_{t=1}^n\int_0^\infty I_{\left\{\widehat{\varepsilon}_{t,X}>U_0(n/[kx]),\ \widehat{\varepsilon}^+_{t,Y}>s\right\}}\,\mathrm{d}s\\
&\quad= \frac{1}{k}\sum_{t=1}^n\int_0^{\widehat{\varepsilon}^+_{t,Y}} I_{\left\{\widehat{\varepsilon}_{t,X}>U_0(n/[kx])\right\}}\,\mathrm{d}s\\
&\quad= \frac{1}{k}\sum_{t=1}^n \widehat{\varepsilon}^+_{t,Y} I_{\left\{\widehat{\varepsilon}_{t,X}>U_0(n/[kx])\right\}}\\
&\quad= x\widehat{\theta}^*_{kx/n}.
\end{aligned}$$

By similar arguments,

$$x\widetilde{\theta}^*_{kx/n} = -U_1(n/k)\int_0^\infty T_n\left\{x, s_n(y)\right\}\mathrm{d}y^{-\gamma_1}. \tag{D.10}$$

Thus, in view of (D.9), we only have to show that

$$\sup_{x\in[1/2,2]}\left|\sqrt{k}\int_0^\infty\left[\widehat{T}_n\left\{x,s_n(y)\right\} - T_n\left\{x,s_n(y)\right\}\right]\mathrm{d}y^{-\gamma_1}\right| = o_{\mathrm{P}}(1). \tag{D.11}$$

By our choice $\delta_n = n^{\iota-\xi}$ and Proposition 2, the following inequality holds w.p.a. 1, as $n \to \infty$,

$$\begin{aligned}
\widehat{T}_n\{x, s_n(y)\} &= \frac{1}{k}\sum_{t=1}^n I_{\left\{\widehat{\varepsilon}_{t,X}>U_0\left(\frac{n}{kx}\right),\ \widehat{\varepsilon}^+_{t,Y}>U_+\left(\frac{n}{ks_n(y)}\right)\right\}}\\
&\le \frac{1}{k}\sum_{t=1}^n I_{\left\{\varepsilon_{t,X}>(1-\delta_n)U_0\left(\frac{n}{kx}\right),\ \varepsilon^+_{t,Y}>(1-\delta_n)U_+\left(\frac{n}{ks_n(y)}\right)\right\}}\\
&= \frac{1}{k}\sum_{t=1}^n I_{\left\{\varepsilon_{t,X}>U_0\left(\frac{n}{kr_n^-(x)}\right),\ \varepsilon^+_{t,Y}>U_+\left(\frac{n}{ks_n^-(y)}\right)\right\}}\\
&= T_n\{r_n^-(x),\ s_n^-(y)\}.
\end{aligned}$$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

Similarly,

$$\widehat{T}_n\{x, s_n(y)\} \geq T_n\{r_n^+(x), s_n^+(y)\}$$

w.p.a. 1, as $n \to \infty$. Fix some arbitrary $\varepsilon > 0$ and define

$$\mathcal{W}_2 := \Big\{ T_n\big\{r_n^+(x), s_n^+(y)\big\} - T_n\big\{x, s_n(y)\big\} \leq \widehat{T}_n\big\{x, s_n(y)\big\} - T_n\big\{x, s_n(y)\big\} \\ \leq T_n\big\{r_n^-(x), s_n^-(y)\big\} - T_n\big\{x, s_n(y)\big\} \Big\}.$$

Then, by the above, $\mathrm{P}\{\mathcal{W}_2\} > 1 - \varepsilon$ for sufficiently large n. Thus, to prove (D.11), it suffices to show that

$$\sup_{x \in [1/2, 2]} \left| \sqrt{k} \int_0^\infty \Big[T_n\big\{r_n^\pm(x), s_n^\pm(y)\big\} - T_n\big\{x, s_n(y)\big\} \Big] \mathrm{d}y^{-\gamma_1} \right| = o_{\mathrm{P}}(1).$$

We only do so for $T_n\{r_n^+(x), s_n^+(y)\}$, as the claim for $T_n\{r_n^-(x), s_n^-(y)\}$ can be established analogously. Decompose

$$\begin{aligned}
&\sup_{x \in [1/2, 2]} \left| \sqrt{k} \int_0^\infty \Big[T_n\big\{r_n^+(x), s_n^+(y)\big\} - T_n\big\{x, s_n(y)\big\} \Big] \mathrm{d}y^{-\gamma_1} \right| \\
&\leq \sup_{x \in [1/2, 2]} \left| \int_0^\infty \sqrt{k} \Big[T_n\big\{r_n^+(x), s_n^+(y)\big\} - R_n\big\{r_n^+(x), s_n^+(y)\big\} \Big] - W_R(x, y)\, \mathrm{d}y^{-\gamma_1} \right| \\
&\quad + \sup_{x \in [1/2, 2]} \left| \int_0^\infty \sqrt{k} \Big[T_n\big\{x, s_n(y)\big\} - R_n\big\{x, s_n(y)\big\} \Big] - W_R(x, y)\, \mathrm{d}y^{-\gamma_1} \right| \\
&\quad + \sup_{x \in [1/2, 2]} \left| \int_0^\infty \sqrt{k} \Big[R_n\big\{r_n^+(x), s_n^+(y)\big\} - R(x, y) \Big] \mathrm{d}y^{-\gamma_1} \right| \\
&\quad + \sup_{x \in [1/2, 2]} \left| \int_0^\infty \sqrt{k} \Big[R_n\big\{x, s_n(y)\big\} - R(x, y) \Big] \mathrm{d}y^{-\gamma_1} \right| \\
&= C_1 + C_2 + C_3 + C_4.
\end{aligned}$$

We show in turn that $C_1, \ldots, C_4$ are asymptotically negligible.

By (D.10) and the fact that similarly $x\theta_{kx/n} = -U_1(n/k) \int_0^\infty R_n\{x, s_n(y)\}\, \mathrm{d}y^{-\gamma_1}$,

$$C_2 = \sup_{x \in [1/2, 2]} \left| \frac{\sqrt{k}}{U_1(n/k)} \big[x\widetilde{\theta}^*_{kx/n} - x\theta_{kx/n} \big] - \int_0^\infty W_R(x, y)\, \mathrm{d}y^{-\gamma_1} \right|. \tag{D.12}$$

Thus, $C_2 = o_{\mathrm{P}}(1)$ follows from Proposition 2 in Cai et al. (2015).

Carefully reading the proof of that proposition reveals that $s_n(y)$ in (D.12) (appearing in both $x\widetilde{\theta}^*_{kx/n}$ via (D.10), and $x\theta_{kx/n} = -U_1(n/k) \int_0^\infty R_n\{x, s_n(y)\}\, \mathrm{d}y^{-\gamma_1}$) can be replaced by $s_n^+(y)$ without changing the conclusion that the term is $o_{\mathrm{P}}(1)$. This exploits the fact that $\sup_{y \in (0,1]} s_n(y)/y^{(\gamma_1 + \eta_1)} \to 0$ for $\eta_1 > \gamma_1$ and $\sup_{y \in (0,T]} |s_n(y) - y| \to 0$ continue to hold for $s_n^+(y)$. Additionally, using Lemma 4, we see that x in (D.12) can be replaced by $r_n^+(x)$, such that

$$\sup_{x \in [1/2, 2]} \left| \int_0^\infty \sqrt{k} \big[T_n\{r_n^+(x), s_n^+(y)\} - R_n\{r_n^+(x), s_n^+(y)\} \big] - W_R\{r_n^+(x), y\}\, \mathrm{d}y^{-\gamma_1} \right| = o_{\mathrm{P}}(1).$$

Hence, $C_1 = o_{\mathrm{P}}(1)$ follows if we can show that

$$\sup_{x\in[1/2,2]}\left|\int_0^{\infty} W_R(x,y) - W_R\{r_n^+(x),y\}\,\mathrm{d}y^{-\gamma_1}\right| = o_{\mathrm{P}}(1). \tag{D.13}$$

By Corollary 1.11 of Adler (1990), $(x,y)\mapsto W_R(x,y)$ is continuous on $[1/2,2]\times(0,\infty)$. This implies that

$$[1/2,2]\ni x\mapsto \int_0^{\infty} W_R(x,y)\,\mathrm{d}y^{-\gamma_1}$$

is continuous and, hence, uniformly continuous on the bounded interval $[1/2,2]$. Thus, (D.13) follows from Lemma 4, showing that $C_1 = o_{\mathrm{P}}(1)$.

The fact that $C_4 = o(1)$ follows directly from Cai et al. (2015, p. 439).

To prove $C_3 = o(1)$, consider the bound

$$\begin{aligned}
C_3 \le{}& \sup_{x\in[1/2,2]}\left|\int_0^{\infty}\sqrt{k}\Big[R_n\big\{r_n^+(x),s_n^+(y)\big\} - R\big\{r_n^+(x),y(1+\delta_n)^{-1/\gamma_1}\big\}\Big]\,\mathrm{d}y^{-\gamma_1}\right| \\
&+ \sup_{x\in[1/2,2]}\left|\int_0^{\infty}\sqrt{k}\Big[R\big\{r_n^+(x),y(1+\delta_n)^{-1/\gamma_1}\big\} - R\big\{r_n^+(x),y\big\}\Big]\,\mathrm{d}y^{-\gamma_1}\right| \\
&+ \sup_{x\in[1/2,2]}\left|\int_0^{\infty}\sqrt{k}\Big[R\big\{r_n^+(x),y\big\} - R(x,y)\Big]\,\mathrm{d}y^{-\gamma_1}\right| \\
={}& C_{31}+C_{32}+C_{33}.
\end{aligned}$$

Recalling that $s_n^+(y) = s_n\{y(1+\delta_n)^{-1/\gamma_1}\}$, we obtain by a change of variables that

$$\begin{aligned}
C_{31} &= \sup_{x\in[1/2,2]}\left|\int_0^{\infty}\sqrt{k}\Big[R_n\big\{r_n^+(x),s_n^+(y)\big\} - R\big\{r_n^+(x),y(1+\delta_n)^{-1/\gamma_1}\big\}\Big]\,\mathrm{d}y^{-\gamma_1}\right| \\
&= (1+\delta_n)^{-1}\sup_{x\in[1/2,2]}\left|\int_0^{\infty}\sqrt{k}\Big[R_n\big\{r_n^+(x),s_n(y)\big\} - R\big\{r_n^+(x),y\big\}\Big]\,\mathrm{d}y^{-\gamma_1}\right| \\
&= o(1),
\end{aligned}$$

where we used the fact that $C_4 = o(1)$ together with Lemma 4 in the final step. For C_{32}, we again use a change of variables to obtain

$$\int_0^{\infty} R\big\{r_n^+(x),y(1+\delta_n)^{-1/\gamma_1}\big\}\,\mathrm{d}y^{-\gamma_1} = (1+\delta_n)^{-1}\int_0^{\infty} R\big\{r_n^+(x),y\big\}\,\mathrm{d}y^{-\gamma_1}.$$

Thus, since $(1+\delta_n)^{-1} = 1-\delta_n+o(\delta_n)$ from a Taylor expansion,

$$C_{32} \le K\sqrt{k}\delta_n \sup_{x\in[1/2,2]}\int_0^{\infty} R\{r_n^+(x),y\}\,\mathrm{d}y^{-\gamma_1} = o(1).$$

Finally, Schmidt and Stadtmüller (2006, Thm. 1(ii)) establish homogeneity of the R-function, i.e., $R(sx,sy) = sR(x,y)$ for all $s>0$ and $x,y\ge 0$. Using this and a change of variables,

$$\begin{aligned}
&\int_0^{\infty}\sqrt{k}\Big[R\big\{r_n^+(x),y\big\} - R(x,y)\Big]\,\mathrm{d}y^{-\gamma_1} \\
&\qquad = \sqrt{k}\Big[\int_0^{\infty} R\big\{r_n^+(x),y\big\}\,\mathrm{d}y^{-\gamma_1} - \int_0^{\infty} R(x,y)\,\mathrm{d}y^{-\gamma_1}\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{k}\left[\int_0^\infty r_n^+(x) R\{1, y/r_n^+(x)\}\, \mathrm{d}y^{-\gamma_1} - \int_0^\infty x R(1, y/x)\, \mathrm{d}y^{-\gamma_1}\right] \\
&= \sqrt{k}\left[r_n^+(x)^{1-\gamma_1} - x^{1-\gamma_1}\right] \int_0^\infty R(1, y)\, \mathrm{d}y^{-\gamma_1} \\
&= o(1)
\end{aligned}
$$

uniformly in $x \in [1/2, 2]$ by Lemma 4. Thus, $C_{33} = o(1)$, and the conclusion follows. □

LEMMA 6. *Under the conditions of Theorem 1, we have that, as $n \to \infty$,*

$$\sqrt{k}(e_n - 1) \xrightarrow{\text{P}} -W_R(1, \infty), \tag{D.14}$$

$$\sqrt{k}(\widehat{e}_n - 1) \xrightarrow{\text{P}} -W_R(1, \infty), \tag{D.15}$$

where e_n and $\widehat{e}_n$ are defined in (C.2) and (C.3), respectively.

Proof. From (C.1),

$$\sup_{x \in (0, T]} \left| \sqrt{k}\left[\frac{1}{k}\sum_{t=1}^{n} I_{\left\{\varepsilon_{t,X} > U_0(n/[kx])\right\}} - x\right] - W_R(x, \infty)\right| \xrightarrow{a.s.} 0. \tag{D.16}$$

Put $T_{n,1}(x) = \frac{1}{k}\sum_{t=1}^{n} I_{\left\{\varepsilon_{t,X} > U_0(n/[kx])\right\}}$ for short. The proof draws heavily on Example A.0.3 in de Haan and Ferreira (2006). Define

$$T_{n,1}^{\leftrightsquigarrow}(x) := \frac{n}{k}\left[1 - F_0\left(\varepsilon_{(\lfloor kx \rfloor + 1), X}\right)\right].$$

Then,

$$\sup_{x \in (0, T]} \left| T_{n,1}\{T_{n,1}^{\leftrightsquigarrow}(x)\} - x\right| = \sup_{x \in (0, T]} \left| \frac{\lfloor kx \rfloor}{k} - x\right| \le \frac{1}{k},$$

which implies

$$\sup_{x \in (0, T]} \left| T_{n,1}^{\leftrightsquigarrow}(x) - T_{n,1}^{\leftarrow}(x)\right| = o\left(1/\sqrt{k}\right). \tag{D.17}$$

From (D.16), we obtain via Vervaat's (1972) lemma that

$$\sup_{x \in [1/2, 2]} \left| \sqrt{k}\left[T_{n,1}^{\leftarrow}(x) - x\right] + W_R(x, \infty)\right| \overset{a.s.}{=} o(1),$$

where "←" denotes the left-continuous inverse. Due to (D.17), this implies

$$\sup_{x \in [1/2, 2]} \left| \sqrt{k}\left[T_{n,1}^{\leftrightsquigarrow}(x) - x\right] + W_R(x, \infty)\right| \overset{a.s.}{=} o(1).$$

Putting $x = 1$ in that expression, (D.14) follows.

Arguing similarly as in the proof of Lemma 2, we may show that (C.1) remains valid if the $\varepsilon_{t,X}$ in $T_n(x, \infty)$ are replaced by $\widehat{\varepsilon}_{t,X}$. Hence, (D.15) follows as before. □

Now, we are in a position to prove Lemma 3.

Proof of Lemma 3. Recall that $\widehat{e}_n\widehat{\theta}^*_{k\widehat{e}_n/n}=\widehat{\theta}_{k/n}$ and $e_n\widetilde{\theta}^*_{ke_n/n}=\widetilde{\theta}_{k/n}$. With this, write

$$\begin{aligned}\frac{\sqrt{k}}{U_1(n/k)}\Big[\widehat{\theta}_{k/n}-\widetilde{\theta}_{k/n}\Big]&=\frac{\sqrt{k}}{U_1(n/k)}\Big[\widehat{e}_n\widehat{\theta}^*_{k\widehat{e}_n/n}-e_n\widetilde{\theta}^*_{ke_n/n}\Big]\\&=\frac{\sqrt{k}}{U_1(n/k)}\Big[\widehat{e}_n\widehat{\theta}^*_{k\widehat{e}_n/n}-\widehat{e}_n\widetilde{\theta}^*_{k\widehat{e}_n/n}+\widehat{e}_n\widetilde{\theta}^*_{k\widehat{e}_n/n}-e_n\widetilde{\theta}^*_{ke_n/n}\Big]\\&=o_{\mathrm{P}}(1)+\frac{\sqrt{k}}{U_1(n/k)}\Big[\widehat{e}_n\widetilde{\theta}^*_{k\widehat{e}_n/n}-e_n\widetilde{\theta}^*_{ke_n/n}\Big],\end{aligned}$$

where the final line follows from Lemmas 5 and 6. Decompose the remaining term as follows:

$$\begin{aligned}\frac{\sqrt{k}}{U_1(n/k)}\Big[\widehat{e}_n\widetilde{\theta}^*_{k\widehat{e}_n/n}-e_n\widetilde{\theta}^*_{ke_n/n}\Big]&=\frac{\sqrt{k}}{U_1(n/k)}e_n\Big[\widetilde{\theta}^*_{k\widehat{e}_n/n}-\widetilde{\theta}^*_{ke_n/n}\Big]+\frac{\sqrt{k}}{U_1(n/k)}\big[\widehat{e}_n-e_n\big]\widetilde{\theta}^*_{k\widehat{e}_n/n}\\&=:D_1+D_2.\end{aligned}$$

Consider D_1 and D_2 separately. For D_1, we get that

$$\begin{aligned}D_1&=e_n\left\{\frac{\sqrt{k}}{U_1(n/k)}\big[\widetilde{\theta}^*_{k\widehat{e}_n/n}-\theta_{k\widehat{e}_n/n}\big]+\int_0^\infty W_R(1,s)\,\mathrm{d}s^{-\gamma_1}\right\}\\&\quad-e_n\left\{\frac{\sqrt{k}}{U_1(n/k)}\big[\widetilde{\theta}^*_{ke_n/n}-\theta_{ke_n/n}\big]+\int_0^\infty W_R(1,s)\,\mathrm{d}s^{-\gamma_1}\right\}\\&\quad+e_n\frac{\sqrt{k}}{U_1(n/k)}\big[\theta_{k\widehat{e}_n/n}-\theta_{ke_n/n}\big]\\&=e_n\frac{\sqrt{k}}{U_1(n/k)}\big[\theta_{k\widehat{e}_n/n}-\theta_{ke_n/n}\big]+o_{\mathrm{P}}(1),\end{aligned}\tag{D.18}$$

where we have used that $e_n=1+o_{\mathrm{P}}(1)$ from Lemma 6 and $\frac{\sqrt{k}}{U_1(n/k)}\big[\widetilde{\theta}^*_{ke_n/n}-\theta_{ke_n/n}\big]\xrightarrow{\mathrm{P}}-\int_0^\infty W_R(1,s)\,\mathrm{d}s^{-\gamma_1}$ from the proof of Proposition 3 in Cai et al. (2015). The fact that this convergence also holds when replacing e_n with $\widehat{e}_n$ follows similarly, since e_n and $\widehat{e}_n$ are asymptotically equivalent by Lemma 6. For the remaining term, we need the following result from Cai et al. (2015, p. 439):

$$e_n\theta_{ke_n/n}=e_n^{1-\gamma_1}\theta_{k/n}+o_{\mathrm{P}}\{U_1(n/k)/\sqrt{k}\}.$$

Again from Lemma 6, it follows that this continues to hold upon replacing e_n with $\widehat{e}_n$, such that

$$\widehat{e}_n\theta_{k\widehat{e}_n/n}=\widehat{e}_n^{1-\gamma_1}\theta_{k/n}+o_{\mathrm{P}}\{U_1(n/k)/\sqrt{k}\}.$$

Using these two results, we deduce that

$$\begin{aligned}e_n\frac{\sqrt{k}}{U_1(n/k)}\big[\theta_{k\widehat{e}_n/n}-\theta_{ke_n/n}\big]&=\frac{\sqrt{k}}{U_1(n/k)}\Big[\frac{e_n}{\widehat{e}_n}\{\widehat{e}_n\theta_{k\widehat{e}_n/n}\}-\theta_{k/n}\Big]\\&\quad-\frac{\sqrt{k}}{U_1(n/k)}\Big[e_n\theta_{ke_n/n}-\theta_{k/n}\Big]\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sqrt{k}}{U_1(n/k)}\Big[\frac{e_n}{\widehat{e}_n}\Big\{\widehat{e}_n^{1-\gamma_1}\theta_{k/n} + o_{\mathrm{P}}\{U_1(n/k)/\sqrt{k}\}\Big\} - \theta_{k/n}\Big] \\
&\qquad - \frac{\sqrt{k}}{U_1(n/k)}\Big[e_n^{1-\gamma_1}\theta_{k/n} + o_{\mathrm{P}}\{U_1(n/k)/\sqrt{k}\} - \theta_{k/n}\Big] \\
&= \frac{\sqrt{k}}{U_1(n/k)}\Big[\frac{e_n}{\widehat{e}_n}\big(\widehat{e}_n^{1-\gamma_1} - 1\big)\theta_{k/n}\Big] - \frac{\sqrt{k}}{U_1(n/k)}\Big[\frac{e_n}{\widehat{e}_n} - 1\Big]\theta_{k/n} \\
&\qquad - \frac{\sqrt{k}}{U_1(n/k)}\Big[\big(e_n^{1-\gamma_1} - 1\big)\theta_{k/n}\Big] + o_{\mathrm{P}}(1) \\
&= \frac{e_n}{\widehat{e}_n}\cdot\sqrt{k}\big(\widehat{e}_n^{1-\gamma_1} - 1\big)\cdot\frac{\theta_{k/n}}{U_1(n/k)} - \frac{1}{\widehat{e}_n}\cdot\sqrt{k}\big[e_n - \widehat{e}_n\big]\cdot\frac{\theta_{k/n}}{U_1(n/k)} \\
&\qquad - \sqrt{k}\big(e_n^{1-\gamma_1} - 1\big)\cdot\frac{\theta_{k/n}}{U_1(n/k)} + o_{\mathrm{P}}(1) \\
&= o_{\mathrm{P}}(1), \qquad\qquad \textbf{(D.19)}
\end{aligned}
$$

where the final step follows from

$$
\sqrt{k}\big(\widehat{e}_n^{1-\gamma_1} - 1\big) \xrightarrow{\mathrm{P}} (\gamma_1 - 1)W_R(1,\infty),
$$

$$
\sqrt{k}\big(e_n^{1-\gamma_1} - 1\big) \xrightarrow{\mathrm{P}} (\gamma_1 - 1)W_R(1,\infty)
$$

(as a consequence of the delta method applied to Lemma 6) and $\theta_{k/n}/U_1(n/k) \to \int_0^\infty R(1, s^{-1/\gamma_1})\,\mathrm{d}s$ (from Prop. 1 of Cai et al., 2015). From (D.18) and (D.19), $D_1 = o_{\mathrm{P}}(1)$.

For D_2, write

$$
\begin{aligned}
D_2 &= \big[\sqrt{k}(e_n - 1) - \sqrt{k}(\widehat{e}_n - 1)\big]\cdot\frac{\theta_{k/n}}{U_1(n/k)}\cdot\frac{1}{\widehat{e}_n}\cdot\frac{\widehat{e}_n\widetilde{\theta}^*_{k\widehat{e}_n/n}}{\theta_{k/n}} \\
&= \big[-W_R(1,\infty) + W_R(1,\infty) + o_{\mathrm{P}}(1)\big]\cdot\Big[\int_0^\infty R(1, s^{-1/\gamma_1})\,\mathrm{d}s + o(1)\Big]\cdot\big[1 + o_{\mathrm{P}}(1)\big] \\
&\qquad \cdot\Big[\frac{1}{\sqrt{k}}\frac{U_1(n/k)}{\theta_{k/n}}\frac{\sqrt{k}}{U_1(n/k)}\big(\widehat{e}_n\widetilde{\theta}^*_{k\widehat{e}_n/n} - \theta_{k/n}\big) + 1\Big] \\
&= o_{\mathrm{P}}(1)\cdot\Big[\frac{1}{\sqrt{k}}\Big\{\int_0^\infty R(1, s^{-1/\gamma_1})\,\mathrm{d}s + o(1)\Big\}^{-1}\Theta\int_0^\infty R(1, s^{-1/\gamma_1})\,\mathrm{d}s + 1\Big] \\
&= o_{\mathrm{P}}(1),
\end{aligned}
$$

where we used in the second-to-last step that (6) holds and that

$$
\frac{\sqrt{k}}{U_1(n/k)}\big(\widehat{e}_n\widetilde{\theta}^*_{k\widehat{e}_n/n} - \theta_{k/n}\big) \xrightarrow{\mathrm{P}} \Theta\int_0^\infty R(1, s^{-1/\gamma_1})\,\mathrm{d}s,
$$

which follows from the proof of Proposition 3 in Cai et al. (2015) (together with Lemma 6). The conclusion of the lemma follows. □

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

E. Proof of Theorem 2

Proof of Theorem 2. It follows similarly as in the proof of Theorem 1 that

$$\min\left\{\sqrt{k},\sqrt{k_d}/\log(d_n)\right\}\log\left(\frac{\widehat{\theta}_{n,p,d}}{\theta_{n,p,d}}\right)=\min\left\{\sqrt{k},\sqrt{k_d}/\log(d_n)\right\}\log\left(\frac{\widehat{\theta}_{p,d}}{\theta_{p,d}}\right)+o_{\mathrm{P}}(1).$$

From analogous arguments used in the proof of Proposition 3 (see, in particular, (C.4) and (C.5)), we obtain for all $d=1,\dots,D$ that

$$\min\left\{\sqrt{k},\sqrt{k_d}/\log(d_n)\right\}\log\left(\frac{\widehat{\theta}_{p,d}}{\theta_{p,d}}\right)=\sqrt{k_d}(\widehat{\gamma}_d-\gamma_d)+o_{\mathrm{P}}(1)=\Gamma_d+o_{\mathrm{P}}(1),$$

where $\Gamma_d\sim N(0,\gamma_d^2)$. Note that this exploits the assumption that $r_d=\lim_{n\to\infty}\sqrt{k}\log(d_n)/\sqrt{k_d}=\infty$. Therefore, the proof is finished if we show that, as $n\to\infty$,

$$\left(\sqrt{k_1}\{\widehat{\gamma}_1-\gamma_1\},\dots,\sqrt{k_D}\{\widehat{\gamma}_D-\gamma_D\}\right)'\xrightarrow{d}(\Gamma_1,\dots,\Gamma_D)',$$

where $\mathrm{Cov}(\Gamma_i,\Gamma_j)=\sigma_{i,j}$ for $i,j=1,\dots,D$. To that end, we apply similar arguments as used in the proof of Proposition 3 in Hoga (2018). Specifically, we have to verify his conditions (M1)–(M4). Note that we cannot directly apply his Proposition 3, because it derives the joint limit of the Hill estimates only for a common intermediate sequence, whereas we allow for possibly distinct $k_1,\dots,k_D$.

Since the $\boldsymbol{\varepsilon}_t$ are i.i.d., the β-mixing condition (M1) is immediate for any $r_n\to\infty$ with $r_n=o\big(\sqrt{k}/\log^2(k)\big)$ in the notation of Hoga (2018). In the following, we let $r_n=o\big(\min\{\sqrt{k}/\log^2(k),n/k\}\big)$.

For (M2), note by independence of the $\boldsymbol{\varepsilon}_t$ that

$$\begin{aligned}
&\frac{n}{r_nk}\mathrm{Cov}\left(\sum_{t=1}^{r_n}I_{\left\{\varepsilon_{t,Y_i}>U_i(\frac{n}{kx})\right\}},\sum_{t=1}^{r_n}I_{\left\{\varepsilon_{t,Y_j}>U_j(\frac{n}{ky})\right\}}\right)\\
&\quad=\frac{n}{r_nk}\sum_{t=1}^{r_n}\mathrm{Cov}\left(I_{\left\{\varepsilon_{t,Y_i}>U_i(\frac{n}{kx})\right\}},I_{\left\{\varepsilon_{t,Y_j}>U_j(\frac{n}{ky})\right\}}\right)\\
&\quad=\frac{n}{r_nk}\sum_{t=1}^{r_n}\left[\mathrm{P}\left\{\varepsilon_{t,Y_i}>U_i\Big(\frac{n}{kx}\Big),\ \varepsilon_{t,Y_j}>U_j\Big(\frac{n}{ky}\Big)\right\}-\frac{kx}{n}\frac{ky}{n}\right]\\
&\quad=\frac{n}{k}\mathrm{P}\left\{\varepsilon_{t,Y_i}>U_i\Big(\frac{n}{kx}\Big),\ \varepsilon_{t,Y_j}>U_j\Big(\frac{n}{ky}\Big)\right\}+\frac{n}{k}\frac{k^2xy}{n^2}\\
&\quad=R_{i,j}(x,y)+o(1)+o(k/n)\underset{(n\to\infty)}{\longrightarrow}R_{i,j}(x,y),
\end{aligned}$$

where the final line uses Assumption 4*. This establishes (M2).

For (M3), note that since the $\boldsymbol{\varepsilon}_t$ are i.i.d., their ρ-mixing coefficients are trivially zero, such that, by Lemma 2.3 of Shao (1993) (for $q=4$ in his notation),

$$\begin{aligned}
&\frac{n}{r_nk}\mathrm{E}\left[\sum_{t=1}^{r_n}I_{\left\{U_d(\frac{n}{ky})<\varepsilon_{t,Y_d}\le U_d(\frac{n}{kx})\right\}}\right]^4\\
&\le\frac{n}{r_nk}K\left\{r_n^2\mathrm{E}^2\left[I^2_{\left\{U_d(\frac{n}{ky})<\varepsilon_{t,Y_d}\le U_d(\frac{n}{kx})\right\}}\right]+r_n\mathrm{E}\left[I^4_{\left\{U_d(\frac{n}{ky})<\varepsilon_{t,Y_d}\le U_d(\frac{n}{kx})\right\}}\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{n}{r_n k} K\left\{ r_n^2 \mathrm{P}^2\left\{ U_d\Big(\frac{n}{ky}\Big) < \varepsilon_{t,Y_d} \le U_d\Big(\frac{n}{kx}\Big)\right\} + r_n \mathrm{P}\left\{ U_d\Big(\frac{n}{ky}\Big) < \varepsilon_{t,Y_d} \le U_d\Big(\frac{n}{kx}\Big)\right\}\right\} \\
&= \frac{n}{r_n k} K\left\{ r_n^2 \frac{k^2}{n^2}(y-x)^2 + r_n \frac{k}{n}(y-x)\right\} \\
&\le K\frac{r_n k}{n}(y-x)^2 + K(y-x) \underset{(n\to\infty)}{\longrightarrow} K(y-x),
\end{aligned}
$$

since $r_n = o(n/k)$. This implies (M3).

Finally, condition (M4) follows from Assumption 5*.

Having verified the conditions of Proposition 3 of Hoga (2018), we may follow the steps in that proof to derive that for any $\eta \in [0,1/2)$,

$$
\sup_{y\in(0,T]} y^{-\eta} \left\| \sqrt{k}\begin{pmatrix} T_{n,1}(y)-y \\ \vdots \\ T_{n,D}(y)-y \end{pmatrix} - \begin{pmatrix} W_1(y) \\ \vdots \\ W_D(y) \end{pmatrix} \right\| \xrightarrow{a.s.} 0, \tag{E.1}
$$

where $T_{n,d}(y) := \frac{1}{k}\sum_{t=1}^{n} I_{\{\varepsilon_{t,Y_d} > U_d(n/[ky])\}}$, and $\boldsymbol{W}(y) = \big(W_1(y),\ldots,W_D(y)\big)'$ is a D-variate continuous, zero-mean Gaussian process with covariance function

$$
\operatorname{Cov}\big(\boldsymbol{W}(y_1),\boldsymbol{W}(y_2)\big) = \left[\frac{R_{i,j}(y_1,y_2)+R_{i,j}(y_2,y_1)}{2}\right]_{i,j=1,\ldots,D};
$$

see, in particular, Lemma 2 of Hoga (2018). Equation (E.1) is the analog of (D.2) in the proof of Lemma 2.

Applying the steps in the proof of Lemma 2 to each of the components in (E.1) gives, as $n\to\infty$,

$$
\sqrt{k_d}(\widehat{\gamma}_d - \gamma_d) \xrightarrow{\mathrm{P}} \gamma_d q_d^{-1/2}\left[\int_0^1 u^{-1} W_d(q_d u)\,\mathrm{d}u - W_d(q_d)\right] = \Gamma_d
$$

for each $d = 1,\ldots,D$. Since $\mathrm{E}[\Gamma_d] = 0$, we obtain for $i,j = 1,\ldots,D$ using Fubini's theorem that

$$
\begin{aligned}
\operatorname{Cov}(\Gamma_i,\Gamma_j) &= \mathrm{E}[\Gamma_i\Gamma_j] \\
&= \frac{\gamma_i\gamma_j}{\sqrt{q_iq_j}}\mathrm{E}\left[\left\{\int_0^1 u^{-1}W_i(q_iu) - W_i(q_i)\,\mathrm{d}u\right\}\left\{\int_0^1 v^{-1}W_j(q_jv) - W_j(q_j)\,\mathrm{d}v\right\}\right] \\
&= \frac{\gamma_i\gamma_j}{\sqrt{q_iq_j}}\mathrm{E}\left[\int_0^1\int_0^1 \big\{u^{-1}W_i(q_iu) - W_i(q_i)\big\}\big\{v^{-1}W_j(q_jv) - W_j(q_j)\big\}\,\mathrm{d}u\,\mathrm{d}v\right] \\
&= \frac{\gamma_i\gamma_j}{\sqrt{q_iq_j}}\int_0^1\int_0^1 \mathrm{E}\Big[\big\{u^{-1}W_i(q_iu) - W_i(q_i)\big\}\big\{v^{-1}W_j(q_jv) - W_j(q_j)\big\}\Big]\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{\gamma_i\gamma_j}{\sqrt{q_iq_j}}\int_0^1\int_0^1 \frac{\mathrm{E}\big[W_i(q_iu)W_j(q_jv)\big]}{uv} - \frac{\mathrm{E}\big[W_i(q_iu)W_j(q_j)\big]}{u} \\
&\qquad\qquad - \frac{\mathrm{E}\big[W_i(q_i)W_j(q_jv)\big]}{v} + \mathrm{E}\big[W_i(q_i)W_j(q_j)\big]\,\mathrm{d}u\,\mathrm{d}v
\end{aligned}
$$

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

$$= \frac{\gamma_i\gamma_j}{\sqrt{q_iq_j}} \int_0^1\int_0^1 \frac{R_{i,j}(q_iu,q_jv)+R_{i,j}(q_jv,q_iu)}{2uv} - \frac{R_{i,j}(q_iu,q_j)+R_{i,j}(q_j,q_iu)}{2u} - \frac{R_{i,j}(q_i,q_jv)+R_{i,j}(q_jv,q_i)}{2v} + \frac{R_{i,j}(q_i,q_j)+R_{i,j}(q_j,q_i)}{2}\,\mathrm{d}u\,\mathrm{d}v.$$

Due to Theorem 1(ii) of Schmidt and Stadtmüller (2006), the $R_{i,j}(\cdot,\cdot)$-function is homogeneous (i.e., $R_{i,j}(sx,sy) = sR_{i,j}(x,y)$ for all $s>0, x,y\geq 0$). Hence,

$$\begin{aligned}
&\int_0^1\int_0^1 \frac{R_{i,j}(q_iu,q_jv)}{2uv}\,\mathrm{d}u\,\mathrm{d}v\\
&\quad= \int_0^1\left[\int_0^v \frac{R_{i,j}(q_iu,q_jv)}{2uv}\,\mathrm{d}u\right]\mathrm{d}v + \int_0^1\left[\int_0^u \frac{R_{i,j}(q_iu,q_jv)}{2uv}\,\mathrm{d}v\right]\mathrm{d}u\\
&\quad= \int_0^1\left[\int_0^v \frac{R_{i,j}(q_iu/v,q_j)}{2u}\,\mathrm{d}u\right]\mathrm{d}v + \int_0^1\left[\int_0^u \frac{R_{i,j}(q_i,q_jv/u)}{2v}\,\mathrm{d}v\right]\mathrm{d}u\\
&\quad= \int_0^1\left[\int_0^1 \frac{R_{i,j}(q_iu,q_j)}{2u}\,\mathrm{d}u\right]\mathrm{d}v + \int_0^1\left[\int_0^1 \frac{R_{i,j}(q_i,q_jv)}{2v}\,\mathrm{d}v\right]\mathrm{d}u\\
&\quad= \int_0^1 \frac{R_{i,j}(q_iu,q_j)}{2u}\,\mathrm{d}u + \int_0^1 \frac{R_{i,j}(q_i,q_jv)}{2v}\,\mathrm{d}v.
\end{aligned}$$

Similar arguments yield that

$$\int_0^1\int_0^1 \frac{R_{i,j}(q_jv,q_iu)}{2uv}\,\mathrm{d}u\,\mathrm{d}v = \int_0^1 \frac{R_{i,j}(q_ju,q_i)}{2u}\,\mathrm{d}u + \int_0^1 \frac{R_{i,j}(q_j,q_iv)}{2v}\,\mathrm{d}v.$$

Therefore,

$$\mathrm{Cov}(\Gamma_i,\Gamma_j) = \frac{\gamma_i\gamma_j}{\sqrt{q_iq_j}}\frac{R_{i,j}(q_i,q_j)+R_{i,j}(q_j,q_i)}{2},$$

as claimed. This finishes the proof. □

F. Proof of Proposition 1

Here, we only show that $\widehat{R}_{i,j}(q_i,q_j) \xrightarrow{\mathrm{P}} R_{i,j}(q_i,q_j)$, since the other claim of the proposition can be established analogously. The proof requires the preliminary Lemmas 7–10 for which we have to introduce some additional notation.

Fix $i,j\in\{1,\ldots,D\}$, and define

$$\widetilde{\varphi}_{k/n}(x,y) = \frac{1}{k}\sum_{t=1}^{n} I_{\left\{\varepsilon_{t,Y_i}>\varepsilon_{(\lfloor kx\rfloor+1),Y_i},\ \varepsilon_{t,Y_j}>\varepsilon_{(\lfloor ky\rfloor+1),Y_j}\right\}},$$

$$\widetilde{\varphi}^{*}_{k/n}(x,y) = \frac{1}{k}\sum_{t=1}^{n} I_{\left\{\varepsilon_{t,Y_i}>U_i(n/[kx]),\ \varepsilon_{t,Y_j}>U_j(n/[ky])\right\}},$$

$$\widehat{\varphi}_{k/n}(x,y) = \frac{1}{k}\sum_{t=1}^{n} I_{\left\{\widehat{\varepsilon}_{t,Y_i} > \widehat{\varepsilon}_{(\lfloor kx\rfloor+1),Y_i},\ \widehat{\varepsilon}_{t,Y_j} > \widehat{\varepsilon}_{(\lfloor ky\rfloor+1),Y_j}\right\}},$$

$$\widehat{\varphi}^{*}_{k/n}(x,y) = \frac{1}{k}\sum_{t=1}^{n} I_{\left\{\widehat{\varepsilon}_{t,Y_i} > U_i(n/[kx]),\ \widehat{\varepsilon}_{t,Y_j} > U_j(n/[ky])\right\}},$$

and

$$e_{n,i} = \frac{n}{k}\left[1 - F_i\left(\varepsilon_{(k_i+1),Y_i}\right)\right],$$
$$\widehat{e}_{n,i} = \frac{n}{k}\left[1 - F_i\left(\widehat{\varepsilon}_{(k_i+1),Y_i}\right)\right].$$

In analogy to $R_n(x,y)$ in Appendix C, we also define

$$R_{n,i,j}(x,y) = \frac{n}{k}\mathrm{P}\left\{F_i(\varepsilon_{t,Y_i}) > 1 - kx/n,\ F_j(\varepsilon_{t,Y_j}) > 1 - ky/n\right\}.$$

We have the following analog to Lemma 1, which again follows from Proposition 3.1 of Einmahl et al. (2006).

LEMMA 7. *Suppose that (12) holds. Then, for any $\eta \in [0,1/2)$ and $T > 0$, it holds that, as $n \to \infty$,*

$$\sup_{x,y\in(0,T]} y^{-\eta}\left|\sqrt{k}\left\{\widetilde{\varphi}^{*}_{k/n}(x,y) - R_{n,i,j}(x,y)\right\} - W_{i,j}(x,y)\right| \xrightarrow{a.s.} 0,$$
$$\sup_{x\in(0,T]} x^{-\eta}\left|\sqrt{k}\left\{\widetilde{\varphi}^{*}_{k/n}(x,\infty) - x\right\} - W_{i,j}(x,\infty)\right| \xrightarrow{a.s.} 0, \tag{F.1}$$

where $W_{i,j}(\cdot,\cdot)$ is a zero-mean Gaussian process with covariance structure given by

$$\mathrm{E}\left[W_{i,j}(x_1,y_1)W_{i,j}(x_2,y_2)\right] = R_{i,j}(x_1 \wedge x_2, y_1 \wedge y_2).$$

The first step is to prove the following analog of Lemma 6.

LEMMA 8. *Under the conditions of Theorem 2, we have that, as $n \to \infty$,*

$$\sqrt{k}\left(e_{n,i} - \frac{k_i}{k}\right) \xrightarrow{\mathrm{P}} -W_{i,j}(q_i,\infty),$$
$$\sqrt{k}\left(\widehat{e}_{n,i} - \frac{k_i}{k}\right) \xrightarrow{\mathrm{P}} -W_{i,j}(q_i,\infty).$$

Proof. From (F.1),

$$\sup_{x\in(0,T]} x^{-\eta}\left|\sqrt{k}\left\{\frac{1}{k}\sum_{t=1}^{n} I_{\left\{\varepsilon_{t,Y_i} > U_i(n/[kx])\right\}} - x\right\} - W_{i,j}(x,\infty)\right| \xrightarrow{a.s.} 0.$$

Following the steps leading up to (D.3), this implies

$$\sup_{x\in(0,T]} x^{-\eta}\left|\sqrt{k_i}\left\{\frac{1}{k_i}\sum_{t=1}^{n} I_{\left\{\varepsilon_{t,Y_i} > U_i(n/[k_i x])\right\}} - x\right\} - q^{-1/2}W_{i,j}(q_i x,\infty)\right| \xrightarrow{a.s.} 0.$$

From this relation, we may argue as in the proof of Lemma 6 to obtain that

$$\sqrt{k_i}\Big(\frac{k}{k_i}e_{n,i}-1\Big)\stackrel{\mathrm{P}}{\longrightarrow}-q_i^{-1/2}W_{i,j}(q_i,\infty)$$

and, because $k_i/k\to q_i$,

$$\sqrt{k}\Big(e_{n,i}-\frac{k_i}{k}\Big)\stackrel{\mathrm{P}}{\longrightarrow}-W_{i,j}(q_i,\infty).$$

The claim for $\widehat{e}_{n,i}$ follows similarly by showing (as in the proof of Lemma 2) that (F.1) remains valid upon replacing ε_{t,Y_i} in $\widetilde{\varphi}^*_{k/n}(x,\infty)$ with the $\widehat{\varepsilon}_{t,Y_i}$ (cf. the proof of Lemma 6). □

LEMMA 9. *Under the conditions of Theorem 2, it holds that, for any* $\varepsilon\in(0,1)$,

$$\sup_{x,y\in[\varepsilon,\varepsilon^{-1}]}\left|\widehat{\varphi}^*_{k/n}(x,y)-\widetilde{\varphi}^*_{k/n}(x,y)\right|=o_{\mathrm{P}}(1).$$

Proof. In analogy to $r_n^{\pm}(x)$ from (D.7), we define $r^{\pm}_{n,i}(x)=\frac{n}{k}\Big[1-F_i\big\{(1\pm\delta_n)U_i(n/[kx])\big\}\Big]$. Set $\delta_n=n^{\iota-\xi}$ for sufficiently small $\iota>0$ to ensure that $\sqrt{k}\delta_n=o(1)$. Then, (a straightforward analog of) Proposition 2 allows us to deduce that w.p.a. 1, as $n\to\infty$,

$$\begin{aligned}\widehat{\varphi}^*_{k/n}(x,y)&\le\frac{1}{k}\sum_{t=1}^{n}I_{\left\{\varepsilon_{t,Y_i}>(1-\delta_n)U_i(n/[kx]),\ \varepsilon_{t,Y_j}>(1-\delta_n)U_j(n/[ky])\right\}}\\&=\frac{1}{k}\sum_{t=1}^{n}I_{\left\{\varepsilon_{t,Y_i}>U_i(n/[kr^-_{n,i}(x)]),\ \varepsilon_{t,Y_j}>(1-\delta_n)U_j(n/[kr^-_{n,j}(y)])\right\}}\\&=\widetilde{\varphi}^*_{k/n}\big(r^-_{n,i}(x),r^-_{n,j}(y)\big).\end{aligned}$$

Similarly, $\widetilde{\varphi}^*_{k/n}\big(r^+_{n,i}(x),r^+_{n,j}(y)\big)\le\widehat{\varphi}^*_{k/n}(x,y)$. Thus, the event

$$\begin{aligned}\mathcal{W}_3:=\Big\{\widetilde{\varphi}^*_{k/n}\big(r^+_{n,i}(x),r^+_{n,j}(y)\big)-\widetilde{\varphi}^*_{k/n}(x,y)&\le\widehat{\varphi}^*_{k/n}(x,y)-\widetilde{\varphi}^*_{k/n}(x,y)\\&\le\widetilde{\varphi}^*_{k/n}\big(r^-_{n,i}(x),r^-_{n,j}(y)\big)-\widetilde{\varphi}^*_{k/n}(x,y)\Big\}\end{aligned}$$

occurs w.p.a. 1, as $n\to\infty$. To prove the lemma, it therefore suffices to show that

$$\sup_{x,y\in[\varepsilon,\varepsilon^{-1}]}\left|\widetilde{\varphi}^*_{k/n}\big(r^{\pm}_{n,i}(x),r^{\pm}_{n,j}(y)\big)-\widetilde{\varphi}^*_{k/n}(x,y)\right|=o_{\mathrm{P}}(1).$$

To task this, write

$$\begin{aligned}&\sup_{x,y\in[\varepsilon,\varepsilon^{-1}]}\left|\widetilde{\varphi}^*_{k/n}\big(r^{\pm}_{n,i}(x),r^{\pm}_{n,j}(y)\big)-\widetilde{\varphi}^*_{k/n}(x,y)\right|\\&\quad\le\sup_{x,y\in[\varepsilon,\varepsilon^{-1}]}\left|\widetilde{\varphi}^*_{k/n}\big(r^{\pm}_{n,i}(x),r^{\pm}_{n,j}(y)\big)-R_{n,i,j}\big(r^{\pm}_{n,i}(x),r^{\pm}_{n,j}(y)\big)\right|\\&\qquad+\sup_{x,y\in[\varepsilon,\varepsilon^{-1}]}\left|\widetilde{\varphi}^*_{k/n}(x,y)-R_{n,i,j}(x,y)\right|\end{aligned}$$

$$
\begin{aligned}
&+ \sup_{x,y\in[\varepsilon,\varepsilon^{-1}]} \left| R_{n,i,j}\big(r^{\pm}_{n,i}(x), r^{\pm}_{n,j}(y)\big) - R_{i,j}\big(r^{\pm}_{n,i}(x), r^{\pm}_{n,j}(y)\big) \right| \\
&+ \sup_{x,y\in[\varepsilon,\varepsilon^{-1}]} \left| R_{n,i,j}(x,y) - R_{i,j}(x,y) \right| \\
&+ \sup_{x,y\in[\varepsilon,\varepsilon^{-1}]} \left| R_{i,j}\big(r^{\pm}_{n,i}(x), r^{\pm}_{n,j}(y)\big) - R_{i,j}(x,y) \right| \\
=: {}& E_1 + E_2 + E_3 + E_4 + E_5.
\end{aligned}
$$

The fact that $E_1 = o_{\mathrm{P}}(1)$ and $E_2 = o_{\mathrm{P}}(1)$ follows from Lemma 7 together with $\sup_{x\in[\varepsilon,\varepsilon^{-1}]} |r^{\pm}_{n,i}(x) - x| = o(1)$ for all $i = 1,\ldots,D$ from (a straightforward generalization of) Lemma 4. That $E_3 = o(1)$ and $E_4 = o(1)$ follows from Assumption 4*, where the convergence in (12) is uniform by Theorem 1(v) of Schmidt and Stadtmüller (2006). Finally, $E_5 = o(1)$ follows from $\sup_{x\in[\varepsilon,\varepsilon^{-1}]} |r^{\pm}_{n,i}(x) - x| = o(1)$ and the (Lipschitz) continuity of $R_{i,j}(\cdot,\cdot)$ by Theorem 1(iii) of Schmidt and Stadtmüller (2006). $\square$

LEMMA 10. *Under the conditions of Theorem 2, it holds that, as $n \to \infty$,*

$$
\widehat{\varphi}_{k/n}(k_i/k, k_j/k) - \widetilde{\varphi}_{k/n}(k_i/k, k_j/k) \xrightarrow{\mathrm{P}} 0.
$$

Proof. Write

$$
\begin{aligned}
\widehat{\varphi}_{k/n}(k_i/k, k_j/k) - \widetilde{\varphi}_{k/n}(k_i/k, k_j/k) &= \widehat{\varphi}^{*}_{k/n}(\widehat{e}_{n,i}, \widehat{e}_{n,j}) - \widetilde{\varphi}^{*}_{k/n}(e_{n,i}, e_{n,j}) \\
&= \big[\widehat{\varphi}^{*}_{k/n}(\widehat{e}_{n,i}, \widehat{e}_{n,j}) - \widetilde{\varphi}^{*}_{k/n}(\widehat{e}_{n,i}, \widehat{e}_{n,j})\big] \\
&\quad + \big[\widetilde{\varphi}^{*}_{k/n}(\widehat{e}_{n,i}, \widehat{e}_{n,j}) - R_{n,i,j}(\widehat{e}_{n,i}, \widehat{e}_{n,j})\big] \\
&\quad + \big[R_{n,i,j}(\widehat{e}_{n,i}, \widehat{e}_{n,j}) - R_{i,j}(\widehat{e}_{n,i}, \widehat{e}_{n,j})\big] \\
&\quad + \big[R_{i,j}(\widehat{e}_{n,i}, \widehat{e}_{n,j}) - R_{i,j}(e_{n,i}, e_{n,j})\big] \\
&\quad + \big[R_{i,j}(e_{n,i}, e_{n,j}) - R_{n,i,j}(e_{n,i}, e_{n,j})\big] \\
&\quad + \big[R_{n,i,j}(e_{n,i}, e_{n,j}) - \widetilde{\varphi}^{*}_{k/n}(e_{n,i}, e_{n,j})\big] \\
&=: F_1 + F_2 + F_3 + F_4 + F_5 + F_6.
\end{aligned}
$$

By Lemmas 8 and 9, $F_1 = o_{\mathrm{P}}(1)$ follows. By Lemmas 7 and 8, $F_2 = o_{\mathrm{P}}(1)$ and $F_6 = o_{\mathrm{P}}(1)$. That $F_3 = o_{\mathrm{P}}(1)$ and $F_5 = o_{\mathrm{P}}(1)$ follows from the fact that the convergence in (12) is uniform (by Theorem 1(v) of Schmidt and Stadtmüller, 2006) together with Lemma 8. Finally, $F_4 = o_{\mathrm{P}}(1)$ follows from Lemma 8 and the continuity of $R_{i,j}(\cdot,\cdot)$ (Schmidt and Stadtmüller, 2006, Thm. 1(iii)). $\square$

Proof of Proposition 1. Write

$$
\begin{aligned}
\widehat{R}_{i,j}(q_i, q_j) - R_{i,j}(q_i, q_j) &= \widehat{\varphi}_{k/n}(k_i/k, k_j/k) - R_{i,j}(q_i, q_j) \\
&= \big[\widehat{\varphi}_{k/n}(k_i/k, k_j/k) - \widetilde{\varphi}_{k/n}(k_i/k, k_j/k)\big] \\
&\quad + \big[\widetilde{\varphi}_{k/n}(k_i/k, k_j/k) - R_{n,i,j}(k_i/k, k_j/k)\big] \\
&\quad + \big[R_{n,i,j}(k_i/k, k_j/k) - R_{i,j}(k_i/k, k_j/k)\big] \\
&\quad + \big[R_{i,j}(k_i/k, k_j/k) - R_{i,j}(q_i, q_j)\big] \\
&=: G_1 + G_2 + G_3 + G_4.
\end{aligned}
$$

We show that each term is asymptotically negligible. By Lemma 10, $G_1 = o_{\mathrm{P}}(1)$.

https://doi.org/10.1017/S0266466623000233 Published online by Cambridge University Press

To show that $G_2 = o_{\mathrm{P}}(1)$, we write

$$\begin{aligned}
G_2 &= \widetilde{\varphi}^*_{k/n}(e_{n,i}, e_{n,j}) - R_{n,i,j}(k_i/k, k_j/k) \\
&= \left[\widetilde{\varphi}^*_{k/n}(e_{n,i}, e_{n,j}) - R_{n,i,j}(e_{n,i}, e_{n,j})\right] \\
&\qquad + \left[R_{n,i,j}(e_{n,i}, e_{n,j}) - R_{i,j}(e_{n,i}, e_{n,j})\right] \\
&\qquad + \left[R_{i,j}(e_{n,i}, e_{n,j}) - R_{i,j}(k_i/k, k_j/k)\right] \\
&\qquad + \left[R_{i,j}(k_i/k, k_j/k) - R_{n,i,j}(k_i/k, k_j/k)\right] \\
&=: G_{21} + G_{22} + G_{23} + G_{24}.
\end{aligned}$$

Together, Lemmas 7 and 8 imply that $G_{21} = o_{\mathrm{P}}(1)$. That the remaining terms are also $o_{\mathrm{P}}(1)$ can be established similarly as for the proof of $F_3 = o_{\mathrm{P}}(1)$, $F_4 = o_{\mathrm{P}}(1)$, and $F_5 = o_{\mathrm{P}}(1)$ in the proof of Lemma 10. Hence, $G_2 = o_{\mathrm{P}}(1)$.

From Assumption 4* and the uniform convergence in (12) implied by Schmidt and Stadtmüller (2006, Thm. 1(v)), $G_3 = o(1)$.

That $G_4 = o(1)$ follows from the continuity of $R_{i,j}(\cdot,\cdot)$ (Schmidt and Stadtmüller, 2006, Thm. 1(iii)) and $k_i/k \to q_i$, $k_j/k \to q_j$, as $n \to \infty$.

Overall, the conclusion follows. □

DATA AVAILABILITY STATEMENT

R codes and replication files are available for download from the webpage at https://doi.org/10.5281/zenodo.8081876.

REFERENCES

Acharya, V.V., L.H. Pedersen, T. Philippon, & M. Richardson (2017) Measuring systemic risk. *Review of Financial Studies* 30(1), 2–47.

Adler, R.J. (1990) *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Institute of Mathematical Statistics.

Adrian, T. & M.K. Brunnermeier (2016) CoVaR. *American Economic Review* 106(7), 1705–1741.

Allen, L., T.G. Bali, & Y. Tang (2012) Does systemic risk in the financial sector predict future economic downturns? *Review of Financial Studies* 25(10), 3000–3036.

Bali, T.G. (2007) A generalized extreme value approach to financial risk management. *Journal of Money, Credit and Banking* 39(7), 1613–1649.

Bao, Y., T.-H. Lee, & B. Saltoğlu (2006) Evaluating predictive performance of value-at-risk models in emerging markets: A reality check. *Journal of Forecasting* 25, 101–128.

Basel Committee on Banking Supervision (2019) Basel Framework. Bank for International Settlements, Basel. http://www.bis.org/basel_framework/index.htm?export=pdf.

Beutner, E., A. Heinemann, & S. Smeekes (2021) A justification of conditional confidence intervals. *Electronic Journal of Statistics* 15(1), 2517–2565.

Billio, M., M. Getmansky, A.W. Lo, & L. Pelizzon (2012) Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104(3), 535–559.

Bollerslev, T. (1987) A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.

Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics* 74(3), 498–505.

Breymann, W., A. Dias, & P. Embrechts (2003) Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance* 3, 1–14.

Brownlees, C. & R.F. Engle (2017) SRISK: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies* 30(1), 48–79.

Cai, J.-J., J.H.J. Einmahl, L. de Haan, & C. Zhou (2015) Estimation of the marginal expected shortfall: The mean when a related variable is extreme. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77(2), 417–442.

Cai, J.-J. & E. Musta (2020) Estimation of the marginal expected shortfall under asymptotic independence. *Scandinavian Journal of Statistics* 47(1), 56–83.

Chan, N.H., S.-J. Deng, L. Peng, & Z. Xia (2007) Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics* 137, 556–576.

Chen, C., G. Iyengar, & C.C. Moallemi (2013) An axiomatic approach to systemic risk. *Management Science* 59(6), 1373–1388.

Christoffersen, P. & S. Gonçalves (2005) Estimation risk in financial risk management. *Journal of Risk* 7, 1–28.

Conrad, C. & M. Karanasos (2010) Negative volatility spillovers in the unrestricted ECCC–GARCH model. *Econometric Theory* 26(3), 838–862.

de Haan, L. & A. Ferreira (2006) *Extreme Value Theory*. Springer.

Di Bernardino, E. & C. Prieur (2018) Estimation of the multivariate conditional tail expectation for extreme risk levels: Illustration on environmental data sets. *Environmetrics* 29(7), 1–22.

Drees, H. (2008) Some aspects of extreme value statistics under serial dependence. *Extremes* 11, 35–53.

Drees, H., A. Janßen, S.I. Resnick, & T. Wang (2020) On a minimum distance procedure for threshold selection in tail analysis. *SIAM Journal on Mathematics of Data Science* 2(1), 75–102.

Einmahl, J.H.J., L. de Haan, & D. Li (2006) Weighted approximations of tail copula processes with application to testing the bivariate extreme value condition. *Annals of Statistics* 34(4), 1987–2014.

Engle, R.F. (2002) Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20(3), 339–350.

Fougères, A.-L., L. De Haan, & C. Mercadier (2015) Bias correction in multivariate extremes. *The Annals of Statistics* 43(2), 903–934.

Francq, C., M.D. Jiménez-Gamero, & S.G. Meintanis (2017) Test for conditional ellipticity in multivariate GARCH models. *Journal of Econometrics* 196, 305–319.

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

Francq, C. & J.-M. Zakoïan (2015) Risk-parameter estimation in volatility models. *Journal of Econometrics* 184, 158–173.

Francq, C. & J.-M. Zakoïan (2016) Estimating multivariate GARCH models equation by equation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78(3), 613–635.

FSB (2021) 2021 list of global systemically important banks (G-SIBs). Technical report. https://www.fsb.org/wp-content/uploads/P231121.pdf (accessed May 2022).

Gao, F. & F. Song (2008) Estimation risk in GARCH VaR and ES estimates. *Econometric Theory* 24, 1404–1424.

Giglio, S., B. Kelly, & S. Pruitt (2016) Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics* 119(3), 457–471.

Girardi, G. & A. Tolga Ergün (2013) Systemic risk measurement: Multivariate GARCH estimation of CoVaR. *Journal of Banking & Finance* 37(8), 3169–3180.

Gupta, A. & B. Liang (2005) Do hedge funds have enough capital? A value-at-risk approach. *Journal of Financial Economics* 77, 219–253.

Hafner, C.M., H. Herwartz, & S. Maxand (2022) Identification of structural multivariate GARCH models. *Journal of Econometrics* 227, 212–227.

Hall, P. & Q. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.

He, C. & T. Teräsvirta (2004) An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory* 20, 904–926.

Heffernan, J.E. (2000) A directory of coefficients of tail dependence. *Extremes* 3(3), 279–290.

Hill, B. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3, 1163–1174.

Hoga, Y. (2017) Change point tests for the tail index of β-mixing random variables. *Econometric Theory* 33, 915–954.

Hoga, Y. (2018) Detecting tail risk differences in multivariate time series. *Journal of Time Series Analysis* 39, 665–689.

Hoga, Y. (2019) Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics* 37, 613–624.

Hoga, Y. (2022) Limit theory for forecasts of extreme distortion risk measures and expectiles. *Journal of Financial Econometrics* 20, 18–44.

Hua, L. & H. Joe (2011) Second order regular variation and conditional tail expectation of multiple risks. *Insurance: Mathematics and Economics* 49, 537–546.

IMF/BIS/FSB (2009) Guidance to assess the systemic importance of financial institutions, markets and instruments: Initial considerations. Technical report, IMF. https://www.imf.org/external/np/g20/pdf/100109.pdf (accessed May 2022).

Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.

Kuester, K., S. Mittnik, & M.S. Paolella (2006) Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* 4(1), 53–89.

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

Li, S., L. Peng, & X. Song (2023) Simultaneous confidence bands for conditional value-at-risk and expected shortfall. *Econometric Theory*, 1–35. https://doi.org/10.1017/S0266466622000275.

Martins-Filho, C., F. Yao, & M. Torero (2018) Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory* 34(1), 23–67.

McNeil, A.J. & R. Frey (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7, 271–300.

Nakatani, T. & T. Teräsvirta (2009) Testing for volatility interactions in the constant conditional correlation GARCH model. *The Econometrics Journal* 12, 147–163.

Qin, X. & C. Zhou (2021) Systemic risk allocation using the asymptotic marginal expected shortfall. *Journal of Banking & Finance* 126, 1–16.

Schmidt, R. & U. Stadtmüller (2006) Nonparametric estimation of tail dependence. *Scandinavian Journal of Statistics* 33(2), 307–335.

Shao, Q.-M. (1993) Almost sure invariance principles for mixing sequences of random variables. *Stochastic Processes and their Applications* 48(2), 319–334.

Vervaat, W. (1972) Functional central limit theorems for processes with positive drift and their inverses. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 23, 245–253.

Weissman, I. (1978) Estimation of parameters and large quantiles based on the *k* largest observations. *Journal of the American Statistical Association* 73, 812–815.