Consistent Backtesting Systemic Risk Measures

Soon Heng Leong*

October, 2025

Abstract

This paper offers two novel backtests to evaluate the adequacy of well-known systemic risk measures such as CoVaR, MES, SES, and SRISK. Both the new backtests are robust to estimation risk, that is, their null distributions remain invariant in the presence of estimation risk. While existing backtest is consistent against divergence from the null hypothesis up to a finite order, the paper shows that the new backtests are fully consistent. The real-world implications brought by the new backtests are economically significant as they reveal significantly more cases of inadequate systemic risk modeling among the major financial institutions.

*Soon Leong, sleong@escp.eu, ESCP Business School, Department of Finance. This paper benefited from comments and feedback provided by the editor George Pennacchi and two anonymous reviewers.

I. Introduction

In his recent keynote speech as the chairperson of the European Central Bank's supervisory board, Andrea Enria concluded that the recent pandemic and geopolitical events have turned into macro-level shocks that could alarmingly put the stability of financial system at risk (Enria, 2022). It is therefore crucial that financial supervisors and regulators have access to reliable measures describing and anticipating risk with systemic effects, i.e., the systemic risk measures. To examine the reliability and accuracy of these measures calls for proper validation tools. In today's fast-paced world, the financial system and its participants are responding and acting according to the latest news and events. Thus, methods for testing the reliability of systemic risk measures must adapt accordingly, by placing more attention to recent observations.

Existing literature on backtesting systemic risk is scarce, with the pioneering work contributed by Girardi and Ergün (2013), who propose to backtest the conditional value-at-risk (CoVaR) measure of Adrian and Brunnermeier (2011, 2016) using likelihood ratio method in the vein of Kupiec (1995) and Christoffersen (1998). A major contribution to backtesting systemic risk measures is given by Banulescu-Radu, Hurlin, Leymarie, and Scaillet (2021), hereafter also referred to as BHLS, who provide methods to backtest multiple systemic risk measures considered as important and prominent by the literature, i.e., the delta CoVaR (ΔCoVaR) of Adrian and Brunnermeier (2016), the conditional capital shortfall measure of systemic risk (SRISK) of Acharya, Engle, and Richardson (2012) and Brownlees and Engle (2017), the marginal expected shortfall (MES) and the systemic expected shortfall (SES) of Acharya, Pedersen, Philippon, and Richardson (2017). To this objective, BHLS propose non-likelihood-based statistic such as the t-statistic (to backtest e.g., CoVaR) and the Box-Pierce statistic (to backtest e.g., MES), while taking into consideration the impact of estimation risk.

In this paper, we provide two classes of backtests for checking the adequacy of prominent systemic risk measures. We begin with backtesting the CoVaR and MES measures. We exploit as with existing literature the martingale difference sequence (mds) property of risk indicators derived under the definitions of CoVaR and MES, with a complementary observation. In particular, we show, without imposing additional conditions to existing works, how we can obtain a general procedure, i.e., one that can be used to detect inadequate systemic risk estimated using either CoVaR or MES. This puts our backtesting philosophy in line with the "better-safe-than-sorry" philosophy recently advocated by major regulatory body such as the European Central Bank (see, e.g., Enria, 2022; Rehn, 2020). This follows because in practice, the CoVaR and MES of a financial institution are usually forecasted by the same risk model, and by backtesting CoVaR and MES jointly, we require the risk model to adequately produce both estimates. Our philosophy imposes a more stringent requirement on institutional risk modeling which in turn promotes a safer empirical practice.

By construction, the paper's statistical hypotheses take into account simultaneously both the unconditional and independence conditions commonly examined by the literature. We formulate the first backtest by analogy with the kernel generalization of Box and Pierce (1970) test in Hong (1996), and the additional technicality here lies with treating the impact of estimation risk which stems from forecasting out-sample systemic risk using in-sample parameter estimates. Because our first backtest tackles estimation risk using its standardizing factors, we refer to it as the "standardized" backtest. As the effect of estimation risk can also be asymptotically treated by whitening the test statistic, we obtain the second class of our backtest — the "whitened" backtest. Under regularity conditions, we show both the proposed backtests converge in distribution to $\mathrm{N}(0,1)$ under the null hypothesis.

We also explicitly study and establish the consistency of the new backtests. While the Box-Pierce-type statistic developed in BHLS focuses on divergence from the null hypothesis up to the first m lag order, we show that our backtests are fully consistent to capture a wide range of empirical alternatives. By extension, our backtests also generalize the methodology in Girardi and Ergün (2013) as it focuses on the first-order effect. Intuitively, such generalization is achieved by making m the testing scope of our test statistics and then letting $m \to \infty$. This also appealingly makes the asymptotic distribution of our backtests robust to any finite sample adjustments in m such as those discussed in Breusch and Pagan (1980).

Additionally, we show how the new backtests can also be applied to examine the adequacy of other prominent systemic risk measures. In particular, we show how the new backtests also simultaneously examine the adequacy of the SES and SRISK measures. While existing backtest for SES and SRISK are based on simulated test statistics, the new backtests need no modification. This highlights the generality of the new backtests as they can be conveniently used to detect inadequate systemic risk estimated using either CoVaR, MES, SES, or SRISK. We also discuss how our backtests can be slightly tweaked from focusing on the distress system condition to examining other system level of interest such as the median, which allows backtesting ΔCoVaR since it also evaluates CoVaR at the median system condition.

The Box-Pierce-type statistic, upon allowing its testing scope to grow, can be viewed as a special case of our backtests under the truncated uniform kernel. Both procedures assign equal weight to each lag order. We offer downward weighting kernels to better match the new kernel-based backtests with financial market stylized fact, since market participants are usually more influenced by recent news than by remote past events. Simulation evidence highlights the optimality of our backtests. First, the new backtests demonstrate detection power against alternatives that are of dynamic (i.e., time-dependent) nature and of time-invariant (i.e., unconditional) nature. A growing m improves power against time-invariant alternatives but the opposite is observed against time-dependent alternatives. In this aspect, our downward weighting backtests stand out over existing equally weighted backtest as they give better tradeoff between the two types of alternatives. Simulation evidence also highlights the complementary finite-sample capacities of the new backtests. Our standardized backtest, which preserves data dynamics more effectively, tends to yield better power against alternatives driven by dynamic misspecification. Under the time-invariant alternatives whitened backtest gives slightly better power than its standardized counterpart.

In addition to using simulated data, we also study the performance of the new backtests using real-world data. We focus on the largest financial institutions as their failures (due to e.g., inadequate risk evaluation) are often detrimental to the greater financial system. More precisely, as in Brownlees and Engle (2017) and Banulescu-Radu et al. (2021) we study US institutions with market capitalization greater than 5 billion dollars as of the end of June 2007. We collect data from early 2000 until late 2022 to study also the more recent pandemic-induced financial market crashes. Our findings confirm the empirical improvement brought by the new backtests. For instance, compared to existing backtest the new backtests demonstrate better empirical power as they reveal thousands more cases of inadequate systemic risk modeling over the sample period. Such improvement is also of considerable economic significance as it implies that the new backtests would have detected a larger scale of inadequate systemic risk modeling among the major financial institutions. These additional cases could play an important role in calling for a system-wide revision of existing risk modeling, which in turn would improve the stability of the greater financial system. Our empirical evidence also confirms the complementary features between the new backtests. Thus, we propose, in accordance with the "better-safe-than-sorry" philosophy, that the new backtests are applied together to detect as many indications about large-scale systemic risk inadequacies as possible.

In the spirit of Gourieroux and Zakoïan (2013) and Lazar and Zhang (2019), BHLS propose a system indicator by adjusting the forecasting errors in a forecasted systemic risk series under an objective function. In particular, the objective function in BHLS aims to adjust each forecast with the same magnitude. This may not be optimal because the forecasting errors across a forecasted series are generally nonuniform as, for instance, the next-period forecast is expected to be more accurate with a smaller margin of error than the forecast multiple periods ahead. By reasonings similar to the construction of the new backtests, we improve upon existing objective function such that it aims to adjust with smaller magnitude the immediate forecasts and appropriately with greater magnitude the forecasts multiple steps ahead. This is expected to give an indicator that reflects more accurately the state of the financial system. Indeed, compared with existing indicator, the new indicator shows warning signals that are more noticeable and alarming prior to both the 2008 subprime-induced and the 2020 pandemic-induced stock market crashes.

The remainder of this paper is organized as follows. Section II spells out the framework and statistical theory of the new backtests. Section III provides Monte Carlo evidence highlighting the improvement brought by the new backtests over existing backtest. Section IV confirms the improvement using empirical data. Section V presents an early warning indicator for the financial system. Section VI concludes.

II. Method and statistical theory

A. Preliminaries and definitions

We begin with some preliminaries, and we consider the following notations. Given two entities $\boldsymbol{Y} = (Y_1, Y_2)'$, we let $\boldsymbol{Y}_t \equiv (Y_{1,t}, Y_{2,t})'$ collect their returns at time t. In the context of backtesting systemic risk, $Y_{1,t}$ is the stock return of a financial institution, whereas $Y_{2,t}$ is the system return which is usually proxied by the return of a market index. We let Ω_{t-1} collect the information available at time $t-1$, and we denote by $F_Y(\cdot|\Omega_{t-1})$ the joint distribution function of $\boldsymbol{Y}_t$ given Ω_{t-1}, such that we have for any $\boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^2$, $F_Y(\boldsymbol{y}|\Omega_{t-1}) = \mathbb{P}(Y_{1,t} \leq y_1, Y_{2,t} \leq y_2|\Omega_{t-1})$. Similarly, we let $F_{Y2}(\cdot|\Omega_{t-1})$ denote the marginal distribution of $Y_{2,t}$ given Ω_{t-1}, such that we have for any $y_2 \in \mathbb{R}$, $F_{Y2}(y_2|\Omega_{t-1}) = \mathbb{P}(Y_{2,t} \leq y_2|\Omega_{t-1})$. For notational simplicity, we drop "almost surely" in all equalities involving random variables.

We now introduce the definitions of systemic risk measures MES and CoVaR. Formally, for a given $\alpha \in [0, 1]$, the MES of an institution at time t is defined as

$$MES_{Y1,t}(\alpha) = \mathbb{E}(Y_{1,t}|Y_{2,t} \leq VaR_{Y2,t}(\alpha); \Omega_{t-1}), \tag{1}$$

where $VaR_{Y2,t}(\alpha)$ verifies $VaR_{Y2,t}(\alpha) = F_{Y2}^{-1}(\alpha|\Omega_{t-1})$. Because α is typically specified at a small value such as 5%, the event $Y_{2,t} \leq VaR_{Y2,t}(\alpha)$ captures the left tail distressed system condition. Thus, $MES_{Y1,t}(\alpha)$ quantifies the expected performance of a financial institution given that the system is in distress. The CoVaR definition shares the similarity that it also emphasizes the distress system condition. For a given $\beta \in [0, 1]$, the CoVaR of an institution at time t is the quantity $CoVaR_{Y1,t}(\beta, \alpha)$ verifying the following equation

$$\mathbb{P}(Y_{1,t} \leq CoVaR_{Y1,t}(\beta, \alpha)|Y_{2,t} \leq VaR_{Y2,t}(\alpha); \Omega_{t-1}) = \beta. \tag{2}$$

The CoVaR definition (2) differs slightly from that in Adrian and Brunnermeier (2016), where the main difference is the use of an inequality (instead of an equality) for the conditioning event, that is

$Y_{2,t} \leq VaR_{Y2,t}(\alpha)$. Our definition allows for losses beyond $VaR_{Y2,t}(\alpha)$ and is in line with the backtesting literature (see, e.g., Girardi and Ergün, 2013; Banulescu-Radu et al., 2021). Besides, we note $CoVaR_{Y1,t}(\beta, \alpha)$ verifies $CoVaR_{Y1,t}(\beta, \alpha) = F^{-1}_{Y_1|Y_2 \leq VaR_{Y2}(\alpha)}(\beta|\Omega_{t-1})$, where $F_{Y_1|Y_2 \leq VaR_{Y2}(\alpha)}(\cdot|\Omega_{t-1})$ is the conditional distribution of $Y_{1,t}$ given $Y_{2,t} \leq VaR_{Y2,t}(\alpha)$ and Ω_{t-1}. In other words, $CoVaR_{Y1,t}(\beta, \alpha)$ measures an institution's β-quantile return given that the system is in distress.

In practice, systemic risk measures are typically estimated and forecasted from a parametric model specified by academics, risk managers, or regulators. For instance, Brownlees and Engle (2017) and Banulescu-Radu et al. (2021) consider the dynamic conditional correlation (DCC) specification to model and forecast systemic risk. Therefore we suppose, along with other risk measures of interest, $VaR_{Y2,t}(\alpha) \equiv VaR_{Y2,t}(\alpha, \boldsymbol{\theta}_0) = F^{-1}_{Y2}(\alpha|\Omega_{t-1}, \boldsymbol{\theta}_0)$ and $CoVaR_{Y1,t}(\beta, \alpha) \equiv CoVaR_{Y1,t}(\beta, \alpha, \boldsymbol{\theta}_0) = F^{-1}_{Y_1|Y_2 \leq VaR_{Y2}(\alpha)}(\beta|\Omega_{t-1}, \boldsymbol{\theta}_0)$, are computed from a parametric model with true but unknown p-dimensional parameter vector $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0 \subseteq \boldsymbol{\Theta} \subset \mathbb{R}^p$.

To backtest systemic risk measures, BHLS define a violation risk indicator based on the ex-post risk sequence $\{CoVaR_{Y1,t}(\beta, \alpha, \boldsymbol{\theta}_0), VaR_{Y2,t}(\alpha, \boldsymbol{\theta}_0)\}$ evaluated at a specified (α, β) level, namely

$$h_t(\alpha, \beta, \boldsymbol{\theta}_0) \equiv \mathbb{1}[(Y_{1,t} \leq CoVaR_{Y1,t}(\beta, \alpha, \boldsymbol{\theta}_0)) \cap (Y_{2,t} \leq VaR_{Y2,t}(\alpha, \boldsymbol{\theta}_0))].$$

Next, by identifying the MES-CoVaR relation, that is, $MES_{Y1,t}(\alpha) = MES_{Y1,t}(\alpha, \boldsymbol{\theta}_0) = \int_0^1 CoVaR_{Y1,t}(\beta, \alpha, \boldsymbol{\theta}_0)\mathrm{d}\beta$, we obtain as with BHLS the risk indicator for MES

$$
\begin{aligned}
H_t(\alpha, \boldsymbol{\theta}_0) &\equiv \int_0^1 h_t(\alpha, \beta, \boldsymbol{\theta}_0) \mathrm{d}\beta \\
&= \mathbb{1}[F_{Y_2}(Y_{2,t}|\Omega_{t-1}, \boldsymbol{\theta}_0) \leq \alpha] \times [1 - F_{Y_1|Y_2 \leq VaR_{Y2}(\alpha, \boldsymbol{\theta}_0)}(Y_{1,t}|\Omega_{t-1}, \boldsymbol{\theta}_0)],
\end{aligned}
$$

where the second equality follows by probability integral transform. The idea of backtesting risk measures using their "respective" risk indicators is standard in the literature. For instance, $h_t(\alpha, \beta, \boldsymbol{\theta}_0)$ and $H_t(\alpha, \boldsymbol{\theta}_0)$ are analogous to risk indicators used in backtesting, respectively, the value-at-risk measure (see, e.g., Berkowitz, Christoffersen, and Pelletier, 2011) and the expected shortfall measure (see, e.g., Du and Escanciano, 2017). By the same reasoning, BHLS propose to use $h_t(\alpha, \beta, \boldsymbol{\theta}_0)$ to backtest CoVaR at a specified (α, β) level, and to use $H_t(\alpha, \boldsymbol{\theta}_0)$ to backtest MES at a specified α level.

To backtest risk measures the literature generally relies on exploiting the martingale difference sequence (mds) property of risk indicators. The mds property of $h_t(\alpha, \beta, \boldsymbol{\theta}_0)$ and $H_t(\alpha, \boldsymbol{\theta}_0)$ can be noted as follows. First, by Bayes theorem we have $\mathbb{P}(h_t(\alpha, \beta, \boldsymbol{\theta}_0) = 1|\Omega_{t-1}) = \alpha\beta$. It follows from definition (2) that, $h_t(\alpha, \beta, \boldsymbol{\theta}_0)$ follows the Bernoulli$(\alpha\beta)$ distribution, and that the ex-post centered indicator $\{h_t(\alpha, \beta, \boldsymbol{\theta}_0) - \alpha\beta\}$ is a mds at the specified (α, β) level. Secondly, by Fubini's theorem we note that the mds property of $\{h_t(\alpha, \beta, \boldsymbol{\theta}_0) - \alpha\beta\}$ is preserved by integration, and that the mean of $H_t(\alpha, \boldsymbol{\theta}_0)$ is given by $\alpha/2$. Therefore, $\{H_t(\alpha, \boldsymbol{\theta}_0) - \alpha/2\}$ is also a mds at the specified α level, that is

$$
\mathbb{E}(H_t(\alpha, \boldsymbol{\theta}_0) - \alpha/2|\Omega_{t-1}) = 0. \tag{3}
$$

Notice that by definition of $H_t(\alpha, \boldsymbol{\theta}_0)$ and Fubini's theorem we have $\mathbb{E}(H_t(\alpha, \boldsymbol{\theta}_0)|\Omega_{t-1}) = \int_0^1 \mathbb{E}(h_t(\alpha, \beta, \boldsymbol{\theta}_0)|\Omega_{t-1})\mathrm{d}\beta$. It follows that for equation (3) to hold the CoVaR indicator $h_t(\alpha, \beta, \boldsymbol{\theta}_0)$ must be properly modeled not just at a specified β but instead for all $\beta \in [0, 1]$. This is a complementary observation to existing literature. It implies that by examining the MES measure using $H_t(\alpha, \boldsymbol{\theta}_0)$ at a specified α level, we are simultaneously examining the CoVaR measure using $h_t(\alpha, \beta, \boldsymbol{\theta}_0)$, at the same α level but for all $\beta \in [0, 1]$. Thus, we propose to examine MES and CoVaR jointly based on the mds property of $H_t(\alpha, \boldsymbol{\theta}_0)$, which gives a general procedure, i.e., one that can be conveniently used to detect inadequate systemic risk estimated using either CoVaR or MES. Our procedure is in line with the "better-safe-than-sorry" philosophy advocated by major regulatory body such as the European Central Bank.[1] This is because in practice, the CoVaR and MES of a financial institution are usually estimated and forecasted by the same risk model (see, e.g., Banulescu-Radu et al., 2021). By backtesting MES and CoVaR jointly, we require the risk model to adequately produce both estimates. This imposes a more stringent requirement on institutional risk modeling which in turn promotes a safer empirical practice.

B. The main hypotheses

Therefore, to backtest MES and CoVaR we focus on the mds $\{H_t(\alpha, \boldsymbol{\theta}_0) - \alpha/2\}$ to obtain as with the broader backtesting literature the zero mean condition, i.e., $\mathbb{E}(H_t(\alpha, \boldsymbol{\theta}_0)) - \alpha/2 = 0$ and the zero autocovariance condition, i.e., $\mathbb{E}[(H_t(\alpha, \boldsymbol{\theta}_0) - \mathbb{E}(H_t(\alpha, \boldsymbol{\theta}_0))) \times (H_{t-j}(\alpha, \boldsymbol{\theta}_0) - \mathbb{E}(H_t(\alpha, \boldsymbol{\theta}_0)))] = 0$ for all $j \geq 1$, which are sometimes referred to as the unconditional hypothesis and the independence

[1] We note existing literature follows another (less stringent) philosophy which suggests backtesting CoVaR and MES separately. This will allow for empirical cases where the MES estimate of an institution is inadequate, but the institution's CoVaR estimate (at a particular β level) is deemed adequate and can be relied upon, despite both measures are estimated by the same risk model.

hypothesis, respectively (see, e.g., Christoffersen, 1998; Du and Escanciano, 2017; Banulescu-Radu et al., 2021).

To ensure asymptotic full detection power against both hypotheses, we construct a statistical framework that considers both the unconditional and the independence conditions. This is achieved by first applying $\mathbb{E}(H_t(\alpha, \boldsymbol{\theta}_0)) = \alpha/2$, for all t, in the autocovariance $\gamma(j)$ and autocorrelation $\rho(j)$ functions

$$\gamma(j) \equiv \mathbb{E}\left[(H_t(\alpha, \boldsymbol{\theta}_0) - \alpha/2) \times (H_{t-j}(\alpha, \boldsymbol{\theta}_0) - \alpha/2)\right], \quad \rho(j) = \gamma(j)/\gamma(0). \tag{4}$$

Then, based on $\rho(j)$ we specify the main hypotheses of the paper

$$\mathbb{H}_0 : \rho(j) = 0, \text{ for all } j \geq 1, \quad \mathbb{H}_1 : \rho(j) \neq 0 \text{ for any } j \geq 1.$$

By construction, our hypotheses take into account jointly both the zero mean condition and the zero autocovariance condition. A statistical test consistent against divergence from $\mathbb{H}_0$, that is, with asymptotic unit power under $\mathbb{H}_1$, would be consistent against both conditions. This is confirmed using the new backtests in our simulation study.

In practice, we do not observe $H_t(\alpha, \boldsymbol{\theta}_0)$ and $\rho(j)$ but we can estimate them using the observed sample $\{Y_{1,t}, Y_{2,t}\}_{t=1}^{T+n}$. In this aspect, we note that the fundamental notion of backtesting is about examining the quality of a sequence of forecasts (using e.g., $\mathbb{H}_0$). This calls for a forecasting scheme. For meaningful comparison we use as with BHLS the "fixed forecasting estimation scheme". That is, we first use an in-sample return data from $t = 1$ to $t = T$ (i.e., $\{Y_{1,t}, Y_{2,t}\}_{t=1}^{T}$) to estimate $\boldsymbol{\theta}_0$, by e.g., $\hat{\boldsymbol{\theta}}_T$, where $\hat{\boldsymbol{\theta}}_T$ is a consistent estimator for $\boldsymbol{\theta}_0$, with the asymptotic variance of $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$ given by $\boldsymbol{\Sigma}$.

Then, using $\hat{\boldsymbol{\theta}}_T$ and the out-sample return data observed from $t = T + 1$ to $t = T + n$ (i.e., $\{Y_{1,t}, Y_{2,t}\}_{t=T+1}^{T+n}$), we forecast and compute $\{H_t(\alpha, \hat{\boldsymbol{\theta}}_T)\}_{t=T+1}^{t=T+n}$. Therefore, the sample counterpart of $\rho(j)$ is given by

$$\hat{\rho}_n(j) = \hat{\gamma}_n(j)/\hat{\gamma}_n(0), \tag{5}$$

where $\hat{\gamma}_n(j)$ is the sample autocovariance

$$\hat{\gamma}_n(j) = \frac{1}{n-j} \sum_{t=T+1+j}^{T+n} (H_t(\alpha, \hat{\boldsymbol{\theta}}_T) - \alpha/2) \times (H_{t-j}(\alpha, \hat{\boldsymbol{\theta}}_T) - \alpha/2). \tag{6}$$

A backtest can be readily constructed based on the Box-Pierce (Box and Pierce, 1970) test statistic, namely

$$S = n \sum_{j=1}^{m} \hat{\rho}_n^2(j). \tag{7}$$

However, in the presence of estimation risk (i.e., $n/T \to \omega$, and $0 < \omega < \infty$), it has been shown that S does not follow a standard distribution. Besides, as S checks $\mathbb{H}_0$ up to a finite order m, it may not be consistent against some higher order alternatives.

C. Test statistics and asymptotic theory

In this paper we propose for the context of backtesting systemic risk two classes of consistent $\mathrm{N}(0,1)$ backtests. The construction of our first backtest is analogous to the kernel generalization of

Box-Pierce test in Hong (1996), which takes into account all available $\hat{\rho}_n(j)$

$$n \sum_{j=1}^{n-1} k^2(j/m)\hat{\rho}_n^2(j), \tag{8}$$

where $k(\cdot)$ is a kernel function with kernel parameter m which is a positive integer. Some commonly used $k(\cdot)$ are the Bartlett (BAR) kernel, the Daniell (DAN) kernel, the Parzen (PAR) kernel, the Quadratic-Spectral (QS) kernel, and the truncated uniform (TRU) kernel

$$\text{BAR: } k(z) = \begin{cases} 1 - |z|, & |z| \leq 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$\text{DAN: } k(z) = \frac{\sin(\pi z)}{\pi z}, \quad -\infty < z < \infty,$$

$$\text{PAR: } k(z) = \begin{cases} 1 - 6z^2 + 6|z|^3, & |z| \leq 0.5, \\ 2(1 - |z|)^3, & 0.5 < |z| \leq 1, \\ 0, & \text{otherwise}, \end{cases}$$

$$\text{QS: } k(z) = \frac{25}{12\pi^2 z^2}\left(\frac{\sin(6\pi z/5)}{6\pi z/5} - \cos(6\pi z/5)\right), \quad -\infty < z < \infty,$$

$$\text{TRU: } k(z) = \begin{cases} 1, & |z| \leq 1, \\ 0, & \text{otherwise}. \end{cases}$$

The BAR, PAR, and TRU kernels have compact support, and m can be viewed as the truncation parameter for these kernels, i.e., $k(z) = 0$ for $|z| > 1$.[2] All the given kernels satisfy the assumptions on $k(\cdot)$ in the online appendix. Except for the TRU kernel, all kernels discount the weights of higher order

[2]Notably, Newey and West (1987) propose to use the BAR kernel in the context of estimating Heteroskedasticity and Autocorrelation Consistent (HAC) covariance matrix.

lags. Using downward weighting kernel gives better match to the well-known stylized fact that financial market is more influenced by recent news than by remote past events, and is thus expected to give better empirical rejection power. As with S, expression (8) may not follow a standard asymptotic distribution in the presence of estimation risk. We quantify and account for the impact of estimation risk in our test statistics accordingly, and our first backtest is the standardized version of expression (8), namely

$$Q^0 = \frac{n \sum_{j=1}^{n-1} k^2(j/m)\hat{\rho}_n^2(j) - M_{1n}(k)}{V_{1n}(k)^{1/2}},$$

where

$$\begin{aligned}
M_{1n}(k) &= \sum_{j=1}^{n-1} k^2(j/m)[1 + \omega \boldsymbol{R}(j)' \boldsymbol{\Sigma} \boldsymbol{R}(j)], \\
V_{1n}(k) &= 2 \sum_{j=1}^{n-1} k^4(j/m)[1 + \omega \boldsymbol{R}(j)' \boldsymbol{\Sigma} \boldsymbol{R}(j)]^2 + 2 \sum_{j=1}^{n-1} \sum_{\substack{i=1 \\ i \neq j}}^{n-1} k^2(i/m) k^2(j/m) [\omega \boldsymbol{R}(i)' \boldsymbol{\Sigma} \boldsymbol{R}(j)]^2,
\end{aligned}$$

and $\boldsymbol{R}(j)$ is given by

$$\boldsymbol{R}(j) = \frac{1}{\alpha(1/3 - \alpha/4)} \mathbb{E}\left\{(H_{t-j}(\alpha, \boldsymbol{\theta}_0) - \alpha/2) \nabla_{\boldsymbol{\theta}} [H_t(\alpha, \boldsymbol{\theta}_0)]\right\}, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

The quantities $M_{1n}(k)$ and $V_{1n}(k)$ are approximately the mean and variance of expression (8). The additional technicality here with respect to the Hong (1996) test lies with treating the impact of estimation risk, which stems from estimating out-sample risk measures based on in-sample parameter estimates. Intuitively, we quantify estimation risk by first measuring the sensitivity of the out-sample risk indicator with respect to the in-sample parameter in the form of $\boldsymbol{R}(j)$. The sensitivity is then scaled according to the asymptotic variance of the parameter estimates $\boldsymbol{\Sigma}$. Based on the calculated estimation risk we then quantify its impact on the asymptotic distribution of our test statistic, and the impact is subsequently adjusted by $M_{1n}(k)$ and $V_{1n}(k)$. This makes the asymptotic distribution of Q^0 robust to estimation risk. In the absence of estimation risk (i.e., $\omega = 0$), $M_{1n}(k)$ and $V_{1n}(k)$ reduce to $M_{1n}(k) = \sum_{j=1}^{n-1} k^2(j/m)$ and $V_{1n}(k) = 2\sum_{j=1}^{n-1} k^4(j/m)$, respectively.

In practice a feasible test statistic Q requires consistent estimators for $\boldsymbol{R}(j)$ and $\boldsymbol{\Sigma}$, which we denote by $\hat{\boldsymbol{R}}(j)$ and $\hat{\boldsymbol{\Sigma}}$, respectively. Therefore, the feasible test statistic is given by

$$Q = \frac{n\sum_{j=1}^{n-1} k^2(j/m)\hat{\rho}_n^2(j) - \hat{M}_{1n}(k)}{\hat{V}_{1n}(k)^{1/2}},$$

where $(\hat{M}_{1n}(k), \hat{V}_{1n}(k))$ are defined as $(M_{1n}(k), V_{1n}(k))$ with $(\boldsymbol{R}(j), \boldsymbol{\Sigma})$ replaced by $(\hat{\boldsymbol{R}}(j), \hat{\boldsymbol{\Sigma}})$.

We refer to Q as the "standardized" backtest as it tackles estimation risk using its standardizing factors. Because the effect of estimation risk can also be asymptotically addressed by whitening (i.e., decorrelating) the $\{\hat{\rho}_n(j)\}$ sequence, we obtain the second class of our backtest.[3] Put $\{\hat{\rho}_n(j)\}_{j=1}^{n-1}$ in the vector $\hat{\boldsymbol{\rho}}_n$, and let $\hat{\rho}_{nr}(j)$ denote the asymptotically whitened version of $\hat{\rho}_n(j)$. In particular, $\hat{\rho}_{nr}(j)$ can be, but not limited to, the j-th element of the vector $\hat{\boldsymbol{\Delta}}^{-1/2}\hat{\boldsymbol{\rho}}_n$, where $\hat{\boldsymbol{\Delta}}$ is a consistent estimator of matrix $\boldsymbol{\Delta}$ whose (i,j)-th element is given by $\boldsymbol{\Delta}(i,j) = \delta(i,j) + \omega\boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)$, with $\delta(i,j)$ the Kronecker delta which takes the value of 1 if $i = j$ and 0 if otherwise. Similarly, by putting $\{\rho(j)\}_{j=1}^{n-1}$ in $\boldsymbol{\rho}$, we have the j-th element of the vector $\boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho}$, which we denote by $\rho_r(j)$, being the true counterpart of $\hat{\rho}_{nr}(j)$. The robust Box-Pierce-type backtest proposed in BHLS is based on the whitened sequence

$$S_r = n\sum_{j=1}^{m} \hat{\rho}_{nr}^2(j),$$

[3]Notably, the core concept of Generalized Least Squares (GLS) is based on the same whitening technique.

and it has been shown that $S_r \xrightarrow{d} \chi^2(m)$ under regularity conditions.

Our second class of N$(0,1)$ backtest is also based on $\{\hat{\rho}_{nr}(j)\}$, namely

$$Q_r = \frac{n\sum_{j=1}^{n-1} k^2(j/m)\hat{\rho}_{nr}^2(j) - \sum_{j=1}^{n-1} k^2(j/m)}{\left\{2\sum_{j=1}^{n-1} k^4(j/m)\right\}^{1/2}}.$$

The standardizing factors in Q_r are those in Q when there is no estimation risk ($\omega = 0$). Intuitively, estimation risk has been adjusted by whitening $\{\hat{\rho}_n(j)\}$, and thus the standardizing factors of Q_r need no further adjustment. Therefore, we refer to Q_r as the "whitened" backtest. When the TRU kernel is used, we obtain

$$Q_{r\text{TRU}} = \frac{n\sum_{j=1}^{m} \hat{\rho}_{nr}^2(j) - m}{\{2m\}^{1/2}},$$

which can be viewed as the generalized BHLS statistic that allows m to grow. As with Q we propose downward weighting kernel for Q_r to give a closer match to financial market stylized facts. This is expected to improve the detection power of our backtests. Indeed, $Q_{r\text{TRU}}$ and S_r show similar power in our simulation study, but are generally less optimal than the new backtests with downward weighting kernels.

To ensure consistency against a wide range of empirical alternatives, we let the kernel parameter m — which can be viewed as the testing scope parameter of Q and Q_r — grow with the size of the sample under examination at a proper rate, that is, $m \to \infty$ as $n \to \infty$ with $m/n \to 0$. From a theoretical perspective, passing $m \to \infty$ also makes the asymptotic distribution of a backtest robust to any finite sample adjustment in m. Indeed, if $H_t(\alpha, \boldsymbol{\theta}_0)$ exhibits autoregressive behavior, the degree of freedom in the asymptotic χ^2 distribution of Box-Pierce-type test may require adjustment from m to

$m - c$, where $c < \infty$ is the order of the autoregression (see, e.g., Breusch and Pagan, 1980). This becomes asymptotically irrelevant when a backtest allows $m \to \infty$, as with Q and Q_r.

We now give the asymptotic properties of Q and Q_r.

Theorem 1. Suppose Assumptions 1–4 in the online appendix hold, and let $T \to \infty$, $n \to \infty$, $n/T \to \omega < \infty$, $m \to \infty$, $m/n \to 0$. Then $Q \xrightarrow{d} \mathrm{N}(0, 1)$ under $\mathbb{H}_0$.

Proposition 1. Suppose the conditions of Theorem 1 hold. Then $Q_r \xrightarrow{d} \mathrm{N}(0, 1)$ under $\mathbb{H}_0$.

Theorem 2. Suppose Assumptions 1–6 in the online appendix hold, and let $T \to \infty$, $n \to \infty$, $n/T \to \omega < \infty$, $m \to \infty$, $m/n \to 0$. Then

$$\frac{m^{1/2}}{n} Q \xrightarrow{p} \sum_{j=1}^{\infty} \rho^2(j) \bigg/ \left[2 \int_0^{\infty} k^4(z) \mathrm{d}z \right]^{1/2}.$$

Proposition 2. Suppose the conditions of Theorem 2 hold. Then

$$\frac{m^{1/2}}{n} Q_r \xrightarrow{p} \sum_{j=1}^{\infty} \rho_r^2(j) \bigg/ \left[2 \int_0^{\infty} k^4(z) \mathrm{d}z \right]^{1/2}.$$

The detailed mathematical proof for the above results is provided in Online Appendix B–E. In Theorem 1 and Proposition 1 we show under regularity conditions that both our backtests Q and Q_r follow the asymptotic $\mathrm{N}(0, 1)$ distribution under $\mathbb{H}_0$, even in the presence of estimation risk. In Theorem 2 and Proposition 2 we show that whenever $\rho(j) \neq 0$ for any $j \geq 1$, which implies $\rho_r(j) \neq 0$ for some $j \geq 1$, we have $m^{1/2}/nQ$ and $m^{1/2}/nQ_r$ converge to a finite positive quantity. This shows that Q and Q_r diverge to positive infinity at rate $n/m^{1/2}$ to yield asymptotic full power. That is, for some constants $K_n = o(n/m^{1/2})$, we have $\mathbb{P}(Q > K_n) \to 1$ and $\mathbb{P}(Q_r > K_n) \to 1$ under $\mathbb{H}_1$. Thus, the new backtests are consistent, providing a valuable addition to existing literature. Because Q and Q_r go to positive infinity under $\mathbb{H}_1$, they are upper-tailed backtests; and the upper-tailed critical value of $\mathrm{N}(0, 1)$ at the 5% significance level is approximately 1.645. Finally, although the new backtests are asymptotically consistent, in small sample they may not give similar power since Q is a standardized statistic while Q_r is a whitened statistic. This is studied using simulation experiments.

D. Generalized backtests for other measures

In this subsection, we discuss how (Q, Q_r), which backtest CoVaR and MES, can also be applied to backtest other prominent systemic risk measures, that is, the conditional capital shortfall measure of systemic risk (SRISK) of Acharya et al. (2012) and Brownlees and Engle (2017), the systemic expected shortfall (SES) of Acharya et al. (2017), and the ΔCoVaR measure of Adrian and Brunnermeier (2016). First, we show how backtesting SES or SRISK is equivalent to backtesting (CoVaR, MES) using the new backtests. Secondly, we discuss how our backtests can be easily modified to examine systemic risk at other level of system condition such as the median level, which concerns the ΔCoVaR measure.

1. Backtesting SRISK and SES

We begin as usual with the definitions. Following Acharya et al. (2012), the SRISK of institution Y_1 at time t is given by

$$SRISK_{Y1,t}(\alpha) = gL_{1,t} - (1 - g)W_{1,t}[\exp(18 \times MES_{Y1,t}(\alpha))], \tag{9}$$

where $g \in [0, 1]$ is the prudential capital fraction; $L_{1,t}$ is the debt of institution Y_1; and $W_{1,t} > 0$ is the market value of institution Y_1. Besides, in Acharya et al. (2017) the association between SES and MES is given by

$$(10) \qquad SES_{Y1,t}(\alpha) = [gLV_{1,t} - 1 - \Pi MES_{Y1,t}(\alpha) + \Lambda]W_{1,t},$$

where $LV_{1,t} \equiv (L_{1,t} + W_{1,t})/W_{1,t}$ is the leverage ratio of institution Y_1; Λ and $\Pi > 0$ are constant terms. By taking the first derivatives we note that $SRISK_{Y1,t}(\alpha)$ and $SES_{Y1,t}(\alpha)$ are decreasing in $MES_{Y1,t}(\alpha)$ monotonically. More generally, we denote by $M_t(MES_{Y1,t}(\alpha))$ any monotonic functions of $MES_{Y1,t}(\alpha)$, such as $SRISK_{Y1,t}(\alpha)$ and $SES_{Y1,t}(\alpha)$. Recall that to backtest CoVaR and MES we use the $H_t(\alpha)$ indicator, which we can rewrite as

$$H_t(\alpha) = \mathbb{1}\{(Y_{2,t} \leq VaR_{Y2,t}(\alpha))\} \int_0^1 \mathbb{1}\{[Y_{1,t} \leq CoVaR_{Y1,t}(\beta, \alpha)]\} \mathrm{d}\beta,$$

with $\mathbb{1}\{[Y_{1,t} \leq CoVaR_{Y1,t}(\beta, \alpha)]\}$ being the integrand over $\beta \in [0, 1]$. The risk indicator associated with $M_t(MES_{Y1,t}(\alpha))$ is naturally given by

$$\tilde{H}_t(\alpha) \equiv \mathbb{1}\{(Y_{2,t} \leq VaR_{Y2,t}(\alpha))\} \int_0^1 \mathbb{1}\{[M_t(Y_{1,t}) \geq M_t(CoVaR_{Y1,t}(\beta, \alpha))]\} \mathrm{d}\beta,$$

with $\mathbb{1}\{[M_t(Y_{1,t}) \geq M_t(CoVaR_{Y1,t}(\beta, \alpha))]\}$ being now the integrand over $\beta \in [0, 1]$. We could therefore propose to backtest SES or SRISK using $\tilde{H}_t(\alpha)$, but notice that $H_t(\alpha)$ and $\tilde{H}_t(\alpha)$ are actually identical as their integrands take the value of 1 at the same point in time. Computing any of our backtests using $H_t(\alpha)$ or $\tilde{H}_t(\alpha)$ gives the same value and the same conclusion. On this basis, we could conveniently apply the proposed test statistics (Q, Q_r) to backtest SES or SRISK. This gives the new backtests general capacities as they can be used to detect inadequate systemic risk estimated using either CoVaR, MES, SES, or SRISK.

2. Backtesting at other system conditions

We have thus far focused on backtesting at the distress system condition, which is the emphasis of most systemic risk measures such as CoVaR, MES, SES, and SRISK. This is achieved by evaluating $H_t(\alpha)$ at the left tail, e.g., by setting $\alpha = 5\%$. To backtest ΔCoVaR, we first note that it is given by the difference between CoVaR evaluated at the distressed system condition and CoVaR evaluated at the median system condition. It thus remains to show how our backtesting procedure can be modified to examine systemic risk measures (such as CoVaR and SES) at various other system conditions such as the median. This is achieved by generalizing the $H_t(\alpha)$ indicator such that it can consider any level of system condition specified by the range (α_l, α_u), that is

$$
\begin{aligned}
&H_t(\alpha_l, \alpha_u) \\
&\equiv \mathbb{1}\{(VaR_{Y2,t}(\alpha_l) \leq Y_{2,t} \leq VaR_{Y2,t}(\alpha_u))\} \int_0^1 \mathbb{1}\{[Y_{1,t} \leq CoVaR_{Y1,t}(\beta, \alpha_l, \alpha_u)]\} \mathrm{d}\beta,
\end{aligned}
$$

where as with the definition of $CoVaR_{Y1,t}(\beta, \alpha)$ in equation (2), we have $CoVaR_{Y1,t}(\beta, \alpha_l, \alpha_u)$ verifying $\mathbb{P}(Y_{1,t} \leq CoVaR_{Y1,t}(\beta, \alpha_l, \alpha_u) | VaR_{Y2,t}(\alpha_l) \leq Y_{2,t} \leq VaR_{Y2,t}(\alpha_u); \Omega_{t-1}) = \beta$. The mean of $H_t(\alpha_l, \alpha_u)$ is $(\alpha_u - \alpha_l)/2$, and we verify $\{H_t(\alpha_l, \alpha_u) - (\alpha_u - \alpha_l)/2\}$ retains the mds property. Moreover, using $\{H_t(\alpha_l, \alpha_u) - (\alpha_u - \alpha_l)/2\}$ does not affect the asymptotic results of (Q, Q_r). When $\alpha_l = 0$ and $\alpha_u = \alpha$, $H_t(\alpha_l, \alpha_u)$ is equivalent to $H_t(\alpha)$. The $H_t(\alpha_l, \alpha_u)$ indicator is useful when we concern examining systemic risk measures at other levels of system condition. To backtest at the median condition, we can set e.g., $(\alpha_l, \alpha_u) = (45\%, 55\%)$.

III. Monte Carlo evidence

In this section we study the finite sample performance of the proposed backtests using simulation experiments. Because a distressed financial system tends to drive broader economic problems, it is repeatedly emphasized and quantified by the systemic risk literature. This is reflected by the formulas of prominent systemic risk measures such as ΔCoVaR, MES, SES, and SRISK. Therefore, we use throughout $H_t(\alpha_l, \alpha_u)$ with $(\alpha_l, \alpha_u) = (0, 5\%)$, which is equivalent to using $H_t(\alpha)$ with $\alpha = 5\%$. We work with the DCC-GARCH data generating process (DGP) of Engle (2002) and Aielli (2013). More precisely, $\boldsymbol{Y}_t$ is generated according to the following set of equations

$$
\begin{cases}
Y_{i,t} = \alpha_{i0} + \alpha_{i1} Y_{i,t-1} + \eta_{i,t}, \quad i = 1, 2, \quad t = 1, 2, ..., T, \\
\boldsymbol{\eta}_t = (\eta_{1,t}, \eta_{2,t})' = \boldsymbol{\sigma}_t^{1/2} \boldsymbol{z}_t, \quad \boldsymbol{\sigma}_t = \boldsymbol{D}_t^{1/2} \boldsymbol{R}_t \boldsymbol{D}_t^{1/2}, \\
\boldsymbol{D}_t = [(\sigma_{1,t}, 1)', (1, \sigma_{2,t})'], \quad \sigma_{i,t} = \beta_{i0} + \beta_{i1} \eta_{i,t-1}^2 + \beta_{i2} \eta_{i,t-1}^2 \mathbb{1}(\eta_{i,t-1} < 0) + \beta_{i3} \sigma_{i,t-1}, \\
\boldsymbol{R}_t = \text{Diag}(\boldsymbol{Q}_t)^{-1/2} \boldsymbol{Q}_t \text{Diag}(\boldsymbol{Q}_t)^{-1/2}, \quad \boldsymbol{Q}_t \equiv [(q_{11,t}, q_{21,t})', (q_{12,t}, q_{22,t})'], \\
\boldsymbol{Q}_t = (1 - a - b) \boldsymbol{S} + a (\text{Diag}(\boldsymbol{Q}_{t-1})^{1/2} \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' \text{Diag}(\boldsymbol{Q}_{t-1})^{1/2}) + b \boldsymbol{Q}_{t-1}, \\
\boldsymbol{S} = [(1, s)', (s, 1)'], \quad \boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})' = (\eta_{1,t}/\sigma_{1,t}^{1/2}, \eta_{2,t}/\sigma_{2,t}^{1/2})', \\
q_{ii,t} = (1 - a - b) + a \varepsilon_{i,t-1}^2 q_{ii,t-1} + b q_{ii,t-1}.
\end{cases}
\tag{11}
$$

The selected DGP aims to simulate data that reflects empirical stylized facts. We use first-order dynamic as it reflects the tendency of market participants to respond and react to the latest news and events. The data is generated with volatilities $\sigma_{i,t}$ following the GARCH process in Glosten, Jagannathan, and Runkle (1993), to allow for the well-known asymmetric market reactions. We generate separately for each time period t the innovation term $\boldsymbol{z}_t$ from the standardized Student-t distribution with degree of freedom ν, i.e., $\boldsymbol{z}_t \sim \text{t}_\nu(\boldsymbol{0}, \boldsymbol{I}_2)$, where $\boldsymbol{I}_2$ is the identity matrix of order 2. This allows for the heavy tail feature of financial data. Moreover, we calibrate the parameters of DGP (11) using observed data spanning from 3-Jan-2000 to 31-Dec-2021. We use the daily security log-returns (in percentage) of Bank of America as $Y_{1,t}$; we proxy $Y_{2,t}$ using the log-returns of the Center for Research in Security Prices (CRSP) value-weighted market index. Based on the estimated parameters, we set

$$(12) \qquad \begin{cases} (\alpha_{10}, \alpha_{11}, \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}) = (0.0634, 0.0015, 0.0268, 0.0302, 0.0228, 0.9160) \\ (\alpha_{20}, \alpha_{21}, \beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}) = (0.0708, 0.0010, 0.0184, 0.0159, 0.0489, 0.9002) \\ (a, b, s, \nu) = (0.0318, 0.9407, 0.7721, 7) \end{cases}$$

The calibrated parameters reflect financial market stylized facts. For instance, the positive signs in the asymmetry parameters (β_{12}, β_{22}) reflect the stronger market reactions to negative news than the positive counterparts. The relatively large GARCH parameters (β_{13}, β_{23}) with respect to the ARCH parameters (β_{11}, β_{21}) highlight the persistent feature in the volatility of financial series. The positive s parameter captures the expected positive comovement between Bank of America and the financial market. A relatively small ν parameter reflects the heavy tails feature of financial returns. Also, the set of parameters (12) satisfies stationary requirements.

We study the impact of different kernel functions by considering all the kernels discussed previously. To study the capacity of our backtests, we use a range of values for the kernel parameter $m = \{5, 10, 20, 30\}$. We consider several combinations of in-sample size and out-sample size $(T, n) = \{(250, 250), (250, 500), (500, 500), (500, 250)\}$. For each combination we simulate $T + n + 500$ observations, and then we discard the first 500 observations to work with the latest $T + n$ data points. The objective here is to stabilize the generated data by reducing the effects of the initial values in the DGP, which we set as $(Y_{1,0}, Y_{2,0}, \sigma_{1,0}, \sigma_{2,0}) = [\alpha_{10}, \alpha_{20}, \beta_{10}/(1 - \beta_{11} - 0.5\beta_{12} - \beta_{13}),$ $\beta_{20}/(1 - \beta_{21} - 0.5\beta_{22} - \beta_{23})]$.

Let $\boldsymbol{\phi}$ collects the parameters in DGP (11) with true values $\boldsymbol{\phi}_0$. To estimate $\boldsymbol{\phi}$ we use the maximum likelihood method, and therefore we note that the Student-t likelihood function (in logarithm) of DGP (11) for observation at time t is given by

$$
\begin{aligned}
l_t(\boldsymbol{\phi}) = \log\Big[\Gamma\Big(\frac{2\eta+1}{2\eta}\Big)\Big] - \log\Big[\Gamma\Big(\frac{1}{2\eta}\Big)\Big] - \log\Big(\frac{1-2\eta}{\eta}\Big) - \log(\pi) \\
- 1/2\log(|\boldsymbol{\sigma}_t|) - \frac{2\eta+1}{2\eta}\log\Big(1 + \frac{\eta}{1-2\eta}\boldsymbol{\varepsilon}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\varepsilon}_t\Big),
\end{aligned}
\tag{13}
$$

where $\Gamma(\cdot)$ is the gamma function and $\eta = \nu^{-1}$ is the inverse of the degree of freedom parameter, which is constrained at $\eta < 0.5$ to ensure the existence of second moment. Parameter estimation is executed by maximizing the sum of likelihood $\sum_t l_t(\boldsymbol{\phi})$, and we denote the estimated parameter vector by $\hat{\boldsymbol{\phi}}_T$. With $\hat{\boldsymbol{\phi}}_T$ we compute $H_t(\alpha, \hat{\boldsymbol{\phi}}_T)$, and we note that the consistency and asymptotic normality of $\hat{\boldsymbol{\phi}}_T$ have been established for a general class of multivariate GARCH models (see, e.g., Fiorentini, Sentana, and Calzolari, 2003, p. 535). That is, we have

$$
\sqrt{T}(\hat{\boldsymbol{\phi}}_T - \boldsymbol{\phi}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),
$$

where $\boldsymbol{\Sigma}/T$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\phi}}_T$, which we estimate consistently using the Hessian matrix obtained from the estimation procedure. To compute the test statistics it remains to evaluate $\hat{\boldsymbol{R}}(j)$ which we give in Online Appendix F. In this aspect, we face as with BHLS the challenge of differentiability, which we address using the procedures suggested therein to obtain a consistent estimate.

Finally, we compute Q and Q_r to examine $\mathbb{H}_0$. For comparison we also compute the Box-Pierce-type backtests, i.e., S and S_r, where backtestings are performed based on the $\chi^2(m)$ limiting distribution (BHLS, Corollary 2). Although this is generally not accurate for S, we include it to highlight the effect of estimation risk. Throughout, testing significance level is 5%.

We begin with the size study, that is, backtesting systemic risk forecasts issued by the correctly specified risk model (11). We refer to this study as experiment NULL, and we report the rejection rates in Table 1. As expected, the size of the nonwhitened S backtest is severely distorted whereas the whitened S_r backtest shows reasonable rejection rates. This is due to S having nonstandard distribution in the presence of estimation risk, which is accounted for by whitening the test statistic as in S_r. By the same reasoning, our whitened backtest Q_r also displays reasonable sizes across the combinations of (T, n). In contrast to S, our nonwhitened backtest Q is reasonably sized as it has considered estimation risk in its standardizing factors explicitly. While Q over rejects slightly when the sample size is small, it shows stability as soon as (T, n) increases to 500. We also note the usual bootstrap procedure can be readily applied if a better size is desired for small sample. Overall, both the proposed Q_r and Q backtests show robustness to estimation risk. Besides, the size of both backtests are stable under different kernel functions along with a range of values for the kernel parameter m. In other words, the asymptotic N$(0, 1)$ distribution of Q_r and Q established in Theorem 1 and Proposition 1 is well approximated in small sample, even when the generated data exhibits heavier tails than the normal distribution.

[Insert Table 1 approximately here]

We now study the ability of our backtests to detect alternative cases in which we have inadequacy of the systemic risk measures. Although under regularity conditions both Q and Q_r enjoy asymptotic unit detection power against divergence from the null hypothesis of measure adequacy, their finite sample powers are not known and thus will be studied and compared in the following. We continue the comparison with existing method, but we focus on the S_r backtest since comparison with S would be inaccurate as it is not a robust test. We design a series of experiments.

In the first alternative experiment, we simulate the scenario in which a risk manager specifies an adequate model to estimate and forecast systemic risk, but based on a misspecified risk level. In other words, the risk indicator is forecasted at risk level α_1 but the intended risk level is α_2, with $\alpha_2 \neq \alpha_1$. The desired statistical effect is simulated by inducing risk level mismatch between the risk indicator and its mean in the autocovariance computation, that is

$$\hat{\gamma}_n(j) = \frac{1}{n-j} \sum_{t=T+1+j}^{T+n} (H_t(\alpha_1, \hat{\boldsymbol{\phi}}_T) - \alpha_2/2) \times (H_{t-j}(\alpha_1, \hat{\boldsymbol{\phi}}_T) - \alpha_2/2).$$

We refer to this experiment as ALTER1, and we note that it also allows us to examine the power of our backtests against the unconditional hypothesis, that is, the condition $\mathbb{E}(H_t(\alpha, \boldsymbol{\phi}_0)) = \alpha/2$. We set $(\alpha_1, \alpha_2) = (5\%, 10\%)$; simulation results are reported in Table 2.

[Insert Table 2 approximately here]

We observe the nontrivial rejection power by our backtests, and the power generally increases with the sample size. This is in line with Theorem 2 and Proposition 2. We also observe an increase in rejection rates as m increases. This is due to the time-invariant nature of the misspecification in this experiment as the level of the risk indicator is incorrectly specified across each time period. Because the time scope of all backtests increases in m, with that they naturally capture more evidence against alternatives that persists in time. This results in a higher finite sample rejection rates. We observe that when the TRU kernel is used, Q_r yields similar rejection power as S_r. This is expected because the S_r backtest can be viewed as a special case of our Q_r backtest when the TRU kernel is used, as both backtests assign equal weight to each lag order. We also observe the comparable performance across kernels as $m > 20$. Overall, this experiment highlights that the new backtests possess detection power against divergence from the condition $\mathbb{E}(H_t(\alpha, \boldsymbol{\theta}_0)) = \alpha/2$, with Q_r having slightly better power than Q. The results also highlight that by letting m grow, the detection power of a backtest could increase as it may capture more evidence against divergence from the null.

In the second alternative experiment, we simulate the scenario in which the distribution of the systemic risk model is incorrectly specified. More precisely, systemic risk measures are estimated and forecasted based on risk model (11) but the actual innovation of the data is contaminated as follows

$$\boldsymbol{z}_t \sim [0.4 \times \mathrm{t}_\nu(\mathbf{0} - 1.5, \boldsymbol{I}_2 \times \sqrt{1.75}) + 0.6 \times \chi^2(\mathbf{1})].$$

We refer to this experiment as ALTER2, and we report the rejection rates in Table 3. As is expected, both Q and Q_r display rejection power against experiment ALTER2. Because the data is contaminated across each time period, we also observe that the rejection rates increase in m. The equally weighted backtest S_r continues to yield comparable power to Q_r when the TRU kernel is applied. The performance across kernel functions remains similar for this experiment. Overall, the experiment shows that our backtests possess detection power against inadequate systemic risk forecasts due to distributional misspecification, with Q_r having slightly better power than Q.

[Insert Table 3 approximately here]

In the third alternative experiment, we simulate the scenario in which we have data dynamics being incorrectly specified. More precisely, systemic risk measures are estimated and forecasted based on risk model (11) but the actual dynamics of $Y_{i,t}$ follow the threshold autoregressive process

$$
\begin{aligned}
&Y_{i,t} = a_{i,t} \times Y_{i,t-1} + \eta_{i,t}, \quad i = 1, 2, \\
&a_{i,t} = 0.9 \times \mathbb{1}(\eta_{i,t-1} \leq -1).
\end{aligned}
$$

We refer to this experiment as ALTER3, and we report the rejection rates in Table 4. We observe our backtests continue to display rejection power that generally increases with the sample size. In contrast to experiments ALTER1 and ALTER2, we observe the rejection rates of all backtests decrease in m. This is because, unlike experiments ALTER1 and ALTER2 in which we have time-invariant misspecification, under experiment ALTER3 we have time-dependent alternative with a particular focus on the first lag in the data dynamics. As a result, increasing the testing scope (i.e., by setting a large m) may not capture more evidence against such alternative, and instead will reduce the overall evidence captured by the backtests due to the averaging effect across the testing scope. We continue to observe S_r yielding comparable power to Q_r when the TRU kernel is applied, but their powers are lower than Q_r with the downward weighting kernels. This is because a test that assigns more weights to observations at lower lag order is expected to be more efficient against ALTER3, in which we have first-order dynamic misspecification. Across the combinations of (T, n), we observe that using the downward weighting kernels alleviates power reduction in m. Interestingly, the finite sample power of our Q backtest is notably higher than that of our Q_r backtest. This could be attributed to the fact that Q is not a whitened statistic, and due to that Q preserves data dynamics more effectively and thus it tends to give better power against alternatives driven by dynamic misspecification. To further examine such effect, we carry out another dynamic driven alternative experiment.

[Insert Table 4 approximately here]

In the final experiment, we simulate the scenario in which a risk manager correctly specifies an autoregressive process for the $Y_{i,t}$ series, but misses some higher order and delayed effect. More precisely, systemic risk measures are estimated and forecasted based on risk model (11) but the actual autoregressive dynamics of $Y_{i,t}$ admit a delayed effect

$$Y_{i,t} = \alpha_{i0} + \alpha_{i1} Y_{i,t-1} + 0.75 Y_{i,t-2} + \eta_{i,t}, \quad i = 1, 2.$$

We refer to this experiment as ALTER4, and we report the rejection rates in Table 5. Overall, the power pattern under ALTER4 is similar to that under ALTER3, as we have dynamic driven alternatives in both experiments. We continue to observe the optimality of downward weighting backtests as they generally outperform equally weighted backtests and they alleviate power reduction in m. For instance, in the $(T, n) = (500, 500)$ panel, the power given by Q_r with downward weighting kernels for $m = 10$ and $m = 20$ is very comparable, whereas the power given by S_r shows a clear decline as m increases from 10 to 20.

[Insert Table 5 approximately here]

Summing up the simulation study, the proposed Q and Q_r backtests show robustness against estimation risk to display reasonable finite sample sizes, and they demonstrate nontrivial detection power against alternatives of different natures. Setting a large m yields better power against alternatives that are of time-invariant nature but the opposite is observed against alternatives that are of time-dependent nature. In this aspect, downward weighting backtests such as Q and Q_r are optimal as they alleviate power reduction in m. Thus, the power improvement offered by the new backtests is more significant under the time-dependent alternatives. Against alternatives driven by dynamic misspecification, we find that the standardized backtest Q — which preserves data dynamics more effectively — tends to yield better power than the whitened backtest Q_r. We also find that under the time-invariant alternatives whitened backtest gives slightly better power than standardized backtest. Overall the new backtests demonstrate complementary capacities.

IV. Empirical evidence

A. Data

While the performance of the new backtests (Q and Q_r) has been demonstrated using simulated data, it is equally important to verify their performance using real-world data. In this section, we apply the new backtests to examine the reliability of systemic risk forecasts issued by major financial institutions. For consistency with existing literature we focus on large US institutions, that is, we follow Brownlees and Engle (2017) and Banulescu-Radu et al. (2021) to use institutions with market capitalization greater than 5 billion dollars as of the end of June 2007. The names and CRSP tickers of the selected institutions are tabulated in Online Appendix G. We collect institutional observations dating from early 2000 to late 2022, so that we also study the more recent pandemic-induced financial market crashes. More precisely, we use daily observations from 3-Jan-2000 to 1-Dec-2022. Our dataset is naturally unbalanced because not every institution has been trading continuously over the study period. For instance, in 2008 Bear Stearns and Lehman Brothers collapsed over the subprime financial crisis. We compute daily log-returns and market capitalization for the selected institutions using CRSP database, and we consider as market return the daily log-return of the CRSP value-weighted market index.

At any given time period over the full sample, we examine for all available institutions the reliability of their systemic risk forecasts provided the required out-sample data is available. When moving our analysis across time periods, we use expanding window scheme to make full use of in-sample data availability. More precisely, we use two years of daily observations as the initial in-sample (i.e., from Jan-2000 to Jan-2002), and the sample length increases as we move forward with our analysis on a monthly basis. Thus, our in-sample size T contains at least two years of data which is approximately 500 observations.

We use testing parameters guided by our simulation study. The statistical significance level of all backtests is kept at 5%. We maintain $\alpha = 5\%$. For any given institution, we model its in-sample return series using the same risk model and the same estimation procedure implemented in the simulation study. This yields parameter estimates of the risk model $\hat{\boldsymbol{\phi}}_T$, which then allows us to forecast the risk indicator series $H_t(\alpha, \hat{\boldsymbol{\phi}}_T)$, for $t = T+1, ..., T+n$. For this purpose we need to select the out-sample size n, and we note that a value too large for this parameter would reduce the period available for examination. For instance, to backtest one (three) year(s) of out-sample forecasts in 2021 would require observed data up to 2022 (2025). We use $n = 250$, which is about one year of out-sample data. We then compute our backtests to examine if the generated systemic risk forecasts are reliable. Both Q and Q_r backtests are used as we learn about their complementary capacities from the simulation study. As a benchmark for existing methodology, we include the Box-Pierce-typed backtest by computing S_r. We use downward weighting kernel for our backtests, and for natural comparison with S_r we use the Bartlett kernel because it belongs to the class of truncated kernel. Regarding kernel parameter, we use both the lowest and the highest values covered in the simulation study, i.e., $m = (5, 30)$. In a typical monthly analysis, we backtest the systemic risk forecasts issued by each institution so long as the required data is available. Accordingly, in each monthly analysis we apply Bonferroni correction to the backtesting significance level based on the number of backtested institutions for that month.[4]

B. Backtesting results

Figure 1 plots on a monthly basis the percentage of institutions with inadequate systemic risk forecasts detected by the Q backtest (depicted in blue), the Q_r backtest (depicted in black), and the S_r backtest (depicted in red). We observe that compared with S_r both the new backtests have higher detection frequencies over inadequate systemic risk forecasts as the blue and black bars dominate their red counterparts on many occasions. This finding holds for both m. We also observe several sizeable clusters in the plot, which we highlight and discuss in the following because these clusters imply that the systemic risk forecasts of many major institutions are inaccurate and unreliable.

[Insert Figure 1 approximately here]

In the cluster between Jan-2003 and Jan-2004, the performance of all backtests is rather similar. More precisely, the monthly average rejection frequencies by Q, Q_r and S_r are, respectively, 64%, 56% and 54%, at $m = 5$. When $m = 30$, the average rejection rates are 75%, 70% and 72%, for Q, Q_r and S_r, respectively. In this cluster we observe for all three backtests an increase in rejection frequencies as m increases. This is consistent with our expectation that by increasing the m parameter, the testing

[4]As highlighted in Bajgrowicz, Scaillet, and Treccani (2016), the Bonferroni method is a relatively conservative approach. Because the main objective of this empirical study is to examine the performance of the new backtests with respect to a benchmark, a conservative approach is appropriate as it helps to ensure that any improvement from the new backtests is not due to chance. For a liberal alternative to the Bonferroni method, one could refer to Barras, Scaillet, and Wermers (2010). We thank a reviewer for sharing this discussion.

scope of a backtest increases, which in turn may lead to the backtest detecting a more diverse forms of systemic risk inadequacy. Economically, the tall 2003–2004 cluster suggests that the risk forecasts of many institutions for the period 2004–2005 were highly inadequate. This reflects the 2004–2005 credit deterioration of the US market, which at its height saw the credit deterioration of several major US entities such as General Motors to the junk status. Besides, the finding could also be attributed to the fact that the DCC-GARCH specification is simply not a suitable model for systemic risk as far as this period is concerned. In practice, risk managers should consider revising the model or using alternative specifications.

In the cluster between Jan-2004 and Jan-2006, the monthly average rejection frequencies by Q, Q_r and S_r are, respectively, 12%, 10% and 13%, at $m = 5$. When $m = 30$, the rejection rates are 27%, 26% and 20% for Q, Q_r and S_r, respectively. Similar to the previous cluster, we observe an increase in rejection frequencies across all backtests when m is large. When $m = 5$, S_r has slightly better performance over Q and Q_r, which reverses when $m = 30$. The reason as to why the detection rate of S_r does not catch up with our backtests is likely due to it losing some power when checking a larger scope, a phenomenon too observed in our simulation study. By contrast, the rejection rates of the new backtests increase and outperform S_r. This highlights that the empirical alternatives captured with a large m could offset the potential loss in power when the backtest is properly weighted according to empirical stylized facts. We obtain similar conclusions for the cluster between Jan-2011 and July-2011.

We now discuss the remaining major clusters of model inadequacy, that is, the cluster between July-2014 and July-2015, the 2017 cluster, and the cluster beyond Jan-2019. The precise monthly average rejection frequencies by (Q, Q_r, S_r) in the July-2014–July-2015 cluster are (40%, 24%, 3%) at $m = 5$, and (16%, 9%, 1%) at $m = 30$. In the 2017 cluster, the rejection frequencies by (Q, Q_r, S_r) are (68%, 59%, 52%) at $m = 5$, and (62%, 51%, 8%) at $m = 30$. The tall 2017 cluster suggests that the

DCC-GARCH specification may not be a suitable model to forecast systemic risk for 2018, which could be due to the fact that this period witnessed several major international political events such as the US-China tradewar and the UK's Brexit. In the cluster beyond Jan 2019, the rejection frequencies by (Q, Q_r, S_r) are (44%, 32%, 20%) at $m = 5$, and (37%, 29%, 8%) at $m = 30$. Given that the cluster beyond Jan-2019 coincides with the run-up to the pandemic-induced market crashes, we highlight the nontrivial under detection of S_r compared with the new backtests. In particular, S_r under detects 24% (12%) compared with Q (Q_r) at $m = 5$, and 29% (21%) compared with Q (Q_r) at $m = 30$. The prominence of the new backtests over S_r holds in the other two clusters regardless of m. Although all three backtests suffer some power reduction in m, it is minimal for the new backtests. This further highlights the optimality of our backtests in the tradeoff between losing detection power and detecting a wider alternatives as the testing scope increases.

To see the economic implications brought by the new backtests on a greater scale, we note that in any given month over the full examination period we backtest an average of 74 financial institutions. We further note the monthly average rejection rates by (Q, Q_r, S_r) over the full examination period are (18%, 15%, 11%) at $m = 5$, and (18%, 16%, 9%) at $m = 30$. Compared with Q, S_r under detects 7% per month $\times$ 228 months $\times$ 74 institutions $= 1181$ cases over the course of our empirical study when $m = 5$. When $m = 30$, S_r under rejects 1518 (1181) cases compared with Q (Q_r). We note from these numbers that the improvement of our backtests over S_r is of considerable economic significance, since, if the new backtests are used instead of S_r, a larger scale of inadequate systemic risk forecasts would have been detected among major financial institutions. These additional cases could play a critical role in prompting regulatory attention, which could call for a full-scale revision of existing systemic risk modeling in the wider economy, which in turn would improve financial stability.

Although the average findings may suggest Q regularly outperforms Q_r, we highlight that in some instances such as the Jan-2013–July-2013 period, Q_r has detected alternatives not captured by Q. This is in line with our simulation findings. The complementary performance of Q_r and Q can be attributed to them generally not capturing the same dynamics about systemic risk inaccuracy, since the former is a whitened test whereas the latter is not. Thus, when examining the overall health of the financial system in practice, regulators could apply the formula $\max(Q, Q_r)$ to detect as many indications about large-scale measure inadequacies of systemic risk as possible. This provides a "better-safe-than-sorry" approach which is recently advocated by the European Central Bank (see, e.g., Rehn, 2020; Enria, 2022).

One could argue that the empirical performance of the new backtests is only evident for a relatively small n, and that the performance of all backtests would be very comparable when n increases. To examine the sensitivity of our main findings, we repeat the backtesting analysis with a larger $n = 500$, and we plot the results in Figure 2. Due to the more demanding out-sample size, the time dimension of Figure 2 naturally reduces slightly compared with that of Figure 1. By the same reasoning, the average number of institutions qualified for examination reduces to 73. Compared with Figure 1, we observe an increase in rejection frequencies across all backtests. This is in line with Theorem 2 and Proposition 2 that the power of our backtests is expected to increase in n. We also observe that the overall rejection pattern for all backtests is similar to that in Figure 1, with the new backtests continue to outperform existing backtest. More precisely, the monthly average rejection rates over the full sample by (Q, Q_r, S_r) are (29%, 24%, 17%) at $m = 5$, and (30%, 26%, 13%) at $m = 30$. The economic implications offered by the new backtests are now more evident, as for instance, S_r under rejects 2681 (2050) cases compared with Q (Q_r) at $m = 30$. Unlike S_r, we note the rejection rates of Q and Q_r are similar across m, which highlights once again the optimality of our backtests as they give optimal tradeoff in m.

[Insert Figure 2 approximately here]

To ensure that the empirical performance of the new backtests is not driven by a specific choice of kernel function, we repeat every backtesting analysis using the QS kernel, which, unlike the BAR kernel, belongs to the class of kernel with unbounded support. Figure 3 plots the analysis results for every empirical setting considered thus far. The results suggest that our main findings are highly robust as the prominence of Q and Q_r over existing backtest remains highly evident. This highlights, as with our simulation evidence, that the choice of kernel does not have a strong impact over the performance of our backtests so long as the kernel function is downward weighting.

[Insert Figure 3 approximately here]

Finally, we perform a robustness analysis by whitening $\hat{\rho}_n(j)$ up to the same order as the kernel parameter, that is, we have $\hat{\rho}_n(j)$ being whitened up to order m, which we denote by $\hat{\rho}_{nrm}(j)$, $j = 1, ..., m$. Such partial whitening does not affect the computation of truncated whitened backtests such as S_r and Q_r with a truncated kernel. In particular, $\hat{\rho}_{nrm}(j)$ is the j-th element of the vector $\hat{\boldsymbol{\Delta}}_m^{-1/2}\hat{\boldsymbol{\rho}}_n^{(m)}$, where $\hat{\boldsymbol{\rho}}_n^{(m)}$ collects $\{\hat{\rho}_n(j)\}_{j=1}^m$, and $\hat{\boldsymbol{\Delta}}_m$ is a $m \times m$ matrix with the same (i, j)-th element as that in $\hat{\boldsymbol{\Delta}}$. The Box-Pierce-type statistic based on $\hat{\rho}_{nrm}(j)$ is given by $S_r = n\sum_{j=1}^m \hat{\rho}_{nrm}^2(j)$. For proper comparison we also use $\hat{\rho}_{nrm}$ in our whitened backtest Q_r with the Bartlett kernel. We plot the results in Figure 4, and we can see that the rejection patterns are very similar to those in Figure 1 and Figure 2. The improvement brought by the new backtests remains highly evident and robust.

[Insert Figure 4 approximately here]

C. Summary and regulatory implications

In summary, using real-world data generated by the major US financial institutions, we find that, compared with existing approach, the new backtests Q and Q_r offer better empirical power as they reveal thousands more cases of systemic risk inadequacies. Such improvement is also of considerable economic significance. Our findings remain robust under a series of sensitivity analyses including using alternative kernel parameters, alternative sample sizes, and alternative kernel classes. Because of their complementary features, Q and Q_r can be applied together to detect as many indications about large-scale measure inadequacies of systemic risk as possible. This provides a useful "better-safe-than-sorry" approach for financial supervisors and regulators.

We acknowledge that despite being estimated from an inadequate risk model, systemic risk measures may still be useful in e.g., ranking financial institutions by their systemic relevance. This can be attributed to ranking operations being impartial to effects that are uniform across all institutions such as inaccuracy from using the same (inadequate) risk model. Using the same risk model for a group of institutions is also fairly standard in the academic literature (see, e.g., Brownlees and Engle, 2017; Banulescu-Radu et al., 2021). In reality, however, each financial institution usually has its own proprietary risk model, which is regularly tested for accuracy and is subject to ongoing refinement. In this aspect, Q and Q_r serve as new statistical tools for checking, with better power, whether a risk model can be further improved to produce more accurate systemic risk estimates. With more accurate risk estimates across all financial institutions, financial stability will improve as a collective result.

V. System warning indicator

In the spirit of Gourieroux and Zakoïan (2013) and Lazar and Zhang (2019), BHLS propose a system indicator by solving the following quadratic program

$$\alpha_c = \underset{\tilde{\alpha}\in]0,1[}{\operatorname{argmin}} \left[\frac{1}{n}\sum_{j=1}^{n}(H_{T+j}(\tilde{\alpha},\hat{\boldsymbol{\phi}}_T) - \alpha/2)\right]^2 . \tag{14}$$

The idea is based on correcting the n forecasting errors in the forecasted risk series $H_{T+j}(\alpha, \hat{\boldsymbol{\phi}}_T)$, for $j = 1, ..., n$. Because the margin of error across a forecasted series is generally nonuniform — in particular, the next-period forecast is expected to be more accurate with a smaller margin of error than the forecast multiple periods ahead — assigning equal weight to correct each of the n forecasts in equation (14) regardless of the magnitude of their expected errors may not be optimal. Therefore, an improved objective function is one that assigns nonuniform weights for the error corrections. Using the kernel tools introduced so far, we can conveniently generalize equation (14) as follows

$$\alpha_c = \underset{\tilde{\alpha}\in]0,1[}{\operatorname{argmin}} \left[\frac{\sum_{j=1}^{n} k^2(1-j/n)\times(H_{T+j}(\tilde{\alpha},\hat{\boldsymbol{\phi}}_T) - \alpha/2)}{\sum_{j=1}^{n} k^2(1-j/n)}\right]^2 . \tag{15}$$

When k is set as the TRU kernel we retrieve the program objective in equation (14). We propose to use here downward weighting kernels such as the BAR function. When the systemic risk model is perfectly specified, the objective function (15) gives $\alpha_c = \alpha$ regardless of the kernel functions used. When otherwise, the uniformly weighted function gives α_c that aims to adjust with the same magnitude each of the n forecasts. Using the recommended kernel, the objective function yields α_c that aims to adjust with smaller magnitude the immediate forecasts and appropriately with greater magnitude the forecasts n steps ahead. The α_c obtained under our approach gives a closer match in correcting the expected forecasting errors, and thus an indicator based on our α_c is expected to reflect more accurately the state of the financial system.

To build a warning indicator we quantify as with BHLS the deviation between the forecasted measure $MES_{Y1,t+1}(\alpha, \hat{\boldsymbol{\phi}}_T)$ and the adjusted measure $MES_{Y1,t+1}(\alpha_c, \hat{\boldsymbol{\phi}}_T)$, i.e.,

$$EWI_t(\alpha, \alpha_c) = MES_{Y1,t+1}(\alpha, \hat{\boldsymbol{\phi}}_T) - MES_{Y1,t+1}(\alpha_c, \hat{\boldsymbol{\phi}}_T). \tag{16}$$

To interpret $EWI_t(\alpha, \alpha_c)$ we note that MES is typically negative valued because an institution is expected to make a loss rather than a profit whenever the broader system is in distress. Hence, with the convention that risk is associated with losses rather than profits, a positive $EWI_t(\alpha, \alpha_c)$ indicates an unexpected increase in systemic risk, whereas a negative $EWI_t(\alpha, \alpha_c)$ indicates an unexpected decrease in systemic risk. As with BHLS we use $n = 250$.

To illustrate on the system-wide scale the empirical unexpected deviation in systemic risk, we plot in Figure 5 the $\overline{EWI}_t(\alpha, \alpha_c)$ indicator, which is calculated by summing the individual $EWI_t(\alpha, \alpha_c)$ of each institution studied previously. The sum is then scaled by the number of institutions to ensure meaningful comparison across the time dimension. For comparison we plot in black the indicator based on the BAR kernel, and in red the indicator based on the TRU kernel. As discussed, the latter retrieves the program objective in BHLS.

[Insert Figure 5 approximately here]

We observe for the most part that both indicators are stable with very similar values. This is in line with our expectation that weighting functions do not play a major role in times where the net unexpected systemic risk among all institutions is close to zero. We observe the expected positive spikes before both the 2008 subprime-induced market crash and the 2020 pandemic-induced market crash. The warning indicators capture the large-scale unexpected increase in systemic risk among the major financial institutions. In other words, the number of institutions with unexpected increase in systemic risk far exceeds that with unexpected decrease in systemic risk. More notably, our indicator yields warning signals that are more profound by reacting more acutely and sharply prior to both crises. In particular, compared with existing indicator the warning signals transmitted by our indicator prior to both 2008 and 2020 market crashes are about 40% more alarming! This finding shows that by matching more closely with the expected forecasting errors, our indicator yields warning signals that are more noticeable to the regulators. This finding is especially significant in hindsight because the US financial system indeed witnessed two major and unprecedented crises following steep spikes in the proposed indicator. For instance, the subprime-induced system instability in 2008 saw the collapse of major institutions such as Bear Stearns and Lehman Brothers; whereas the pandemic-induced system instability in 2020 saw the US stock market facing a series of crashes and large-scale trading halts.

It could be noted that the pattern of model inadequacies in Section IV is not fully aligned with the pattern of the $\overline{EWI}_t(\alpha, \alpha_c)$ indicator in Figure 5. For instance, the $\overline{EWI}_t(\alpha, \alpha_c)$ indicator is rather tranquil during the 2003–2004 credit crisis, whereas the 2007–2008 subprime crisis is not associated with inadequate modeling. First, we note that the adjustment in equation (15) is only based on a particular condition of systemic risk inadequacy that new backtests could detect. The measure inadequacies reflected in Section IV could be due to other statistical properties being violated such as serial correlation. Secondly, the $\overline{EWI}_t(\alpha, \alpha_c)$ indicator is purely a mathematical operation that aggregates a series of deviations described by equation (16), where each individual deviation need not be statistically significant to be considered. In summary, while the new backtests focus on gauging the adequacy of individual risk modeling in the statistical sense, when monitoring the overall financial health, both statistical-based measures (e.g., the proposed backtests) and mathematical-based measures (e.g., the proposed warning indicator) should be consulted.

Finally, we perform several sensitivity analyses to ensure that our main findings are robust. We confirm that setting a larger out-sample size (e.g., $n = 500$) and that using an alternative class of kernel (e.g., the QS function), do not alter the main findings.

VI. Conclusion

This paper offered two classes of novel backtests for checking the adequacy of well-known systemic risk measures. The asymptotic properties of the new backtests were established under regularity conditions. The paper demonstrated the consistency of the new backtests, and showed that both backtests follow the standard normal null distribution even in the presence of estimation risk. The paper conducted extensive simulation and empirical exercises to verify the improvement brought by the new backtests over existing method. Simulation evidence confirmed the optimality of the new backtests as they gave optimal tradeoff between losing detection power and detecting a wider alternatives as the testing scope increases. The empirical study highlighted the economic significance brought by the new backtests as the evidence suggested that the new backtests would have detected a larger scale of inadequate systemic risk modeling among the major financial institutions. Finally, the paper offered an early warning indicator which, when compared with existing indicator, yielded more profound warning signals prior to both the 2008 subprime-induced and the 2020 pandemic-induced market crashes.

References

Acharya, V.; R. Engle; and M. Richardson. "Capital shortfall: A new approach to ranking and regulating systemic risks." *American Economic Review*, 102 (2012), 59–64.

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

Adrian, T., and M. K. Brunnermeier. "CoVaR." National Bureau of Economic Research (2011).

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

Aielli, G. P. "Dynamic conditional correlation: On properties and estimation." *Journal of Business & Economic Statistics*, 31 (2013), 282–299.

Bajgrowicz, P.; O. Scaillet; and A. Treccani. "Jumps in high-frequency data: Spurious detections, dynamics, and news." *Management Science*, 62 (2016), 2198–2217.

Banulescu-Radu, D.; C. Hurlin; J. Leymarie; and O. Scaillet. "Backtesting marginal expected shortfall and related systemic risk measures." *Management Science*, 67 (2021), 5730–5754.

Barras, L.; O. Scaillet; and R. Wermers. "False discoveries in mutual fund performance: Measuring luck in estimated alphas." *Journal of Finance*, 65 (2010), 179–216.

Berkowitz, J.; P. Christoffersen; and D. Pelletier. "Evaluating value-at-risk models with desk-level data." *Management Science*, 57 (2011), 2213–2227.

Box, G. E., and D. A. Pierce. "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models." *Journal of the American statistical Association*, 65 (1970), 1509–1526.

Breusch, T. S., and A. R. Pagan. “The Lagrange multiplier test and its applications to model specification in econometrics.” *Review of Economic Studies*, 47 (1980), 239–253.

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

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

Du, Z., and J. C. Escanciano. “Backtesting expected shortfall: Accounting for tail risk.” *Management Science*, 63 (2017), 940–958.

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

Enria, A. “Better safe than sorry: banking supervision in the wake of exogenous shocks.” (2022).

Fiorentini, G.; E. Sentana; and G. Calzolari. “Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student t innovations.” *Journal of Business & Economic Statistics*, 21 (2003), 532–546.

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

Glosten, L. R.; R. Jagannathan; and D. E. Runkle. “On the relation between the expected value and the volatility of the nominal excess return on stocks.” *Journal of Finance*, 48 (1993), 1779–1801.

Gourieroux, C., and J.-M. Zakoïan. “Estimation adjusted VaR.” *Econometric Theory*, 29 (2013), 735–770.

Hong, Y. “Consistent testing for serial correlation of unknown form.” *Econometrica*, 64 (1996), 837–864.

Kupiec, P. “Techniques for verifying the accuracy of risk measurement models.” *Journal of Derivatives*, 3 (1995), 73–94.

Lazar, E., and N. Zhang. “Model risk of expected shortfall.” *Journal of Banking & Finance*, 105 (2019), 74–93.

Newey, W. K., and K. D. West. “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55 (1987), 703–708.

Rehn, O. “ECB is better safe than sorry with easy policy, Rehn says.” (2020).

TABLE 1

Rejection rates under experiment NULL

This table reports experimental rejection rates (in %) of Q, Q_r, S and S_r, at the 5% significance level. The subscripts BAR, DAN, PAR, QS and TRU denote, respectively, the Bartlett kernel, the Daniell kernel, the Parzen kernel, the Quadratic-Spectral kernel and the Truncated Uniform kernel. The integer m denotes the kernel parameter. Number of replications = 1000.

m	5	10	20	30	m	5	10	20	30
				$T = 250$, $n = 250$					
Q_{BAR}	5.8	7.1	7.0	7.1	$Q_{r\text{BAR}}$	4.5	5.4	5.5	5.4
Q_{PAR}	5.2	6.7	6.9	7.0	$Q_{r\text{PAR}}$	4.3	5.2	5.7	5.8
Q_{QS}	6.3	7.1	7.2	7.4	$Q_{r\text{QS}}$	5.0	5.7	5.7	5.8
Q_{DAN}	5.8	7.0	7.6	8.0	$Q_{r\text{DAN}}$	4.6	6.1	6.0	6.0
Q_{TRU}	7.1	8.1	8.2	7.9	$Q_{r\text{TRU}}$	6.4	6.8	6.6	6.7
S	9.0	11.5	13.8	14.6	S_r	5.1	5.8	6.2	6.1
				$T = 250$, $n = 500$					
Q_{BAR}	3.9	5.7	6.7	7.3	$Q_{r\text{BAR}}$	3.9	5.0	6.2	6.3
Q_{PAR}	3.3	4.3	6.5	7.3	$Q_{r\text{PAR}}$	3.7	4.1	5.6	5.9
Q_{QS}	4.0	6.0	7.3	7.9	$Q_{r\text{QS}}$	3.7	5.3	5.6	6.6
Q_{DAN}	4.1	5.8	7.5	8.2	$Q_{r\text{DAN}}$	3.7	5.5	6.1	7.0
Q_{TRU}	6.2	7.8	8.8	8.0	$Q_{r\text{TRU}}$	5.5	6.0	7.3	7.4
S	8.5	11.8	15.9	15.6	S_r	4.6	5.7	6.6	7.1
				$T = 500$, $n = 250$					
Q_{BAR}	6.2	7.3	7.0	6.9	$Q_{r\text{BAR}}$	4.8	6.0	6.7	6.0
Q_{PAR}	4.7	6.7	6.9	6.9	$Q_{r\text{PAR}}$	4.0	5.2	6.4	7.0
Q_{QS}	6.5	7.6	7.3	7.3	$Q_{r\text{QS}}$	4.9	6.5	6.6	6.7
Q_{DAN}	6.1	7.9	8.3	8.4	$Q_{r\text{DAN}}$	5.0	6.3	7.3	7.0
Q_{TRU}	7.9	8.3	7.8	8.5	$Q_{r\text{TRU}}$	6.3	6.3	6.5	7.0
S	8.9	9.1	11.0	12.4	S_r	5.8	5.6	5.9	6.7
				$T = 500$, $n = 500$					
Q_{BAR}	5.0	5.3	5.7	6.4	$Q_{r\text{BAR}}$	5.4	5.3	5.1	4.8
Q_{PAR}	4.9	5.1	6.2	5.6	$Q_{r\text{PAR}}$	4.6	5.4	5.3	4.9
Q_{QS}	5.3	5.9	6.1	6.3	$Q_{r\text{QS}}$	5.1	5.3	4.7	5.0
Q_{DAN}	5.4	5.8	5.9	7.3	$Q_{r\text{DAN}}$	5.6	5.4	5.2	5.2
Q_{TRU}	6.4	6.6	7.2	6.7	$Q_{r\text{TRU}}$	4.7	4.5	5.4	5.4
S	7.7	9.3	11.8	12.5	S_r	4.2	4.0	4.9	4.9

TABLE 2

Rejection rates under experiment ALTER1

This table reports experimental rejection rates (in %, size-adjusted) of Q, Q_r and S_r, at the 5% significance level. The subscripts BAR, DAN, PAR, QS and TRU denote, respectively, the Bartlett kernel, the Daniell kernel, the Parzen kernel, the Quadratic-Spectral kernel and the Truncated Uniform kernel. The integer m denotes the kernel parameter. Number of replications = 1000.

m	5	10	20	30	m	5	10	20	30
					$T = 250$, $n = 250$				
Q_{BAR}	22.1	23.9	29.3	30.5	$Q_{r\text{BAR}}$	24.5	26.9	31.3	32.8
Q_{PAR}	20.1	21.5	25.6	29.1	$Q_{r\text{PAR}}$	23.9	24.6	28.6	30.8
Q_{QS}	21.7	24.7	30.1	32.0	$Q_{r\text{QS}}$	24.8	28.0	32.6	33.6
Q_{DAN}	23.1	25.5	29.7	32.9	$Q_{r\text{DAN}}$	27.7	28.0	32.2	34.6
Q_{TRU}	23.9	27.4	31.9	34.2	$Q_{r\text{TRU}}$	26.2	30.1	33.5	37.0
					S_r	26.2	30.1	33.5	37.0
					$T = 250$, $n = 500$				
Q_{BAR}	30.6	31.0	35.4	37.7	$Q_{r\text{BAR}}$	35.6	37.6	40.6	44.8
Q_{PAR}	30.2	32.2	32.6	35.8	$Q_{r\text{PAR}}$	32.7	35.8	38.6	40.4
Q_{QS}	31.5	32.0	36.8	38.2	$Q_{r\text{QS}}$	36.0	38.2	42.2	44.7
Q_{DAN}	32.6	32.5	36.7	38.0	$Q_{r\text{DAN}}$	37.0	37.7	43.1	45.3
Q_{TRU}	32.5	34.8	38.4	41.1	$Q_{r\text{TRU}}$	36.3	40.1	43.1	46.2
					S_r	36.3	40.1	43.1	46.2
					$T = 500$, $n = 250$				
Q_{BAR}	19.6	22.0	26.8	30.4	$Q_{r\text{BAR}}$	22.7	24.4	28.1	30.9
Q_{PAR}	20.9	20.4	23.7	27.4	$Q_{r\text{PAR}}$	22.7	23.6	24.9	28.6
Q_{QS}	19.3	23.2	28.2	32.3	$Q_{r\text{QS}}$	23.6	24.8	29.3	33.5
Q_{DAN}	21.2	23.6	26.4	30.1	$Q_{r\text{DAN}}$	24.2	24.5	29.2	34.2
Q_{TRU}	22.0	25.5	33.0	33.9	$Q_{r\text{TRU}}$	21.2	28.0	33.3	33.3
					S_r	21.2	28.0	33.3	33.3
					$T = 500$, $n = 500$				
Q_{BAR}	28.8	35.9	39.9	41.7	$Q_{r\text{BAR}}$	27.3	34.2	40.7	43.3
Q_{PAR}	23.3	31.3	36.1	40.3	$Q_{r\text{PAR}}$	25.7	31.8	36.6	40.9
Q_{QS}	30.0	35.2	40.3	44.5	$Q_{r\text{QS}}$	30.5	36.2	43.3	44.5
Q_{DAN}	30.4	35.8	42.3	41.7	$Q_{r\text{DAN}}$	30.6	35.9	43.0	45.3
Q_{TRU}	33.1	38.6	42.8	44.4	$Q_{r\text{TRU}}$	36.0	41.8	45.1	47.0
					S_r	36.0	41.8	45.1	47.0

TABLE 3

Rejection rates under experiment ALTER2

This table reports experimental rejection rates (in %, size-adjusted) of Q, Q_r and S_r, at the 5% significance level. The subscripts BAR, DAN, PAR, QS and TRU denote, respectively, the Bartlett kernel, the Daniell kernel, the Parzen kernel, the Quadratic-Spectral kernel and the Truncated Uniform kernel. The integer m denotes the kernel parameter. Number of replications = 1000.

m	5	10	20	30	m	5	10	20	30
					$T = 250$, $n = 250$				
Q_{BAR}	43.5	46.2	52.0	53.4	$Q_{r\text{BAR}}$	42.8	45.5	51.8	53.9
Q_{PAR}	41.6	43.7	48.0	51.6	$Q_{r\text{PAR}}$	42.6	43.9	48.6	51.4
Q_{QS}	43.4	46.6	52.9	55.8	$Q_{r\text{QS}}$	43.9	47.4	53.2	55.3
Q_{DAN}	45.4	48.1	52.4	55.3	$Q_{r\text{DAN}}$	46.0	47.7	52.1	55.0
Q_{TRU}	45.4	50.4	54.5	58.1	$Q_{r\text{TRU}}$	45.3	50.2	54.7	58.3
					S_r	45.3	50.2	54.7	58.3
					$T = 250$, $n = 500$				
Q_{BAR}	54.6	56.3	62.3	64.0	$Q_{r\text{BAR}}$	55.2	58.1	65.0	70.1
Q_{PAR}	53.8	56.5	59.3	62.7	$Q_{r\text{PAR}}$	51.0	56.2	60.7	65.1
Q_{QS}	56.2	58.1	62.6	65.0	$Q_{r\text{QS}}$	55.8	59.6	67.0	71.5
Q_{DAN}	57.5	58.8	63.9	65.5	$Q_{r\text{DAN}}$	58.2	60.6	68.6	71.6
Q_{TRU}	57.6	60.8	64.1	66.3	$Q_{r\text{TRU}}$	57.9	63.5	68.4	70.6
					S_r	57.9	63.5	68.4	70.6
					$T = 500$, $n = 250$				
Q_{BAR}	44.0	48.4	55.1	58.5	$Q_{r\text{BAR}}$	45.9	48.7	53.6	58.1
Q_{PAR}	43.6	45.7	50.7	55.4	$Q_{r\text{PAR}}$	44.4	47.4	48.8	54.8
Q_{QS}	44.7	50.0	57.0	59.9	$Q_{r\text{QS}}$	47.5	48.8	56.6	60.8
Q_{DAN}	46.2	50.5	55.1	58.3	$Q_{r\text{DAN}}$	47.2	49.5	56.0	60.9
Q_{TRU}	46.7	51.9	59.9	62.1	$Q_{r\text{TRU}}$	45.6	52.5	59.5	60.0
					S_r	45.6	52.5	59.5	60.0
					$T = 500$, $n = 500$				
Q_{BAR}	55.8	62.6	67.8	70.1	$Q_{r\text{BAR}}$	51.8	61.1	69.5	73.8
Q_{PAR}	52.1	58.4	64.4	67.9	$Q_{r\text{PAR}}$	50.4	57.0	64.8	69.8
Q_{QS}	56.6	63.1	68.1	72.2	$Q_{r\text{QS}}$	55.4	63.7	72.9	75.1
Q_{DAN}	58.1	63.3	69.8	70.1	$Q_{r\text{DAN}}$	56.0	64.1	72.6	76.3
Q_{TRU}	60.0	67.1	70.5	74.2	$Q_{r\text{TRU}}$	61.4	70.0	74.8	75.8
					S_r	61.4	70.0	74.8	75.8

TABLE 4

Rejection rates under experiment ALTER3

This table reports experimental rejection rates (in %, size-adjusted) of Q, Q_r and S_r, at the 5% significance level. The subscripts BAR, DAN, PAR, QS and TRU denote, respectively, the Bartlett kernel, the Daniell kernel, the Parzen kernel, the Quadratic-Spectral kernel and the Truncated Uniform kernel. The integer m denotes the kernel parameter. Number of replications = 1000.

m	5	10	20	30	m	5	10	20	30
				$T = 250, n = 250$					
Q_{BAR}	20.0	14.8	12.1	10.2	$Q_{r\text{BAR}}$	16.0	10.6	8.7	7.2
Q_{PAR}	23.8	17.0	11.8	10.7	$Q_{r\text{PAR}}$	20.7	12.8	8.8	7.8
Q_{QS}	18.4	12.0	10.3	8.3	$Q_{r\text{QS}}$	14.2	9.3	7.5	5.7
Q_{DAN}	19.2	13.4	9.6	9.1	$Q_{r\text{DAN}}$	15.3	9.7	7.2	5.8
Q_{TRU}	11.0	8.1	5.8	5.8	$Q_{r\text{TRU}}$	7.3	6.9	4.2	3.9
					S_r	7.3	6.9	4.2	3.9
				$T = 250, n = 500$					
Q_{BAR}	31.1	21.1	15.0	12.2	$Q_{r\text{BAR}}$	23.5	14.2	10.0	9.0
Q_{PAR}	37.7	29.4	17.6	14.1	$Q_{r\text{PAR}}$	27.0	18.8	11.0	9.2
Q_{QS}	30.2	18.2	12.0	8.5	$Q_{r\text{QS}}$	20.4	11.7	8.0	7.4
Q_{DAN}	30.7	18.3	12.0	8.1	$Q_{r\text{DAN}}$	21.4	11.9	8.4	7.6
Q_{TRU}	17.5	11.0	6.8	5.9	$Q_{r\text{TRU}}$	10.7	6.8	6.5	5.9
					S_r	10.7	6.8	6.5	5.9
				$T = 500, n = 250$					
Q_{BAR}	24.5	17.8	13.7	11.8	$Q_{r\text{BAR}}$	17.7	9.7	6.5	6.3
Q_{PAR}	31.4	20.6	14.6	12.4	$Q_{r\text{PAR}}$	24.5	14.1	6.6	6.2
Q_{QS}	21.4	14.6	11.1	10.3	$Q_{r\text{QS}}$	15.7	6.9	6.1	5.6
Q_{DAN}	23.0	15.6	8.9	8.3	$Q_{r\text{DAN}}$	15.6	7.0	6.0	5.7
Q_{TRU}	12.8	8.6	9.1	8.7	$Q_{r\text{TRU}}$	6.0	5.9	5.4	4.6
					S_r	6.0	5.9	5.4	4.6
				$T = 500, n = 500$					
Q_{BAR}	39.9	32.3	23.3	20.4	$Q_{r\text{BAR}}$	18.6	12.0	10.0	8.7
Q_{PAR}	42.4	36.5	26.4	22.2	$Q_{r\text{PAR}}$	24.4	16.8	10.5	9.4
Q_{QS}	38.0	27.3	18.9	17.4	$Q_{r\text{QS}}$	17.6	10.7	8.8	6.7
Q_{DAN}	37.9	28.8	20.4	13.9	$Q_{r\text{DAN}}$	17.1	10.4	8.4	7.1
Q_{TRU}	23.9	16.4	12.1	11.6	$Q_{r\text{TRU}}$	10.5	8.0	5.8	5.7
					S_r	10.5	8.0	5.8	5.7

TABLE 5

Rejection rates under experiment ALTER4

This table reports experimental rejection rates (in %, size-adjusted) of Q, Q_r and S_r, at the 5% significance level. The subscripts BAR, DAN, PAR, QS and TRU denote, respectively, the Bartlett kernel, the Daniell kernel, the Parzen kernel, the Quadratic-Spectral kernel and the Truncated Uniform kernel. The integer m denotes the kernel parameter. Number of replications = 1000.

m	5	10	20	30	m	5	10	20	30
					$T = 250$, $n = 250$				
Q_{BAR}	36.3	33.4	31.0	26.5	$Q_{r\text{BAR}}$	20.8	20.7	21.0	17.9
Q_{PAR}	29.0	34.4	30.7	30.3	$Q_{r\text{PAR}}$	16.1	21.8	21.7	20.2
Q_{QS}	35.9	30.8	28.5	26.1	$Q_{r\text{QS}}$	22.8	21.5	20.0	15.8
Q_{DAN}	37.3	32.6	27.7	26.2	$Q_{r\text{DAN}}$	23.4	21.0	18.9	16.8
Q_{TRU}	30.4	25.7	21.7	18.5	$Q_{r\text{TRU}}$	20.5	17.9	13.8	13.2
					S_r	20.5	17.9	13.8	13.2
					$T = 250$, $n = 500$				
Q_{BAR}	57.8	52.3	43.0	37.3	$Q_{r\text{BAR}}$	29.1	26.5	22.4	20.7
Q_{PAR}	52.2	59.0	47.9	42.3	$Q_{r\text{PAR}}$	19.5	29.4	24.3	21.7
Q_{QS}	58.5	49.6	38.5	31.6	$Q_{r\text{QS}}$	30.3	25.2	20.0	17.3
Q_{DAN}	58.6	49.6	39.0	32.0	$Q_{r\text{DAN}}$	30.1	25.1	20.6	17.1
Q_{TRU}	49.2	37.5	26.8	25.2	$Q_{r\text{TRU}}$	24.6	18.1	14.1	11.9
					S_r	24.6	18.1	14.1	11.9
					$T = 500$, $n = 250$				
Q_{BAR}	37.4	37.1	35.8	34.2	$Q_{r\text{BAR}}$	25.8	24.9	21.5	19.8
Q_{PAR}	32.1	39.1	36.5	35.9	$Q_{r\text{PAR}}$	17.4	27.7	23.0	21.8
Q_{QS}	37.9	37.1	33.9	31.1	$Q_{r\text{QS}}$	27.7	23.8	20.3	18.6
Q_{DAN}	38.6	37.4	31.8	28.2	$Q_{r\text{DAN}}$	26.7	23.5	20.3	18.2
Q_{TRU}	35.2	29.1	26.3	21.6	$Q_{r\text{TRU}}$	21.6	18.7	14.3	12.2
					S_r	21.6	18.7	14.3	12.2
					$T = 500$, $n = 500$				
Q_{BAR}	58.5	60.5	55.2	51.6	$Q_{r\text{BAR}}$	28.9	31.9	30.0	27.6
Q_{PAR}	47.4	60.3	58.4	54.8	$Q_{r\text{PAR}}$	18.7	32.7	32.5	29.9
Q_{QS}	60.5	58.8	51.7	48.7	$Q_{r\text{QS}}$	32.4	33.1	29.2	23.9
Q_{DAN}	59.3	59.3	53.1	43.4	$Q_{r\text{DAN}}$	31.1	32.4	28.8	24.5
Q_{TRU}	55.4	49.1	37.5	33.2	$Q_{r\text{TRU}}$	32.6	26.6	19.0	16.9
					S_r	32.6	26.6	19.0	16.9

FIGURE 1

Empirical rejection rates

This figure plots on a monthly basis the percentage of institutions with inadequate systemic risk forecasts detected using the Q backtest (depicted in blue), the Q_r backtest (depicted in black), and the S_r backtest (depicted in red). The out-sample size is denoted by n. Backtesting significance level is 5% with Bonferroni correction applied for every monthly analysis. The Bartlett kernel is used for Q and Q_r with kernel parameter m.

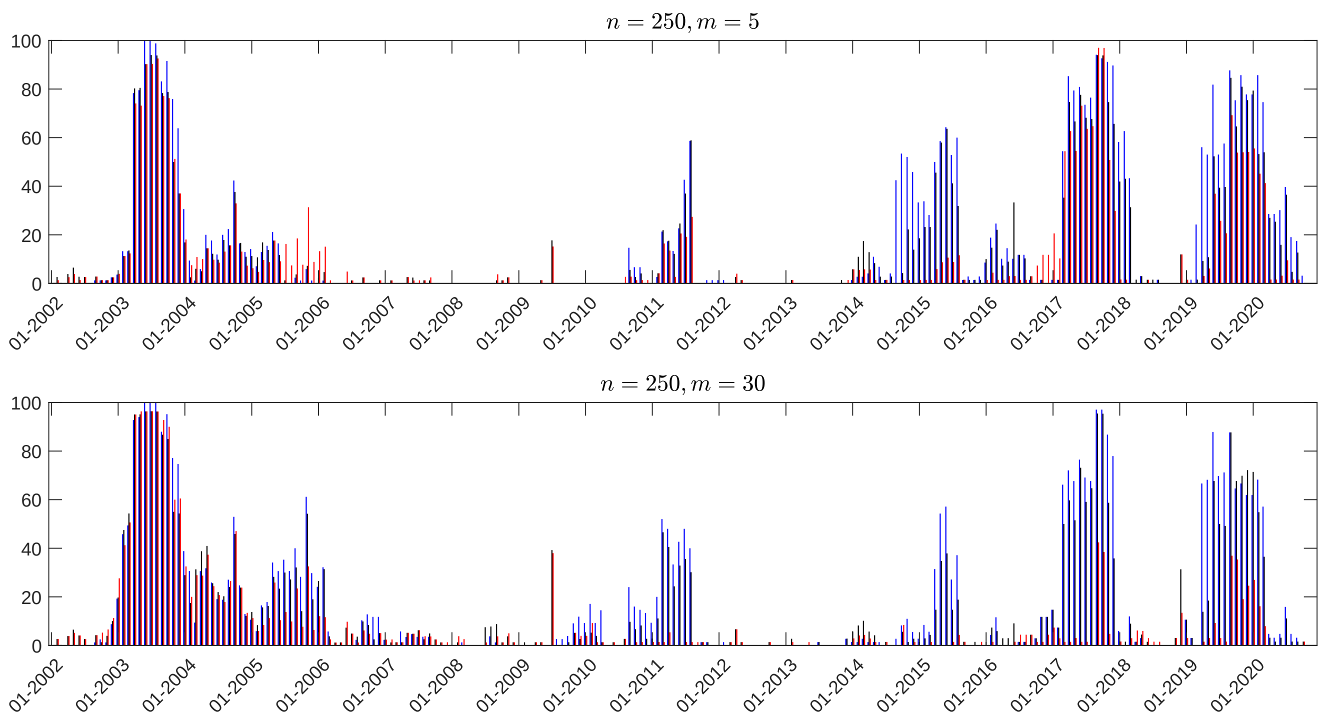


FIGURE 2

Empirical rejection rates

This figure plots on a monthly basis the percentage of institutions with inadequate systemic risk forecasts detected using the Q backtest (depicted in blue), the Q_r backtest (depicted in black), and the S_r backtest (depicted in red). The out-sample size is denoted by n. Backtesting significance level is 5% with Bonferroni correction applied for every monthly analysis. The Bartlett kernel is used for Q and Q_r with kernel parameter m.

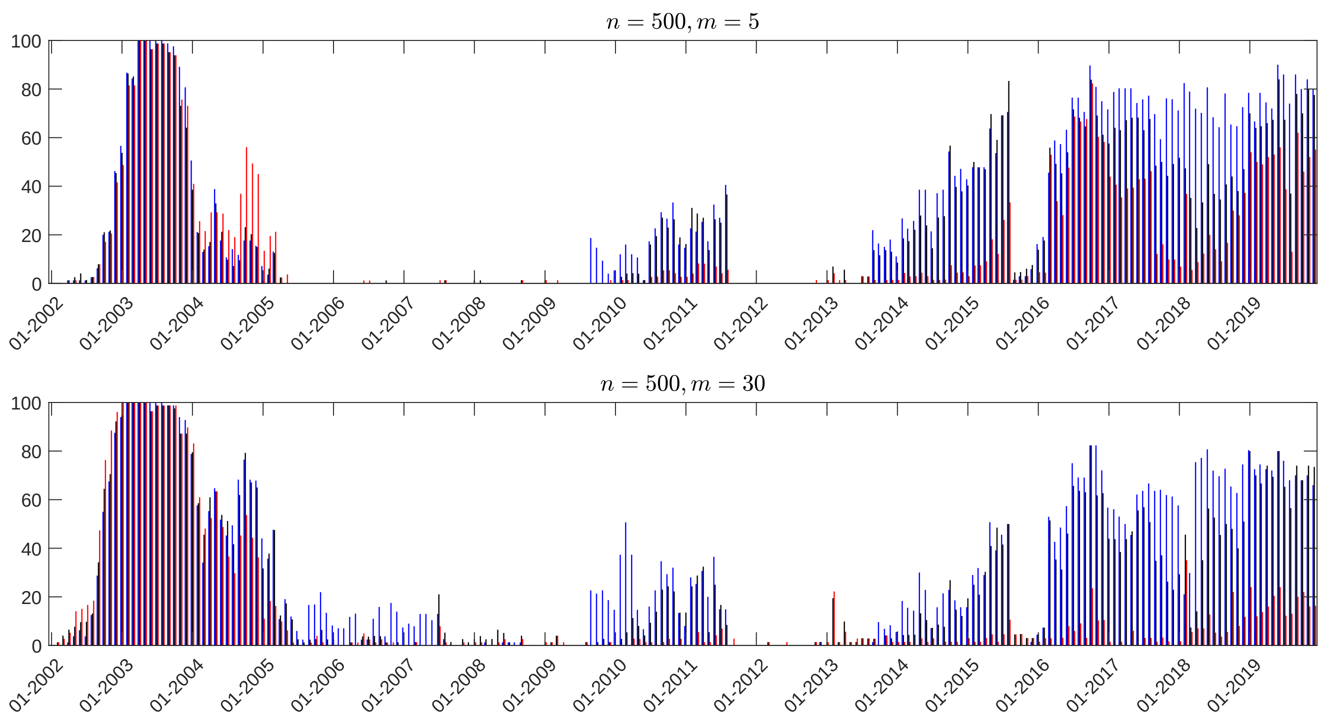


FIGURE 3

Empirical rejection rates

This figure plots on a monthly basis the percentage of institutions with inadequate systemic risk forecasts detected using the Q backtest (depicted in blue), the Q_r backtest (depicted in black), and the S_r backtest (depicted in red). The out-sample size is denoted by n. Backtesting significance level is 5% with Bonferroni correction applied for every monthly analysis. The Quadratic-Spectral kernel is used for Q and Q_r with kernel parameter m.

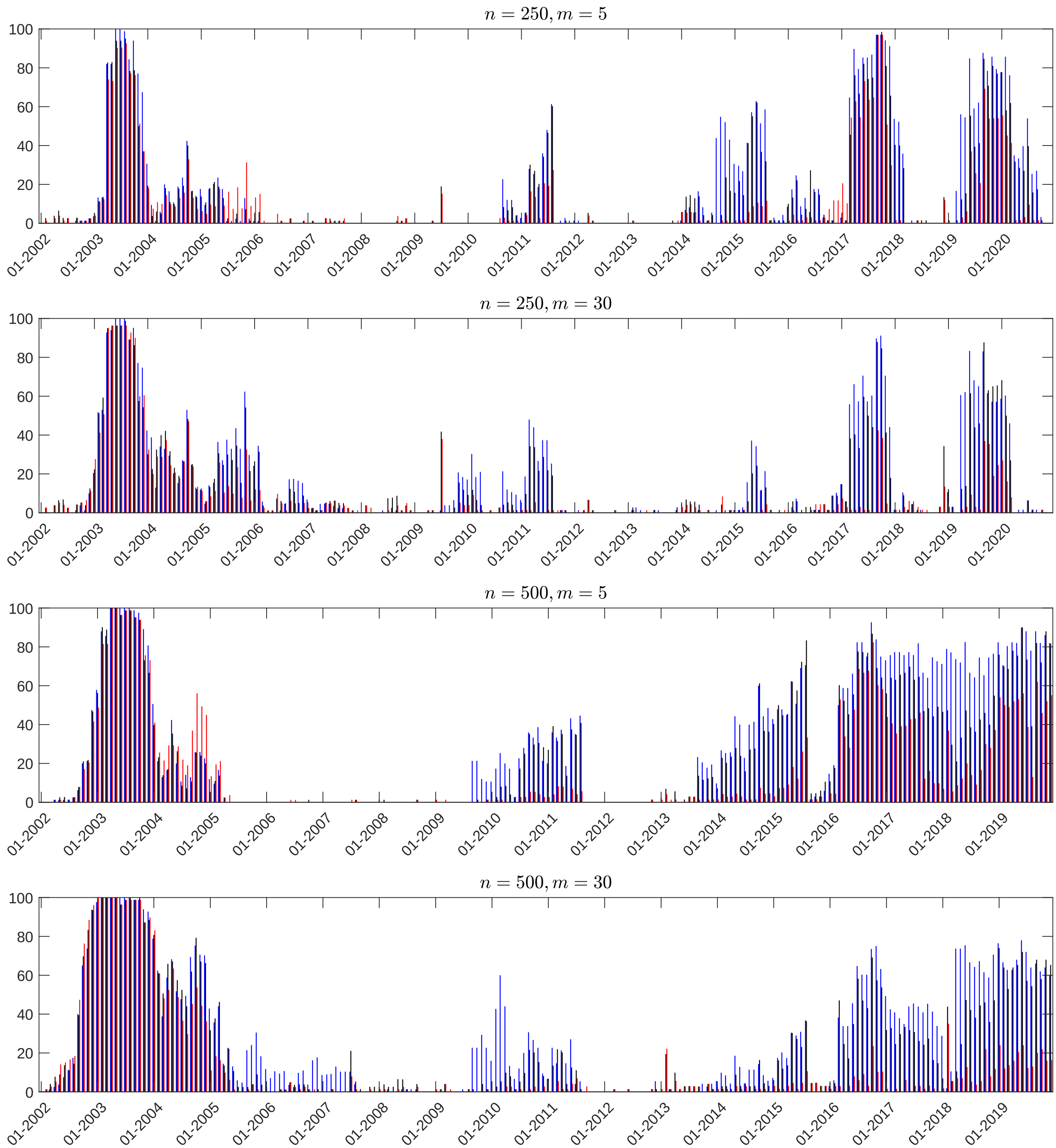


FIGURE 4

Empirical rejection rates

This figure plots on a monthly basis the percentage of institutions with inadequate systemic risk forecasts detected using the Q backtest (depicted in blue), the Q_r backtest (depicted in black), and the S_r backtest (depicted in red). The out-sample size is denoted by n. Backtesting significance level is 5% with Bonferroni correction applied for every monthly analysis. The Bartlett kernel is used for Q and Q_r with kernel parameter m. Q_r and S_r are computed using $\hat{\rho}_{nrm}(j)$.

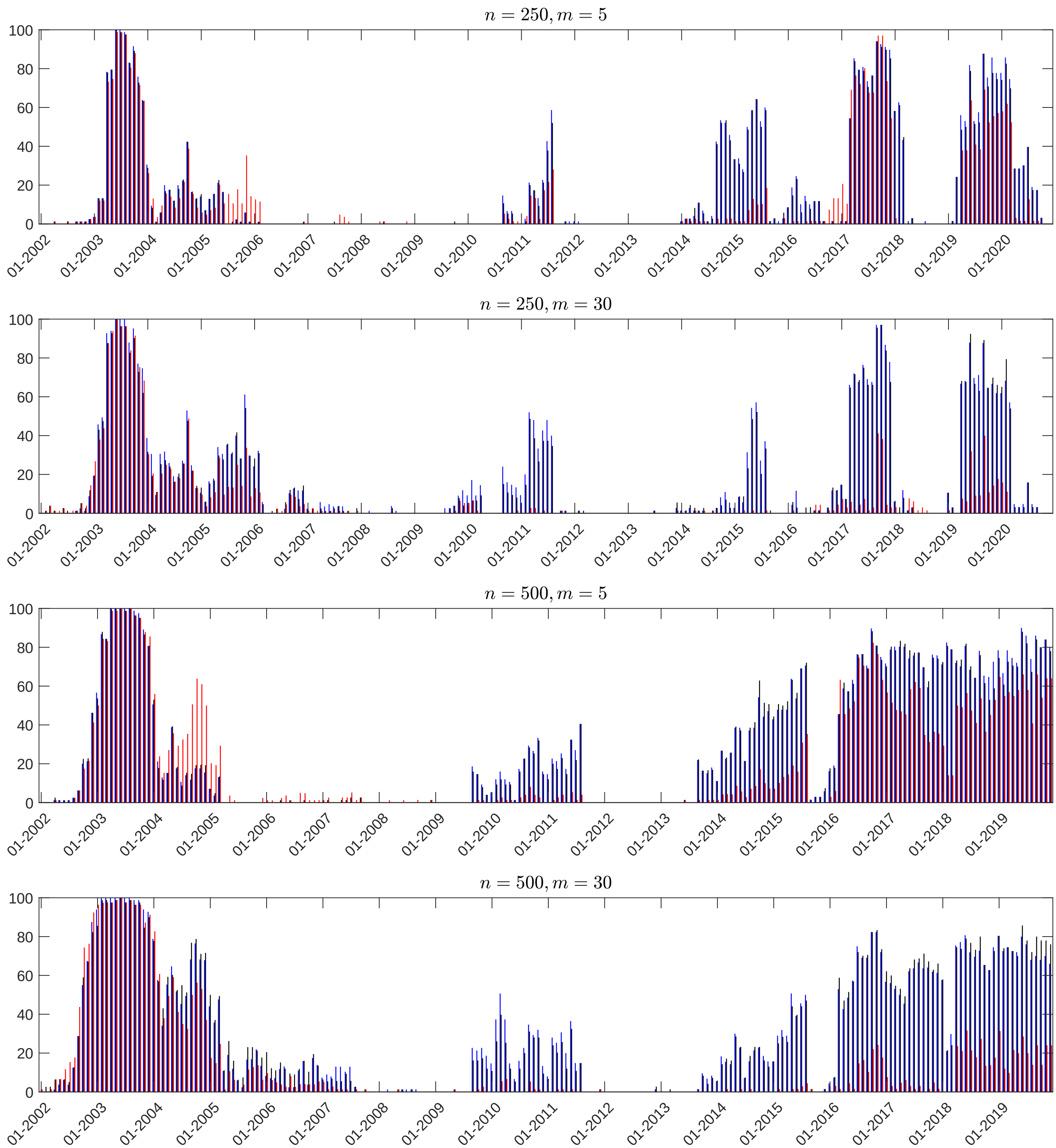
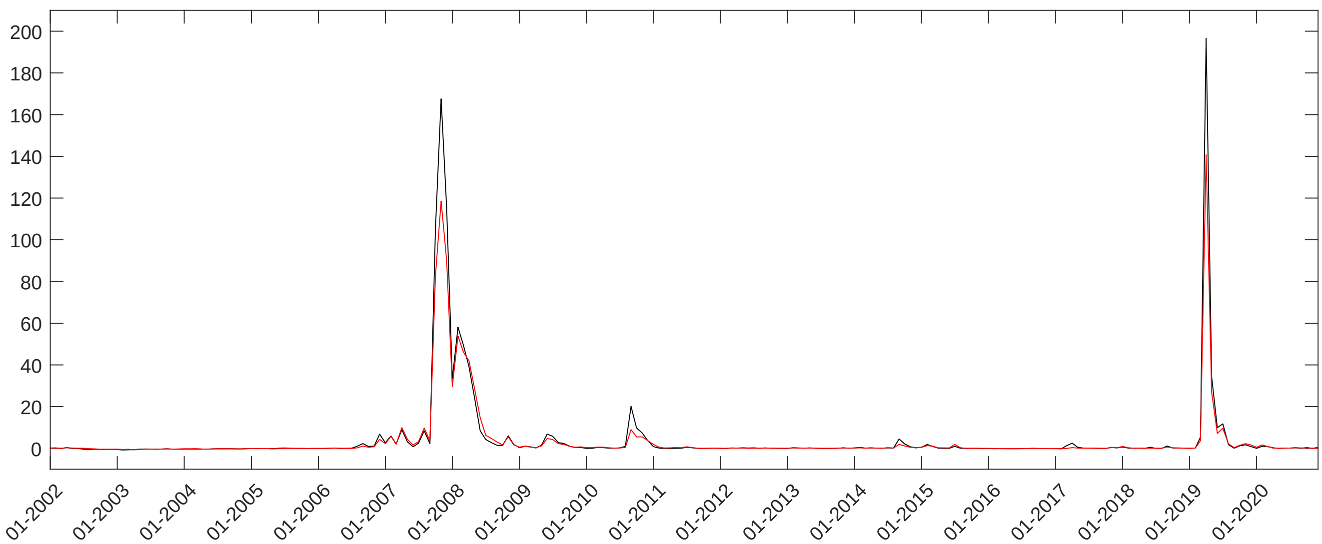


FIGURE 5

System warning indicator

This figure plots the $\overline{EWI}_t(\alpha, \alpha_c)$ indicator on a monthly basis. The black (red) line is based on α_c obtained by solving the quadratic program (15) using the Bartlett (Truncated Uniform) kernel.



Online Appendix
“Consistent Backtesting Systemic Risk Measures”

Soon Leong

October 2025

Summary and notations

The mathematical proof sketched here is inspired by Hong (1996) and Banulescu-Radu, Hurlin, Leymarie and Scaillet (2021). The main added technicality with respect to Hong (1996) is treating the impacts of estimation risk in our backtests, whereas that with respect to Banulescu-Radu et al. (2021) is showing that both the new backtests are fully consistent. Throughout the appendices, the following notations are used. The Euclidean norm is denoted using $||\cdot||$. The notations O_p and o_p are the usual order in probability notations. The scalar c denotes a positive finite generic constant that may differ from place to place. Let $(\rho_n(j), \gamma_n(j))$ be defined as $(\hat{\rho}_n(j), \hat{\gamma}_n(j))$ with $\hat{\boldsymbol{\theta}}_T$ replaced by $\boldsymbol{\theta}_0$.

Appendix A Assumptions

Assumption 1. The stochastic process $\boldsymbol{Y}_t = (Y_{1,t}, Y_{2,t})'$ is strictly stationary and ergodic, with unknown twice continuously differentiable conditional distribution function $F_Y(\boldsymbol{y}|\Omega_{t-1}) = \mathbb{P}(Y_{1,t} \leq y_1, Y_{2,t} \leq y_2|\Omega_{t-1})$, where $\boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^2$.

Assumption 2. The estimator $\hat{\boldsymbol{\theta}}_T$ is consistent for $\boldsymbol{\theta}_0$, that is $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a positive definite bounded matrix. Besides, there exists $\hat{\boldsymbol{\Sigma}}$ such that $\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}} = o_p(1)$.

Assumption 3. The quantity $\boldsymbol{R}(j)$ is summable with bounded smoothing estimator $\boldsymbol{R}_n(j, \boldsymbol{\theta})$ that is consistent at the optimal rate $n^{-2/5}$, and $\boldsymbol{R}_n(j, \hat{\boldsymbol{\theta}}_T) = \hat{\boldsymbol{R}}(j)$. That is, (i) $\sum_{j=0}^{\infty} \|\boldsymbol{R}(j)\| < \infty$, (ii) $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \mathbb{E}\,\|\nabla_{\boldsymbol{\theta}}[\boldsymbol{R}_n(j, \boldsymbol{\theta})]\| = O_p(1),\ 0 \leq j \leq n-1$, and (iii) $\boldsymbol{R}_n(j, \boldsymbol{\theta}_0) - \boldsymbol{R}(j) = O_p(n^{-2/5})$.

Assumption 4. The kernel function $k : \mathbb{R} \to [-1, 1]$ is a symmetric function that is continuous at 0 and all points except a finite number of points on $\mathbb{R}$, with $k(0) = 1$ and $\int_{-\infty}^{\infty} k^2(z)\mathrm{d}z < \infty$.

Assumption 5. The autocorrelation statistics have bounded derivatives. That is, for $0 \leq j \leq n-1$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \mathbb{E}\|\nabla_{\boldsymbol{\theta}}[\rho(j)]\|^2 = O(1)$, and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0} \mathbb{E}\|\nabla_{\boldsymbol{\theta}}[\rho_n(j)] - \nabla_{\boldsymbol{\theta}}[\rho(j)]\|^2 = O(1)$.

Assumption 6. $\{H_t(\alpha,\boldsymbol{\theta}_0)\}$ is a fourth order stationary process with $\sum_{j=1}^{\infty}\gamma^2(j)<\infty$ and $\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|\kappa_1(j,k,l)|<\infty$, where $\kappa_1(j,k,l)$ is the fourth order cumulant of the joint distribution of $\{H_t(\alpha,\boldsymbol{\theta}_0),H_{t+j}(\alpha,\boldsymbol{\theta}_0),H_{t+k}(\alpha,\boldsymbol{\theta}_0),H_{t+l}(\alpha,\boldsymbol{\theta}_0)\}$.

Appendix B Proof of Theorem 1

To begin we recall the Q formula

$$Q=\frac{n\sum_{j=1}^{n-1}k^2(j/m)\hat{\rho}_n^2(j)-\hat{M}_{1n}(k)}{\hat{V}_{1n}(k)^{1/2}},$$

where

$$\hat{M}_{1n}(k)=\sum_{j=1}^{n-1}k^2(j/m)[1+\omega\hat{\boldsymbol{R}}(j)'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{R}}(j)],$$
$$\hat{V}_{1n}(k)=2\sum_{j=1}^{n-1}k^4(j/m)[1+\omega\hat{\boldsymbol{R}}(j)'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{R}}(j)]^2+2\sum_{j=1}^{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n-1}k^2(i/m)k^2(j/m)[\omega\hat{\boldsymbol{R}}(i)'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{R}}(j)]^2.$$

Proof of Theorem 1. Recall the theorem

Theorem 1. Suppose Assumptions 1–4 in the online appendix hold, and let $T\to\infty$, $n\to\infty$, $n/T\to\omega<\infty$, $m\to\infty$, $m/n\to 0$. Then $Q\xrightarrow{d}\mathrm{N}(0,1)$ under $\mathbb{H}_0$.

The theorem can be given by the following lemmas

Lemma B.1. Suppose the conditions of Theorem 1 hold. Then

$$\frac{n\sum_{j=1}^{n-1}k^2(j/m)\left\{\rho_n(j)+(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2-M_{1n}(k)}{V_{1n}(k)^{1/2}}\xrightarrow{d}\mathrm{N}(0,1),$$

where

$$M_{1n}(k)=\sum_{j=1}^{n-1}k^2(j/m)[1+\omega\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)],$$
$$V_{1n}(k)=2\sum_{j=1}^{n-1}k^4(j/m)[1+\omega\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2+2\sum_{j=1}^{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n-1}k^2(i/m)k^2(j/m)[\omega\boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2.$$

Lemma B.2. Suppose the conditions of Theorem 1 hold. Then

$$\frac{n\sum_{j=1}^{n-1} k^2(j/m)\left\{\hat{\rho}_n^2(j) - \left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2\right\}}{V_{1n}(k)^{1/2}} = o_p(1).$$

Lemma B.3. Suppose the conditions of Theorem 1 hold. Then

$$\frac{\sum_{j=1}^{n-1} k^2(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)] - \sum_{j=1}^{n-1} k^2(j/m)[1+\omega \hat{\boldsymbol{R}}(j)'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{R}}(j)]}{V_{1n}(k)^{1/2}} = o_p(1).$$

Lemma B.4. Suppose the conditions of Theorem 1 hold. Then

$$\frac{V_{1n}(k)^{1/2}}{\hat{V}_{1n}(k)^{1/2}} \xrightarrow{p} 1.$$

□

Proof of Lemma B.1. It is useful to first calculate the conditional distribution of $u_{2,t}(\boldsymbol{\theta}_0)$ and $u_{12,t}(\boldsymbol{\theta}_0)$. We recall the expression of $H_t(\alpha, \boldsymbol{\theta}_0)$, and to reduce notational burden dependence on the fixed quantity $\boldsymbol{\theta}_0$ is dropped

$$\begin{aligned} H_t(\alpha, \boldsymbol{\theta}_0) = H_t(\alpha) &= \mathbb{1}[F_{Y_2}(Y_{2,t}|\Omega_{t-1}) \leq \alpha] \times [1 - F_{Y_1|Y_2 \leq VaR_2(\alpha)}(Y_{1,t}|\Omega_{t-1})] \\ &\equiv \mathbb{1}(u_{2,t} \leq \alpha) \times (1 - u_{12,t}). \end{aligned}$$

In the following we calculate the conditional distribution of $u_{2,t}$ and $u_{12,t}$. In particular, we shall show using probability integral transform that

$$u_{2,t}|\Omega_{t-1} \sim \mathrm{U}[0,1], \tag{B.1}$$
$$u_{12,t}|\{\Omega_{t-1}, u_{2,t} \leq \alpha\} \sim \mathrm{U}[0,1]. \tag{B.2}$$

We begin with $u_{2,t}$, which by definition $u_{2,t} = F_{Y_2}(Y_{2,t}|\Omega_{t-1})$. We let $F_{u_2}(\cdot|\Omega_{t-1})$ denote the conditional distribution of $u_{2,t}$ given Ω_{t-1}. We have for any $u \in \mathbb{R}$

$$\begin{aligned} F_{u_2}(u|\Omega_{t-1}) &= \mathbb{P}(u_{2,t} \leq u|\Omega_{t-1}) \\ &= \mathbb{P}(F_{Y_2}(Y_{2,t}|\Omega_{t-1}) \leq u|\Omega_{t-1}) \\ &= \mathbb{P}(Y_{2,t} \leq F_{Y_2}^{-1}(u|\Omega_{t-1})|\Omega_{t-1}) \\ &= F_{Y_2}(F_{Y_2}^{-1}(u|\Omega_{t-1})|\Omega_{t-1}) \\ &= u. \end{aligned}$$

This implies $u_{2,t}|\Omega_{t-1} \sim \mathrm{U}[0,1]$. We now focus on $u_{12,t} = F_{Y_1|Y_2 \leq VaR_2(\alpha)}(Y_{1,t}|\Omega_{t-1})$, which can be written explicitly as $u_{12,t} = F_{Y_1|Y_2 \le VaR_2(\alpha)}(Y_{1,t}|\{Y_{2,t} \le VaR_{2,t}(\alpha), \Omega_{t-1}\}) = F_{Y_1|Y_2 \le VaR_2(\alpha)}(Y_{1,t}|\{u_{2,t} \le \alpha, \Omega_{t-1}\})$. By similar reasoning, we let $F_{u_{12}|u_2 \le \alpha}(\cdot|\{u_{2,t} \le \alpha, \Omega_{t-1}\})$ denote the conditional distribution of $u_{12,t}$ given $u_{2,t} \le \alpha$ and Ω_{t-1}. We have for any $u \in \mathbb{R}$

$$
\begin{aligned}
F_{u_{12}}(u|\{u_{2,t} \le \alpha, \Omega_{t-1}\}) &= \mathbb{P}(u_{12,t} < u|\{u_{2,t} \le \alpha, \Omega_{t-1}\}) \\
&= \mathbb{P}(F_{Y_1|Y_2 \le VaR_2(\alpha)}(Y_{1,t}|\{u_{2,t} \le \alpha, \Omega_{t-1}\}) < u|\{u_{2,t} \le \alpha, \Omega_{t-1}\}) \\
&= \mathbb{P}(Y_{1,t} < F^{-1}_{Y_1|Y_2 \le VaR_2(\alpha)}(u|\{u_{2,t} \le \alpha, \Omega_{t-1}\})|\{u_{2,t} \le \alpha, \Omega_{t-1}\}) \\
&= F_{Y_1|Y_2 \le VaR_2(\alpha)}\left(F^{-1}_{Y_1|Y_2 \le VaR_2(\alpha)}(u|\{u_{2,t} \le \alpha, \Omega_{t-1}\})\Big|\{u_{2,t} \le \alpha, \Omega_{t-1}\}\right) \\
&= u.
\end{aligned}
$$

This implies $u_{12,t}|\{\Omega_{t-1}, u_{2,t} \le \alpha\} \sim \mathrm{U}[0,1]$, and verifies (B.1)–(B.2).

Put $\{\rho_n(j)\}_{j=1}^m$ in the vector $\boldsymbol{\rho}_n^{(m)}$. Under $\mathbb{H}_0$, we have for any (possibly countably infinite) collection of $\boldsymbol{\rho}_n^{(m)}$ that

$$
\boldsymbol{K}_m \sqrt{n} \boldsymbol{\rho}_n^{(m)} \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{K}_m \boldsymbol{K}_m), \tag{B.3}
$$

where $\boldsymbol{K}_m$ is a diagonal matrix collecting the kernel sequence $\{k(j/m)\}_{j=1}^m$. The result follows directly from Proposition 2.9 of Hayashi (2000) that $\sqrt{n}\boldsymbol{\rho}_n^{(m)} \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_m)$ under $\mathbb{H}_0$, where $\boldsymbol{I}_m$ is the identity matrix of order m. The result holds provided (i) $H_t(\alpha)$ can be decomposed as $\varepsilon_t + \mu$, where μ is a constant and ε_t is a martingale difference sequence (mds); (ii) ε_t has constant and finite conditional second moment, that is, $\mathbb{E}(\varepsilon_t^2|\Omega_{t-1}) = \sigma^2 < \infty$. To verify the conditions we let $\mu = \alpha/2$, and we let $\varepsilon_t = (H_t(\alpha) - \alpha/2)$. The proof of (i) completes by noting the mds property of $\{H_t(\alpha) - \alpha/2\}$, that is we have for all t, $\mathbb{E}(\varepsilon_t|\Omega_{t-1}) = \mathbb{E}(H_t(\alpha) - \alpha/2|\Omega_{t-1}) = 0$. The result also implies $\mathbb{E}((H_t(\alpha) - \alpha/2)^2|\Omega_{t-1}) = \mathbb{E}(H_t(\alpha)^2|\Omega_{t-1}) - \alpha^2/4$. Thus, to verify (ii) it suffices to show that $\mathbb{E}(H_t(\alpha)^2|\Omega_{t-1}) = c < \infty$. We write

$$
\begin{aligned}
\mathbb{E}(H_t(\alpha)^2|\Omega_{t-1}) &= \mathbb{E}\left\{\mathbb{E}\left[\mathbb{1}(u_{2,t} \le \alpha) \times (1 - u_{12,t})^2|\{\Omega_{t-1}, u_{2,t} \le \alpha\}\right]|\Omega_{t-1}\right\} \\
&= \mathbb{E}\left\{\mathbb{1}(u_{2,t} \le \alpha)\mathbb{E}\left[(1 - u_{12,t})^2|\{\Omega_{t-1}, u_{2,t} \le \alpha\}\right]|\Omega_{t-1}\right\} \\
&= \alpha/3 < \infty.
\end{aligned} \tag{B.4}
$$

The first equality follows by the tower property that $\mathbb{E}(X|F_1) = \mathbb{E}(\mathbb{E}(X|F_2)|F_1)$ for $F_1 \subseteq F_2$. The last equality makes use of the results in (B.1)–(B.2) that $u_{2,t}|\Omega_{t-1} \sim \mathrm{U}[0,1]$ and $u_{12,t}|\{\Omega_{t-1}, u_{2,t} \le \alpha\} \sim \mathrm{U}[0,1]$. This implies, for instance, that $\mathbb{1}(u_{2,t} \le \alpha)|\Omega_{t-1} \sim$ Bernoulli(α) and that $\mathbb{E}[u_{12,t}^2|\{\Omega_{t-1}, u_{2,t} \le \alpha\}] = 1/3$.

We now evaluate the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]$. To this we first evaluate the probability limit of $\nabla_{\boldsymbol{\theta}}[\rho_n(j)]$. By the quotient rule, we have

$$\nabla_{\boldsymbol{\theta}}[\rho_n(j)] = \gamma_n(0)^{-1}\nabla_{\boldsymbol{\theta}}[\gamma_n(j)] - \gamma_n(0)^{-1}\rho_n(j)\nabla_{\boldsymbol{\theta}}[\gamma_n(0)]. \tag{B.5}$$

The second term in (B.5) is $o_p(1)$ by noting $\rho_n(j) \xrightarrow{p} 0$, which is implied by $\gamma_n(j) \xrightarrow{p} 0$, which in turns follows by Markov's inequality and $\mathbb{E}(\gamma_n(j)) = 0$ under $\mathbb{H}_0$. For the first term, we have the quantity $\gamma_n(0)^{-1} \xrightarrow{p} [\alpha(1/3 - \alpha/4)]^{-1}$, which follows by continuous mapping theorem, Markov's inequality and

$$\begin{aligned}
\mathbb{E}[\gamma_n(0)] &= \mathbb{E}[\mathbb{E}[\gamma_n(0)|\Omega_{t-1}]] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}[(H_t(\alpha) - \alpha/2)^2|\Omega_{t-1}]\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{t=T+1}^{T+n}\mathbb{E}[H_t(\alpha)^2 - \alpha^2/4|\Omega_{t-1}]\right] \\
&= \alpha(1/3 - \alpha/4),
\end{aligned} \tag{B.6}$$

where we make use of the results in (B.4). We now focus on the quantity $\nabla_{\boldsymbol{\theta}}[\gamma_n(j)]$. By the law of iterated expectation, we have $\mathbb{E}[\nabla_{\boldsymbol{\theta}}[\gamma_n(j)]] = \mathbb{E}[\mathbb{E}[\nabla_{\boldsymbol{\theta}}[\gamma_n(j)]|\Omega_{t-1}]]$. Thus we evaluate

$$\begin{aligned}
&\mathbb{E}[\nabla_{\boldsymbol{\theta}}[\gamma_n(j)]|\Omega_{t-1}] \\
&= \frac{1}{n-j}\sum_{t=T+1+j}^{T+n}\mathbb{E}\{\nabla_{\boldsymbol{\theta}}[(H_t(\alpha) - \alpha/2) \times (H_{t-j}(\alpha) - \alpha/2)]|\Omega_{t-1}\} \\
&= \frac{1}{n-j}\sum_{t=T+1+j}^{T+n}\mathbb{E}\{(H_{t-j}(\alpha) - \alpha/2)\nabla_{\boldsymbol{\theta}}[H_t(\alpha)]|\Omega_{t-1}\} + \nabla_{\boldsymbol{\theta}}[H_{t-j}(\alpha)]\mathbb{E}\{(H_t(\alpha) - \alpha/2)|\Omega_{t-1}\} \\
&= \frac{1}{n-j}\sum_{t=T+1+j}^{T+n}\mathbb{E}\{(H_{t-j}(\alpha) - \alpha/2)\nabla_{\boldsymbol{\theta}}[H_t(\alpha)]|\Omega_{t-1}\},
\end{aligned}$$

where the second equality makes use of the product rule, and the last equality follows by $\mathbb{E}\{(H_t(\alpha)-\alpha/2)|\Omega_{t-1}\} = 0$. Therefore, we have $\nabla_{\boldsymbol{\theta}}[\gamma_n(j)] \xrightarrow{p} \frac{1}{n-j}\sum_{t=T+1+j}^{T+n}\mathbb{E}\{(H_{t-j}(\alpha)-\alpha/2)\nabla_{\boldsymbol{\theta}}[H_t(\alpha)]\}$, which follows by Markov's inequality and

$$\mathbb{E}[\nabla_{\boldsymbol{\theta}}[\gamma_n(j)]] = \mathbb{E}[\mathbb{E}[\nabla_{\boldsymbol{\theta}}[\gamma_n(j)]|\Omega_{t-1}]] = \mathbb{E}\{(H_{t-j}(\alpha) - \alpha/2)\nabla_{\boldsymbol{\theta}}[H_t(\alpha)]\}.$$

Therefore we have

$$\nabla_{\boldsymbol{\theta}}[\rho_n(j)] \xrightarrow{p} \frac{1}{\alpha(1/3 - \alpha/4)}\mathbb{E}\{(H_{t-j}(\alpha) - \alpha/2)\nabla_{\boldsymbol{\theta}}[H_t(\alpha)]\} = \boldsymbol{R}(j).$$

Given $\nabla_{\boldsymbol{\theta}}[\rho_n(j)] \xrightarrow{p} \boldsymbol{R}(j)$, $\sqrt{n} \to \sqrt{\omega}\sqrt{T}$, and Assumption 2, we have by Slutsky's theorem

$$k(j/m)\sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)] \xrightarrow{d} \mathrm{N}(0, k^2(j/m)\omega\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)), \tag{B.7}$$

under the null. Since $\nabla_{\boldsymbol{\theta}}[\rho_n(j)]$ converges to the constant $\boldsymbol{R}(j)$, Slutsky's theorem also implies that the statistical randomness of the term $k(j/m)\sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]$ is induced by the $\hat{\boldsymbol{\theta}}_T$ estimate which depends on the in-sample. By contrast, the randomness of $k(j/m)\sqrt{n}\rho_n(j)$ is induced by the out-sample iid generalized errors (see, e.g., Rosenblatt, 1952). Given the statistical independence, we have

$$k(j/m)\sqrt{n}\left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\} \xrightarrow{d} \mathrm{N}(0, k^2(j/m)[1 + \omega\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]). \tag{B.8}$$

Next, we study the dependence structure across time. Although $\{k(j/m)\sqrt{n}\rho_n(j)\}_{j=1}^m$ converges to a sequence of independent random variables as shown in (B.3), that is generally not the case for $\{k(j/m)\sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\}_{j=1}^m$. However, we have the sum of squares

$$n\sum_{j=1}^{m} k^2(j/m)\left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2,$$

which we denote in short by $n\sum_{j=1}^m X_n^2(j) \equiv n\boldsymbol{X}_n^{(m)\prime}\boldsymbol{X}_n^{(m)}$, converging to a sum of independent sequence. This follows by first observing that we have from (B.7)–(B.8)

$$\sqrt{n}\boldsymbol{X}_n^{(m)} \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Delta}_k), \tag{B.9}$$

where the (i,j)-th element of $\boldsymbol{\Delta}_k$ is given by

$$\boldsymbol{\Delta}_k(i,j) = k(i/m)k(j/m)[\delta(i,j) + \omega\boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)], \tag{B.10}$$

with $\delta(i,j)$ being the Kronecker delta. Secondly, because the real symmetric matrix $\boldsymbol{\Delta}_k$ can be decomposed into $\boldsymbol{\Delta}_k = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}'$, where $\boldsymbol{V}$ is an orthogonal matrix and $\boldsymbol{D}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\Delta}_k$, we have

$$\boldsymbol{V}'\sqrt{n}\boldsymbol{X}_n^{(m)} \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{D}). \tag{B.11}$$

The diagonal matrix $\boldsymbol{D}$ shows that $\boldsymbol{V}'\sqrt{n}\boldsymbol{X}_n^{(m)}$ converges to a random vector of independent variables. This implies $n\boldsymbol{X}_n^{(m)\prime}\boldsymbol{X}_n^{(m)} = n\sum_{j=1}^m k^2(j/m)\left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2$, which is a sum of the squared variables in $\boldsymbol{V}'\sqrt{n}\boldsymbol{X}_n^{(m)}$, converges to a sum of variables

$\sum_{j=1}^{m} Z^2(j)$, where $\{Z(j)\}_{j=1}^{m}$ is a sequence of independent Gaussian variables. Let $D(j)$ denote the j-th element of $\boldsymbol{D}$. Now, noting that $\sum_{j=1}^{m} D(j)$ is equal to the trace of $\boldsymbol{\Delta}_k$ and that $\sum_{j=1}^{m} D^2(j)$ is equal to the trace of $\boldsymbol{\Delta}_k \times \boldsymbol{\Delta}_k$, we calculate using the Gaussian moment generating function

$$\mathbb{E}\left[\sum_{j=1}^{m} Z^2(j)\right] = \sum_{j=1}^{m} k^2(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)],$$
$$\mathbb{V}\left[\sum_{j=1}^{m} Z^2(j)\right] = 2\sum_{j=1}^{m} k^4(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2 + 2\sum_{j=1}^{m}\sum_{\substack{i=1\\ i\neq j}}^{m} k^2(i/m)k^2(j/m)[\omega \boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2.$$

Finally, as $m \to \infty$, which implies $n \to \infty$, we have by the Lindeberg central limit theorem that

$$\frac{n\sum_{j=1}^{n-1} k^2(j/m)\left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2 - M_{1n}(k)}{V_{1n}(k)^{1/2}} \xrightarrow{d} \mathrm{N}(0,1),$$

where

$$M_{1n}(k) = \sum_{j=1}^{n-1} k^2(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)],$$
$$V_{1n}(k) = 2\sum_{j=1}^{n-1} k^4(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2 + 2\sum_{j=1}^{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n-1} k^2(i/m)k^2(j/m)[\omega \boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2.$$

This completes the proof of Lemma B.1.

□

Proof of Lemma B.2. Recall Lemma B.2

$$\frac{n\sum_{j=1}^{n-1} k^2(j/m)\left\{\hat{\rho}_n^2(j) - \left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2\right\}}{V_{1n}(k)^{1/2}} = o_p(1).$$

First, we show that the denominator $V_{1n}(k)^{1/2} = O(m^{1/2})$. It suffices to show $V_{1n}(k)/2 = O(m)$. We write

$$\begin{aligned} V_{1n}(k)/2 = &\sum_{j=1}^{n-1} k^4(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2 \\ &+ \sum_{j=1}^{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n-1} k^2(i/m)k^2(j/m)[\omega \boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2. \end{aligned} \tag{B.12}$$

We begin by showing that the first term in (B.12) is $O(m)$. We write

$$\sum_{j=1}^{n-1} k^4(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2$$
$$= \sum_{j=1}^{n-1} k^4(j/m) + 2\omega \sum_{j=1}^{n-1} k^4(j/m)\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j) + \omega^2 \sum_{j=1}^{n-1} k^4(j/m)[\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2. \quad \text{(B.13)}$$

By Assumption 4 and as $m \to \infty$, $n \to \infty$, $m/n \to 0$, we have $m^{-1}\sum_{j=1}^{n-1} k^4(j/m) \to \int_0^\infty k^4(z)\mathrm{d}z < \infty$. Therefore, the first term in (B.13) is $O(m)$. In the following we show that the second term is $O(1)$

$$\begin{aligned}
&2\omega \sum_{j=1}^{n-1} k^4(j/m)\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j) \\
&\leq 2\omega\|\boldsymbol{\Sigma}\| \sum_{j=1}^{n-1} k^4(j/m)\|\boldsymbol{R}(j)\|^2 \\
&= 2\omega\|\boldsymbol{\Sigma}\| \left\{ \sum_{j=1}^{n-1}\|\boldsymbol{R}(j)\|^2 + \sum_{j=1}^{n-1}[k^4(j/m)-1]\|\boldsymbol{R}(j)\|^2 \right\} \\
&= O(1), \qquad \text{(B.14)}
\end{aligned}$$

where we make use of Assumption 2, Assumption 3(i), dominated convergence theorem, $\lim_{m\to\infty} k^4(j/m) - 1 \to 0$ and $|k^4(j/m)-1| \leq 1$. It remains to show that the third term in (B.13) is $O(1)$. By Cauchy-Schwarz inequality we write

$$\begin{aligned}
&\omega^2 \sum_{j=1}^{n-1} k^4(j/m)[\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2 \\
&\leq \omega^2\|\boldsymbol{\Sigma}\|^2 \left\{ \sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2 \right\} \left\{ \sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2 \right\} \\
&= O(1) \left\{ \sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2 \right\}^2 \\
&= O(1) \left\{ \left[\sum_{j=1}^{n-1}\|\boldsymbol{R}(j)\|^2 + \sum_{j=1}^{n-1}[k^2(j/m)-1]\|\boldsymbol{R}(j)\|^2 \right] \right\}^2 \\
&= O(1), \qquad \text{(B.15)}
\end{aligned}$$

where we make use of Assumption 2, Assumption 3(i), dominated convergence theorem, $\lim_{m\to\infty} k^2(j/m) - 1 \to 0$ and $|k^2(j/m)-1| \leq 1$. We shall now show that the second term in (B.12) is $O(m)$. By Cauchy-Schwarz inequality, we write

$$\begin{aligned}
&\sum_{j=1}^{n-1}\sum_{\substack{i=1\\ i\neq j}}^{n-1} k^2(i/m)k^2(j/m)[\omega \boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2\\
&\leq \omega^2 \sum_{j=1}^{n-1}\sum_{i=1}^{n-1} k^2(i/m)k^2(j/m)[\boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2\\
&\leq \omega^2\|\boldsymbol{\Sigma}\| \sum_{j=1}^{n-1}\sum_{i=1}^{n-1} k^2(i/m)\|\boldsymbol{R}(i)\|^2 k^2(j/m)\|\boldsymbol{R}(j)\|^2\\
&= \omega^2\|\boldsymbol{\Sigma}\| \left\{\sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2\right\}\left\{\sum_{i=1}^{n-1} k^2(i/m)\|\boldsymbol{R}(i)\|^2\right\}\\
&= O(1)\left\{\sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2\right\}^2\\
&= O(1),
\end{aligned}$$

where the last equality makes use of the result in (B.15). This verifies $V_{1n}(k)^{1/2} = O(m^{1/2})$.

We now focus on the numerator. We have by Taylor's theorem

$$\hat{\rho}_n(j) = \rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)] + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)' f_1(\hat{\boldsymbol{\theta}}_T),\ 1 \leq j \leq n-1, \tag{B.16}$$

where $f_1(\hat{\boldsymbol{\theta}}_T)$ is such that $\lim_{\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta}_0} f_1(\hat{\boldsymbol{\theta}}_T) = 0$. Given Assumption 2 that $(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) = O_p(T^{-1/2}) = O_p(n^{-1/2})$ and that $f_1(\hat{\boldsymbol{\theta}}_T)$ is bounded by $(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$, we have

$$\hat{\rho}_n(j) = \rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)] + O_p(1/n). \tag{B.17}$$

Multiplying both sides by $\sqrt{n}$ we get

$$\sqrt{n}\hat{\rho}_n(j) = \sqrt{n}\rho_n(j) + \sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)] + O_p(n^{-1/2}). \tag{B.18}$$

Squaring both sides gives

$$\begin{aligned}
&n\hat{\rho}_n^2(j) - n\left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2\\
&= 2\{\sqrt{n}\rho_n(j) + \sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\}O_p(n^{-1/2}) + O_p(n^{-1})\\
&= O_p(n^{-1/2}).
\end{aligned} \tag{B.19}$$

The last equality makes use of (B.8) that for all $1 \leq j \leq n-1$, $\sqrt{n}\rho_n(j) + \sqrt{n}(\hat{\boldsymbol{\theta}}_T -$

$\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)] = O_p(1)$. Therefore we have

$$
\begin{aligned}
& n\sum_{j=1}^{n-1} k^2(j/m)\left\{\hat{\rho}_n^2(j) - \left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2\right\} \\
& \leq m \sup_{1\leq j\leq n-1}\left\{n\hat{\rho}_n^2(j) - n\left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2\right\}\left\{m^{-1}\sum_{j=1}^{n-1} k^2(j/m)\right\} \\
& = O_p(mn^{-1/2}). \qquad \text{(B.20)}
\end{aligned}
$$

The last equality follows by Assumption 4. The proof completes by observing that the expression of Lemma B.2 is $O_p(mn^{-1/2})/O_p(m^{1/2}) = O_p(m^{1/2}/n^{1/2}) = o_p(1)$, where the last equality follows by $m/n \to 0$.

□

Proof of Lemma B.3. Because $V_{1n}(k)^{1/2} = O(m^{1/2})$ as shown in the proof of Lemma B.2, it thus suffices to show

$$
\sum_{j=1}^{n-1} k^2(j/m)[1 + \omega\hat{\boldsymbol{R}}(j)'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{R}}(j)] - \sum_{j=1}^{n-1} k^2(j/m)[1 + \omega\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)] = o_p(m^{1/2}). \qquad \text{(B.21)}
$$

We first work out the consistency of $\hat{\boldsymbol{R}}(j)$ for $\boldsymbol{R}(j)$. In particular we shall show

$$
\|\hat{\boldsymbol{R}}(j) - \boldsymbol{R}(j)\| = O_p(n^{-2/5}). \qquad \text{(B.22)}
$$

By the triangle inequality we have

$$
\begin{aligned}
\|\hat{\boldsymbol{R}}(j) - \boldsymbol{R}(j)\| &= \|\boldsymbol{R}_n(j, \hat{\boldsymbol{\theta}}_T) - \boldsymbol{R}(j)\| \\
&\leq \|\boldsymbol{R}_n(j, \hat{\boldsymbol{\theta}}_T) - \boldsymbol{R}_n(j, \boldsymbol{\theta}_0)\| + \|\boldsymbol{R}_n(j, \boldsymbol{\theta}_0) - \boldsymbol{R}(j)\|. \qquad \text{(B.23)}
\end{aligned}
$$

We write for the first term

$$
\begin{aligned}
\|\boldsymbol{R}_n(j, \hat{\boldsymbol{\theta}}_T) - \boldsymbol{R}_n(j, \boldsymbol{\theta}_0)\| &\leq \sup_{\tilde{\boldsymbol{\theta}}\in\boldsymbol{\Theta}_0}\left\|\nabla_{\boldsymbol{\theta}}[\boldsymbol{R}_n(j, \tilde{\boldsymbol{\theta}})]\right\| \times \left\|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right\| \\
&= O_p(T^{-1/2}) = O_p(n^{-1/2}), \qquad \text{(B.24)}
\end{aligned}
$$

which follows by the mean value theorem, Markov's inequality, Assumption 2 and Assumption 3(ii). Besides, with Assumption 3(iii) the second term in (B.23) is $O_p(n^{-2/5})$. This verifies (B.22).

We now verify (B.21). For $a, b = 1, 2, ..., p$, let $R_a(j)$ ($\hat{R}_a(j)$) denote the a-th element of $\boldsymbol{R}(j)$ ($\hat{\boldsymbol{R}}(j)$), and let S_{ab} ($\hat{S}_{ab}$) denote the (a,b)-th element of $\boldsymbol{\Sigma}$ ($\hat{\boldsymbol{\Sigma}}$). We write

$$
\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/m)[1+\omega\hat{\boldsymbol{R}}(j)'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{R}}(j)] - \sum_{j=1}^{n-1} k^2(j/m)[1+\omega\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)] \\
&= \omega \sum_{j=1}^{n-1} k^2(j/m) \sum_{a=1}^{p}\sum_{b=1}^{p} \hat{R}_a(j)\hat{R}_b(j)\hat{S}_{ab} - R_a(j)R_b(j)S_{ab} \\
&= \omega \sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j)\hat{R}_b(j) - R_a(j)R_b(j))S_{ab} \\
&\quad + \omega \sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{j=1}^{n-1} k^2(j/m)(\hat{S}_{ab} - S_{ab})R_a(j)R_b(j) \\
&\quad + \omega \sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j)\hat{R}_b(j) - R_a(j)R_b(j))(\hat{S}_{ab} - S_{ab}) \\
&= \omega \sum_{a=1}^{p}\sum_{b=1}^{p} S_{ab} \sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j)\hat{R}_b(j) - R_a(j)R_b(j)) \\
&\quad + \omega \sum_{a=1}^{p}\sum_{b=1}^{p} (\hat{S}_{ab} - S_{ab}) \sum_{j=1}^{n-1} k^2(j/m)R_a(j)R_b(j) \\
&\quad + \omega \sum_{a=1}^{p}\sum_{b=1}^{p} (\hat{S}_{ab} - S_{ab}) \sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j)\hat{R}_b(j) - R_a(j)R_b(j)) \\
&= \omega \sum_{a=1}^{p}\sum_{b=1}^{p} \mathcal{A}_1 + \omega \sum_{a=1}^{p}\sum_{b=1}^{p} \mathcal{A}_2 + \omega \sum_{a=1}^{p}\sum_{b=1}^{p} \mathcal{A}_3, \text{ say.}
\end{aligned}
\tag{B.25}
$$

To complete the proof we shall show $\mathcal{A}_i = o_p(m^{1/2})$ for $i = 1,2,3$. We begin with $\mathcal{A}_1$ and $\mathcal{A}_3$. Because by Assumption 2 we have $S_{ab} = O_p(1)$ and $\hat{S}_{ab} - S_{ab} = o_p(1) = O_p(1)$, it suffices to show $\sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j)\hat{R}_b(j) - R_a(j)R_b(j)) = o_p(m^{1/2})$. We have

$$
\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j)\hat{R}_b(j) - R_a(j)R_b(j)) \\
&= \sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j) - R_a(j))(\hat{R}_b(j) - R_b(j)) + \sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j) - R_a(j))R_b(j) \\
&\quad + \sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_b(j) - R_b(j))R_a(j) \\
&= \mathcal{A}_{11} + \mathcal{A}_{12} + \mathcal{A}_{13}, \text{ say.}
\end{aligned}
\tag{B.26}
$$

By Cauchy-Schwarz inequality, we write for $\mathcal{A}_{11}$

$$\begin{aligned}
|\mathcal{A}_{11}| &\leq \left\{\sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j) - R_a(j))^2\right\}^{1/2} \left\{\sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_b(j) - R_b(j))^2\right\}^{1/2} \\
&= O_p(m^{1/2}/n^{2/5}) \times O_p(m^{1/2}/n^{2/5}) \\
&= O_p(m/n^{4/5}) \\
&= O_p(m^{1/2}) \times O_p(m^{1/2}/n^{4/5}) = o_p(m^{1/2}), \qquad \text{(B.27)}
\end{aligned}$$

which makes use of $m/n \to 0$, $\sum_{j=1}^{n-1} k^2(j/m) = O(m)$ and (B.22). We now focus on $\mathcal{A}_{12}$. By Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|\mathcal{A}_{12}| &\leq \left\{\sum_{j=1}^{n-1} k^2(j/m)(\hat{R}_a(j) - R_a(j))^2\right\}^{1/2} \left\{\sum_{j=1}^{n-1} k^2(j/m)R_b^2(j)\right\}^{1/2} \\
&= O_p(m^{1/2}/n^{2/5}) \times \left\{\sum_{j=1}^{n-1} R_b^2(j) + \sum_{j=1}^{n-1}[k^2(j/m) - 1]R_b^2(j)\right\}^{1/2} \\
&= O_p(m^{1/2}/n^{2/5}) \times O_p(1) \\
&= O_p(m^{2/5}/n^{2/5}) \times O_p(m^{1/10}) = o_p(m^{1/2}) \qquad \text{(B.28)}
\end{aligned}$$

where the first equality makes use of the result in (B.27), the second equality follows by Assumption 3(i), dominated convergence theorem, $\lim_{m\to\infty} k^2(j/m) - 1 \to 0$ and $|k^2(j/m) - 1| \leq 1$. By the same reasonings, we also have $\mathcal{A}_{13} = o_p(m^{1/2})$.

We now focus on $\mathcal{A}_2$. Given $\hat{S}_{ab} - S_{ab} = o_p(1)$ it suffices to show $\sum_{j=1}^{n-1} k^2(j/m)R_a(j)R_b(j) = O_p(1)$. We have by Cauchy-Schwarz inequality

$$\begin{aligned}
\sum_{j=1}^{n-1} k^2(j/m)R_a(j)R_b(j) &\leq \left\{\sum_{j=1}^{n-1} k^2(j/m)R_a^2(j)\right\}^{1/2} \left\{\sum_{j=1}^{n-1} k^2(j/m)R_b^2(j)\right\}^{1/2} \\
&= O_p(1), \qquad \text{(B.29)}
\end{aligned}$$

which makes use of the result in (B.28) This completes the proof.

□

Proof of Lemma B.4. It suffices to show

$$\hat{\boldsymbol{R}}(j)'\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{R}}(j) \xrightarrow{p} \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j).$$

By Assumption 2 we have $\hat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}$. The proof completes by the continuous mapping theorem and by noting that we have $\hat{\boldsymbol{R}}(j) \xrightarrow{p} \boldsymbol{R}(j)$ from the result in (B.22).

□

Appendix C Proof of Proposition 1

Recall the formula of Q_r

$$Q_r = \frac{n\sum_{j=1}^{n-1} k^2(j/m)\hat{\rho}_{nr}^2(j) - \sum_{j=1}^{n-1} k^2(j/m)}{\left\{2\sum_{j=1}^{n-1} k^4(j/m)\right\}^{1/2}}.$$

Recall the proposition

Proposition 1. Suppose the conditions of Theorem 1 hold. Then $Q_r \xrightarrow{d} \mathrm{N}(0,1)$ under $\mathbb{H}_0$.

Proof of Proposition 1. Recall that $\hat{\rho}_n(j)$ is the j-th element of $\hat{\boldsymbol{\rho}}_n$ and that $\hat{\rho}_{nr}(j)$ is the j-th element of $\hat{\boldsymbol{\Delta}}^{-1/2}\hat{\boldsymbol{\rho}}_n$. Let $\boldsymbol{r}_n$ collect $\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\}_{j=1}^{n-1}$. Let $\boldsymbol{r}_{nr} \equiv \boldsymbol{\Delta}^{-1/2}\boldsymbol{r}_n$, and let $r_{nr}(j)$ denote the j-th element of $\boldsymbol{r}_{nr}$. The proposition can be given by the following lemmas

Lemma C.1. Suppose the conditions of Proposition 1 hold. Then

$$\frac{n\sum_{j=1}^{n-1} k^2(j/m)r_{nr}^2(j) - \sum_{j=1}^{n-1} k^2(j/m)}{\left\{2\sum_{j=1}^{n-1} k^4(j/m)\right\}^{1/2}} \xrightarrow{d} \mathrm{N}(0,1).$$

Lemma C.2. Suppose the conditions of Proposition 1 hold. Then

$$\frac{n\sum_{j=1}^{n-1} k^2(j/m)\left\{\hat{\rho}_{nr}^2(j) - r_{nr}^2(j)\right\}}{\left\{2\sum_{j=1}^{n-1} k^4(j/m)\right\}^{1/2}} = o_p(1).$$

□

Proof of Lemma C.1. The result is immediate from the proof of Lemma B.1. The key is to note that we have from (B.8)–(B.10) $\sqrt{n}\boldsymbol{r}_n \xrightarrow{d} N(0,\boldsymbol{\Delta})$, where the (i,j)-th element of $\boldsymbol{\Delta}$ is given by $\boldsymbol{\Delta}(i,j) = \delta(i,j) + \omega\boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)$. Repeating the rest of the proof of Lemma B.1 gives the desired result. □

Proof of Lemma C.2. Put $\{k^2(j/m)\}_{j=1}^{n-1}$ in the vector $\boldsymbol{K}_{n2}$. We have from the proof of Lemma B.2 that

$$m^{-1/2}n\hat{\boldsymbol{\rho}}_n'\boldsymbol{K}_{n2}\hat{\boldsymbol{\rho}}_n - m^{-1/2}n\boldsymbol{r}_n'\boldsymbol{K}_{n2}\boldsymbol{r}_n = o_p(1).$$

Next, the proof of Lemma B.4 implies that $\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta} = o_p(1)$. By the continuous mapping theorem

$$m^{-1/2} n \hat{\boldsymbol{\rho}}_n' \hat{\boldsymbol{\Delta}}^{-1/2} \boldsymbol{K}_{n2} \hat{\boldsymbol{\Delta}}^{-1/2} \hat{\boldsymbol{\rho}}_n - m^{-1/2} n \boldsymbol{r}_n' \boldsymbol{\Delta}^{-1/2} \boldsymbol{K}_{n2} \boldsymbol{\Delta}^{-1/2} \boldsymbol{r}_n = o_p(1).$$

Given $\sum_{j=1}^{n-1} k^4(j/m) = O(m)$, we have

$$\frac{n \hat{\boldsymbol{\rho}}_n' \hat{\boldsymbol{\Delta}}^{-1/2} \boldsymbol{K}_{n2} \hat{\boldsymbol{\Delta}}^{-1/2} \hat{\boldsymbol{\rho}}_n - n \boldsymbol{r}_n' \boldsymbol{\Delta}^{-1/2} \boldsymbol{K}_{n2} \boldsymbol{\Delta}^{-1/2} \boldsymbol{r}_n}{\{2 \sum_{j=1}^{n-1} k^4(j/m)\}^{1/2}} = o_p(1).$$

This completes the proof. □

Appendix D Proof of Theorem 2

Recall the result

Theorem 2. Suppose Assumptions 1–6 in the online appendix hold, and let $T \to \infty$, $n \to \infty$, $n/T \to \omega < \infty$, $m \to \infty$, $m/n \to 0$. Then

$$\frac{m^{1/2}}{n} Q \xrightarrow{p} \sum_{j=1}^{\infty} \rho^2(j) \Big/ \left[2 \int_0^{\infty} k^4(z) \mathrm{d}z \right]^{1/2}.$$

Proof of Theorem 2. Given Lemma B.4, to verify the theorem it suffices to show the following lemmas

Lemma D.1. Suppose the conditions of Theorem 2 hold. Then

$$\sum_{j=1}^{n-1} k^2(j/m) \left\{ \rho(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)' \nabla_{\boldsymbol{\theta}} [\rho(j)] \right\}^2 \xrightarrow{p} \sum_{j=1}^{\infty} \rho^2(j).$$

Lemma D.2. Suppose the conditions of Theorem 2 hold. Then

$$\sum_{j=1}^{n-1} k^2(j/m) \hat{\rho}_n^2(j) - \left\{ \rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)' \nabla_{\boldsymbol{\theta}} [\rho_n(j)] \right\}^2 = o_p(1).$$

Lemma D.3. Suppose the conditions of Theorem 2 hold. Then

$$\sum_{j=1}^{n-1} k^2(j/m) \left\{ \rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)' \nabla_{\boldsymbol{\theta}} [\rho_n(j)] \right\}^2 - \left\{ \rho(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)' \nabla_{\boldsymbol{\theta}} [\rho(j)] \right\}^2 = o_p(1).$$

Lemma D.4. Suppose the conditions of Theorem 2 hold. Then

$$n^{-1}\sum_{j=1}^{n-1} k^2(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)] = o_p(1).$$

Lemma D.5. Suppose the conditions of Theorem 2 hold. Then

$$m^{-1}V_{1n}(k) \xrightarrow{p} 2\int_0^\infty k^4(z)\mathrm{d}z.$$

$\square$

Proof of Lemma D.1. We write

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/m)\left\{\rho(j)+(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\right\}^2\\
&=\sum_{j=1}^{n-1} k^2(j/m)\rho^2(j)+\sum_{j=1}^{n-1} k^2(j/m)\{(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\}^2\\
&\quad+2\sum_{j=1}^{n-1} k^2(j/m)\rho(j)(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\\
&=\mathcal{B}_{1n}+\mathcal{B}_{2n}+\mathcal{B}_{3n},\ \text{say.}
\end{aligned}$$

To complete the proof we shall show $\mathcal{B}_{1n} \xrightarrow{p} \sum_{j=1}^{\infty}\rho^2(j)$, $\mathcal{B}_{2n} \xrightarrow{p} 0$ and $\mathcal{B}_{3n} \xrightarrow{p} 0$. We begin with $\mathcal{B}_{2n}$. By Cauchy-Schwarz inequality we have

$$\begin{aligned}
\mathcal{B}_{2n} &= \sum_{j=1}^{n-1} k^2(j/m)\{(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\}^2\\
&\le \|\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\|^2\sum_{j=1}^{n-1} k^2(j/m)\|\nabla_{\boldsymbol{\theta}}[\rho(j)]\|^2\\
&= O_p(m/T)=o_p(1).
\end{aligned}$$

The last equality follows by Markov's inequality, Assumptions 2, 4 and 5. We now focus on $\mathcal{B}_{1n}$

$$\begin{aligned}\mathcal{B}_{1n} &= \sum_{j=1}^{n-1} k^2(j/m)\rho^2(j) \\ &= \sum_{j=1}^{\infty}\rho^2(j) - \sum_{j=n}^{\infty}\rho^2(j) + \sum_{j=1}^{n-1}[k^2(j/m)-1]\rho^2(j) \\ &\xrightarrow{p} \sum_{j=1}^{\infty}\rho^2(j).\end{aligned}$$

The result follows by noting the summability condition in Assumption 6, which implies $\sum_{j=1}^{\infty}\rho^2(j) = \gamma^{-2}(0)\sum_{j=1}^{\infty}\gamma^2(j) < \infty$. This implies that the second term goes to zero, and that together with dominated convergence theorem, $\lim_{m\to\infty} k^2(j/m) - 1 \to 0$ and $|k^2(j/m)-1| \leq 1$, the third term goes to zero.

Finally, we write for $\mathcal{B}_{3n}$

$$\begin{aligned}\mathcal{B}_{3n} &= 2\sum_{j=1}^{n-1} k^2(j/m)\rho(j)(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)] \\ &\leq 2\left\{\sum_{j=1}^{n-1} k^2(j/m)\rho^2(j)\right\}^{1/2} \times \left\{\sum_{j=1}^{n-1} k^2(j/m)\{(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\}^2\right\}^{1/2} \\ &= 2\mathcal{B}_{1n}^{1/2} \times \mathcal{B}_{2n}^{1/2} \\ &= O(1) \times o_p(1) = o_p(1).\end{aligned}$$

This completes the proof.

□

Proof of Lemma D.2. By reasonings similar to that in the proof of Lemma B.2, we have for $1 \leq j \leq n-1$

$$\begin{aligned}&n\hat{\rho}_n^2(j) - n\left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2 \\ &= 2\{\sqrt{n}\rho_n(j) + \sqrt{n}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\}O_p(n^{-1/2}) + O_p(n^{-1}) \\ &= O_p(1), \end{aligned} \tag{D.1}$$

which implies

$$\hat{\rho}_n^2(j) - \left\{\rho_n(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2 = O_p(1/n). \tag{D.2}$$

Therefore we have

$$
\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/m)\left\{\hat{\rho}_n^2(j)-\left\{\rho_n(j)+(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2\right\}\\
&\leq m\sup_{1\leq j\leq n-1}\left\{\hat{\rho}_n^2(j)-\left\{\rho_n(j)+(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2\right\}\left\{m^{-1}\sum_{j=1}^{n-1}k^2(j/m)\right\}\\
&=O_p(m/n)=o_p(1). \qquad \text{(D.3)}
\end{aligned}
$$

This completes the proof.

□

Proof of Lemma D.3. We write

$$
\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/m)\left\{\rho_n(j)+(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho_n(j)]\right\}^2-\left\{\rho(j)+(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\right\}^2\\
&=\sum_{j=1}^{n-1}k^2(j/m)[\rho_n^2(j)-\rho^2(j)]+\sum_{j=1}^{n-1}k^2(j/m)\{(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'[\nabla_{\boldsymbol{\theta}}[\rho_n(j)]-\nabla_{\boldsymbol{\theta}}[\rho(j)]]\}^2\\
&\quad+2\sum_{j=1}^{n-1}k^2(j/m)\rho(j)(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'[\nabla_{\boldsymbol{\theta}}[\rho_n(j)]-\nabla_{\boldsymbol{\theta}}[\rho(j)]]\\
&=\mathcal{C}_{1n}+\mathcal{C}_{2n}+\mathcal{C}_{3n},\ \text{say.}
\end{aligned}
$$

To complete the proof we shall show $\mathcal{C}_{in}\xrightarrow{p}0$, for $i=1,2,3$. We begin with $\mathcal{C}_{2n}$. By Cauchy-Schwarz inequality we have

$$
\begin{aligned}
\mathcal{C}_{2n}&=\sum_{j=1}^{n-1}k^2(j/m)\{(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)'[\nabla_{\boldsymbol{\theta}}[\rho_n(j)]-\nabla_{\boldsymbol{\theta}}[\rho(j)]]\}^2\\
&\leq\|\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\|^2\sum_{j=1}^{n-1}k^2(j/m)\|\nabla_{\boldsymbol{\theta}}[\rho_n(j)]-\nabla_{\boldsymbol{\theta}}[\rho(j)]\|^2\\
&=O_p(m/T)=o_p(1).
\end{aligned}
$$

The last equality follows by Markov's inequality, Assumptions 2, 4 and 5. We now focus on $\mathcal{C}_{1n}$

$$
\begin{aligned}
|\mathcal{C}_{1n}|&=\left|\sum_{j=1}^{n-1}k^2(j/m)\rho^2(j)-\sum_{j=1}^{n-1}k^2(j/m)\rho_n^2(j)\right|\\
&=\left|\gamma^{-2}(0)\sum_{j=1}^{n-1}k^2(j/m)\gamma^2(j)-\gamma_n^{-2}(0)\sum_{j=1}^{n-1}k^2(j/m)\gamma_n^2(j)\right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left|\gamma^{-2}(0) - \gamma_n^{-2}(0)\right| \sum_{j=1}^{n-1} k^2(j/m)\gamma^2(j) + \gamma_n^{-2}(0) \left\{ \sum_{j=1}^{n-1} k^2(j/m) \left|\gamma^2(j) - \gamma_n^2(j)\right| \right\} \\
&= \left|\gamma^{-2}(0) - \gamma_n^{-2}(0)\right| \mathcal{C}_{11n} + \gamma_n^{-2}(0)\mathcal{C}_{12n}, \text{ say.}
\end{aligned}
$$

We have $|\gamma^{-2}(0) - \gamma_n^{-2}(0)|\,\mathcal{C}_{11n} = o_p(1)$, which follows by Markov's inequality, $\mathbb{E}[\gamma(0) - \gamma_n(0)] = 0$, and $\mathcal{C}_{11n} = \sum_{j=1}^{n-1} k^2(j/m)\gamma^2(j) \xrightarrow{p} \sum_{j=1}^{\infty} \gamma^2(j) \leq \infty$ as implied in the proof of Lemma D.1. Because $\gamma_n(0) = O(1)$, to show $\mathcal{C}_{1n} = o_p(1)$ it remains to show that $\mathcal{C}_{12n}$ vanishes in probability. Since $|\gamma^2(j) - \gamma_n^2(j)| = [\gamma_n(j) - \gamma(j)]^2 + 2\,|\gamma(j)|\,|\gamma_n(j) - \gamma(j)|$, we have by Cauchy-Schwarz inequality

$$
\begin{aligned}
\mathcal{C}_{12n} &= \sum_{j=1}^{n-1} k^2(j/m)[\gamma(j) - \gamma_n(j)]^2 + 2\sum_{j=1}^{n-1} k^2(j/m)\,|\gamma(j)|\,|\gamma_n(j) - \gamma(j)| \\
&\leq \sum_{j=1}^{n-1} k^2(j/m)[\gamma(j) - \gamma_n(j)]^2 + 2\mathcal{C}_{11n}^{1/2} \left\{ \sum_{j=1}^{n-1} k^2(j/m)[\gamma(j) - \gamma_n(j)]^2 \right\}^{1/2}.
\end{aligned}
$$

By observing that we have $\sup_{1\leq j\leq n-1} \mathbb{V}[\gamma_n(j)] \leq cn^{-1}$ given Assumption 6 (see, e.g., Hannan, 1970, p. 209), we therefore obtain $\mathcal{C}_{12n} = O_p(m^{1/2}/n^{1/2}) = o_p(1)$. Finally, given $\sum_{j=1}^{n-1} k^2(j/m)\rho^2(j) \xrightarrow{p} \sum_{j=1}^{\infty} \rho^2(j) \leq \infty$, and $\mathcal{C}_{2n} = o_p(1)$, we have $\mathcal{C}_{3n} = o_p(1)$, since

$$
\mathcal{C}_{3n} \leq 2\mathcal{C}_{2n}^{1/2} \left\{ \sum_{j=1}^{n-1} k^2(j/m)\rho^2(j) \right\}^{1/2}.
$$

This completes the proof.

□

Proof of Lemma D.4. Because $m/n \to 0$, it suffices to show $\sum_{j=1}^{n-1} k^2(j/m)[1+\omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)] = O(m)$. We write

$$
\sum_{j=1}^{n-1} k^2(j/m)[1 + \omega \boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)] = \sum_{j=1}^{n-1} k^2(j/m) + \omega \sum_{j=1}^{n-1} k^2(j/m)\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j). \tag{D.4}
$$

The first term in (D.4) is $O(m)$ by Assumption 4. We write for the second term

$$
\begin{aligned}
&\omega \sum_{j=1}^{n-1} k^2(j/m)\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j) \\
&\leq \omega \|\boldsymbol{\Sigma}\| \sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2
\end{aligned}
$$

$$
\begin{aligned}
&= \omega \|\mathbf{\Sigma}\| \left\{ \sum_{j=1}^{n-1} \|\boldsymbol{R}(j)\|^2 + \sum_{j=1}^{n-1} [k^2(j/m) - 1] \|\boldsymbol{R}(j)\|^2 \right\} \\
&= O(1), \qquad \text{(D.5)}
\end{aligned}
$$

where we make use of Assumption 2, Assumption 3(i), dominated convergence theorem, $\lim_{m\to\infty} k^2(j/m) - 1 \to 0$ and $|k^2(j/m) - 1| \leq 1$. This completes the proof. □

Proof of Lemma D.5. We write

$$
\begin{aligned}
m^{-1} V_{1n}(k) = {} & 2m^{-1} \sum_{j=1}^{n-1} k^4(j/m)[1 + \omega \boldsymbol{R}(j)' \mathbf{\Sigma} \boldsymbol{R}(j)]^2 \\
& + 2m^{-1} \sum_{j=1}^{n-1} \sum_{\substack{i=1 \\ i \neq j}}^{n-1} k^2(i/m) k^2(j/m) [\omega \boldsymbol{R}(i)' \mathbf{\Sigma} \boldsymbol{R}(j)]^2. \qquad \text{(D.6)}
\end{aligned}
$$

To complete the proof we show that the first term converges to $2\int_0^\infty k^4(z)\mathrm{d}z$, and that the second term vanishes. We write for the first term

$$
\begin{aligned}
& 2m^{-1} \sum_{j=1}^{n-1} k^4(j/m)[1 + \omega \boldsymbol{R}(j)' \mathbf{\Sigma} \boldsymbol{R}(j)]^2 \\
& = 2m^{-1} \sum_{j=1}^{n-1} k^4(j/m) + 4m^{-1}\omega \sum_{j=1}^{n-1} k^4(j/m) \boldsymbol{R}(j)' \mathbf{\Sigma} \boldsymbol{R}(j) + 2m^{-1}\omega^2 \sum_{j=1}^{n-1} k^4(j/m) [\boldsymbol{R}(j)' \mathbf{\Sigma} \boldsymbol{R}(j)]^2. \qquad \text{(D.7)}
\end{aligned}
$$

For the first term in (D.7), we have $2m^{-1}\sum_{j=1}^{n-1} k^4(j/m) \to 2\int_0^\infty k^4(z)\mathrm{d}z$ given $m \to \infty$ as $n \to \infty$, $m/n \to 0$. In the following we show that the second and third terms in (D.7) are $o(1)$. We write for the second term

$$
\begin{aligned}
& 4m^{-1}\omega \sum_{j=1}^{n-1} k^4(j/m) \boldsymbol{R}(j)' \mathbf{\Sigma} \boldsymbol{R}(j) \\
& \leq 4m^{-1}\omega \|\mathbf{\Sigma}\| \sum_{j=1}^{n-1} k^4(j/m) \|\boldsymbol{R}(j)\|^2 \\
& = 4m^{-1}\omega \|\mathbf{\Sigma}\| \left\{ \sum_{j=1}^{n-1} \|\boldsymbol{R}(j)\|^2 + \sum_{j=1}^{n-1} [k^4(j/m) - 1] \|\boldsymbol{R}(j)\|^2 \right\} \\
& = O(m^{-1}) = o(1), \qquad \text{(D.8)}
\end{aligned}
$$

where we make use of Assumption 2, Assumption 3(i), dominated convergence theorem, $\lim_{m\to\infty} k^4(j/m) - 1 \to 0$ and $|k^4(j/m) - 1| \leq 1$. By Cauchy-Schwarz inequality we write for the third term in (D.7)

$$
\begin{aligned}
&2m^{-1}\omega^2 \sum_{j=1}^{n-1} k^4(j/m)[\boldsymbol{R}(j)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2 \\
&\leq 2m^{-1}\omega^2 \|\boldsymbol{\Sigma}\|^2 \left\{\sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2\right\}\left\{\sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2\right\} \\
&\leq O(m^{-1})\left\{\sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2\right\}^2 \\
&= O(m^{-1})\left\{\left[\sum_{j=1}^{n-1}\|\boldsymbol{R}(j)\|^2 + \sum_{j=1}^{n-1}[k^2(j/m)-1]\|\boldsymbol{R}(j)\|^2\right]\right\}^2 \\
&= O(m^{-1}) \times O(1) = o(1),
\end{aligned} \tag{D.9}
$$

where we make use of Assumption 2, Assumption 3(i), dominated convergence theorem, $\lim_{m\to\infty} k^2(j/m) - 1 \to 0$ and $|k^2(j/m) - 1| \leq 1$.

To complete the proof it remains to show that the second term in (D.6) is $o(1)$. By Cauchy-Schwarz inequality, we write

$$
\begin{aligned}
&2m^{-1} \sum_{j=1}^{n-1} \sum_{\substack{i=1 \\ i\neq j}}^{n-1} k^2(i/m)k^2(j/m)[\omega\boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2 \\
&\leq 2m^{-1}\omega^2 \sum_{j=1}^{n-1}\sum_{i=1}^{n-1} k^2(i/m)k^2(j/m)[\boldsymbol{R}(i)'\boldsymbol{\Sigma}\boldsymbol{R}(j)]^2 \\
&\leq 2m^{-1}\omega^2\|\boldsymbol{\Sigma}\| \sum_{j=1}^{n-1}\sum_{i=1}^{n-1} k^2(i/m)\|\boldsymbol{R}(i)\|^2 k^2(j/m)\|\boldsymbol{R}(j)\|^2 \\
&= 2m^{-1}\omega^2\|\boldsymbol{\Sigma}\| \left\{\sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2\right\}\left\{\sum_{i=1}^{n-1} k^2(i/m)\|\boldsymbol{R}(i)\|^2\right\} \\
&= O(m^{-1})\left\{\sum_{j=1}^{n-1} k^2(j/m)\|\boldsymbol{R}(j)\|^2\right\}^2 \\
&= o(1),
\end{aligned}
$$

where the last equality makes use of the result in (D.9). This completes the proof.

□

Appendix E Proof of Proposition 2

Proof of Proposition 2. To ease the proof we introduce the following notations. Let $\boldsymbol{r}$ collect $\{\rho(j) + (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\}_{j=1}^{n-1}$. Let $\boldsymbol{\rho}$ collect $\{\rho(j)\}_{j=1}^{n-1}$, let $\boldsymbol{\rho}_{\boldsymbol{\nabla}}$ collect $\{(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)'\nabla_{\boldsymbol{\theta}}[\rho(j)]\}_{j=1}^{n-1}$. Let $\boldsymbol{r}_r \equiv \boldsymbol{\Delta}^{-1/2}\boldsymbol{r}$, and let $r_r(j)$ denote the j-th element of $\boldsymbol{r}_r$. Let $\boldsymbol{\rho}_r \equiv \boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho}$, and let $\rho_r(j)$ denote the j-th element of $\boldsymbol{\rho}_r$. Recall the result

Proposition 2. Suppose the conditions of Theorem 2 hold. Then

$$\frac{m^{1/2}}{n} Q_r \xrightarrow{p} \sum_{j=1}^{\infty} \rho_r^2(j) \Big/ \left[2\int_0^{\infty} k^4(z)\mathrm{d}z\right]^{1/2}.$$

The theorem is given by the following lemmas

Lemma E.1. Suppose the conditions of Proposition 2 hold. Then

$$\sum_{j=1}^{n-1} k^2(j/m) r_r^2(j) \xrightarrow{p} \sum_{j=1}^{\infty} \rho_r^2(j)$$

Lemma E.2. Suppose the conditions of Proposition 2 hold. Then

$$\sum_{j=1}^{n-1} k^2(j/m)\hat{\rho}_{nr}^2(j) - r_{nr}^2(j) = o_p(1)$$

Lemma E.3. Suppose the conditions of Proposition 2 hold. Then

$$\sum_{j=1}^{n-1} k^2(j/m) r_{nr}^2(j) - r_r^2(j) = o_p(1)$$

□

Proof of Lemma E.1. To begin we show $\sum_{j=1}^{n-1} k^2(j/m) r_r^2(j) \xrightarrow{p} \sum_{j=1}^{n-1} k^2(j/m)\rho_r^2(j)$. To that we note $\boldsymbol{\rho}_{\boldsymbol{\nabla}} = O_p(T^{-1/2})$ and $\boldsymbol{\Delta} = O(1)$, which implies $\boldsymbol{r}_r = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho} + o_p(1)$. With this we have

$$\begin{aligned}
&\sum_{j=1}^{n-1} k^2(j/m) r_r^2(j) \\
&= \boldsymbol{r}_r' K_{n2} \boldsymbol{r}_r \\
&= [\boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho} + \boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho}_{\nabla}]' K_{n2} [\boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho} + \boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho}_{\nabla}] \\
&\xrightarrow{p} [\boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho}]' K_{n2} [\boldsymbol{\Delta}^{-1/2}\boldsymbol{\rho}]
\end{aligned}$$

$$= \sum_{j=1}^{n-1} k^2(j/m)\rho_r^2(j).$$

In the following we shall show $\sum_{j=1}^{n-1} k^2(j/m)\rho_r^2(j) \xrightarrow{p} \sum_{j=1}^{\infty}\rho_r^2(j)$. We write

$$\begin{aligned}\sum_{j=1}^{n-1} k^2(j/m)\rho_r^2(j) &= \sum_{j=1}^{\infty}\rho_r^2(j) - \sum_{j=n}^{\infty}\rho_r^2(j) + \sum_{j=1}^{n-1}[k^2(j/m)-1]\rho_r^2(j)\\ &= \mathcal{D}_{1n} - \mathcal{D}_{2n} + \mathcal{D}_{3n}, \text{ say.}\end{aligned}$$

To complete the proof we shall show $\mathcal{D}_{2n} \to 0$ and $\mathcal{D}_{3n} \to 0$. We first show that $\mathcal{D}_{1n} = \sum_{j=1}^{\infty}\rho_r^2(j) < \infty$. Denote by λ_{max} the largest eigenvalues of $\boldsymbol{\Delta}^{-1}$, and we note that $0 < \lambda_{max} \leq 1$ because $\boldsymbol{\Delta}$ is given by the sum of an identity matrix and a positive semidefinite matrix. Therefore we have $\sum_{j=1}^{n-1}\rho_r^2(j) = \boldsymbol{\rho}'\boldsymbol{\Delta}^{-1}\boldsymbol{\rho} \leq \lambda_{max}\|\boldsymbol{\rho}\|^2 \leq \|\boldsymbol{\rho}\|^2 = \boldsymbol{\rho}'\boldsymbol{\rho} = \sum_{j=1}^{n-1}\rho^2(j)$, which, as $n \to \infty$, yields $\sum_{j=1}^{\infty}\rho_r^2(j) \leq \sum_{j=1}^{\infty}\rho^2(j) < \infty$. The last inequality follows from the result of Lemma D.1. This implies $\mathcal{D}_{2n} \to 0$. Finally, we have $\mathcal{D}_{3n} \to 0$ by dominated convergence theorem, $\lim_{m\to\infty} k^2(j/m) - 1 \to 0$, $|k^2(j/m)-1| \leq 1$ and $\mathcal{D}_{1n} < \infty$. This completes the proof.

□

Proof of Lemma E.2. We have shown in Lemma D.2 that

$$\hat{\boldsymbol{\rho}}_n'\boldsymbol{K}_{n2}\hat{\boldsymbol{\rho}}_n - \boldsymbol{r}_n'\boldsymbol{K}_{n2}\boldsymbol{r}_n = o_p(1).$$

We also have $\hat{\boldsymbol{\Delta}} - \boldsymbol{\Delta} = o_p(1)$ from the proof of Lemma B.4. Therefore, by the continuous mapping theorem

$$\hat{\boldsymbol{\rho}}_n'\hat{\boldsymbol{\Delta}}^{-1/2}\boldsymbol{K}_{n2}\hat{\boldsymbol{\Delta}}^{-1/2}\hat{\boldsymbol{\rho}}_n - \boldsymbol{r}_n'\boldsymbol{\Delta}^{-1/2}\boldsymbol{K}_{n2}\boldsymbol{\Delta}^{-1/2}\boldsymbol{r}_n = o_p(1).$$

This completes the proof.

□

Proof of Lemma E.3. The proof is immediate by noting that we have from Lemma D.3

$$\boldsymbol{r}_n'\boldsymbol{K}_{n2}\boldsymbol{r}_n - \boldsymbol{r}'\boldsymbol{K}_{n2}\boldsymbol{r} = o_p(1),$$

which, given $\boldsymbol{\Delta} = O(1)$, implies

$$\boldsymbol{r}_n'\boldsymbol{\Delta}^{-1/2}\boldsymbol{K}_{n2}\boldsymbol{\Delta}^{-1/2}\boldsymbol{r}_n - \boldsymbol{r}'\boldsymbol{\Delta}^{-1/2}\boldsymbol{K}_{n2}\boldsymbol{\Delta}^{-1/2}\boldsymbol{r} = o_p(1).$$

□

Appendix F Computing $\hat{\boldsymbol{R}}(j)$

Recall that $\boldsymbol{R}(j)$ is given by

$$\boldsymbol{R}(j) = \frac{1}{\alpha(1/3 - \alpha/4)} \mathbb{E}\left\{(H_{t-j}(\alpha, \boldsymbol{\theta}_0) - \alpha/2)\nabla_{\boldsymbol{\theta}}[H_t(\alpha, \boldsymbol{\theta}_0)]\right\},$$

and that $H_t(\alpha, \boldsymbol{\theta}_0)$ is given by

$$\begin{aligned} H_t(\alpha, \boldsymbol{\theta}_0) &= \mathbb{1}[F_{Y_2}(Y_{2,t}|\Omega_{t-1}, \boldsymbol{\theta}_0) \leq \alpha] \times [1 - F_{Y_1|Y_2 \leq VaR_2(\alpha)}(Y_{1,t}|\Omega_{t-1}, \boldsymbol{\theta}_0)] \\ &= \mathbb{1}(u_{2,t}(\boldsymbol{\theta}_0) \leq \alpha) \times (1 - u_{12,t}(\boldsymbol{\theta}_0)). \end{aligned}$$

Because the indicator function $\mathbb{1}(u_{2,t}(\boldsymbol{\theta}_0) \leq \alpha)$ in $H_t(\alpha, \boldsymbol{\theta}_0)$ is not differentiable, we follow Banulescu-Radu et al. (2021) to evaluate $\nabla_{\boldsymbol{\theta}}[H_t(\alpha, \boldsymbol{\theta}_0)]$ using $\nabla_{\boldsymbol{\theta}}[H_t^{\oplus}(\alpha, \boldsymbol{\theta}_0)]$, where the quantity $H_t^{\oplus}(\alpha, \boldsymbol{\theta}_0)$ is analogous to $H_t(\alpha, \boldsymbol{\theta}_0)$, but with the indicator function $\mathbb{1}(\cdot)$ replaced by a continuously differentiable estimator. To ensure comparability we use the same consistent smoothing estimator as that given in Banulescu-Radu et al. (2021, Appendix E). Therefore, the consistent estimator for $\boldsymbol{R}(j)$ is given by

$$\hat{\boldsymbol{R}}(j) = \boldsymbol{R}_n(j, \hat{\boldsymbol{\theta}}_T) = \frac{1}{\alpha(1/3 - \alpha/4)} \frac{1}{n-j} \sum_{t=T+1+j}^{T+n} \left\{(H_{t-j}(\alpha, \hat{\boldsymbol{\theta}}_T) - \alpha/2)\nabla_{\boldsymbol{\theta}}[H_t^{\oplus}(\alpha, \hat{\boldsymbol{\theta}}_T)]\right\}.$$

Similarly, $\tilde{\boldsymbol{R}}(j)$ is the analogous counterpart of $\hat{\boldsymbol{R}}(j)$ with true parameter $\boldsymbol{\theta}_0$

$$\tilde{\boldsymbol{R}}(j) = \boldsymbol{R}_n(j, \boldsymbol{\theta}_0) = \frac{1}{\alpha(1/3 - \alpha/4)} \frac{1}{n-j} \sum_{t=T+1+j}^{T+n} \left\{(H_{t-j}(\alpha, \boldsymbol{\theta}_0) - \alpha/2)\nabla_{\boldsymbol{\theta}}[H_t^{\oplus}(\alpha, \boldsymbol{\theta}_0)]\right\}.$$

Finally, we allow as with Banulescu-Radu et al. (2021, Appendix F) a small sample factor $\sqrt{1 - j/n}$ for $\boldsymbol{R}(j)$ and its estimators. This does not affect the asymptotic properties of our backtests.

Appendix G Institution tickers and names

Ticker	Institution name	Ticker	Institution name
ABK	AMBAC FINANCIAL GROUP INC	HIG	HARTFORD FINANCIAL SVCS GRP INC
ACAS	AMERICAN CAPITAL STRATEGIES LTD	HUM	HUMANA INC
AET	AETNA INC NEW	ICE	INTERCONTINENTALEXCHANGE INC
AFL	AFLAC INC	KEY	KEYCORP NEW
AGE	EDWARDS A G INC	LEH	LEHMAN BROTHERS HOLDINGS INC
AIG	AMERICAN INTERNATIONAL GROUP INC	LM	LEGG MASON INC
AIZ	ASSURANT INC	LNC	LINCOLN NATIONAL CORP
ALL	ALLSTATE CORP	LTR	LOEWS CORP
AMP	AMERIPRISE FINANCIAL INC	MA	MASTERCARD INC
AMTD	AMERITRADE HOLDING CORP	MBI	M B I A INC
AOC	AON CORP	MER	MERRILL LYNCH & CO INC
ATH	ANTHEM INC	MET	METLIFE INC
AXP	AMERICAN EXPRESS CO	MI	MARSHALL & ILSLEY CORP
BAC	BANK OF AMERICA CORP	MMC	MARSH & MCLENNAN COS INC
BBT	B B & T CORP	MTB	M & T BANK CORP
BEN	FRANKLIN RESOURCES INC	MWD	MORGAN STANLEY DEAN WITTER & CO
BK	BANK NEW YORK INC	NCC	NATIONAL CITY CORP
BKLY	BERKLEY W R CORP	NMX	NYMEX HOLDINGS INC
BLK	BLACKROCK INC	NTRS	NORTHERN TRUST CORP
BOT	C B O T HOLDINGS INC	NYX	N Y S E GROUP INC
BRK	BERKSHIRE HATHAWAY INC DEL	PBCT	PEOPLES BANK BRIDGEPORT
BSC	BEAR STEARNS COMPANIES INC	PFG	PRINCIPAL FINANCIAL GROUP INC
C	CITIGROUP INC	PGR	PROGRESSIVE CORP OH
CB	CHUBB CORP	PNC	P N C BANK CORP
CBG	C B RICHARD ELLIS GROUP INC	PRU	PRUDENTIAL FINANCIAL INC
CBH	COMMERCE BANCORP INC NJ	QCSB	QUEENS COUNTY BANCORP INC
CBSS	COMPASS BANCSHARES INC	RGBK	REGIONS FINANCIAL CORP
CCR	COUNTRYWIDE CREDIT INDS INC	SAFC	SAFECO CORP
CI	C I G N A CORP	SCH	SCHWAB CHARLES CORP NEW
CINF	CINCINNATI FINANCIAL CORP	SEIC	S E I INVESTMENTS CO
CIT	C I T GROUP INC NEW	SLM	S L M HOLDING CORP
CMA	COMERICA INC	SNV	SYNOVUS FINANCIAL CORP
CMB	CHASE MANHATTAN CORP NEW	SPC	ST PAUL COS INC
CME	CHICAGO MERCANTILE EXCH HLDG INC	STI	SUNTRUST BANKS INC
CNA	C N A FINANCIAL CORP	STT	STATE STREET CORP
COF	CAPITAL ONE FINANCIAL CORP	SV	STILWELL FINANCIAL INC
CVTY	COVENTRY HEALTH CARE INC	SVRN	SOVEREIGN BANCORP INC
EGRP	E TRADE GROUP INC	TMK	TORCHMARK CORP
FHS	FOUNDATION HEALTH SYSTEMS INC	TROW	T ROWE PRICE ASSOC INC
FITB	FIFTH THIRD BANCORP	UB	UNIONBANCAL CORP
FNF	FIDELITY NATIONAL FINANCIAL INC	UNH	UNITED HEALTHCARE CORP
FNM	FEDERAL NATIONAL MORTGAGE ASSN	UNM	UNUMPROVIDENT CORP
FRE	FEDERAL HOME LOAN MORTGAGE CORP	USB	U S BANCORP DEL
FTU	FIRST UNION CORP	WFC	WELLS FARGO & CO NEW
GNW	GENWORTH FINANCIAL INC	WM	WASHINGTON MUTUAL INC
GS	GOLDMAN SACHS GROUP INC	WU	WESTERN UNION CO
HBAN	HUNTINGTON BANCSHARES INC	ZION	ZIONS BANCORPORATION
HCBK	HUDSON CITY BANCORP INC		

References

Banulescu-Radu, D., Hurlin, C., Leymarie, J., Scaillet, O., 2021. Backtesting marginal expected shortfall and related systemic risk measures. Management Science 67, 5730–5754.

Hannan, E.J., 1970. *Multiple Time Series*. John Wiley and Sons. New York.

Hayashi, F., 2000. Econometrics. Princeton University Press.

Hong, Y., 1996. Consistent testing for serial correlation of unknown form. Econometrica 64, 837–864.

Rosenblatt, M., 1952. Remarks on a multivariate transformation. Annals of Mathematical Statistics 23, 470–472.