Inference on Dynamic Systemic Risk Measures

Christian Francq and Jean-Michel Zakoïan*

CREST, Palaiseau, France

December 21, 2024

Abstract

Systemic risk measures (SRM) quantify the risk of a system induced by the possible distress of any of its components. Applications in economics and finance are numerous. We define a general dynamic framework for the risk factors, allowing us to obtain explicit expressions of the corresponding dynamic SRM. We deduce an easy-to-implement statistical approach which, based on semi-parametric assumptions, reduces to estimating univariate location-scale models and to computing (static) quantiles of the residuals. We derive a sound asymptotic theory (including confidence intervals, tests, validity of a residual bootstrap) for major SRM, namely the Conditional VaR (CoVaR) and Delta-CoVaR. Our theoretical results are illustrated via Monte-Carlo experiments and real financial and macroeconomic time series.

Keywords: CoVaR, Delta-CoVaR, Marginal expected shortfall, Multivariate risks, Residual bootstrap

JEL Codes: C32, C52, C58

*Corresponding author. E-mail: zakoian@ensae.fr. .

1 Introduction

Systemic risk refers to the risk of an entire system collapsing. The main applications are in Finance and Economics, where the system can correspond to financial institutions or economic entities, but systemic risk concerns many other fields. For example, Haldane and May (2011) draw analogies between banking and ecological systems. In the following, we use financial vocabulary, referring for example to asset returns, which could be replaced by growth rates of macroeconomic series.

The topic of risk often comes to the forefront during periods of major market stress - often called "crises" characterized by a sudden and usually unanticipated loss of asset value, disturbing the long-term trend, and whose consequences can sometimes endanger the very health of a financial institution. The massive losses experienced during the successive crises have refocused attention on tail risks, whose occurrence is rarer but whose consequences are more severe. The best-known measure is the "Value-at-Risk" (VaR) which, together with the "Expected Shortfall" (ES), has gradually gained popularity since the 90s.

However, this univariate approach to risk - which only focuses on the bank/asset under consideration omitting its relationship with other financial institutions/assets - has proven unsatisfactory. This is due to significant co-movements between assets in the same "system", which create a so-called *systemic* risk. The most obvious illustration of this risk was the global financial crisis of 2008 and the risk that a large number of banking organizations would collapse through contagion.[1] Since then, a large number of systemic risk measures (SRM) have emerged, see Giesecke and Kim (2011), Adrian and Brunnermeier (2011, 2016), Acharya et al. (2017), Brownlees and Engle (2017).

Two popular SRM are the Conditional VaR (CoVaR) and ΔCoVaR of Adrian and Brunnermeier (2016). For given risk levels (α, α'), CoVaR is the α'-quantile of an asset (or institution) loss over some horizon (given an information set), conditional on another (reference) asset being distressed, where the distress is defined by reference to the VaR at level α of the latter asset. From a macroscopic viewpoint, the major concern is the distress of the system and the aim is to measure the systemic contribution of particular institutions (taken as the second asset). The reverse micro point of view can also be adopted, in which case the effect of any system distress on a particular institution is what is expected to be measured. ΔCoVaR is the difference between two CoVaRs differing by the conditioning events of the reference asset: distress versus median state. The ΔCoVaR is a real indicator of the potential systemic effect of a factor on the system. It can be interpreted as the incremental VaR for the system caused by an institution being at risk. Alternative risk measures proposed by Acharya, Pedersen, Philippon and Richardson (2017) extend the concept of Expected Shortfall (ES). In particular, the Marginal Expected Shortfall (MES) is defined as the expected loss on an asset or institution's equity conditional on the occurrence of a large loss in the market's return. In the same vein, Brownlees and Engle (2017) introduced the notion of SRISK which measures the capital shortfall of a firm conditional on a severe market decline.

While the study of risk measures originally developed within a static framework, emphasis has been placed in recent years on the estimation of *dynamic* risk measures, as in Francq and Zakoïan (2015) and Patton, Ziegel and Chen (2019). To our knowledge, the estimation of dynamic SRM has not been studied yet, except by Dimitriadis and Hoga (2022). In this article, a joint dynamic model for the VaR and CoVaR is proposed in the spirit of the CAViaR model of Engle and Manganelli (2004). The pair (VaR, CoVaR) is assumed to follow a general semi-parametric dynamic model for given risk levels (α, α'), and a two-step quantile regression estimator is

[1] More recently, the bankruptcy of Silicon Valley Bank, the liquidation of Silvergate Bank and the collapse of Crédit Suisse have raised fears of contagion.

studied. We advocate an alternative approach which, instead of directly modeling SRM for each pair of risk levels, derives their dynamics from assumptions on the underlying risk factors. Such assumptions are mild in the sense that they do not constitute a full model of the joint dynamic of the factors (like for instance a multivariate GARCH).

The first contribution of the paper is to provide a framework for deriving conditional SRM. Our approach is equation-by-equation based, relying on dynamic location-scale models for each risk factor. It turns out that the CoVaR and ΔCoVaR of the two factors under consideration are explicitly related to (i) the (unconditional) CoVaR and ΔCoVaR of the underlying innovations, and (ii) the conditional location and scale of the first factor.[2] By disentangling conditional moments of the risk factors and characteristics of the joint marginal distribution of the innovations in dynamic SRM, our approach paves the way to a simple-to-implement estimation method which, in a first step, estimates the conditional moments parameters, and, in a second step, uses empirical characteristics of the residuals.

Our main contribution is an asymptotic theory for dynamic SRM in a semi-parametric framework in which no strong distributional assumptions on the innovation processes are made. We rely on Quasi-Maximum Likelihood (QML) equation-by-equation estimation for the conditional moment parameters. Such moments are allowed to incorporate present and past values of exogenous variables. The main statistical challenge is to derive the joint asymptotic distribution of the QML estimators and of the empirical characteristics of the innovations. An important technical hurdle to overcome is that such characteristics (involving indicator variables used to define empirical quantiles) are non-smooth functions of the QML estimators. We show consistency and asymptotic normality of our proposed estimator and derive asymptotic Confidence Intervals (CIs) for the dynamic CoVaR and ΔCoVaR. Our asymptotic results also allow us to consider testing hypotheses, such as *risk systemicity*. The latter occurs if the first factor has an increased risk of being in distress given that the second factor is in distress.

In a systemic approach, the level of reserves of "Institution 1" is determined at each date by the value of the CoVaR between this institution and the "system" (Institution 2). On the other hand, the standard "unsystemic" risk measure is the conditional VaR. It is thus of interest to test whether it is worth modifying the reserves to take into account the systemic risk, or if the reserves determined in the standard situation are safe.

Finally, we show the relevance of our methodology for analyzing the systemic risk of financial institutions and European countries through their Credit Default Swap (CDS) spreads on sovereign debts.

Dynamic SRM have links with the concept of extremogram developed and studied by Davis and Mikosh (2009) (see also Davis, Mikosh and Cribben (2012)). The extremogram allows to capture serial dependence in the tails of a stationary time series. The tail dependence coefficient that is often used in extreme value theory and quantitative risk management (see e.g. McNeil, Frey and Embrechts, 2005) can be viewed as a particular case of extremogram. By contrast, the SRM considered in this paper capture, in a dynamic way, the contemporaneous dependence in the tails of two time series. Another difference is that we do not resort to extreme value theory for estimating SRM, as the risk levels we consider are "moderately small".

The rest of the paper is organized as follows. Section 2 introduces the semi-parametric dynamic SRM estimators and provides conditions for their consistency. Asymptotic distributions of the CoVaR and ΔCoVaR estimators are derived in Section 3. Section 4 presents CIs for the dynamic CoVaR, obtained from the asymptotic distribution or using a bootstrap approach. For the sake of comparison, we also introduce a Gaussian procedure for constructing CIs. Hypotheses testing is studied in Section 5. In particular, we develop backtests taking into account the condi-

[2] No such decomposition exists, in general, if one assumes a multivariate GARCH model (DCC, BEKK, ...).

tional nature of our systemic risk measures. We show on a large set of financial returns that the semi-parametric approach outperforms the parametric method based on Gaussian assumptions. Section 6 presents Monte Carlo experiments. Section 7 illustrates the theoretical results on series of financial returns and on CDS spreads. Firstly, we propose a ranking of financial institutions based on their estimated systemic impact. Secondly, we provide dynamic comparison between systemic and individual risks of large banks. Thirdly, using macroeconomic data, we adopt a different perspective by evaluating the impact of different economies on a particular country. Section 8 concludes. Assumptions and proofs of the main results are displayed in an Appendix. In an online supplemental document, we provide additional proofs and complementary results.

2 Semi-parametric SRM

We provide a dynamic framework for deriving explicit formulas for the dynamic CoVaR, ΔCoVaR, MES and their semi-parametric estimators.

2.1 Dynamic SRM

The *dynamic* VaR of a real process $(X_t)_{t\in\mathbb{Z}}$ at risk level $\alpha \in (0,1)$, denoted by $\text{VaR}_t^X(\alpha)$, is defined as the opposite of the α-quantile of the conditional distribution of X_t:

$$\text{VaR}_t^X(\alpha) = -\inf\{z_t : P_{t-1}[X_t \leqslant z_t] \geqslant \alpha\}$$

where P_{t-1} denotes the historical distribution conditional on $\{X_u, u < t\}$. With this definition, $\text{VaR}_t^X(\alpha)$ is generally positive for small values of α.

For bivariate risks, related concepts are the CoVaR and ΔCoVaR introduced by Adrian and Brunnermeier (2011, 2016), see also Girardi and Ergün (2013). The *dynamic* CoVaR of a process (X_t) relative to a process (Y_t) at risk levels $\alpha, \alpha' \in (0,1)$, denoted by $\text{CoVaR}_t^{X|Y}(\alpha,\alpha')$, can be defined as

$$\text{CoVaR}_t^{X|Y}(\alpha,\alpha') = -\inf\{z_t : P_{t-1}[X_t \leqslant z_t | Y_t \leqslant -\text{VaR}_t^Y(\alpha')] \geqslant \alpha\},$$

where P_{t-1} now denotes the historical distribution conditional on $\{(X_u, Y_u), u < t\}$.[3] The *dynamic* ΔCoVaR of (X_t) relative to (Y_t) at risk levels $\alpha, \alpha' \in (0,1)$ and $\alpha'' \in (0, 0.5)$, can be defined as

$$\Delta\text{CoVaR}_t^{X|Y}(\alpha,\alpha',\alpha'') = \text{CoVaR}_t^{X|Y}(\alpha,\alpha') - \text{CoVaR}_t^{X|m_Y}(\alpha,\alpha'')$$

where the latter is a "median-state" CoVaR defined by

$$\text{CoVaR}_t^{X|m_Y}(\alpha,\alpha'') = -\inf\{z_t : P_{t-1}\left[X_t \leqslant z_t \,\middle|\, Y_t \in A_t(\alpha'')\right] \geqslant \alpha\},$$

where $A_t(\alpha'') = (-\text{VaR}_t^Y(50\% - \alpha''), -\text{VaR}_t^Y(50\% + \alpha'')]$. Following Acharya et al. (2012) and Brownlees and Engle (2017), the *dynamic* MES is defined by

$$\text{MES}_t^{X|Y}(\alpha') = -E_{t-1}[X_t | Y_t \leqslant -\text{VaR}_t^Y(\alpha')] = \int_0^1 \text{CoVaR}_t^{X|Y}(\alpha,\alpha')d\alpha,$$

provided the conditional expectation is well defined.

[3] In the original definition of Adrian and Brunnermeier (2016), the conditioning event was written with an equality, $Y_t = -\text{VaR}_t^Y(\alpha')$. Subsequent works have generally preferred to work with an inequality, which is both more interpretable and more convenient for statistical purposes.

The proposed measures are defined for horizon 1, but they can be extended to higher horizons. For instance, the dynamic CoVaR *at horizon* $h \geqslant 1$ of (X_t) relative to (Y_t) can be defined as

$$\text{CoVaR}_{t,h}^{X|Y}(\alpha, \alpha') = -\inf\{z_t : P_{t-1}[X_{t+h-1} \leqslant z_t | Y_{t+h-1} \leqslant -\text{VaR}_{t,h}^{Y}(\alpha')] \geqslant \alpha\},$$

where $\text{VaR}_{t,h}^{Y}(\alpha') = -\inf\{z_t : P_{t-1}[Y_{t+h-1} \leqslant z_t] \geqslant \alpha'\}$. In these formulas, the two processes are considered contemporaneously. However, an asymmetry between the horizons of X and Y could also be introduced. Such measures are however difficult to study when the predictive distribution of (X_t, Y_t) is not explicit at any horizon. For this reason, we will focus in this article on SRMs at horizon 1.

2.2 Dynamic SRM for conditional location-scale models

Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})'$ be a vector of risk factors, whose first two components satisfy

$$\epsilon_{it} = \mu_i(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\theta}_0^{(i)}) + \sigma_i(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \ldots; \boldsymbol{\theta}_0^{(i)})\eta_{it} := \mu_{it} + \sigma_{it}\eta_{it}, \quad t \in \mathbb{Z}, \quad i = 1, 2, \quad (1)$$

where (η_{1t}, η_{2t}) is an iid process with $E\eta_{it}^2 = 1$, $\sigma_{it} > 0$ and $\boldsymbol{\theta}_0^{(i)}$ are unknown parameter vectors belonging to compact parameter sets $\boldsymbol{\Theta}^{(i)} \subset \mathbb{R}^{d_i}$ for some positive integers d_i. Note that the conditional mean, μ_{it}, and the volatility, σ_{it}, of each component may depend on the past of all components of $\boldsymbol{\epsilon}_t$. In this respect, the components for $i > 2$ can be viewed as *exogenous* variables for the dynamics of the first two ones. Assume that the variables η_{it} are independent from the σ-field $\mathcal{F}_{t-1}$ generated by $\{\boldsymbol{\epsilon}_{t-u}, u > 0\}$.

The VaR of ϵ_{1t} conditional on $\mathcal{F}_{t-1}$ at level α is given by $\text{VaR}_t^{\epsilon_1}(\alpha) = -\mu_{1t} - \sigma_{1t}\xi_\alpha^{(1)}$, where $\xi_\alpha^{(i)} = \inf\{x : G^{(i)}(x) \geqslant \alpha\}$ is the α-quantile of the cumulative distribution function (cdf) $G^{(i)}$ of η_{it}. We will call co-cdf the functions

$$F(x|y) = P[\eta_{1t} \leqslant x \mid \eta_{2t} \leqslant y], \quad F^{\Delta}(x|A) = P[\eta_{1t} \leqslant x \mid \eta_{2t} \in A]$$

for $x \in \mathbb{R}$, $y \in \mathbb{R}$ such that $P[\eta_{2t} \leqslant y] \neq 0$, and A a measurable set such that $P[\eta_{2t} \in A] \neq 0$.

Our first result is a straightforward consequence of the definitions of the conditional VaR and CoVaR, and the independence between (η_{1t}, η_{2t}) and the past of $\boldsymbol{\epsilon}_t$.

Proposition 2.1. *We have, under the independence between the variables η_{it} and the σ-field $\mathcal{F}_{t-1}$,*

$$CoVaR_t^{\epsilon_1|\epsilon_2}(\alpha, \alpha') = -\mu_{1t} - \sigma_{1t}u(\alpha, \alpha'), \quad \Delta CoVaR_t^{\epsilon_1|\epsilon_2}(\alpha, \alpha', \alpha'') = -\sigma_{1t}\{u(\alpha, \alpha') - \underline{u}(\alpha, \alpha'')\},$$
$$MES_t^{\epsilon_1|\epsilon_2}(\alpha') = -\mu_{1t} - \sigma_{1t}v(\alpha'),$$

where $u(\alpha, \alpha') = \inf\left\{x : F\left(x \mid \xi_{\alpha'}^{(2)}\right) \geqslant \alpha\right\}$, $\underline{u}(\alpha, \alpha'') = \inf\left\{x : F^{\Delta}\left(x \mid A_{\alpha''}^{(2)}\right) \geqslant \alpha\right\}$, $v(\alpha') = E[\eta_{1t} \mid \eta_{2t} \leqslant \xi_{\alpha'}^{(2)}] = \int_0^1 u(\alpha, \alpha')d\alpha$ *and* $A_{\alpha''}^{(2)} = \left(\xi_{0.5-\alpha''}^{(2)}, \xi_{\alpha''+0.5}^{(2)}\right]$.

This result, expressing the dynamic SRM of ϵ_{1t} relative to ϵ_{2t} is crucial to the simplicity of our approach. Indeed, conditional moments μ_{1t} and σ_{1t} are disentangled from joint characteristics of the innovations in the dynamic SRM.

We will call $u(\alpha, \alpha')$ the co-quantile of η_1 and η_2 at levels (α, α'). Obviously, the CoVaR and VaR sample paths will not cross except if η_{1t} and η_{2t} are independent (in which case the two sample paths coincide). Intuitively, risk sytemicity should occur when the variables η_{1t} and η_{2t} are "positively dependent" (for instance if they are comonotonic). More precisely, we have the following result comparing the VaR and CoVaR of ϵ_{1t} at level α. The proof is given in Section E of the online supplemental document.

Proposition 2.2. *If η_{1t} and η_{2t} have continuous distributions, we have*

$$\begin{aligned} CoVaR_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha') \geqslant VaR_t^{\epsilon_1}(\alpha) &\Longleftrightarrow u(\alpha,\alpha') \leqslant \xi_\alpha^{(1)} \\ &\Longleftrightarrow F^\Delta(u(\alpha,\alpha') \mid (\xi_{\alpha'}^{(2)},\infty)) \leqslant F(u(\alpha,\alpha') \mid \xi_{\alpha'}^{(2)}) \\ &\Longleftrightarrow F^\Delta(\xi_\alpha^{(1)} \mid (\xi_{\alpha'}^{(2)},\infty)) \leqslant F(\xi_\alpha^{(1)} \mid \xi_{\alpha'}^{(2)}). \end{aligned}$$

In particular, the previous inequalities hold when η_{1t} and η_{2t} are right-continuous increasing functions of the same variable U_t (thus η_{1t} and η_{2t} are comonotonic).

Interestingly, a CoVaR that is larger than the VaR for the first component implies a higher probability of distress when the second component is in distress than when it is not. This latter property reflects the intuitive notion of systemic risk. Finally, we note that in the case of positively correlated Gaussian variables this property also holds (see Section F of the online supplemental document).

The form of the dynamic CoVaR, as an affine function of a fixed characteristic of the joint distribution of the innovations with coefficients depending on the first two conditional moments of the first component, suggests a two-step estimation method. As in Amengual et al. (2013) we propose a sequential estimation method but, contrary to this reference, the second step is nonparametric.

2.3 Semi-parametric estimation

Given observations $\boldsymbol{\epsilon}_1,\dots,\boldsymbol{\epsilon}_n$, and using arbitrary initial values $\tilde{\boldsymbol{\epsilon}}_j$ for $j \leqslant 0$, we define for any $\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}$, $\tilde{\mu}_{it}(\boldsymbol{\theta}^{(i)}) = \mu_i(\boldsymbol{\epsilon}_{t-1},\dots,\boldsymbol{\epsilon}_1,\tilde{\boldsymbol{\epsilon}}_0,\tilde{\boldsymbol{\epsilon}}_{-1},\dots;\boldsymbol{\theta}^{(i)})$, $\tilde{\sigma}_{it}(\boldsymbol{\theta}^{(i)}) = \sigma_i(\boldsymbol{\epsilon}_{t-1},\dots,\boldsymbol{\epsilon}_1,\tilde{\boldsymbol{\epsilon}}_0,\tilde{\boldsymbol{\epsilon}}_{-1},\dots;\boldsymbol{\theta}^{(i)})$, which will be used as proxies of $\mu_{it}(\boldsymbol{\theta}^{(i)}) = \mu_i(\boldsymbol{\epsilon}_{t-1},\dots,\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_0,\boldsymbol{\epsilon}_{-1},\dots;\boldsymbol{\theta}^{(i)})$ and $\sigma_{it}(\boldsymbol{\theta}^{(i)}) = \sigma_i(\boldsymbol{\epsilon}_{t-1},\dots,\boldsymbol{\epsilon}_1,\boldsymbol{\epsilon}_0,\boldsymbol{\epsilon}_{-1},\dots;\boldsymbol{\theta}^{(i)})$. The parameters $\boldsymbol{\theta}_0^{(i)}$ in (1) can be estimated equation-by-equation (see Francq and Zakoïan (2016)) using the Gaussian QML approach:

$$\widehat{\boldsymbol{\theta}}^{(i)} = \arg\min_{\boldsymbol{\theta}^{(i)}\in\boldsymbol{\Theta}^{(i)}} \sum_{t=1}^n \frac{\{\epsilon_{it} - \tilde{\mu}_{it}(\boldsymbol{\theta}^{(i)})\}^2}{\tilde{\sigma}_{it}^2(\boldsymbol{\theta}^{(i)})} + \log \tilde{\sigma}_{it}^2(\boldsymbol{\theta}^{(i)}).$$

This approach can be called semi-parametric as it does not require full specification of the dynamics of $\boldsymbol{\epsilon}_t$, nor any assumption concerning the joint distribution of (η_{1t},η_{2t}). Key assumptions for the consistency and asymptotic normality of the estimators $\widehat{\boldsymbol{\theta}}^{(i)}$ are: (a) independence between the innovations η_{it} and the σ-field $\mathcal{F}_{t-1}$; and (b) correct specification of the conditional mean and variance functions. See Escanciano (2008) for a class of joint and marginal spectral diagnostic tests for parametric conditional means and variances in general time series models.

Let the residuals $\widehat{\eta}_{it} = \{\epsilon_{it} - \tilde{\mu}_{it}(\widehat{\boldsymbol{\theta}}^{(i)})\}/\tilde{\sigma}_{it}(\widehat{\boldsymbol{\theta}}^{(i)})$ for $i=1,2$ and $t=1,\dots,n$. Let $\widehat{u}(\alpha,\alpha')$, $\widehat{\underline{u}}(\alpha,\alpha'')$ and $\widehat{v}(\alpha')$ be the estimators of $u(\alpha,\alpha')$, $\underline{u}(\alpha,\alpha'')$ and $v(\alpha')$ respectively, such that

$$\widehat{u}(\alpha,\alpha') = \inf \arg\min_{z\in\mathbb{R}} \sum_{t=1}^n \rho_\alpha(\widehat{\eta}_{1t} - z)\mathbb{1}_{\widehat{\eta}_{2t}\leqslant\widehat{\xi}_{\alpha'}^{(2)}}, \tag{2}$$

$$\widehat{\underline{u}}(\alpha,\alpha'') = \inf \arg\min_{z\in\mathbb{R}} \sum_{t=1}^n \rho_\alpha(\widehat{\eta}_{1t} - z)\mathbb{1}_{\widehat{\eta}_{2t}\in\widehat{A}_{\alpha''}^{(2)}}, \qquad \widehat{v}(\alpha') = \frac{1}{n\alpha'}\sum_{t=1}^n \widehat{\eta}_{1t}\mathbb{1}_{\widehat{\eta}_{2t}\leqslant\widehat{\xi}_{\alpha'}^{(2)}},$$

where $\widehat{A}_{\alpha''}^{(2)} = \left(\widehat{\xi}_{0.5-\alpha''}^{(2)}, \widehat{\xi}_{\alpha''+0.5}^{(2)}\right]$, $\rho_\alpha(z) = z(\alpha - \mathbb{1}_{z<0})$ is the usual "check" function and $\widehat{\xi}_{\alpha'}^{(2)}$ is the α'-quantile of $\widehat{\eta}_{21},\dots,\widehat{\eta}_{2n}$, that is the $\lceil n\alpha'\rceil$-th order statistics of the residuals, where $\lceil x\rceil$ denotes the smallest integer larger than x. Estimators of $\mathrm{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha')$, $\Delta\mathrm{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha',\alpha'')$ and $\mathrm{MES}_t^{\epsilon_1|\epsilon_2}(\alpha')$ are thus

$$\begin{aligned}
\widehat{\mathrm{CoVaR}}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha') &= -\widetilde{\mu}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)}) - \widetilde{\sigma}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)})\widehat{u}(\alpha,\alpha'),\\
\Delta\widehat{\mathrm{CoVaR}}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha',\alpha'') &= -\widetilde{\sigma}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)})\{\widehat{u}(\alpha,\alpha') - \widehat{\underline{u}}(\alpha,\alpha'')\},\\
\widehat{\mathrm{MES}}_t^{\epsilon_1|\epsilon_2}(\alpha') &= -\widetilde{\mu}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)}) - \widetilde{\sigma}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)})\widehat{v}(\alpha').
\end{aligned} \tag{3}$$

Note that these estimators are extremely simple to obtain in practice since they depend on conditional moments estimated equation-by-equation and on explicit characteristics of the residuals.

We now study the consistency of the empirical co-quantiles $\widehat{u}(\alpha,\alpha')$ and $\widehat{\underline{u}}(\alpha,\alpha'')$. Assumptions are displayed in Appendix A.

Proposition 2.3. *Under* **A1**-**A4**, $\mathbf{A5_1}$, $\mathbf{A5_2}(\xi_{\alpha'}^{(2)})$, $\mathbf{A6}(\alpha,\alpha')$, $\mathbf{A}7(\xi_{\alpha'}^{(2)})$, *we have the strong convergence* $\widehat{u}(\alpha,\alpha') \to u(\alpha,\alpha')$ *a.s.*

Proposition 2.4. *Under* **A1**-**A4**, $\mathbf{A5_1}$ *and, for* $\tau \in \{-1,1\}$ *and* $\alpha'' \in (0,1/2)$, $\mathbf{A5_2}(\xi_{0.5+\tau\alpha''}^{(2)})$, $\mathbf{A6}(\alpha, 0.5+\tau\alpha'')$, $\mathbf{A}7(\xi_{0.5+\tau\alpha''}^{(2)})$, *we have* $\widehat{\underline{u}}(\alpha,\alpha'') \to \underline{u}(\alpha,\alpha'')$ *a.s.*

Proofs and extended versions of Propositions 2.3 and 2.4 obtained by relaxing (ii) in **A6** are provided in Section E of the online supplemental document. The consistency of $\widehat{v}(\alpha')$ is established in the next proposition (see also Section E for a proof).

Proposition 2.5. *Under* **A1**-**A5** *and* **A8**-**A9**, *we have* $\widehat{v}(\alpha') \to v(\alpha')$ *a.s.*

3 Asymptotic distribution

We will now establish the asymptotic distributions of the estimators of $\mathrm{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha')$ and $\Delta\mathrm{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha',\alpha'')$. We focus in this section on pure conditional scale models, i.e. we assume that Model (1) reduces to $\epsilon_{it} = \sigma_{it}\eta_{it}$ for $i = 1,2$. Denote by $\mathbf{A2}^*$–$\mathbf{A4}^*$ the assumptions **A2**–**A4** in absence of the conditional means μ_i.

Our strategy to obtain the joint asymptotic distribution of $\widehat{u}(\alpha,\alpha')$ and $\widehat{\underline{u}}(\alpha,\alpha'')$ relies on deriving the asymptotic distribution of the empirical co-cdf $\widehat{F}(x|y)$. The latter is displayed in Appendix B. To deduce the law of the empirical co-quantiles, it is not sufficient to consider the case where (x,y) is fixed, but we need to establish a *stochastic equicontinuity* result (*i.e.* that the limiting distribution is the same when (x,y) is replaced by a random sequence (x_n,y_n) tending to (x,y) in probability).

It is standard to show that assumptions $\mathbf{B1_i} - \mathbf{B5_i}$ in Appendix A ensure the following Bahadur expansion for the QMLE of $\boldsymbol{\theta}_0^{(i)}$ (and hence the asymptotic normality):

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^{(i)} - \boldsymbol{\theta}_0^{(i)}) = \frac{\boldsymbol{J}_i^{-1}}{2\sqrt{n}}\sum_{t=1}^{n}(\eta_{it}^2 - 1)\boldsymbol{D}_{it} + o_{\mathbb{P}}(1), \tag{4}$$

where $\boldsymbol{J}_i = E(\boldsymbol{D}_{it}\boldsymbol{D}_{it}')$ and $\boldsymbol{D}_{it} = \boldsymbol{D}_{it}(\boldsymbol{\theta}_0^{(i)})$.

Assumption **B7** in the appendix requires the volatility to be stable by scaling, up to a change of the volatility parameter. It is satisfied by all commonly used GARCH-type model and entails formidable simplifications in the upcoming asymptotic results.

Under **B6**, denote by $f_1(\cdot \mid y)$ (resp.$f_2(\cdot \mid x)$) the density of η_{1t} (resp. η_{2t}) conditional on $\eta_{2t} \leqslant y$ (resp. $\eta_{1t} \leqslant x$) assuming $G^{(2)}(y) > 0$ (resp. $G^{(1)}(x) > 0$).[4] To alleviate notation, write $\widehat{u} = \widehat{u}(\alpha, \alpha')$, $u = u(\alpha, \alpha')$.

Theorem 3.1. *Let* **A1**, **A2***–**A4***, **B1**-**B6** *and* **B8** *hold. Let* $\boldsymbol{\Omega}_i = E(\boldsymbol{D}_{it})$. *We have*

$$\sqrt{n}\left\{\widehat{u} - u\right\} = \frac{-1}{\sqrt{n}\alpha' f_1(u|\xi^{(2)}_{\alpha'})} \sum_{t=1}^{n} \left(\mathbb{1}_{\eta_{1t} \leqslant u,\, \eta_{2t} \leqslant \xi^{(2)}_{\alpha'}} - \alpha\alpha'\right)$$

$$+ \frac{1}{f_1(u|\xi^{(2)}_{\alpha'})} \frac{G^{(1)}(u)}{\alpha'} \frac{f_2(\xi^{(2)}_{\alpha'}|u)}{g^{(2)}(\xi^{(2)}_{\alpha'})} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{\mathbb{1}_{\eta_{2t} \leqslant \xi^{(2)}_{\alpha'}} - \alpha'\right\} - \frac{u}{2\sqrt{n}} \boldsymbol{\Omega}_1' \boldsymbol{J}_1^{-1} \sum_{t=1}^{n} (\eta_{1t}^2 - 1) \boldsymbol{D}_{1t},$$

up to some $o_{\mathbb{P}}(1)$. *With the additional assumption* $\mathbf{B7_1}$ *we have,*

$$\sqrt{n}\left\{\widehat{u} - u\right\} \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \sigma^2(\alpha, \alpha') = \boldsymbol{\lambda}' \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \boldsymbol{\lambda}\right), \qquad \textit{where}$$

$$\boldsymbol{\lambda}' = \left(\frac{-1}{\alpha' f_1(u|\xi^{(2)}_{\alpha'})}, \frac{1}{\alpha'} \frac{G^{(1)}(u)}{f_1(u|\xi^{(2)}_{\alpha'})} \frac{f_2(\xi^{(2)}_{\alpha'}|u)}{g^{(2)}(\xi^{(2)}_{\alpha'})}, \frac{-u}{2}\right), \; \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} = \begin{pmatrix} \alpha\alpha'(1-\alpha\alpha') & \alpha\alpha'(1-\alpha') & \alpha' \varrho_{\alpha,\alpha'} \\ \alpha\alpha'(1-\alpha') & (1-\alpha')\alpha' & \alpha' \nabla_{\alpha,\alpha'} \\ \alpha' \varrho_{\alpha,\alpha'} & \alpha' \nabla_{\alpha,\alpha'} & \kappa_1 - 1 \end{pmatrix}$$

and $\varrho_{\alpha,\alpha'} = E(\eta_{1t}^2 \mathbb{1}_{\eta_{1t} \leqslant u(\alpha,\alpha')} | \eta_{2t} \leqslant \xi_{\alpha'}) - \alpha$, $\quad \nabla_{\alpha,\alpha'} = E(\eta_{1t}^2 | \eta_{2t} \leqslant \xi_{\alpha'}) - 1$.

Remark 1. The third term in the Bahadur expansion of $\sqrt{n}\left\{\widehat{u} - u\right\}$ reflects the effect of the estimation of the model parameters. The estimation of the volatility of the second component does not matter for the asymptotic accuracy of $\widehat{u}$ which may seem surprising. This can be explained by the fact that the set of indicator variables involved in (2) is intuitively not very sensitive to the accuracy of estimation of the volatility of the second component. It is also worth noticing that the sole assumption $\mathbf{B7_1}$ (without $\mathbf{B7_2}$) is sufficient to ensure that the asymptotic distribution of $\widehat{u}$ is model-free, that is, independent of the volatility specifications. However, estimation matters: replacing residuals with (supposedly observed) innovations would simplify the asymptotic variance.

Remark 2. Under **B7**, estimating the asymptotic variance $\sigma^2(\alpha, \alpha')$ reduces to estimating characteristics of the joint distribution of the innovation components η_{it}. Most of them are standard (e.g. the density $g^{(2)}$ of the second component), and can be estimated by usual nonparametric estimators applied to the residuals $\widehat{\eta}_{it}$. The estimation of $f_i(y \mid x)$, for $i = 1, 2$, is less standard but can be achieved for instance by a straightforward adaptation of the Kernel density estimation. See Section 4.1 for details.

Remark 3. In the case where the two innovations η_{1t} and η_{2t} are independent, u reduces to the $\alpha-$quantile $\xi^{(1)}_{\alpha}$ of η_{1t}. In this case we find that

$$\sqrt{n}\left(\widehat{u} - \xi^{(1)}_{\alpha}\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{\alpha' f^2(\xi^{(1)}_{\alpha})} + \frac{\xi^{(1)}_{\alpha} \varrho(\xi^{(1)}_{\alpha})}{f(\xi^{(1)}_{\alpha})} + \frac{\kappa_1 - 1}{4} (\xi^{(1)}_{\alpha})^2\right).$$

where f denotes the density of η_{1t} and $\varrho(\xi^{(1)}_{\alpha}) = E(\eta_{1t}^2 \mathbb{1}_{\eta_{1t} \leqslant \xi^{(1)}_{\alpha}}) - \alpha$, whereas the asymptotic distribution of the empirical quantile of the residuals $\widehat{\eta}_{1t}$ (obtained for $\alpha' = 1$ in the asymptotic distribution of Theorem 3.1) is

$$\sqrt{n}\left(\widehat{\xi}^{(1)}_{\alpha} - \xi^{(1)}_{\alpha}\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(\xi^{(1)}_{\alpha})} + \frac{\xi^{(1)}_{\alpha} \varrho(\xi^{(1)}_{\alpha})}{f(\xi^{(1)}_{\alpha})} + \frac{\kappa_1 - 1}{4} (\xi^{(1)}_{\alpha})^2\right). \tag{5}$$

[4] We provide in Appendix F a closed form formula for $f_2(\cdot \mid x)$ in the Gaussian case.

Unsurprisingly, the estimator $\widehat{u}$ is asymptotically less accurate (particularly when α' is small), as a price paid for the unnecessary inclusion of the residuals of the second volatility model in the estimation of $\xi_\alpha^{(1)}$. However, the difference affects only the first term in the asymptotic variances, not the second and third terms measuring the impact of the estimation (i.e. the use of residuals instead of innovations).

We are now in a position to derive the joint asymptotic distribution of $\widehat{\boldsymbol{\theta}}^{(1)}$ and $\widehat{u}$ which will be used to obtain CIs for the CoVaR.

Corollary 3.1. *Under the assumptions of Theorem 3.1, including* $\mathbf{B7_1}$*, we have*

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0^{(1)} \\ \widehat{u} - u \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{\mathbf{0}, \boldsymbol{\Sigma}(\alpha,\alpha') := \begin{pmatrix} \frac{\kappa_1 - 1}{4}\boldsymbol{J}_1^{-1} & \frac{1}{2}\boldsymbol{J}_1^{-1}\boldsymbol{\Omega}_1\boldsymbol{e}_3'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{\lambda} \\ \frac{1}{2}\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_3\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1} & \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{\lambda} \end{pmatrix}\right\}$$

with $\boldsymbol{e}_3 = (0,0,1)'$.

Remark 4. Estimating the asymptotic variance $\boldsymbol{\Sigma}(\alpha,\alpha')$ requires not only estimators of characteristics of the innovations distributions, but also of the matrices $\boldsymbol{J}_1$ and $\boldsymbol{\Omega}_1$ related to the volatility model of the first component ϵ_{1t}. Let $\widetilde{\boldsymbol{D}}_{1t}(\boldsymbol{\theta}^{(1)}) = \widetilde{\sigma}_{1t}^{-1}(\boldsymbol{\theta}^{(1)})\partial\widetilde{\sigma}_{1t}(\boldsymbol{\theta}^{(1)})/\partial\boldsymbol{\theta}^{(1)}$ and $\widehat{\boldsymbol{D}}_{1t} = \widetilde{\boldsymbol{D}}_{1t}\left(\widehat{\boldsymbol{\theta}}^{(1)}\right)$. Under **A1**–**A4** and $\mathbf{B1_1}$–$\mathbf{B5_1}$, it can be shown that $\sup_{\boldsymbol{\theta}^{(1)}\in\boldsymbol{\Theta}^{(1)}}\left\|\widetilde{\boldsymbol{D}}_{1t}(\boldsymbol{\theta}^{(1)}) - \boldsymbol{D}_{1t}(\boldsymbol{\theta}^{(1)})\right\| \leqslant K\rho^t u_t$, with (u_t) a positive stationary process such that $Eu_t^4 < \infty$. We deduce that

$$\widehat{\boldsymbol{\Omega}}_1 := \frac{1}{n}\sum_{t=1}^n \widehat{\boldsymbol{D}}_{1t} \to \boldsymbol{\Omega}_1, \qquad \widehat{\boldsymbol{J}}_1 := \frac{1}{n}\sum_{t=1}^n \widehat{\boldsymbol{D}}_{1t}\widehat{\boldsymbol{D}}_{1t}' \to \boldsymbol{J}_1 \text{ a.s.}$$

Remark 5. The joint asymptotic distribution of $\widehat{\boldsymbol{\theta}}^{(1)}$ and $\widehat{\underline{u}}(\alpha,\alpha'')$ can be derived under the same assumptions. See Section G of the online supplemental material. We leave the asymptotic distribution of $\widehat{v}(\alpha')$ for further investigations.

4 Conditional confidence intervals for the CoVaR

The difficulty in defining CIs for the dynamic CoVaR is that it is a random variable and not a fixed parameter. The dynamic CoVaR depends on past values and fixed unknown parameters. By considering those past values as fixed, we use the delta method to deduce a conditional CI. We refer to Beutner, Heinemann and Smeekes (2021) for a justification of such CIs of random objects. We will develop two approaches, one based on the joint asymptotic distribution of the estimator of the volatility parameter and the estimator of $u(\alpha,\alpha')$, and one relying on a bootstrapping technique. In Section F of the online supplemental document we also introduce a parametric CoVaR estimation procedure based on the assumption of Gaussian conditional distributions.

4.1 CI based on the asymptotic distribution

To obtain approximate CIs for $\text{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}$ based on $\widehat{\text{CoVaR}}_{n+1} = -\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}^{(1)})\widehat{u}$, we use the asymptotic joint distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}^{(1)'} - \boldsymbol{\theta}_0^{(1)'}, \widehat{u} - u)'$, which requires estimating the densities $g^{(2)}$ and $f_1(\cdot \mid y)$. For the former density, we use a standard Kernel density estimator. For the latter density, for y such that $\mathbb{P}(\eta_{2t} \leqslant y) > 0$, one can use the kernel density estimator

$\widehat{f}_1(x \mid y) = \frac{\mathbb{1}_{n_2(y)>0}}{n_2(y)b_{n_2(y)}} \sum_{t=1}^n K\left(\frac{x-\widehat{\eta}_{1t}}{b_{n_2(y)}}\right) \mathbb{1}_{\widehat{\eta}_{2t}\leqslant y}$, where $n_2(y)$ is the (assumed positive) number of $\widehat{\eta}_{2t}$'s for $t = 1, \ldots, n$ such as $\widehat{\eta}_{2t} \leqslant y$, the kernel K is a standardized probability density and (b_n) is a sequence of positive numbers. There exists an extensive literature on the choice of the tuning parameters (kernel and bandwidth) for kernel density estimation. Kulperger and Yu (2005) showed that, in the case of standard GARCH models, a kernel density estimator $\widehat{g}^{(2)}$ based on the residuals provides a consistent estimator of the density $g^{(2)}$. We thus rely on existing results on the choice of the tunning parameters. For our numerical illustrations, we used the default values of the R function `density()`.

To further alleviate the notation, write $\widehat{\xi} = \widehat{\xi}^{(2)}_{\alpha'}$ and $\xi = \xi^{(2)}_{\alpha'}$. We do not specify particular estimators for the previous densities, but assume the assumption **B9** in the appendix. Under this assumption and those of Corollary 3.1, by the delta method, an approximate $(1-\alpha_0)$ CI for $\mathrm{CoVaR}^{\epsilon_1|\epsilon_2}_{n+1}$, where $\alpha_0 \in (0,1)$, has bounds given by

$$-\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}^{(1)})\widehat{u} \pm \frac{1}{\sqrt{n}}\Phi^{-1}(1-\alpha_0/2)\left\{\boldsymbol{\delta}'_{n+1}\widehat{\boldsymbol{\Sigma}}(\alpha,\alpha')\boldsymbol{\delta}_{n+1}\right\}^{1/2},$$

where $\widehat{\boldsymbol{\Sigma}}(\alpha,\alpha')$ is a consistent estimator of $\boldsymbol{\Sigma}(\alpha,\alpha')$, and $\boldsymbol{\delta}'_{n+1} = \left[\frac{\partial\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}_n)}{\partial\boldsymbol{\theta}'}\widehat{u} \quad \widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}_n)\right]$.

4.2 Bootstrapped CoVaR

Monte Carlo experiments discussed below reveal that the finite sample behaviors of the CoVaR and ΔCoVaR estimators are not always well approximated by their asymptotic distributions. To better approximate these distributions, various bootstrap procedures can be considered (see for instance Hall and Yao (2003), Shimizu (2010), Cavaliere et al. (2022), Beutner et al. (2024) for references on bootstrap procedures for GARCH-type models). It should be emphasized that caution is advised as bootstrap procedures do not always work (see Andrews (2000), Shimizu (2013), Cavaliere and Georgiev (2020) and references therein). Kreiss et al. (2011) and Shimizu (2013) proposed a technique based on a single Newton-Raphson iteration that significantly speeds up computations. We adapt this trick to our framework, to propose the following resampling algorithm for constructing approximate confidence intervals for the CoVaR. We are assuming $\mathbf{B7_1}$ entailing a simplification in the asymptotic expansion of Theorem 3.1.

Algorithm

1. Given the observations $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$, compute the QMLE $\widehat{\boldsymbol{\theta}}^{(i)}$ and the residuals $\widehat{\eta}_{i1}, \ldots, \widehat{\eta}_{in}$[5] for $i = 1, 2$.

2. Generate vectors $(\eta^*_{1t}, \eta^*_{2t})'$, $t = 1, \ldots, n$ from independent draws with replacement among the paired residuals $\{(\widehat{\eta}_{11}, \widehat{\eta}_{21})', \ldots, (\widehat{\eta}_{1n}, \widehat{\eta}_{2n})'\}$. Generate the Newton-Raphson bootstrap estimates

$$\widehat{\boldsymbol{\theta}}^{(1)*} = \widehat{\boldsymbol{\theta}}^{(1)} + \frac{\widehat{\boldsymbol{J}}_1^{-1}}{2n}\sum_{t=1}^n(\eta^{*2}_{1t} - \widehat{m}_2)\widehat{\boldsymbol{D}}_{1t} \qquad \text{and}$$

$$\begin{aligned}\widehat{u}^* =& \widehat{u} - \frac{1}{n\widehat{\alpha}'\widehat{f}_1(\widehat{u}|\widehat{\xi})}\sum_{t=1}^n\left\{\mathbb{1}_{\eta^*_{1t}\leqslant\widehat{u},\, \eta^*_{2t}\leqslant\widehat{\xi}} - \hat{\alpha}\hat{\alpha}'\right\} \\ &+ \frac{1}{\widehat{f}_1(\widehat{u}|\widehat{\xi})}\frac{\widehat{G}^{(1)}(\widehat{u})}{\hat{\alpha}'}\frac{\widehat{f}_2(\widehat{\xi}|\widehat{u})}{\widehat{g}^{(2)}(\widehat{\xi})}\frac{1}{n}\sum_{t=1}^n\left\{\mathbb{1}_{\eta^*_{2t}\leqslant\widehat{\xi}} - \hat{\alpha}'\right\} - \frac{\widehat{u}}{2n}\sum_{t=1}^n(\eta^{*2}_{1t} - \widehat{m}_2)\end{aligned}$$

[5] As noted by Beutner et al. (2024) it is useless to standardize the residuals, but it can be done.

where, using notations in (14), $\hat{\alpha}' = \widehat{G}^{(2)}(\hat{\xi})$, $\hat{\alpha} = \widehat{H}(\hat{u}, \hat{\xi})/\hat{\alpha}'$ and $\widehat{m}_2 = n^{-1}\sum_{t=1}^{n}\widehat{\eta}_{1t}^2$.[6] The bootstrap estimator of $\mathrm{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}(\alpha, \alpha')$ is $\mathrm{CoVaR}^* = -\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}^{(1)*})\hat{u}^*$.

3. Repeat B times Step 2, and denote by $\mathrm{CoVaR}_1^*, \ldots, \mathrm{CoVaR}_B^*$ the bootstrap estimates of $\mathrm{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}(\alpha, \alpha')$. An approximate $(1-\alpha_0)$ CI for $\mathrm{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}(\alpha, \alpha')$ is $\left[\mathrm{CoVaR}^*_{(\alpha_0/2)},\ \mathrm{CoVaR}^*_{(1-\alpha_0/2)}\right]$ where $\mathrm{CoVaR}^*_{(\alpha_0)}$ denotes the empirical α_0-quantile of the B bootstrap CoVaR estimates.

Remark 6. We used statistics rather than the fixed constants α and α' to have centred sums. For instance, if we replace $\hat{\alpha}'$ by α' in $\frac{1}{n}\sum_{t=1}^{n}\left\{1_{\eta_{2t}^*\leqslant\hat{\xi}} - \hat{\alpha}'\right\}$, the sum is no longer centred with respect to the bootstrap distribution. The aim of the procedure being to mimic the Bahadur expansion obtained with the original variables, centring is essential.

Remark 7. Escanciano and Goh (2014) proposed tests for the correct specification of dynamic VaR based on a multiplier bootstrap procedure after orthogonal projection of the statistics of interest onto the tangent space of the nuisance parameters. The interesting but not obvious extension of such tests to CoVaR specification assessment is left for future research.

Let $q^*_{\alpha_0}$ be the α_0-quantile of the bootstrap distribution of $\sqrt{n}\left\{\mathrm{CoVaR}^* - \widehat{\mathrm{CoVaR}}_{n+1}\right\}$. The CI

$$\left[\mathrm{CoVaR}^*_{(\alpha_0/2)},\ \mathrm{CoVaR}^*_{(1-\alpha_0/2)}\right] = \left[\widehat{\mathrm{CoVaR}}_{n+1} + \frac{1}{\sqrt{n}}q^*_{\alpha_0/2}, \widehat{\mathrm{CoVaR}}_{n+1} + \frac{1}{\sqrt{n}}q^*_{1-\alpha_0/2}\right]$$

is called the reversed-tails (RT) interval by Beutner et al. (2024). The later reference also considers the equal-tailed percentile (EP) interval $\left[\widehat{\mathrm{CoVaR}}_{n+1} - \frac{1}{\sqrt{n}}q^*_{1-\alpha_0/2}, \widehat{\mathrm{CoVaR}}_{n+1} - \frac{1}{\sqrt{n}}q^*_{\alpha_0/2}\right]$ and the symmetric (SY) interval $\left[\widehat{\mathrm{CoVaR}}_{n+1} - \frac{1}{\sqrt{n}}\breve{q}^*_{1-\alpha_0}, \widehat{\mathrm{CoVaR}}_{n+1} + \frac{1}{\sqrt{n}}\breve{q}^*_{1-\alpha_0}\right]$, where $\breve{q}^*_{\alpha_0}$ is the α_0-quantile of the bootstrap distribution of $\sqrt{n}\left|\mathrm{CoVaR}^* - \widehat{\mathrm{CoVaR}}_{n+1}\right|$. Although RT and EP have the same length, Falk and Kaufmann (1991) showed that RT generally has a better finite sample coverage probability than EP, so we used RT in the applications.

The next result establishes the validity of the resampling algorithm.

Theorem 4.1. *Let* **A1**, **A2***–**A4***, *and* **B1**-**B10** *hold. For almost all realization* $(\boldsymbol{\epsilon}_t)$, *as* $n \to \infty$ *we have, given* $(\boldsymbol{\epsilon}_t)$,

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}^{(1)*} - \widehat{\boldsymbol{\theta}}^{(1)} \\ \hat{u}^* - \hat{u} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{\mathbf{0}, \boldsymbol{\Sigma}(\alpha, \alpha')\right\} \tag{6}$$

and when the number of replications B tends to infinity, the coverage probabilities of the RT, EP and SY CIs tend to $(1-\alpha_0)$.

The proof of Theorem 4.1 is given in Section E of the online supplemental document. The resampling algorithm thus provides a way to approximate the distribution in Corollary 3.1, and any statistics depending of this distribution, without having to estimate $\boldsymbol{\Sigma}(\alpha, \alpha')$ directly.

5 Testing hypotheses

In financial risk management, VaR allows to define reserves for a given financial entity. Since CoVaR relies on additional information about a second financial entity, it might help setting more prudential reserves. We first test whether CoVaR induces more reserve than VaR. Then we consider backtests for the CoVaR, individually or jointly with the VaR of the second series.

[6] It is known that, under some mild regularity conditions (see Section 11 in the supplemental document of Francq and Zakoïan (2022) (hereafter FZ)), $\widehat{m}_2$ is exactly equal to 1, for any n.

5.1 Test of systemicity

In a systemic approach, the level of reserves of "Institution 1" is determined at each date by the value of the CoVaR between this institution and the "system" (Institution 2). On the other hand, the standard "unsystemic" risk measure is the conditional VaR. It is thus of interest to test whether it is worth modifying the reserves to take into account the systemic risk, or if the reserves determined in the standard situation are safe. Namely, we wish to test the hypothesis, for specified risk levels α, α',

$$H_0^{\text{sys}}: \quad \text{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha, \alpha') \leqslant \text{VaR}_t^{\epsilon_1}(\alpha) \iff \quad H_0^{\text{sys}}: \quad u(\alpha, \alpha') \geqslant \xi_\alpha^{(1)}.$$

The dynamic hypothesis to be tested thus reduces to a static one. The following proposition derives a test for this assumption. The proof is given in Section H of the online supplemental material.

Proposition 5.1. *Let the assumptions of Theorem 3.1 including* $\mathbf{B7_1}$ *hold, and let* $\hat{s}^2_{\alpha,\alpha'}$ *be a weakly consistent estimator of*

$$\begin{aligned} s^2_{\alpha,\alpha'} =& \frac{\alpha^2(1-\alpha')}{\alpha' f_1^2(\xi_\alpha^{(1)}|\xi_{\alpha'}^{(2)})} \frac{f_2(\xi_{\alpha'}^{(2)}|\xi_\alpha^{(1)})}{g^{(2)}(\xi_{\alpha'}^{(2)})} \left\{ \frac{f_2(\xi_{\alpha'}^{(2)}|\xi_\alpha^{(1)})}{g^{(2)}(\xi_{\alpha'}^{(2)})} - 2 \right\} \\ &+ \frac{\alpha(1-\alpha)}{g^{(1)}(\xi_\alpha^{(1)}) f_1(\xi_\alpha^{(1)}|\xi_{\alpha'}^{(2)})} \left\{ \frac{f_1(\xi_\alpha^{(1)}|\xi_{\alpha'}^{(2)})}{g^{(1)}(\xi_\alpha^{(1)})} - 2 \right\} + \frac{\alpha(1-\alpha\alpha')}{\alpha' f_1^2(\xi_\alpha^{(1)}|\xi_{\alpha'}^{(2)})}. \end{aligned}$$

For all $\alpha_0 \in (0,1)$*, the test of rejection region*

$$C^{\text{sys}} = \{\sqrt{n}(\hat{u}(\alpha,\alpha') - \hat{\xi}_\alpha^{(1)}) < \hat{s}_{\alpha,\alpha'}\Phi^{-1}(1-\alpha_0)\}, \tag{7}$$

has the asymptotic probability of rejection α_0 *under* H_0^{sys} *and* 1 *under* $H_1^{\text{sys}}: u(\alpha,\alpha') < \xi_\alpha^{(1)}$.

It is worth noting that the test in (7) only depends on estimated characteristics of the innovations distribution.

5.2 Backtests

Backtests are commonly used in risk management to assess the validity of risk measures (see e.g. Jorion, 2007, and Christoffersen, 2003). For $\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}$ and $u^{(i)} \in \mathbb{R}$, $i = 1, 2$, let

$$\text{CoVaR}_t(\boldsymbol{\theta}^{(1)}, u^{(1)}) = -\sigma_{1t}(\boldsymbol{\theta}^{(1)})u^{(1)}, \quad \text{VaR}_t(\boldsymbol{\theta}^{(2)}, u^{(2)}) = -\sigma_{2t}(\boldsymbol{\theta}^{(2)})u^{(2)},$$

and, for $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{(1)'}, \boldsymbol{\theta}^{(2)'}\right)'$ and $\boldsymbol{u} = \left(u^{(1)}, u^{(2)}\right)'$, define the joint exceedance process

$$h_t(\boldsymbol{\theta}, \boldsymbol{u}) = \mathbb{1}_{\epsilon_{1t} \leqslant -\text{CoVaR}_t(\boldsymbol{\theta}^{(1)}, u^{(1)}),\ \epsilon_{2t} \leqslant -\text{VaR}_t(\boldsymbol{\theta}^{(2)}, u^{(2)})},$$

and its approximation $\tilde{h}_t(\boldsymbol{\theta}, \boldsymbol{u})$ measurable with respect to the σ-field generated by the observations $\{\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_t\}$. Denote by $\mathcal{B}(\alpha)$ the Bernoulli distribution of parameter α. Consider the null

$$H_0: \ h_1(\boldsymbol{\theta}_0, \boldsymbol{u}_0), \ldots, h_n(\boldsymbol{\theta}_0, \boldsymbol{u}_0) \text{ are independent and } \mathcal{B}(\alpha\alpha') \text{ distributed}$$

for some $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}^{(1)} \times \boldsymbol{\Theta}^{(2)}$ and $\boldsymbol{u}_0 \in \mathbb{R}^2$. Following the notation and terminology of Escanciano and Olmo (2010), we consider the fixed forecasting scheme, and estimate the parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{u}_0$ on the estimation window $t = 1, \ldots, R$ by $\hat{\boldsymbol{\theta}}_R = \left(\hat{\boldsymbol{\theta}}^{(1)'}, \hat{\boldsymbol{\theta}}^{(2)'}\right)'$ and $\hat{\boldsymbol{u}}_R = \left(\hat{u}(\alpha,\alpha'), \hat{\xi}_{\alpha'}^{(2)}\right)'$, replacing n by R in the expressions of $\widehat{\boldsymbol{\theta}}^{(i)}$, $\widehat{u}(\alpha,\alpha')$ and $\widehat{\xi}^{(2)}_{\alpha'}$. We then evaluate $P = n - R$ CoVaR and VaR predictions of the backtesting window $t = R+1,\dots,n$. Let the empirical joint exceedance process

$$\widehat{h}_t = \widetilde{h}_t(\widehat{\boldsymbol{\theta}}_R, \widehat{\boldsymbol{u}}_R) = \mathbb{1}_{\widehat{\eta}_{1t} \leqslant \widehat{u}(\alpha,\alpha'),\, \widehat{\eta}_{2t} \leqslant \widehat{\xi}^{(2)}_{\alpha'}}.$$

We now test two consequences of H_0, by the standard backtests that the unconditional (uc) probability of exceedance is equal to $\alpha\alpha'$ and that the exceedances are independent (ind), see Christoffersen (2003). In our framework, the first backtest employs the statistic of Kupiec (1995) defined by

$$\widetilde{\mathrm{LR}}_{\mathrm{uc}} = 2\log\frac{(n_1/P)^{n_1}\{1-(n_1/P)\}^{n_2}}{(\alpha\alpha')^{n_1}(1-\alpha\alpha')^{n_2}}, \quad n_1 = \sum_{t=R+1}^{n} \widehat{h}_t, \quad n_2 = P - n_1.$$

The second backtest uses the likelihood ratio test statistic $\mathrm{LR}_{\mathrm{ind}}$ of independence against a first-order Markov alternative:

$$\widetilde{\mathrm{LR}}_{\mathrm{ind}} = 2\log\frac{(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}}{(1-\pi_1)^{n_0}\pi_1^{n_1}}, \; n_{ij} = \sum_{t=2+R}^{n} \mathbb{1}_{\widehat{h}_{t-1}=i, \widehat{h}_t=j}, \; n_i = \sum_{j=0}^{1} n_{ji}, \; \pi_{ij} = \frac{n_{ij}}{n_i},$$

and $\pi_1 = \frac{n_{01}+n_{11}}{P-1}$. An exact uc (euc) test is defined with the standardized exceedance mean

$$K_P = \frac{1}{\sqrt{P}} \sum_{t=R+1}^{n} \left(\widehat{h}_t - \alpha\alpha'\right).$$

For $\alpha, \alpha_0 \in (0,1)$, let

$$C(n,\alpha,\alpha_0) = \inf\left\{ c \geqslant 0 : \sum_{i: \left|\frac{i-n\alpha}{\sqrt{n}}\right| > c} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} \leqslant \alpha_0 \right\}.$$

Denote by $\chi^2_k(\alpha)$ the α-quantile of a chi-square distribution with k degrees of freedom. The following proposition is established in Section H of the online supplemental material.

Proposition 5.2. *Let the assumptions of Proposition 2.3,* **B1**-**B6**, $\mathbf{B7_1}$ *and* **B8** *be true. Then the null H_0 is true. For all $\alpha_0 \in (0,1)$, if $R \to \infty$ the test of rejection region*

$$\mathcal{C}^{\mathrm{euc}} = \{|K_P| > C(P, \alpha\alpha', \alpha_0)\} \tag{8}$$

has a probability of rejection less than α_0 and (the law of K_P being discrete) as close to α_0 as possible, for all fixed P. For all $\alpha_0 \in (0,1)$, if $R \to \infty$, $P \to \infty$ and $P/R \to 0$, the tests of rejection regions

$$\mathcal{C}^{\mathrm{uc}} = \{\widetilde{\mathrm{LR}}_{\mathrm{uc}} > \chi^2_1(1-\alpha_0)\}, \qquad \mathcal{C}^{\mathrm{ind}} = \{\widetilde{\mathrm{LR}}_{\mathrm{ind}} > \chi^2_1(1-\alpha_0)\}, \tag{9}$$

have the asymptotic probability of rejection α_0.

Remark 8 ((In)consistency of the UC backtests). With obvious notation, assume a data generating process (DGP) of the form $\epsilon_t = \sigma_t(\boldsymbol{\theta}_0)\eta_t$ and a volatility model $v_t(\boldsymbol{\vartheta})$ that is not necessarily the true one (*i.e.* $v_t(\cdot)$ and $\sigma_t(\cdot)$ are incompatible). Under stationarity and general regularity conditions there exists an estimator $\widehat{\boldsymbol{\vartheta}}$ which tends to a pseudo-true parameter value $\boldsymbol{\vartheta}_0$.

The pseudo-innovations $e_t := \epsilon_t / v_t(\boldsymbol{\vartheta}_0)$ are generally not iid, but under appropriate assumptions their empirical α-quantile converges to the theoretical one $\xi_\alpha(e)$. Note that the two-step semi-parametric VaR based on the wrong volatility specification has the correct asymptotic unconditional violation probability because $P\{\epsilon_t \leqslant v_t(\boldsymbol{\vartheta}_0)\xi_\alpha(e)\} = \alpha$. As formally proved by Escanciano and Pei (2012), it follows that the UC tests are inconsistent. Of course, the same is true for the CoVaR. This does not mean that the UC test is not useful since a rejection of the null may indicate that the length R of the learning window is too small or that the DGP is not stationary. Note also that the UC test is often consistent with purely parametric VaR estimation methods. For example $P\left\{\epsilon_t \leqslant v_t(\boldsymbol{\vartheta}_0)\Phi^{-1}(\alpha)\right\} \neq \alpha$ for some α when η_t is not $\mathcal{N}(0,1)$ distributed.

Escanciano and Olmo (2010) considered other forecasting schemes and provided the–more complicated–asymptotic distributions of K_P and other test statistics when $\lim_{n\to\infty} P/R = \pi \neq 0$. The reader is also referred to Banulescu, Hurlin, Leymarie and Scaillet (2021) for a study of different approaches for backtesting SRM.

Of course, when these backtests reject the null, it is not possible to know if it is because $\text{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha')$ or $\text{VaR}_t^{\epsilon_2}(\alpha')$ is incorrect, or both. It is even possible that CoVaR and VaR are both wrong, but compensate each other so the backtests do not reject. Thus it seems quite natural to focus on the dates $t_j > P$, $j = 1,\dots,P'$, for which $\widehat{\eta}_{2t_j} \leqslant \widehat{\xi}^{(2)}_{\alpha'}$, and to test the null

$$H_0':\ h_1'(\boldsymbol{\theta}_0^{(1)}, u_0^{(1)}),\dots,h_{P'}'(\boldsymbol{\theta}_0^{(1)}, u_0^{(1)}) \text{ are iid with common } \mathcal{B}(\alpha) \text{ distribution}$$

for some $\boldsymbol{\theta}_0^{(1)} \in \boldsymbol{\Theta}^{(1)}$ and some $u_0^{(1)} \in \mathbb{R}$, where $h_j'(\boldsymbol{\theta}^{(1)}, u^{(1)}) = \mathbb{1}_{\epsilon_{1t_j} \leqslant -\text{CoVaR}_{t_j}(\boldsymbol{\theta}^{(1)}, u^{(1)})}$. Let $\widehat{h}_j' = h_j'\left\{\widehat{\boldsymbol{\theta}}^{(1)}, \widehat{u}(\alpha,\alpha')\right\}$ where the parameter estimates are computed on the estimation window. If P' is considered as fixed, it is natural to use the test of rejection region

$$\left\{\left|\frac{\sum_{j=1}^{P'}\widehat{h}_j' - \alpha}{\sqrt{P'}}\right| > C(P',\alpha,\alpha_0)\right\}. \tag{10}$$

However, this test is not fully justified here because P' is random. The reader is referred to Bartholomew (1967) for a general reference on testing problems with random sample size. Taking into account the randomness of P', we obtain the following result.

Proposition 5.3. *Let the assumptions of Proposition 2.3,* **B1**-**B6**, **B7**$_1$ *and* **B8** *be true. For all $\alpha_0 \in (0,1)$, if $R \to \infty$ the test of rejection region*

$$\left\{\left|\frac{\sum_{j=1}^{P'}\widehat{h}_j' - \alpha}{\sqrt{P'}}\right| > C^*(P,\alpha,\alpha',\alpha_0)\right\}, \tag{11}$$

has a probability of rejection less than α_0 and as close to α_0 as possible, for all fixed P, where

$$C^*(P,\alpha,\alpha',\alpha_0) = \inf\left\{c \geqslant 0 : \sum_{n'=1}^{P}\mathbb{P}\left(\left|\frac{\mathcal{B}(n',\alpha)-\alpha}{\sqrt{n'}}\right| > c\right)\mathbb{P}\left\{\mathcal{B}(P,\alpha') = n'\right\} \leqslant \alpha_0\right\}, \tag{12}$$

and $\mathcal{B}(n,p)$ denotes a binomial variable with parameters n and p.

Table 1 shows that, with the naive critical value (10), the probability of type I error[7] can be far from the nominal level when P (and thus P') is small, but the critical values defined by (10) and (12) are almost the same when P is large[8].

[7] Computed as $\sum_{P'=1}^{P}\mathbb{P}\left(\left|\frac{\mathcal{B}(P',\alpha)-\alpha}{\sqrt{P'}}\right| > C\right)\mathbb{P}\left\{\mathcal{B}(P,\alpha') = P'\right\}$, where C is given by $C(P',\alpha,\alpha_0)$ or $C^*(P,\alpha,\alpha',\alpha_0)$.

[8] By construction the probability of first kind error is always smaller than α_0, both with critical value (10) and with (12).

Table 1: Probabilities of type I error for the backtests (10) (P' fixed) and (12) (P' random) at the nominal level $\alpha_0 = 5\%$

$P' \setminus P$	10	20	30	40	50	80	90	100	110	120	980	990	1000	1010	1020
fixed	0.9	1.3	1.7	2.0	2.2	2.6	2.7	2.7	2.8	2.9	4.2	4.2	4.3	4.3	4.3
random	3.6	1.5	4.3	2.8	4.3	4.9	4.9	4.8	4.7	4.8	4.4	4.9	4.8	4.7	4.6

6 Monte Carlo experiments

In this section, we first assess on simulations the validity of the three methods (based respectively on the Gaussian assumption, the asymptotic distribution and the bootstrap) for constructing CIs for the CoVaR. Then, we study the performance of the backtests introduced in Section 5.2.

6.1 Assessing the CI estimation methods

In order to evaluate the validity of the estimation approaches based on the asymptotic distribution or on the bootstrap, we first perform a Monte Carlo experiment with $N = 2,000$ independent simulations of the bivariate GARCH Model

$$\sigma_{it}^2 = \omega_i + \alpha_{ii}\epsilon_{i,t-1}^2 + \alpha_{ij}\epsilon_{j,t-1}^2 + \beta_i\sigma_{i,t-1}^2, \quad i,j = 1,2. \tag{13}$$

with standardized Student innovations with $\nu = 6$ degrees of freedom, and correlation $\rho = 0.6$, $\theta_0^{(1)} = (\omega_1, \alpha_{11}, \alpha_{12}, \beta_1) = (0.001, 0.05, 0.01, 0.9)$ and $\theta_0^{(2)} = (\omega_2, \alpha_{21}, \alpha_{22}, \beta_2) = (0.001, 0.01, 0.1, 0.85)$. Note that the model allows for cross-effects through the coefficients α_{ij}. For the sake of comparison, we will also considered the Gaussian case: when the joint distribution of the innovations is known to be Gaussian with unknown correlation ρ, the coefficient u is entirely determined by ρ and can be estimated in a parametric way. See appendix F for the asymptotic distribution of this parametric estimator and the resulting asymptotic CI.

We look at the coverage rate (CR) of the CIs and their interval width (IW) as defined below:

$$IW(\alpha, \alpha') := \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\widehat{CoVaR}_{i,n+1}^{h}(\alpha,\alpha') - \widehat{CoVaR}_{i,n+1}^{l}(\alpha,\alpha')}{\widehat{CoVaR}_{i,n+1}(\alpha,\alpha')}\right|$$

where $\widehat{CoVaR}_{i,n+1}^{h}(\alpha,\alpha')$ refers to the upper bound of the $(1-\alpha_0)$ CI of the first out-of-sample observation $(n+1)$ of the i^{th} draw and $\widehat{CoVaR}_{i,n+1}^{l}(\alpha,\alpha')$ to the lower bound; $\widehat{CoVaR}_{i,n+1}(\alpha,\alpha')$ refers to the estimator of the CoVaR of the out-of-sample log-return of the i^{th} draw. We calibrate the model over the in-sample period of sizes $n = 1,000$ and $n = 3,000$ respectively, and compute the CR on the observation $n+1$. Doing so, the out-of-sample observations of the N different draws are independent and their coverage rate should therefore converge to the considered risk level α_0.

Table 2 reveals that CIs relying on the naive Gaussian approach work poorly for this model with Student innovations compared to the CIs based on the asymptotic and bootstrap approaches that have coverage rates closer to the confidence level 95%. The IWs of the semi-parametric and bootstrap approaches are equivalent, but the IWs of the Gaussian approach are clearly too small, leading to an overestimation of the accuracy of this CoVaR estimation approach.

6.2 Backtests

We ran a Monte Carlo experiment of $N = 2,000$ independent simulations, consisting in drawing for each simulation a bivariate GARCH model (13), estimating the $CoVaR(\alpha, \alpha')$ for each

Table 2: Coverage rates and interval widths for the 95% CI of the CoVaR(10%, 20%) estimated by the three methods - Data are generated by GARCH models with Student Innovations

	CR (G)	IW (G)	CR (A)	IW (A)	CR (B)	IW (B)
$n = 1,000$	0.759	0.181	0.932	0.343	0.946	0.346
$n = 3,000$	0.709	0.101	0.949	0.201	0.953	0.205
$n = 5,000$	0.641	0.079	0.951	0.158	0.961	0.166

CR stands for *coverage rate* and IW for *interval width*. "Gaussian" (G), "Asymptotic" (A) and "Bootstrap" (B) refer to the methods of Sections F.3 in the online supplemental document, 4.1 and 4.2 respectively

Table 3: Backtesting results on simulated data with bivariate Student distribution of $\nu = 6$ degrees of freedom ρ=0.6, and $\alpha_0 = 5\%$

	UC Test (Naive Gaussian)	UC Test (Semi-Para.)	Ind Test (Naive Gaussian)	Ind Test (Semi-Para.)
$R = 1,000$	0.185	0.086	0.200	0.086
$R = 3,000$	0.161	0.056	0.182	0.055
$R = 5,000$	0.163	0.056	0.183	0.061

simulation and running backtests on each simulation. In this section, we are testing the H_0 hypothesis presented in Section 5.2. Each sample is split into two periods: the *in-sample* period ($R = 1,000$, $R = 3,000$, $R = 5,000$ observations) on which the parameters are estimated, and then the *out-of-sample* period ($P = 250$ observations) on which the backtest is performed at the nominal level $\alpha_0 = 5\%$. Table 3 presents the results of the backtest when $\alpha = 0.05$, $\alpha' = 0.1$ and the innovation follows a bivariate standardized Student model with 6 degrees of freedom, for the (naive) parametric Gaussian and semi-parametric CoVaR estimators. We found that the distributions of the two likelihood ratio test statistics were very far from their χ^2_1 approximation. Therefore, we used the exact law of each statistic under the null (which is a discrete law) with a randomisation procedure to obtain an exact level α_0. Clearly, the Gaussian approach leads to a misspecified estimator (for this Student model).

We ran the same backtest on a similar model replacing Student by Gaussian innovations and obtained close results between the parametric Gaussian and the semi-parametric approaches with respect to their coverage rates.

7 Empirical studies

In this section, we apply our asymptotic results to financial and macroeconomic data and we illustrate the potential usefulness of estimating dynamic SRM for risk assessment.

Table 4: Backtesting results for H_0 (Frequency of Null Rejection over a selection of 426 US Stocks) at the nominal level $\alpha_0 = 5\%$ with the S&P 500 as the conditioning asset

Model	Unconditional Coverage Test	Independence Test
Parametric Gaussian	0.110	0.110
Semi-Parametric	0.066	0.063

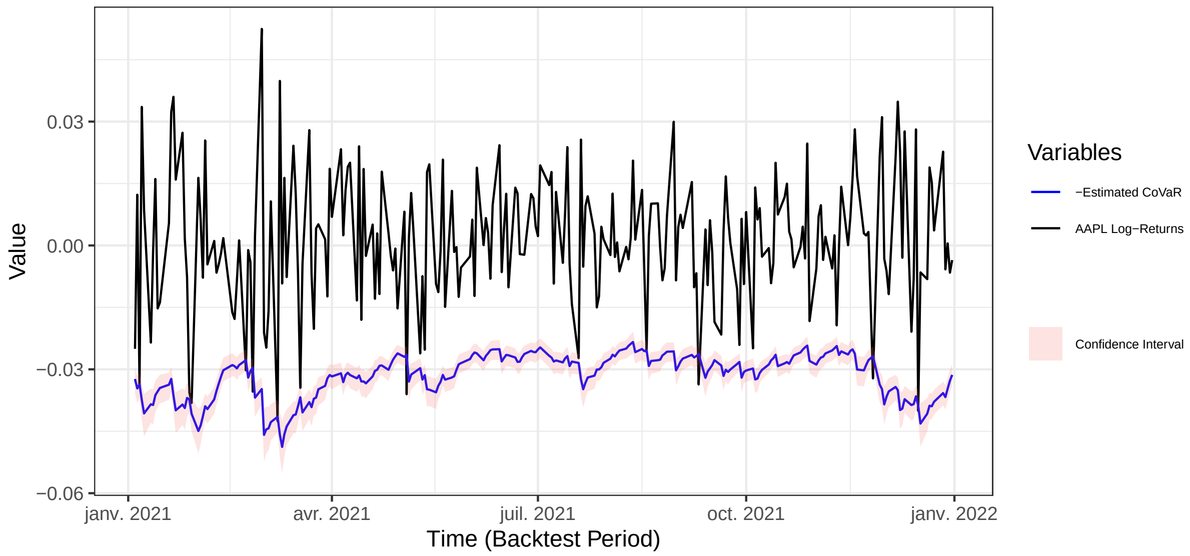


Figure 1: Bootstrapped 90%-CI of AAPL's CoVaR(10%, 20%), conditional on the S&P 500 being at risk

7.1 Financial applications

7.1.1 Daily stock returns

We first compare the CI estimation methods for the series of daily log returns of 426 US stocks over a 13-year period (3,000 in-sample data points vs. 250 out-of-sample observations). The daily log return of the S&P 500 is used as the "conditioning series"[9] to compute the CIs for CoVaR(α, α') of each stock, using the bivariate GARCH model (13). As an illustration, Figure 1 shows the estimated 90% CI obtained using the bootstrap method for the CoVaR(10%, 20%) of Apple Inc.'s (AAPL) daily log returns. We do not present the results for the other two methods and for the other stocks, but it turns out that the bootstrap and asymptotic approaches give very similar results, while the Gaussian approach generally gives smaller confidence intervals, with the risk of too low coverage rates, as shown on simulated data (see Table 2).

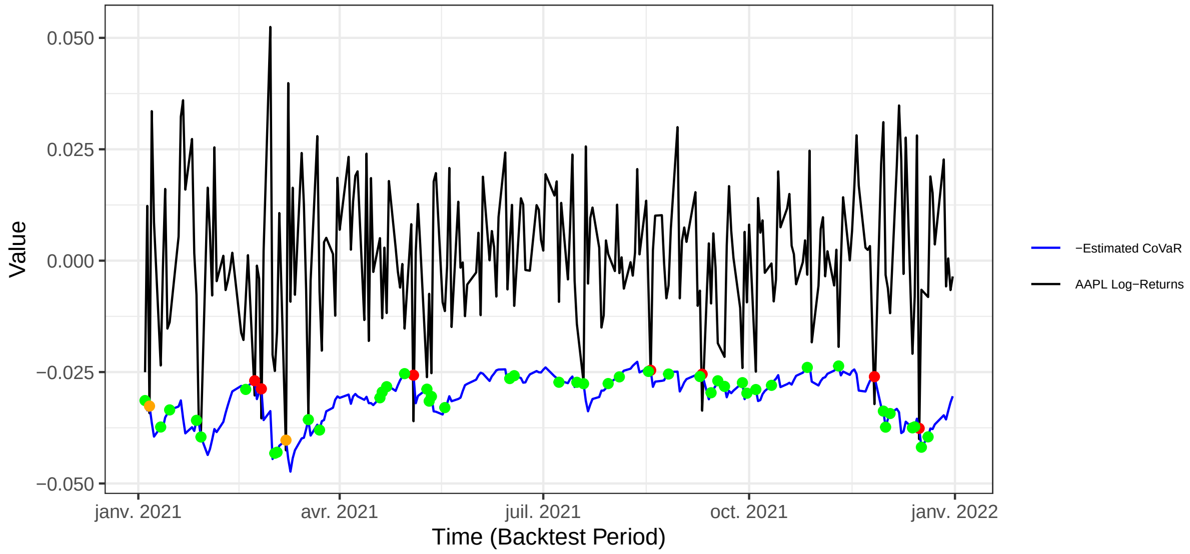
Again for illustrative purposes, Figure 2 shows the exceedances of CoVaR(10%, 20%) and VaR(20%) on the AAPL-S&P 500 example. The observed exceedances are in line with the nominal levels and no clear dependence can be observed.

Now we illustrate the backtesting of the hypothesis H_0 of independent exceedances (see Section 5.2). Table 4 shows that the semi-parametric model performs much better than the Gaussian parametric model for estimating the CoVaR(5%, 10%) of these stocks, which is consistent with the stylized fact described in the literature that log returns are not Gaussian. The H_0' test (results not shown here) shows similar results, with the semi-parametric model outperforming the Gaussian.

7.1.2 Ranking financial institutions by systemic impact

The aim of this application is to show how "systematically important financial institutions" (SIFI) can be identified from the viewpoint of dynamic SRM. For this application, we relied on Model (13) with spillover effects in volatilities. We study the CoVaR of the "financial system"

[9] The practitioner may be interested in knowing the impact of a potential market decline - for which the S&P is a proxy - on a particular stock or portfolio.



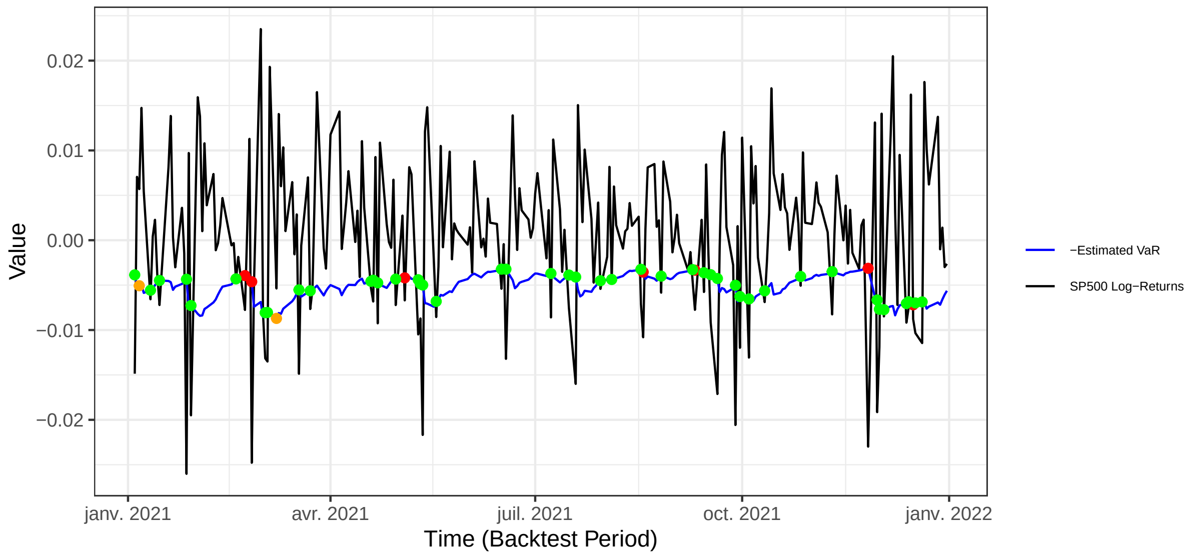


Figure 2: -CoVaR(10%, 20%) of AAPL conditional on S&P 500 daily log-returns (top graph) and -VaR(20%) of S&P 500 (bottom graph). Red: exceedances of CoVaR (process 1) and VaR (process 2); green: exceedances of VaR while no exceedances of CoVaR; orange: exceedances of CoVaR while no exceedances of VaR. The number of violations of the VaR (44+7 = 51 red + green bullets) and CoVaR (7 red bullets) are close to the nominal numbers $T\alpha' = 50$ and $T\alpha\alpha' = 250 \times 0.1 \times 0.2 = 5$, respectively.

Table 5: Ranking of Financial Institutions by Systemic Impact (i.e. $\overline{\Delta CoVaR}$)

Company Name	Region	Rank wrt. ρ	Rank wrt. $\overline{CoVaR}$	Rank wrt. $\overline{\Delta CoVaR}$
Citigroup Incl	US	4	2	1
Morgan Stanleyl	US	2	1	2
Bank of America Corp	US	3	3	3
JPMorgan Chase & Co	US	5	4	4
Wells Fargo & Co	US	7	5	5
BlackRock Inc	US	1	7	6
Goldman Sachs Group Inc	US	9	13	7
Berkshire Hathaway Inc Class B	US	8	10	8
Allianz	Germany	11	9	9
Toronto Dominion Bank	Canada	6	12	10
American Express Co	US	10	11	11
Charles Schwab Corp	US	12	14	12
Royal Bank of Canada	Canada	13	6	13
Marsh & McLennan Inc	US	16	8	14
S&P Global Inc	US	18	17	15
Chubb Ltd	US	14	15	16
Progressive Corp	US	15	16	17
HSBC Holdings PLC	UK	17	18	18
Commonwealth Bank of Australia	Australia	19	19	19

conditional on the distress of some major financial institutions. As a proxy for the financial system, we consider a benchmark Exchange Traded Fund (ETF): the *iShares Global Financials* from Blackrock. We then study the CoVaR of the log-returns of the ETF using for the conditioning events the log-returns of 19 major companies composing the index fund as of Oct. 2022. We consider the average (both in time and for different risk levels) CoVaR and ΔCoVaR as a summary of dynamic SRM [10] as opposed to the naive measure implied by the correlation coefficient between returns. The quantities are computed over the $n = 3,000$ observations period from 2009/02/05 to 2021/01/05. Table 5 provides rankings of 19 financial institutions in terms of their systemic impact, based on the average SRM and the correlation coefficient. It reveals a new picture on systemic risk, by comparison with the usual correlation coefficient (for instance Citigroup is ranked 1^{st} according to $\overline{\Delta CoVaR}$ and 4th according to ρ, Blackrock is ranked 1^{st} according to ρ anf 6th according to $\overline{\Delta CoVaR}$).

7.1.3 Dynamic comparison between systemic and individual risks of large banks

For this application, we considered a dataset consisting of the daily log returns of 18 of the world's largest banks between January 2, 2004 and December 1, 2015 ($n = 3,000$ observations), covering both the global financial crisis and the Eurozone crisis. For the "system" we considered an index built such that the returns are given by the average daily log-returns of the 18 banks abovementioned weighted by their daily market capitalization. This index is therefore a good proxy for the performance of the overall banking industry during the period. The regulator may want to

[10]
$$\overline{CoVaR}^{ETF|SFI_j} := \frac{1}{N_{cov}n} \sum_{(\alpha,\alpha')\in E_{\alpha,\alpha'}} \sum_{t=1}^{n} CoVaR_t^{ETF|SFI_j}(\alpha,\alpha'), \quad j = 1,\dots 19 \quad \text{and}$$
$$\overline{\Delta CoVaR}^{ETF|SFI_j} := \frac{1}{N_{\Delta cov}n} \sum_{(\alpha,\alpha',\alpha'')\in E_{\alpha,\alpha',\alpha''}} \sum_{t=1}^{n} \Delta CoVaR_t^{ETF|SFI_j}(\alpha,\alpha',\alpha'')$$

with $E_{\alpha,\alpha'}$ and $E_{\alpha,\alpha',\alpha''}$ referring to the sets of selected vectors of confidence levels (α,α') and $(\alpha,\alpha',\alpha'')$ with respective cardinals N_{cov} and $N_{\Delta cov}$. We considered $E_{\alpha,\alpha'} = \{(0.5, 0.5), (0.1, 0.2), (0.05, 0.1), (0.01, 0.05)\}$ and $E_{\alpha,\alpha',\alpha''} = \{(0.5, 0.5, 0.25), (0.1, 0.2, 0.25), (0.05, 0.1, 0.25), (0.01, 0.05, 0.25)\}$.

study the $\Delta CoVaR$ of the financial system conditional on the fact that a constituent factor is at risk in order to know which institutions are the most systemic (i.e. to which institutions' distress the log-returns tail distribution of the system is most sensitive). Therefore, it is natural to map financial organizations according to their specific risk (measured by $\mathrm{VaR}_t^{\epsilon_2}(\alpha')$) and to their systemic risk (measured by $\Delta CoVaR_t^{\epsilon_1|\epsilon_2}(\alpha, \alpha', \alpha'')$) to assess their relative risk for the system. It is common to interpret the VaR as a level of capital reserve that a financial organization keeps to face possible losses. The purpose of Figure 3 is thus to identify factors that tend to have a higher systemic risk (higher ΔCoVaR)[11] relative to their reserves calibrated on VaR levels. Similar charts have been presented in Adrian and Brunnermeier (2011) and in Girardi and Ergün (2013). Since the proxy for the system is a portfolio of the major banks weighted by their daily market capitalization, large banks are more likely to have a strong effect on the system, but the size effect does not capture the whole risk. As our model is dynamic, we can draw such scatterplots for every time t in the period. We took for Figure 3 the average VaR and $\Delta CoVaR$ over the period. One can read on these charts that the most risky institutions for the system are located at the top left and the least risky ones relative to their respective VaR level are located at the bottom right. The fact that Credit Suisse, which was rescued from near bankruptcy by rival UBS in March 2023, is not one of the top banks for systemic risk according to Figure 3 may seem like good news. We can also see that US banks are on average more systemic according to this criteria and that the relation between the $\Delta CoVaR$ of the system and the VaR of the institution in a period of market stress is close to linear. To appreciate the dynamics of the model, we look at the ratio ΔCoVaR / $\mathrm{VaR}^{(2)}$ through time with ΔCoVaR referring to the ΔCoVaR of the system conditioning to the selected financial institutions and $VaR^{(2)}$ refers to the VaR of the financial institution. Figure 4 represents a 21-day moving average of this ratio through time during the same period as presented above. The higher the ratio, the larger systemic risk is borne by the financial institution compared to its VaR (i.e. its capital requirements). We notice in particular a strong increase in the ratio for GS, which aligned itself with JPM during the 2008 crisis and which never came down to its initial level, unlike Crédit Agricole, which recovered its pre-crisis level more quickly.

7.2 Systemic sovereign risk estimation using CDS spreads data

Additionally to their obvious financial applications, systemic risk models can provide economic insights (see *e.g.* Morelli, Ottonello and Perez, 2022). They can for example be used to assess the solidity of a country's economy at certain period of time[12] by analyzing CDS spreads on sovereign debt. For more information on the use and interpretation of CDS spreads, we refer to Oh and Patton (2018). As is well known, an increase in returns on CDS spreads on sovereign bonds translates into a rise of the default risk of the underlying country. The objective of this application is to identify which country, in case of distress, has the strongest impact on the default risk of a reference country. We considered the 5-year sovereign CDS spreads of 10 selected Eurozone countries plus the UK between 2013 and 2021 ($n = 3,000$ observations) and got the data from Bloomberg. Since we observed shifts (identified as potential outliers) in the Bloombers' data set, we replaced the original data at these shift points with a moving average of order 5. Many other economic studies measured systemic risk through CDS (see *e.g.* Ang and Longstaff, 2013). In this paper, we focus on Euro sovereign debt posterior to the ban on the so-called "naked CDS". Traditional stationarity tests (ADF, KPSS) ensure the validity of

[11] which have stress periods that are the most correlated with those of the market

[12] precisely the perception of the financial markets about the default risk of a country

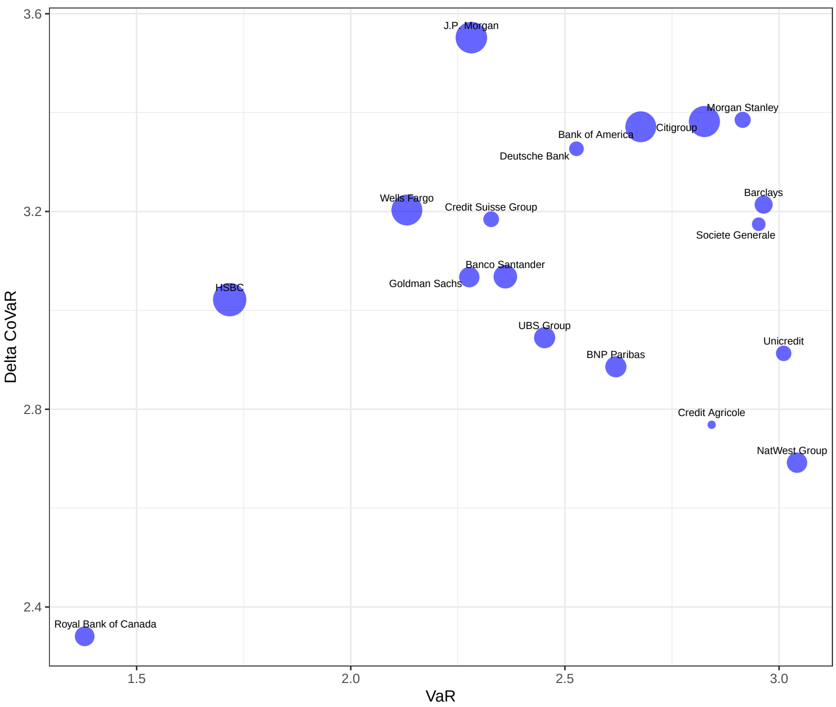


Figure 3: Average ΔCoVaR for the financial system versus average VaR of the conditioning series (major financial institutions) over the considered period

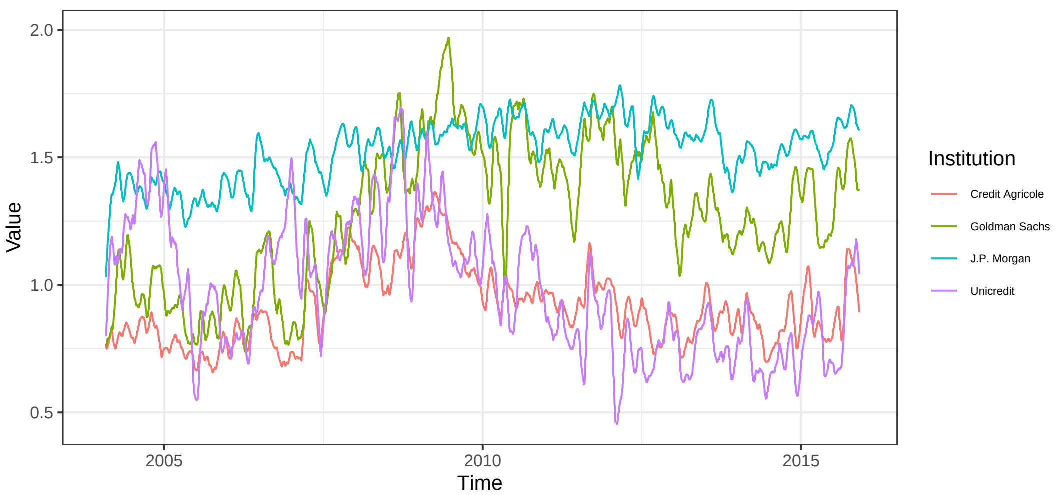


Figure 4: Evolution of the ratio ΔCoVaR of the system / VaR of the institution over time (3,000 observations) | risk levels: $\alpha = 5\%$, $\alpha' = 10\%$, $\alpha'' = 20\%$ - rolling period of 21 days

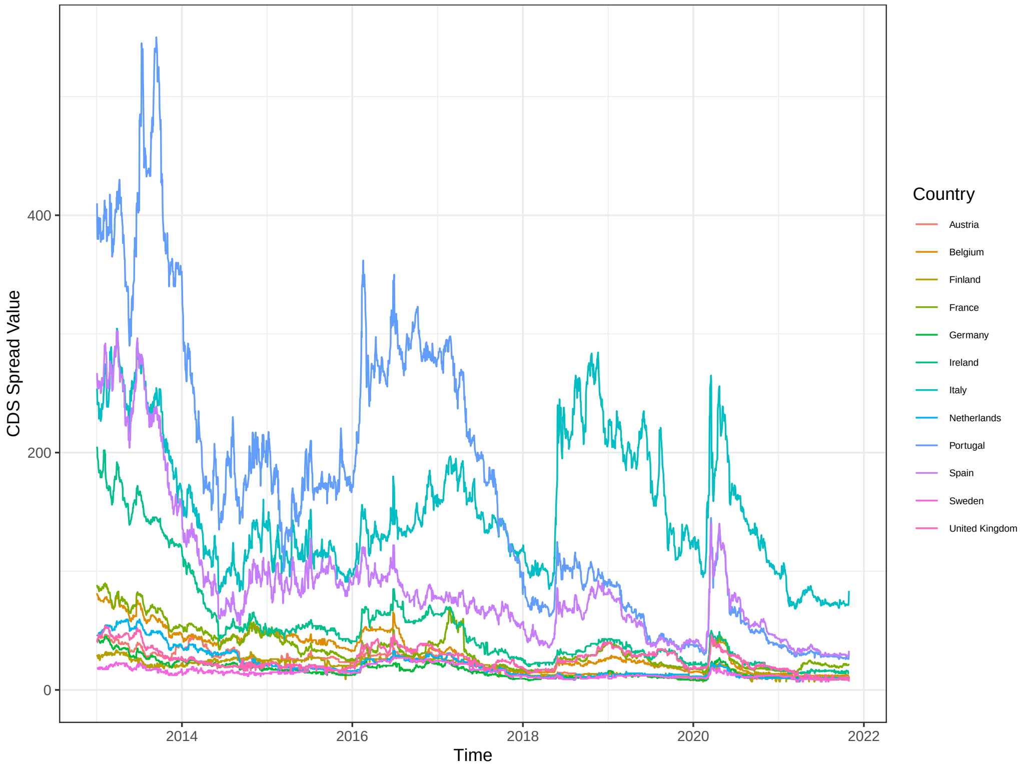


Figure 5: CDS spreads of selected Eurozone countries between 2013 and 2021

the model for CDS log-returns and suggest the need to include an AR component:

$$y_{i,t} \;\; = \;\; \mu_i + \phi_i y_{i,t-1} + \epsilon_{i,t}, \quad \epsilon_{i,t} = \sigma_{it}\eta_{it}, \quad \sigma^2_{it} = \omega_i + \alpha_i \epsilon^2_{i,t-1} + \beta_i \sigma^2_{i,t-1}, \quad i = 1, 2$$

with $y_{i,t}$ referring to the *opposite* daily log-returns of the CDS under consideration. Figure 5 shows the evolution of the CDS spreads of the 5-year sovereign bonds for selected Eurozone countries[13]. As the Portugal CDS spreads have the highest values over time, we illustrate in Figure 6 the effect of a rise in the default risk of some European countries on the increase of the default risk of Portugal. The higher the value of $-u(\cdot,\cdot)$ the more likely it is that the distress of one country accentuates the default risk of the reference country. It is also important to notice that in this application (unlike in the previous one) the volatility of series 1 is only function of its own past values [14] and therefore the quantities $u(\alpha, \alpha')$ computed with respect to each conditioning series are directly comparable. Figure 6 shows the systemic risk with respect to the different countries, as measured by $-\hat{u}(0.5, 0.5)$. We used the asymptotic distribution in Theorem 3.1 to represent interquartile ranges. It allows us to identify three groups of countries whose increased default risk could contribute to Portugal's default risk with different intensities. We observe a strong dependence between the Portuguese economy and the economies of Italy and Spain, which have the highest percentage of public debt to GDP in the Eurozone. France also has a high debt level and is in a group with Ireland and Belgium, while Northern European countries with historically lower debt levels have sovereign CDS spreads log-returns less correlated to those of Portugal.

[13] We excluded Greece and other countries because of the presence of missing data

[14] In the first application it was not the case, due to the presence of the spillover coefficient $\alpha_{i,j}$ in Model (13).

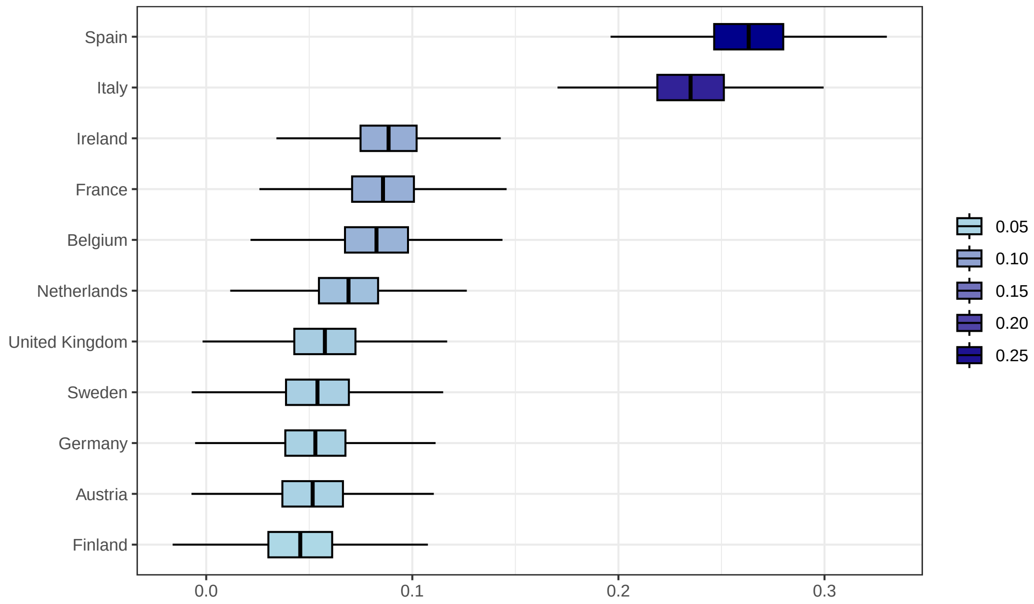


Figure 6: -$\hat{u}(0.5, 0.5)$ as a measure of tail dependence of Portugal 5-year sovereign CDS spreads conditional on the CDS spread log-returns of 10 Eurozone countries (+UK)

8 Concluding remarks

In this paper, we proposed an econometric approach for estimating dynamic SRM in a semi-parametric framework. In this setting, dynamic SRM are explicit functions of the first two conditional moments of one risk factor and simple joint characteristics of the innovations distribution. The derivation of the asymptotic distributions of the QML estimators of the CoVaR and ΔCoVaR was achieved under general assumptions on the volatility processes. We also showed the validity of a residual bootstrap approach for constructing asymptotic CIs for the dynamic SRM, characterizing the *estimation risk* often neglected in empirical studies. We proposed backtests for the CoVaR, generalizing the tests introduced for the VaR and other risk measures. Applying these tests on the S&P 500 components revealed the superiority of semi-parametric over Gaussian CoVaR estimators. We also showed how the dynamic SRM can be used to rank SIFIs. According to our study on the global banking system, American banks carry more systemic risk than European banks. Our empirical study based on the CDS of European countries highlights the importance of dynamic interdependencies between economies, visible through extreme multivariate quantiles and more general SRM.

Acknowledgements

For helpful and constructive comments we thank the Co-Editor, Associate Editor and two anonymous referees. We are grateful to the Agence Nationale de la Recherche (ANR), France, which supported this work via the project MLEforRisk (ANR-21-CE26-0007). We thank Julien Royer and Yujian Chen for providing us data sets used in the empirical studies. Finally, we warmly thank LoÃŕc Cantin who contributed as a co-author to the first version of this work, but was, unfortunately, unavailable for contact during the revision process. His contribution was essential in particular for the numerical parts of the paper. However, these parts had to be completely

reworked for the final version of the paper.[15]

A Assumptions

Let $K > 0$ be a generic constant, or a random variable that is measurable with respect to $\mathcal{F}_0$. Let $\rho \in (0,1)$. Our consistency result requires the following assumptions. For $i = 1, 2$:

A1: $(\boldsymbol{\epsilon}_t)$ is a strictly stationary and ergodic process, and η_{it} is independent from $\{\boldsymbol{\epsilon}_{t-u}, u > 0\}$.

A2: The function $\boldsymbol{\theta}^{(i)} \mapsto (\mu_i(x_1, x_2, \ldots; \boldsymbol{\theta}^{(i)}), \sigma_i(x_1, x_2, \ldots; \boldsymbol{\theta}^{(i)}))$ is continuously differentiable, for any real sequence $(x_j)_{j \geqslant 1}$. Almost surely, $\sigma_{it}(\boldsymbol{\theta}^{(i)}) \in (\underline{\omega}, \infty]$ for any $\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}$ and for some $\underline{\omega} > 0$. Moreover, $\mu_{it}(\boldsymbol{\theta}_0^{(i)}) = \mu_{it}(\boldsymbol{\theta}^{(i)})$ and $\sigma_{it}(\boldsymbol{\theta}_0^{(i)}) = \sigma_{it}(\boldsymbol{\theta}^{(i)})$ $a.s.$ iff $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}_0^{(i)}$.

A3: $\sup_{\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}} (|\mu_{it}(\boldsymbol{\theta}^{(i)}) - \widetilde{\mu}_{it}(\boldsymbol{\theta}^{(i)})| + |\sigma_{it}(\boldsymbol{\theta}^{(i)}) - \widetilde{\sigma}_{it}(\boldsymbol{\theta}^{(i)})|) \leqslant K\rho^t$. Moreover, $E \sup_{\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}} |\mu_{it}(\boldsymbol{\theta}^{(i)})|^r < \infty$ and $E \sup_{\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}} |\sigma_{it}(\boldsymbol{\theta}^{(i)})|^r < \infty$ for some $r > 0$.

A4: For $r > 0$, there exists a neighborhood $V(\boldsymbol{\theta}_0^{(i)})$ of $\boldsymbol{\theta}_0^{(i)}$ such that

$$E \sup_{\boldsymbol{\theta}^{(i)} \in V(\boldsymbol{\theta}_0^{(i)})} \left\{ \left(\frac{\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}{\sigma_{it}(\boldsymbol{\theta}^{(i)})} \right)^r + \left\| \frac{\partial \mu_{it}(\boldsymbol{\theta}^{(i)})}{\partial \boldsymbol{\theta}^{(i)}} \right\|^r + \left| \mu_{it}(\boldsymbol{\theta}_0^{(i)}) - \mu_{it}(\boldsymbol{\theta}^{(i)}) \right|^r \left\| \boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)}) \right\|^r \right\} < \infty,$$

where $\boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)}) = \sigma_{it}^{-1}(\boldsymbol{\theta}^{(i)}) \partial \sigma_{it}(\boldsymbol{\theta}^{(i)}) / \partial \boldsymbol{\theta}^{(i)}$.

The first three conditions ensure the strong consistency of the QML estimator of $\boldsymbol{\theta}_0^{(i)}$, for $i = 1, 2$. Assumption **A1** is standard. The existence of a moment of small order r is automatically satisfied under the strict stationarity condition for standard GARCH models, and many of their extensions, but since we do not assume a specific parametric form for the conditional moments, we need to make the assumption. The last part of **A2** is made for identifiability reasons and can be verified directly on specific models. The requirement of a lower bound for the volatility function is due to the estimation criterion, which introduces the volatility in the denominator and in the log. Assumption **A3** is introduced to neglect asymptotically the effect of the initial values, while Assumption **A4** is used to control the difference between the innovations and residuals. In Section D of the online supplemental document, we show how these assumptions can be simplified in the case of Model (13).

We also introduce the following assumption, for $i = 1, 2$, which is simply denoted $\mathbf{A5_i}$ when it holds for all $x \in \mathbb{R}$, and **A5** when it holds for all i and x.

$\mathbf{A5_i}(x)$: The cdf $G^{(i)}$ of η_{it} is Lipschitz continuous in a neighborhood of x.

Under the following assumption, the quantile function of η_{2t} and the co-quantile function $F^{-}(\cdot \mid \xi_{\alpha'}^{(2)})$ of η_{1t} are right-continuous at α' and α, respectively.

$\mathbf{A6}(\alpha, \alpha')$: For $\alpha, \alpha' \in (0, 1)$, (i) the cdf of η_{2t} satisfies $G^{(2)}(y) > \alpha'$ whenever $y > \xi_{\alpha'}^{(2)}$ and (ii) we have $F(x \mid \xi_{\alpha'}^{(2)}) > \alpha$ whenever $x > u(\alpha, \alpha')$.

The next assumption requires continuity of the co-cdf with respect to the conditioning event, uniformly w.r.t. the first component.

[15] The data sets and R codes can be found at
https://github.com/christianfrancq/Dynamic-Systemic-Risk-Measures.git

A7(y_0): We have $\sup_{x\in\mathbb{R}}|F(x|y)-F(x|y_0)|\to 0$ when $y\to y_0$.

The consistency of $\widehat{v}(\alpha')$ requires the following additional assumptions.

A8: The cdf $H(\cdot,\cdot)$ is continuous, where $H(x,y)=\mathbb{P}\{\eta_{1t}\leqslant x,\eta_{2t}\leqslant y\}$.

A9: There exists a neighborhood $V(\boldsymbol{\theta}_0^{(1)})$ of $\boldsymbol{\theta}_0^{(1)}$, and for $j=1,2$, numbers $p_j>0$, $q_j>0$ and $r_j>0$ with $p_j^{-1}+q_j^{-1}+r_j^{-1}=1$, such that

$$E\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\left\{\left\|\frac{\partial\ln\sigma_{1t}(\boldsymbol{\theta}^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right\|^{p_1}+\left|\frac{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})}\right|^{2q_1}+\left|\frac{\mu_{1t}(\boldsymbol{\theta}_0^{(1)})-\mu_{1t}(\boldsymbol{\theta}^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}\right|^{2r_1}\right\}<\infty,$$

and

$$E\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\left\{\left\|\frac{1}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})}\frac{\partial\mu_{1t}(\boldsymbol{\theta}^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right\|^{p_2}+\left|\frac{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})}\right|^{q_2}+\left|\frac{\mu_{1t}(\boldsymbol{\theta}_0^{(1)})-\mu_{1t}(\boldsymbol{\theta}^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}\right|^{r_2}\right\}<\infty.$$

If, in the formula for $\widehat{v}(\alpha')$, the residuals were replaced by innovations (i.e. η_{1t} instead of $\widehat{\eta}_{1t}$, and $\xi_{\alpha'}^{(2)}$ instead of $\widehat{\xi}_{\alpha'}^{(2)}$) the convergence to $v(\alpha')$ would hold by the strong law of large numbers without any continuity assumption on the joint cdf. However, the following example shows that without Assumption **A8**, the convergence in Proposition 2.5 may not hold. Consider the static scale model, $\epsilon_{it}=\sigma_i\eta_{it},\sigma_i>0,i=1,2$, where the variable η_{2t} has a finite-support distribution. Let $\underline{x}=\min\{x,\mathbb{P}(\eta_{2t}=x)>0\}$ and assume that $E[\eta_{1t}\mathbb{1}_{\eta_{2t}=\underline{x}}]\neq 0$. Let $\widehat{\sigma}_i$ be a consistent estimator of $\sigma_i, i=1,2$. The residuals being given by $\widehat{\eta}_{it}=\frac{\sigma_i}{\widehat{\sigma}_i}\eta_{it}$, we have $\frac{1}{n}\sum_{t=1}^n\widehat{\eta}_{1t}\mathbb{1}_{\widehat{\eta}_{2t}\leqslant x}=\frac{\sigma_1}{\widehat{\sigma}_1}\frac{1}{n}\sum_{t=1}^n\eta_{1t}\mathbb{1}_{\eta_{2t}\leqslant\frac{\widehat{\sigma}_2}{\sigma_2}x}$. The latter sum is equal to 0 when $x=\underline{x}$ and $\sigma_2>\widehat{\sigma}_2$ (which holds with a non-vanishing probability for instance with the QML). Hence, for $x=\underline{x}$, $\frac{1}{n}\sum_{t=1}^n\widehat{\eta}_{1t}\mathbb{1}_{\widehat{\eta}_{2t}\leqslant x}$ does not converge to $E[\eta_{1t}\mathbb{1}_{\eta_{2t}=x}]$ with probability 1 as $n\to\infty$. So, $\widehat{v}(\alpha')\nrightarrow v(\alpha')$ when $\xi_{\alpha'}^{(2)}=\underline{x}$.

The following assumptions are used for the asymptotic normality results of Section 3.

B1$_\mathbf{i}$: $\boldsymbol{\theta}_0^{(i)}$ belongs to the interior of $\boldsymbol{\Theta}^{(i)}$.

B2$_\mathbf{i}$: There exist no non-zero $\boldsymbol{x}\in\mathbb{R}^{d_i}$ such that $\boldsymbol{x}'\frac{\partial\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}{\partial\boldsymbol{\theta}^{(i)}}=0$, $\quad a.s.$

B3$_\mathbf{i}$: The function $\boldsymbol{\theta}^{(i)}\mapsto\sigma(x_1,x_2,\dots;\boldsymbol{\theta}^{(i)})$ has continuous second-order derivatives, and $\sup_{\boldsymbol{\theta}^{(i)}\in\boldsymbol{\Theta}^{(i)}}\left\|\frac{\partial\sigma_{it}(\boldsymbol{\theta}^{(i)})}{\partial\boldsymbol{\theta}^{(i)}}-\frac{\partial\widetilde{\sigma}_t(\boldsymbol{\theta}^{(i)})}{\partial\boldsymbol{\theta}^{(i)}}\right\|\leqslant K\rho^t$.

B4$_\mathbf{i}$: There exists a neighborhood $V(\boldsymbol{\theta}_0^{(i)})$ of $\boldsymbol{\theta}_0^{(i)}$ such that

$$E\sup_{\boldsymbol{\theta}^{(i)}\in V(\boldsymbol{\theta}_0^{(i)})}\left\{\left\|\boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)})\right\|^4+\left\|\frac{1}{\sigma_{it}(\boldsymbol{\theta}^{(i)})}\frac{\partial^2\sigma_{it}(\boldsymbol{\theta}^{(i)})}{\partial\boldsymbol{\theta}^{(i)}\partial\boldsymbol{\theta}^{(i)\prime}}\right\|^2+\left|\frac{\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}{\sigma_{it}(\boldsymbol{\theta}^{(i)})}\right|^4+\left|\frac{\sigma_{it}(\boldsymbol{\theta}^{(i)})}{\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}\right|^4\right\}<\infty.$$

Moreover, $\kappa_i=E(\eta_{it}^4)<\infty$.

B5$_\mathbf{i}$: All the coordinates of $\frac{\partial\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}{\partial\boldsymbol{\theta}^{(i)}}$ are a.s. (strictly) positive.

B6: The vector $(\eta_{1t},\eta_{2t})'$ admits a continuous density with respect to the Lebesgue measure on $\mathbb{R}^2$. Let $g^{(i)}$ denote the density of η_{it}.

It is clear that under **B6**, Assumptions **A5** $-$ **A8** are satisfied.

The following assumption is used to simplify the asymptotic distributions of the VaR and CoVaR estimators.

B7$_\mathbf{i}$: For any $\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}$, for any $c > 0$, and any sequence (x_j), there exists $\boldsymbol{\theta}_c^{(i)}$ such that $c\sigma_i(x_1, x_2, \ldots; \boldsymbol{\theta}^{(i)}) = \sigma_i(x_1, x_2, \ldots; \boldsymbol{\theta}_c^{(i)})$.

To establish the asymptotic distribution of $\widehat{u}(\alpha, \alpha')$, we need the following assumption.

B8 : The function $(x, y) \mapsto F(x \mid y)$ is of class C^1 in a neighborhood of $\left(u(\alpha, \alpha'), \xi_{\alpha'}^{(2)}\right)$ and the density $f_1(\cdot \mid \xi_{\alpha'}^{(2)})$ of η_{1t} given $\eta_{2t} \leqslant \xi_{\alpha'}^{(2)}$ is strictly positive in a neighborhood of $u(\alpha, \alpha')$. Moreover, the density $g^{(2)}$ of η_{2t} is strictly positive in a neighborhood of $\xi_{\alpha'}^{(2)}$.

The next assumption is used to define the CI of Section 4.1.

B9 : The estimates $\widehat{f}_1(\widehat{u}|\widehat{\xi})$, $\widehat{f}_2(\widehat{\xi}|\widehat{u})$ and $\widehat{g}^{(2)}(\widehat{\xi})$ strongly converge to $f_1(u|\xi)$, $f_2(\xi|u)$ and $g^{(2)}(\xi)$ as $n \to \infty$.

The next assumption, used for the validity of the bootstrap, slightly reinforces **B4$_\mathbf{1}$**.

B10: There exists a neighborhood $V(\boldsymbol{\theta}_0^{(1)})$ of $\boldsymbol{\theta}_0^{(1)}$, numbers $p > 0$ and $q > 0$ with $p^{-1} + q^{-1} = 1$, such that

$$E \sup_{\boldsymbol{\theta}^{(1)} \in V(\boldsymbol{\theta}_0^{(1)})} \left\{ \left\| \frac{1}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})} \frac{\partial \sigma_{1t}(\boldsymbol{\theta}^{(1)})}{\partial \boldsymbol{\theta}^{(1)}} \right\|^p + \left| \frac{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})} \right|^{4q} \right\} < \infty.$$

B Asymptotic properties of the empirical co-cdf

Let $\widehat{G}^{(i)}(x) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{it} \leqslant x}$ and let the estimator of $F(x|y)$ defined by

$$\widehat{F}(x|y) = \frac{\widehat{H}(x, y)}{\widehat{G}^{(2)}(y)}, \quad \widehat{H}(x, y) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{1t} \leqslant x, \widehat{\eta}_{2t} \leqslant y}, \quad \widehat{G}^{(2)}(y) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{2t} \leqslant y} \tag{14}$$

provided the denominator is not equal to zero.

The following result is an extension of the Glivenko-Cantelli theorem for residuals and conditional distributions.

Proposition B.1. *Assume* **A1**-**A4**. *Let $x, y, y_0 \in \mathbb{R}$, with $G^{(2)}(y) > 0$ and $G^{(2)}(y_0) > 0$.*

(ii) If **A5$_\mathbf{1}$**(x) *and* **A5$_\mathbf{2}$**(y) *hold, we have* $|\widehat{F}(x|y) - F(x|y)| \to 0$ *a.s.*

(iiii) If **A5$_\mathbf{1}$** *and* **A5$_\mathbf{2}$**(y) *hold, we have* $\sup_{x \in \mathbb{R}} |\widehat{F}(x|y) - F(x|y)| \to 0$ *a.s.*

(iii) If **A5$_\mathbf{1}$** *and* **A5$_\mathbf{2}$**(y_0) *hold, there exists a neighborhood* $V(y_0)$ *of* y_0 *such that* $\sup_{x \in \mathbb{R}, y \in V(y_0)} |\widehat{F}(x|y) - F(x|y)| \to 0$ *a.s.*

Proof of Proposition B.1. Recall that $\widehat{H}(x, y) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{1t} \leqslant x, \widehat{\eta}_{2t} \leqslant y}$ and $\widehat{G}^{(2)}(x) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{2t} \leqslant x}$. Let $H_n(x, y) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\eta_{1t} \leqslant x, \eta_{2t} \leqslant y}$, $G_n^{(2)}(x) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\eta_{2t} \leqslant x}$ and $F_n(x|y) = H_n(x, y)/G_n^{(2)}(y)$.

The strong consistency of the QML estimator of $\boldsymbol{\theta}_0^{(i)}$, for $i = 1, 2$, can be established as follows. Omitting the index i for ease of exposition, write

$$Q_n(\boldsymbol{\theta}) \quad = \quad \frac{1}{n}\sum_{t=1}^n \frac{\{\epsilon_t - \mu_t(\boldsymbol{\theta})\}^2}{\sigma_t^2(\boldsymbol{\theta})} + \log \sigma_t^2(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \ell_t(\boldsymbol{\theta}),$$

$$\widetilde{Q}_n(\boldsymbol{\theta}) \quad = \quad \frac{1}{n}\sum_{t=1}^{n}\frac{\{\epsilon_t-\widetilde{\mu}_t(\boldsymbol{\theta})\}^2}{\widetilde{\sigma}_t^2(\boldsymbol{\theta})}+\log\widetilde{\sigma}_t^2(\boldsymbol{\theta})=\frac{1}{n}\sum_{t=1}^{n}\widetilde{\ell}_t(\boldsymbol{\theta}).$$

Following the lines of proof of Theorem 7.4 in Francq and Zakoian (2019), the strong consistency will be a consequence of the following intermediate results:

$$i)\ \lim_{n\to\infty}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|Q_n(\boldsymbol{\theta})-\widetilde{Q}_n(\boldsymbol{\theta})|=0, \quad a.s. \qquad ii)\ \boldsymbol{\theta}\neq\boldsymbol{\theta}_0 \quad\Longrightarrow\quad E\ell_t(\boldsymbol{\theta})<E\ell_t(\boldsymbol{\theta}_0),$$

$iii)$ any $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$ has a neighborhood $V(\boldsymbol{\theta})$ such that

$$\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}^*\in V(\boldsymbol{\theta})}\widetilde{Q}_n(\boldsymbol{\theta}^*)<\limsup_{n\to\infty}\widetilde{Q}_n(\boldsymbol{\theta}_0), \quad a.s.$$

We have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|Q_n(\boldsymbol{\theta})-\widetilde{Q}_n(\boldsymbol{\theta})|\leqslant\frac{1}{n}\sum_{t=1}^{n}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\epsilon_t^2\left|\frac{\widetilde{\sigma}_t^2(\boldsymbol{\theta})-\sigma_t^2(\boldsymbol{\theta})}{\widetilde{\sigma}_t^2(\boldsymbol{\theta})\sigma_t^2(\boldsymbol{\theta})}\right|+2\epsilon_t\left|\frac{\widetilde{\mu}_t(\boldsymbol{\theta})}{\widetilde{\sigma}_t^2(\boldsymbol{\theta})}-\frac{\mu_t(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})}\right|\right.$$

$$\left.+\left|\frac{\widetilde{\mu}_t^2(\boldsymbol{\theta})}{\widetilde{\sigma}_t^2(\boldsymbol{\theta})}-\frac{\mu_t^2(\boldsymbol{\theta})}{\sigma_t^2(\boldsymbol{\theta})}\right|+\left|\log\frac{\sigma_t^2(\boldsymbol{\theta})}{\widetilde{\sigma}_t^2(\boldsymbol{\theta})}\right|\right|\leqslant\frac{K}{n}\sum_{t=1}^{n}\rho^t\left(1+|\epsilon_t|+\epsilon_t^2+\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\mu_t^2(\boldsymbol{\theta})|\right),$$

which, using the existence of a small moment for $|\epsilon_t|$ (deduced from **A1** and **A3**), establishes *i)*. The proof of *ii)* is a direct consequence of **A2**, using the arguments of the proof of Theorem 7.4 in Francq and Zakoian (2019). Finally, the proof of *iii)* uses standard compactness and ergodicity arguments. Hence the strong consistency of the QMLE of $\boldsymbol{\theta}_0^{(i)}$, for $i=1,2$,

Now, we have

$$|\widehat{F}(x|y)-F_n(x|y)|\leqslant\frac{G_n^{(2)}(y)|\widehat{H}(x,y)-H_n(x,y)|+H_n(x,y)|G_n^{(2)}(y)-\widehat{G}^{(2)}(y)|}{G_n^{(2)}(y)\widehat{G}^{(2)}(y)}. \tag{15}$$

Given that $|G_n^{(2)}(y)-\widehat{G}^{(2)}(y)|\to 0$ a.s. for all y (see FZ), and $G_n^{(2)}(y)\to G^{(2)}(y)>0$ a.s. by the ergodic theorem, to show $i)$ it suffices to prove that $|\widehat{H}(x,y)-H_n(x,y)|\to 0$ a.s. for all x,y. The result is straightforward because $\left|\mathbb{1}_{\widehat{\eta}_{1t}\leqslant x}\mathbb{1}_{\widehat{\eta}_{2t}\leqslant y}-\mathbb{1}_{\eta_{1t}\leqslant x}\mathbb{1}_{\eta_{2t}\leqslant y}\right|\leqslant\left|\mathbb{1}_{\widehat{\eta}_{1t}\leqslant x}-\mathbb{1}_{\eta_{1t}\leqslant x}\right|+\left|\mathbb{1}_{\widehat{\eta}_{2t}\leqslant y}-\mathbb{1}_{\eta_{2t}\leqslant y}\right|$ entails

$$|\widehat{H}(x,y)-H_n(x,y)|\leqslant\frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{1}_{\widehat{\eta}_{1t}\leqslant x}-\mathbb{1}_{\eta_{1t}\leqslant x}\right|+\frac{1}{n}\sum_{t=1}^{n}\left|\mathbb{1}_{\widehat{\eta}_{2t}\leqslant y}-\mathbb{1}_{\eta_{2t}\leqslant y}\right|$$

which goes to 0 a.s. by the proof of Theorem 3.1 in FZ. Result $ii)$ is also a consequence of this theorem. Result $iii)$ follows from the fact that the denominator in (15) is bounded away from 0 on $V(y_0)$. □

Now, we derive the asymptotic distribution of the empirical co-cdf.

Theorem B.1. *Assume* **A1**, **A2***–**A4***, *and* **B1**-**B6**. *For any sequence* (x_n,y_n) *of random vectors converging in probability to* $(x,y)\in\mathbb{R}^2$, *with* $G^{(1)}(x)G^{(2)}(y)\neq 0$, *we have*

$$\sqrt{n}\left(\widehat{F}(x_n\mid y_n)-F(x_n\mid y_n)\right)=\frac{1}{\sqrt{n}G^{(2)}(y)}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}\leqslant x,\,\eta_{2t}\leqslant y}-H(x,y)\}$$

$$+\frac{xf_1(x\mid y)}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^{n}(\eta_{1t}^2-1)\boldsymbol{D}_{1t}+\frac{y\Delta(x,y)}{2\sqrt{n}G^{(2)}(y)}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}$$

$$-\frac{F(x\mid y)}{G^{(2)}(y)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\mathbb{1}_{\eta_{2t}\leqslant y}-G^{(2)}(y)\right\}+o_{\mathbb{P}}(1),$$

with $\Delta(x,y) = f_2(y \mid x)G^{(1)}(x) - g^{(2)}(y)F(x \mid y)$.

If in addition **B7** *holds, we have* $\boldsymbol{\Omega}_i'\boldsymbol{J}_i^{-1}\boldsymbol{D}_{it} = 1$ *a.s. and*

$$\sqrt{n}\left(\widehat{F}(x_n \mid y_n) - F(x_n \mid y_n)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2_{x|y}\right), \qquad \textit{where}$$

$$\begin{aligned}\sigma^2_{x|y} =& \frac{F(x \mid y)\{1 - F(x \mid y)\}}{G^{(2)}(y)} + \frac{\{xf_1(x \mid y)\}^2}{4}(\kappa_1 - 1) + \frac{y^2\Delta^2(x,y)}{4\{G^{(2)}(y)\}^2}(\kappa_2 - 1)\\ &+ \frac{xf_1(x \mid y)}{G^{(2)}(y)}\varrho_1(x,y) + \frac{y\Delta(x,y)}{\{G^{(2)}(y)\}^2}\varrho_2(x,y) + \frac{xyf_1(x \mid y)\Delta(x,y)}{2G^{(2)}(y)}\left\{E(\eta_{1t}^2\eta_{2t}^2) - 1\right\},\end{aligned}$$

with $\varrho_i(x,y) = E(\eta_{it}^2 \mathbb{1}_{\eta_{1t}\leqslant x,\, \eta_{2t}\leqslant y}) - E(\eta_{it}^2 \mathbb{1}_{\eta_{2t}\leqslant y})F(x \mid y)$.

The proof is given in Section C.

C Proof of Theorem B.1

Asymptotic distributions of the empirical cdf of residuals were derived, in the case of GARCH models, by Boldin (1998), Horváth, Kokoszka and Teyssière (2001), Berkes and Horváth (2003), Lee and Taniguchi (2005), and FZ among others. The introduction of a co-cdf constitutes the main technical challenge of the proof of Theorem B.1.

First note that

$$\begin{aligned}&\sqrt{n}\left(\widehat{F}(x_n \mid y_n) - F(x_n \mid y_n)\right)\\ &= \underbrace{\frac{\sqrt{n}\left\{\widehat{H}(x_n,y_n) - H(x_n,y_n)\right\}}{\widehat{G}^{(2)}(y_n)}}_{a_n(x_n,y_n)} + F(x_n|y_n)\underbrace{\frac{\sqrt{n}\left\{G^{(2)}(y_n) - \widehat{G}^{(2)}(y_n)\right\}}{\widehat{G}^{(2)}(y_n)}}_{b_n(y_n)}.\end{aligned}$$

Write

$$\widehat{H}(x,y) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\eta_{1t}\leqslant \widetilde{\chi}^{(1)}_{t,n}x,\, \eta_{2t}\leqslant \widetilde{\chi}^{(2)}_{t,n}y}, \quad \text{with} \quad \widetilde{\chi}^{(i)}_{t,n} = \widetilde{\sigma}_t(\widehat{\boldsymbol{\theta}}^{(i)})/\sigma_t(\boldsymbol{\theta}_0^{(i)}).$$

Let $\chi^{(i)}_{t,n} = \sigma_t(\widehat{\boldsymbol{\theta}}^{(i)})/\sigma_t(\boldsymbol{\theta}_0^{(i)})$, let $H_n(x,y) = n^{-1}\sum_{t=1}^n \mathbb{1}_{\eta_{1t}\leqslant x, \eta_{2t}\leqslant y}$, and let

$$\begin{aligned}\widehat{e}(x,y) &= \sqrt{n}\{\widehat{H}(x,y) - H(x,y)\}, \quad e_n(x,y) = \sqrt{n}\{H_n(x,y) - H(x,y)\},\\ h_n^{(1)}(x,y) &= xf_1(x \mid y)G^{(2)}(y)\left(\frac{1}{n}\sum_{t=1}^n \boldsymbol{D}_{1t}'\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0^{(1)}),\\ h_n^{(2)}(x,y) &= yf_2(y \mid x)G^{(1)}(x)\left(\frac{1}{n}\sum_{t=1}^n \boldsymbol{D}_{2t}'\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}_0^{(2)}).\end{aligned}$$

We have

$$\begin{aligned}\widehat{e}(x,y) =& \underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{1}_{\eta_{1t}\leqslant x\chi^{(1)}_{t,n},\, \eta_{2t}\leqslant y\chi^{(2)}_{t,n}} - H\left(x\chi^{(1)}_{t,n}, y\chi^{(2)}_{t,n}\right)}_{\widehat{e}_1(x,y)}\\ &+ \underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^n H\left(x\chi^{(1)}_{t,n}, y\chi^{(2)}_{t,n}\right) - H(x,y)}_{\widehat{e}_2(x,y)}\end{aligned}$$

$$+\underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{1}_{\eta_{1t}\leqslant\tilde{\chi}_{t,n}^{(1)}x,\,\eta_{2t}\leqslant\tilde{\chi}_{t,n}^{(2)}y}-\mathbb{1}_{\eta_{1t}\leqslant x\chi_{t,n}^{(1)},\,\eta_{2t}\leqslant y\chi_{t,n}^{(2)}}}_{\hat{e}_3(x,y)}.$$

We will show that

$$\begin{aligned}\hat{e}_1(x_n,y_n) &= e_n(x,y)+o_{\mathbb{P}}(1), &(16)\\ \hat{e}_2(x_n,y_n) &= h_n^{(1)}(x,y)+h_n^{(2)}(x,y)+o_{\mathbb{P}}(1), &(17)\\ \hat{e}_3(x_n,y_n) &= o_{\mathbb{P}}(1). &(18)\end{aligned}$$

The last three results, (4), **B6** and the Glivenko-Cantelli result in Proposition B.1 entail

$$\begin{aligned}a_n(x_n,y_n) &= \frac{1}{\sqrt{n}G^{(2)}(y)}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}\leqslant x,\,\eta_{2t}\leqslant y}-H(x,y)\}\\ &+\frac{xf_1(x\mid y)}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^{n}(\eta_{1t}^2-1)\boldsymbol{D}_{1t}+\frac{yf_2(y\mid x)G^{(1)}(x)}{2\sqrt{n}G^{(2)}(y)}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}+o_{\mathbb{P}}(1).\end{aligned}$$

Letting $x_n\to\infty$ in the previous equality, we find that

$$b_n(y_n)=\frac{-1}{G^{(2)}(y)}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{2t}\leqslant y}-G^{(2)}(y)\}+\frac{yg^{(2)}(y)}{2\sqrt{n}}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}\right\}+o_{\mathbb{P}}(1).$$

The conclusion then follows by straightforward but tedious computations. The simplification of the asymptotic variance under **B7** is established below. □

We now prove (16). Let, for $\boldsymbol{a}^{(i)}$ a vector of the same size as $\boldsymbol{\theta}^{(i)}$ (small enough so that $\boldsymbol{\theta}_0^{(i)}+\boldsymbol{a}^{(i)}/\sqrt{n}\in\boldsymbol{\Theta}^{(i)}$) and for $\boldsymbol{a}=(\boldsymbol{a}^{(1)\prime},\boldsymbol{a}^{(2)\prime})'$,

$$e_{n,1}(x,y,\boldsymbol{a})=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\mathbb{1}_{\eta_{1t}\leqslant x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})}-H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)\right\},$$

where $\gamma_{t,n}^{(i)}(\boldsymbol{a}^{(i)})=\frac{\sigma_{it}(\boldsymbol{\theta}_0^{(i)}+\frac{\boldsymbol{a}^{(i)}}{\sqrt{n}})}{\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}$. Note that $\hat{e}_1(x,y)=e_{n,1}(x,y,\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0))$ and $e_{n,1}(x,y,\boldsymbol{0})=e_n(x,y)$. Write

$$e_{n,1}(x,y,\boldsymbol{a})-e_n(x,y)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_{t,n}(x,y,\boldsymbol{a}),\tag{19}$$

where $z_{t,n}(x,y,\boldsymbol{a})$ is equal to

$$\mathbb{1}_{\eta_{1t}\leqslant x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})}-H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)-\{\mathbb{1}_{\eta_{1t}\leqslant x,\eta_{2t}\leqslant y}-H(x,y)\}.$$

We will establish a number of auxiliary lemmas.

Lemma C.1. *For any $u>0$ and sufficiently large n,*

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}z_{t,n}(x,y,\boldsymbol{a})\right|>u\right)\leqslant\frac{K}{u^6}\left\{\frac{|x|^3\|\boldsymbol{a}^{(1)}\|^3+|y|^3\|\boldsymbol{a}^{(2)}\|^3}{n^{3/2}}+\frac{|x|\|\boldsymbol{a}^{(1)}\|+|y|\|\boldsymbol{a}^{(2)}\|}{n^{5/2}}\right\}.$$

Proof. Note that $(z_{t,n}(x,y,\boldsymbol{a}),\mathcal{F}_t)_{1\leqslant t\leqslant n}$ is a martingale difference sequence. Hence by Markov and Rosenthal inequalities

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{t,n}(x,y,\boldsymbol{a})\right|>u\right) &\leqslant \frac{1}{n^3u^6}E\left(\sum_{t=1}^{n} z_{t,n}(x,y,\boldsymbol{a})\right)^6\\ \leqslant &\frac{K}{n^3u^6}\left\{E\left(\sum_{t=1}^{n}E(z_{t,n}^2(x,y,\boldsymbol{a})|\mathcal{F}_{t-1})\right)^3+\sum_{t=1}^{n}Ez_{t,n}^6(x,y,\boldsymbol{a})\right\}. \end{aligned} \tag{20}$$

Note that, conditional on $\mathcal{F}_{t-1}$, the random variable

$$\mathbb{1}_{\eta_{1t}\leqslant x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})}-\mathbb{1}_{\eta_{1t}\leqslant x,\eta_{2t}\leqslant y}$$

takes the values $-1,0,1$ with probabilities p_{-1}, p_0 and p_1. If $x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)})\leqslant x$ and $y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\leqslant y$, then $p_{-1}=H(x,y)-H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)$, $p_0=1-p_{-1}$ and $p_1=0$. If $x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)})\geqslant x$ and $y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\geqslant y$, then $p_{-1}=0$, $p_0=1-p_1$ and $p_1=H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)-H(x,y)$. If $x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)})\leqslant x$ and $y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\geqslant y$, then $p_{-1}=H(x,y)-H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\right)$, $p_0=1-p_{-1}-p_1$ and $p_1=H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)-H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\right)$. If $x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)})\geqslant x$ and $y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\leqslant y$, then $p_{-1}=H(x,y)-H\left(x,y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)$, $p_0=1-p_{-1}-p_1$ and $p_1=H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)-H\left(x,y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)$. It follows that $E[z_{t,n}^2(x,y,\boldsymbol{a})|\mathcal{F}_{t-1}]\leqslant p_{-1}+p_1$. We thus have, using **B6**,

$$\begin{aligned}\sum_{t=1}^{n}E[z_{t,n}^2(x,y,\boldsymbol{a})|\mathcal{F}_{t-1}] &\leqslant \sum_{t=1}^{n}|H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)-H(x,y)|\\ &+2|H\left(x,y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)-H(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y)|+2|H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),y\right)-H(x,y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)}))|\\ \leqslant &\frac{K|x|}{\sqrt{n}}\sum_{t=1}^{n}\left\|\frac{1}{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}\frac{\partial\sigma_{1t}(\boldsymbol{\theta}_t^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right\|\|\boldsymbol{a}^{(1)}\|+\frac{K|y|}{\sqrt{n}}\sum_{t=1}^{n}\left\|\frac{1}{\sigma_{2t}(\boldsymbol{\theta}_0^{(2)})}\frac{\partial\sigma_{2t}(\boldsymbol{\theta}_t^{(2)})}{\partial\boldsymbol{\theta}^{(2)}}\right\|\|\boldsymbol{a}^{(2)}\|,\end{aligned}$$

where $\boldsymbol{\theta}_t=(\boldsymbol{\theta}_t^{(1)'},\boldsymbol{\theta}_t^{(2)'})'$ is between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0+\boldsymbol{a}/\sqrt{n}$. Using the c_r-inequality and **B4**, we thus have

$$\begin{aligned}E\left\{\sum_{t=1}^{n}E[z_{t,n}^2(x,y,\boldsymbol{a})|\mathcal{F}_{t-1}]\right\}^3 &\leqslant \frac{K|x|^3\|\boldsymbol{a}^{(1)}\|^3}{n^{3/2}}\sum_{t_1,t_2,t_3=1}^{n}E\prod_{i=1}^{3}\left\|\frac{1}{\sigma_{1t_i}(\boldsymbol{\theta}_0^{(1)})}\frac{\partial\sigma_{1t_i}(\boldsymbol{\theta}_{t_i}^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right\|\\ &+\frac{K|y|^3\|\boldsymbol{a}^{(2)}\|^3}{n^{3/2}}\sum_{t_1,t_2,t_3=1}^{n}E\prod_{i=1}^{3}\left\|\frac{1}{\sigma_{2t_i}(\boldsymbol{\theta}_0^{(2)})}\frac{\partial\sigma_{2t_i}(\boldsymbol{\theta}_{t_i}^{(2)})}{\partial\boldsymbol{\theta}^{(2)}}\right\|\\ &\leqslant n^{3/2}K\left(|x|^3\|\boldsymbol{a}^{(1)}\|^3+|y|^3\|\boldsymbol{a}^{(2)}\|^3\right).\end{aligned}$$

Since $|z_{t,n}(x,y,\boldsymbol{a})|\leqslant 1$, we have

$$\begin{aligned}&\sum_{t=1}^{n}E[z_{t,n}^6(x,y,\boldsymbol{a})]\leqslant\sum_{t=1}^{n}E\left\{E[z_{t,n}^2(x,y,\boldsymbol{a})|\mathcal{F}_{t-1}]\right\}\\ &\leqslant\frac{K|x|}{\sqrt{n}}\sum_{t=1}^{n}E\left\|\frac{1}{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}\frac{\partial\sigma_{1t}(\boldsymbol{\theta}_t^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right\|\|\boldsymbol{a}^{(1)}\|+\frac{K|y|}{\sqrt{n}}\sum_{t=1}^{n}E\left\|\frac{1}{\sigma_{2t}(\boldsymbol{\theta}_0^{(2)})}\frac{\partial\sigma_{2t}(\boldsymbol{\theta}_t^{(2)})}{\partial\boldsymbol{\theta}^{(2)}}\right\|\|\boldsymbol{a}^{(2)}\|\end{aligned}$$

$$\leqslant K\sqrt{n}(|x|\|\boldsymbol{a}^{(1)}\| + |y|\|\boldsymbol{a}^{(2)}\|)$$

and, in view of (20), the conclusion follows. $\square$

Lemma C.2. *For any compact set* $\mathcal{K} \subset \mathbb{R}$, $\sup_{x,y\in\mathcal{K}} \left|n^{-1/2}\sum_{t=1}^n z_{t,n}(x,y,\boldsymbol{a})\right| = o_{\mathbb{P}}(1)$.

Proof. Fix $\varepsilon > 0$ and let $\mathcal{K} \subset \left[-\frac{N\varepsilon}{\sqrt{n}}, \frac{N\varepsilon}{\sqrt{n}}\right]$ with $N = O(\sqrt{n})$. Define $x_j = y_j = \frac{j\varepsilon}{\sqrt{n}}$ for $j = -N, -N+1, \ldots, N-1, N$. Note that, for all $k \geqslant 0$,

$$\sum_{-N\leqslant i,j\leqslant N} |x_i|^k + |y_j|^k \leqslant \frac{K}{n^{k/2}} \sum_{1\leqslant i,j\leqslant N} i^k + j^k = O\left(\frac{N^{k+2}}{n^{k/2}}\right) = O(n).$$

It follows that, by Lemma C.1, for any $u > 0$, there exists $K = K(u,\boldsymbol{a},\varepsilon)$ such that

$$\mathbb{P}\left(\max_{-N\leqslant i,j\leqslant N}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n z_{t,n}(x_i,y_j,\boldsymbol{a})\right| > u\right) \leqslant \frac{K}{\sqrt{n}}. \tag{21}$$

It therefore suffices to show that

$$\limsup_{n\to\infty} \mathbb{P}\left\{\max_{-N\leqslant i,j\leqslant N-1} \delta(i,j,\boldsymbol{a}) > u\right\} = 0, \tag{22}$$

where $\delta(i,j,\boldsymbol{a}) = \sup_{x\in[x_i,x_{i+1}],y\in[y_j,y_{j+1}]}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n z_{t,n}(x,y,\boldsymbol{a}) - z_{t,n}(x_i,y_j,\boldsymbol{a})\right|$. We have, for $i,j = 0,\ldots,N-1$,

$$\begin{aligned}
&\Delta(i,j,\boldsymbol{a})\\
=:&\sup_{x\in[x_i,x_{i+1}],y\in[y_j,y_{j+1}]}\frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{1}_{\eta_{1t}\leqslant x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})} - \mathbb{1}_{\eta_{1t}\leqslant x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})}\\
&+\sup_{x\in[x_i,x_{i+1}],y\in[y_j,y_{j+1}]}\frac{1}{\sqrt{n}}\sum_{t=1}^n H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right) - H\left(x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)\\
\leqslant&\sup_{x\in[x_i,x_{i+1}],y\in[y_j,y_{j+1}]}\frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{1}_{\eta_{1t}\leqslant x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})} - \mathbb{1}_{\eta_{1t}\leqslant x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})}\\
&+\sup_{x\in[x_i,x_{i+1}],y\in[y_j,y_{j+1}]}\frac{1}{\sqrt{n}}\sum_{t=1}^n H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right) - H\left(x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)\\
\leqslant&\frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{1}_{\eta_{1t}\leqslant x_{i+1}\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y_{j+1}\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})} - \mathbb{1}_{\eta_{1t}\leqslant x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})}\\
&+\frac{1}{\sqrt{n}}\sum_{t=1}^n H\left(x_{i+1}\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_{j+1}\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right) - H\left(x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)\\
\leqslant&\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\mathbb{1}_{\eta_{1t}\leqslant x_{i+1}\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y_{j+1}\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})} - H(x_{i+1}\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_{j+1}\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)}))\right\}\right.\\
&\left.-\left\{\mathbb{1}_{\eta_{1t}\leqslant x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}),\eta_{2t}\leqslant y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})} - H(x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)}))\right\}\right| + 2W_n(i,j,\boldsymbol{a}),
\end{aligned}$$

where

$$W_n(i,j,\boldsymbol{a}) = n^{-1/2}\sum_{t=1}^n\left\{H\left(x_{i+1}\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_{j+1}\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right) - H\left(x_i\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_j\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)\right\}.$$

Therefore, $\Delta(i,j,\boldsymbol{a})$ is smaller than

$$\frac{1}{\sqrt{n}}\left|\sum_{t=1}^{n} z_{t,n}(x_{i+1},y_{j+1},\boldsymbol{a})\right| + \frac{1}{\sqrt{n}}\left|\sum_{t=1}^{n} z_{t,n}(x_i,y_j,\boldsymbol{a})\right| + V_n(i,j) + 2W_n(i,j,\boldsymbol{a}), \tag{23}$$

where

$$V_n(i,j) = n^{-1/2}\left|\sum_{t=1}^{n}\left\{\mathbb{1}_{\eta_{1t}\leqslant x_{i+1},\eta_{2t}\leqslant y_{j+1}} - H(x_{i+1},y_{j+1})\right\} - \left\{\mathbb{1}_{\eta_{1t}\leqslant x_i,\eta_{2t}\leqslant y_j} - H(x_i,y_j)\right\}\right|.$$

By Assumption **B6** and the mean-value theorem, $H(x_{i+1},y_{j+1}) - H(x_i,y_j) \leqslant M\varepsilon/\sqrt{n}$ where $M = \sup_{x,y\in\mathbb{R}} f_1(x|y) + f_2(y|x) < \infty$. Thus $W_n(i,j,\mathbf{0}) \leqslant M\epsilon$ and $\Delta(i,j,\mathbf{0}) \leqslant V_n(i,j) + 2M\epsilon$.

From $\delta(i,j,\boldsymbol{a}) \leqslant \Delta(i,j,\boldsymbol{a}) + \Delta(i,j,\mathbf{0})$, we deduce from (23) that

$$\begin{aligned}
&\max_{0\leqslant i,j\leqslant N-1}\delta(i,j,\boldsymbol{a})\\
&\leqslant \max_{0\leqslant i,j\leqslant N}\frac{2}{\sqrt{n}}\left|\sum_{t=1}^{n} z_{t,n}(x_i,y_j,\boldsymbol{a})\right| + 2\max_{0\leqslant i,j\leqslant N-1} W_n(i,j,\boldsymbol{a}) + 2\max_{0\leqslant i,j\leqslant N-1} V_n(i,j) + 2M\varepsilon.
\end{aligned}$$

By the properties of the modulus of continuity of the empirical process (see Shorack and Wellner (1986), p. 542), under Assumption **B6** we have $\max_{0\leqslant i,j\leqslant N-1} V_n(i,j) = o_{\mathbb{P}}(1)$. We also have

$$\begin{aligned}
&\max_{0\leqslant i,j\leqslant N-1} W_n(i,j,\boldsymbol{a})\\
\leqslant\ &\max_{0\leqslant i,j\leqslant N-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(x_{i+1}-x_i)\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)})M + (y_{j+1}-y_j)\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})M = \varepsilon\times O_{\mathbb{P}}(1).
\end{aligned}$$

Similar arguments can be used when i or/and j belong to $\{-N,\dots,0\}$. Thus (22) is established. $\square$

Lemma C.3. *Let $\mathcal{K}$ be a compact subset of $\mathbb{R}$. For any $A>0$ and $\boldsymbol{A} = [-A,A]^d$ with $d = d_1+d_2$, we have $\sup_{x,y\in\mathcal{K}}\sup_{\boldsymbol{a}\in\boldsymbol{A}} |e_{n,1}(x,y,\boldsymbol{a}) - e_n(x,y)| = o_{\mathbb{P}}(1)$.*

Proof. The proof, which uses the grid technique developed by Horváth, Kokoszka and Teyssière (Lemma 4.5, 2001) in the ARCH case, closely follows that of Lemma 6.4 in FZ. We provide it for completeness. In view of (19), it suffices to prove that

$$\sup_{\boldsymbol{a}\in\boldsymbol{A}} X_n(\boldsymbol{a}) = o_{\mathbb{P}}(1), \quad \text{where} \quad X_n(\boldsymbol{a}) = \sup_{x,y\in\mathcal{K}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{t,n}(x,y,\boldsymbol{a})\right|. \tag{24}$$

Let $\varepsilon>0$ such that $N := 2A/\varepsilon$ is an integer and define $a(k) = -A + k\varepsilon$, for $1\leqslant k\leqslant N$. For any $1\leqslant k_1,k_2,\dots k_d\leqslant N$ let $\boldsymbol{k} = (k_1,\dots,k_d)$ and consider the grid of N^d points $\boldsymbol{a}(\boldsymbol{k}) = (\boldsymbol{a}^{(1)}(k)',\boldsymbol{a}^{(2)}(k)')' = (a(k_1),\dots,a(k_d))'$. Let also $\boldsymbol{A}(\boldsymbol{k}) = \{(\boldsymbol{a}^{(1)'},\boldsymbol{a}^{(2)'})' = (a_1,\dots,a_d)\in\boldsymbol{A} | a(k_i)-\varepsilon\leqslant a_i\leqslant a(k_i)\}$ and $\boldsymbol{a}^*(\boldsymbol{k}) = (a(k_1)-\varepsilon,\dots,a(k_d)-\varepsilon)$. For $j=1,\dots,d_1$ and $a_j\leqslant a(k_j)$, we have

$$\begin{aligned}
&H\left(x\gamma_{t,n}^{(1)}(a_1,\dots,a_{j-1},a(k_j),a_{j+1},\dots,a_{d_1}), y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right) - H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)\\
=\ & f_1\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}_{t,j}^{(1)*})|y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right) G^{(2)}\left(y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)\frac{x}{\sqrt{n}}\{a(k_j)-a_j\}\\
&\times\frac{1}{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}\frac{\partial\sigma_{1t}(\boldsymbol{\theta}_0^{(1)} + n^{-1/2}\boldsymbol{a}_{t,j}^{(1)*})}{\partial\boldsymbol{\theta}^{(1)'}}\boldsymbol{e}_j,
\end{aligned}$$

where $\boldsymbol{e}_j$ is the j-th element of the canonical basis of $\mathbb{R}^{d_1}$, and $\boldsymbol{a}_{t,j}^{(1)*}$ is a point between the arguments of $\gamma_{t,n}^{(1)}$ above. By **B6** and $E|\eta_{1t}| < \infty$, we have $\sup_{x,y} |x| f_1(x|y) < \infty$. The latter difference is thus bounded, uniformly in $x \in \mathbb{R}$ and $a_j \in [a(k_j) - \varepsilon, a(k_j)]$, by $K \frac{\varepsilon}{\sqrt{n}} \frac{1}{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})} \frac{\partial \sigma_{1t}(\boldsymbol{\theta}_0^{(1)} + \frac{\boldsymbol{a}_{t,j}^*}{\sqrt{n}})}{\partial \boldsymbol{\theta}^{(1)'}} \boldsymbol{e}_j$. A similar bound holds for $H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^1), y\gamma_{t,n}^{(2)}(a_{d_1+1}, \ldots, a_{j-1}, a(k_j), a_{j+1}, \ldots, a_d)\right) - H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right)$ where $j = d_1 + 1, \ldots, d$. Therefore, for n large enough,

$$\sup_{\boldsymbol{a} \in \boldsymbol{A}(\boldsymbol{k})} \sup_{x,y \in \mathbb{R}} \sum_{t=1}^n \left| H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})\right) - H\left(x\gamma_{t,n}^{(1)}\{\boldsymbol{a}^{(1)}(\boldsymbol{k})\}, y\gamma_{t,n}^{(2)}\{\boldsymbol{a}^{(2)}(\boldsymbol{k})\}\right) \right|$$
$$\leqslant K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \max_{i=1,2} \sup_{\boldsymbol{\theta}^{(i)} \in V(\boldsymbol{\theta}_0^{(i)})} \left\| \frac{1}{\sigma_{it}(\boldsymbol{\theta}^{(i)})} \frac{\partial \sigma_{it}(\boldsymbol{\theta}^{(i)})}{\partial \boldsymbol{\theta}^{(i)}} \right\|,$$

and thus, because the $\gamma_{t,n}^{(i)}(\cdot)$ are increasing functions of their arguments by **B5**,

$$\sup_{\boldsymbol{a} \in \boldsymbol{A}(\boldsymbol{k})} \sup_{x,y \in \mathcal{K}} \left| \sum_{t=1}^n z_{t,n}(x,y,\boldsymbol{a}) - z_{t,n}(x,y,\boldsymbol{a}(\boldsymbol{k})) \right| \leqslant K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \max_{i=1,2} \sup_{\boldsymbol{\theta}^{(i)} \in V(\boldsymbol{\theta}_0^{(i)})} \left\| \boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)}) \right\|$$
$$+ \sup_{x \in \mathcal{K}} \left| \sum_{t=1}^n \mathbb{1}_{\eta_{1t} \leqslant x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}(\boldsymbol{k})), \eta_{2t} \leqslant y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)}(\boldsymbol{k}))} - \mathbb{1}_{\eta_{1t} \leqslant x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)*}(\boldsymbol{k})), \eta_{2t} \leqslant y\gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)*}(\boldsymbol{k}))} \right|$$
$$\leqslant 2K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \max_{i=1,2} \sup_{\boldsymbol{\theta}^{(i)} \in V(\boldsymbol{\theta}_0^{(i)})} \left\| \boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)}) \right\| + \sup_{x \in \mathcal{K}} \left| \sum_{t=1}^n z_{t,n}(x,y,\boldsymbol{a}(\boldsymbol{k})) \right| + \sup_{x \in \mathcal{K}} \left| \sum_{t=1}^n z_{t,n}(x,y,\boldsymbol{a}^*(\boldsymbol{k})) \right|.$$

Note that $\sup_{\boldsymbol{a} \in \boldsymbol{A}} X_n(\boldsymbol{a})$ is less than

$$\max_{\boldsymbol{k} \in \{1,\ldots,N\}^d} \sup_{\boldsymbol{a} \in \boldsymbol{A}(\boldsymbol{k})} \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n [z_{t,n}(x,y,\boldsymbol{a}) - z_{t,n}(x,y,\boldsymbol{a}(\boldsymbol{k}))] \right| + \max_{\boldsymbol{k} \in \{1,\ldots,N\}^d} X_n(\boldsymbol{a}(\boldsymbol{k}))$$
$$\leqslant \frac{2K\varepsilon}{n} \sum_{t=1}^n \max_{i=1,2} \sup_{\boldsymbol{\theta}^{(i)} \in V(\boldsymbol{\theta}_0^{(i)})} \left\| \boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)}) \right\| + 2 \max_{\boldsymbol{k} \in \{1,\ldots,N\}^d} X_n(\boldsymbol{a}(\boldsymbol{k})) + \max_{\boldsymbol{k} \in \{1,\ldots,N\}^d} X_n(\boldsymbol{a}^*(\boldsymbol{k})).$$

By the ergodic theorem and **B4**, the first term in the r.h.s. is almost surely less than a constant times ε when n is large. The two other terms tend to zero in probability because $X_n(\boldsymbol{a}) = o_{\mathbb{P}}(1)$ by Lemma C.2 and the maxima are over a finite number of points. □

Lemma C.4. *Let (x_n, y_n) be a random sequence tending to $(x_0, y_0) \in \mathbb{R}^2$ in probability. If, for $i = 1, 2$, $G^{(i)}$ has a bounded density $g^{(i)}$ then $e_n(x_n, y_n) - e_n(x_0, y_0) = o_{\mathbb{P}}(1)$.*

Proof. Letting $U_t = G^{(1)}(\eta_{1t})$ and $V_t = G^{(2)}(\eta_{2t})$ we have

$$e_n(x,y) \quad = \quad \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{U_t \leqslant G^{(1)}(x), V_t \leqslant G^{(2)}(y)} - \mathbb{P}(U_t \leqslant G^{(1)}(x), V_t \leqslant G^{(2)}(y)).$$

Write $e_n(x,y) = Y_n(G^{(1)}(x), G^{(2)}(x))$ where $Y_n(u,v) = n^{-1/2} \sum_{t=1}^n \mathbb{1}_{U_t \leqslant u, V_t \leqslant v} - \mathbb{P}(U_t \leqslant u, V_t \leqslant v)$. Billingsley (1968) studied the modulus of continuity of $\{Z_n(u), u \in [0,1]\}$ where $Z_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{U_t \leqslant u} - u$ and showed in his formula (22.13) that, for each $\varepsilon > 0$ and $\eta > 0$, there exists $\tau \in (0,1]$ such that for n large enough, $\mathbb{P}\left(\sup_{|u-v|<\tau} |Z_n(u) - Z_n(v)| \geqslant \varepsilon\right) \leqslant \eta$. We will extend this inequality in (28) below to the sequence $\{Y_n(u,v), u, v \in [0,1]\}$.

The sequence (U_t, V_t) being iid, we have

$$E\,|Y_n(u,v) - Y_n(u^*,v^*)|^4 = 3\mu_2^2 + \frac{\mu_4 - 3\mu_2^2}{n}$$

where $\mu_i = \mu_i(u,v,u^*,v^*)$ is i-th central moment of $\mathbb{1}_{U_t \leqslant u, V_t \leqslant v} - \mathbb{1}_{U_t \leqslant u^*, V_t \leqslant v^*}$. Since $\mathbb{P}(U_t \in (u, u+\varepsilon_1], V_t \in (v, v+\varepsilon_2]) \leqslant \varepsilon_1 \wedge \varepsilon_2$, we have

$$E\,|Y_n(u,v) - Y_n(u^*,v^*)|^4 \leqslant 3|u-u^*|^2 \wedge |v-v^*|^2 + \frac{|u-u^*| \wedge |v-v^*|}{n}.$$

If $\frac{\varepsilon}{n} \leqslant |u-u^*| \wedge |v-v^*|$ for $\varepsilon \in (0,1)$, we then have

$$E\,|Y_n(u,v) - Y_n(u^*,v^*)|^4 \leqslant \frac{4}{\varepsilon}|u-u^*|^2 \wedge |v-v^*|^2 \leqslant \frac{2}{\varepsilon}\left(|u-u^*|^2 + |v-v^*|^2\right). \tag{25}$$

Let ι be a number such that $\varepsilon/n \leqslant \iota$ and m a positive integer such that $u + m\iota \leqslant 1$ and $v + m\iota \leqslant 1$. Define a sequence of random variables $\xi_1, \xi_2, \ldots, \xi_{(m+1)^2}$ by $\xi_1 = 0$,

$$\xi_{(i-1)(m+1)+j+1} = Y_n(u+(i-1)\iota, v+(j-1)\iota) - Y_n(u+(i-1)\iota, v+j\iota),$$

for $i = 1, \ldots, m+1$, $j = 1, \ldots, m$ and $\xi_{i(m+1)+1} = Y_n(u+(i-1)\iota, v+m\iota) - Y_n(u+i\iota, v)$, $\quad i = 1, \ldots, m$. Note that

$$\max_{0 \leqslant i,j \leqslant m} |Y_n(u,v) - Y_n(u+i\iota, v+j\iota)| = \max_{1 \leqslant k \leqslant (m+1)^2} |S_k|, \quad S_k = \xi_1 + \cdots + \xi_k.$$

In view of (25), for $0 \leqslant i \leqslant i^* \leqslant m$ and $1 \leqslant j, j^* \leqslant m+1$ such that $i(m+1)+j \leqslant i^*(m+1)+j^*$, we have $E\left|S_{i(m+1)+j} - S_{i^*(m+1)+j^*}\right|^4 \leqslant \frac{2\iota^2}{\varepsilon}\left\{(i^*-i)^2 + |j^*-j|^2\right\}$. This is of the form of inequality (12.42) in Billingsley (1968), with $\gamma = 4$, $\alpha = 2$,

$$u_1 = \cdots = u_m = u_{m+1} = u_{i(m+1)+j} = \sqrt{\frac{2}{\varepsilon}}\iota, \quad 1 \leqslant i \leqslant m,\ 2 \leqslant j \leqslant m+1,$$

and $u_{m+2} = u_{2(m+1)+1} = \cdots = u_{m(m+1)+1} = \sqrt{\frac{2}{\varepsilon}}m\iota$. By Theorem 12.2 of Billingsley (1968), we have

$$\mathbb{P}\left(\max_{0 \leqslant i,j \leqslant m} |Y_n(u,v) - Y_n(u+i\iota, v+j\iota)| \geqslant \tau\right) \leqslant \frac{8}{\varepsilon\tau^4}m^4\iota^2. \tag{26}$$

Now, for $\underline{u} \leqslant \tilde{u} \leqslant \underline{u} + \iota$ and $\underline{v} \leqslant \tilde{v} \leqslant \underline{v} + \iota$, we have

$$\begin{aligned}\mathbb{1}_{U_t \leqslant \tilde{u}, V_t \leqslant \tilde{v}} - \mathbb{P}(U_t \leqslant \tilde{u}, V_t \leqslant \tilde{v}) \leqslant{}& \mathbb{1}_{U_t \leqslant \underline{u}+\iota, V_t \leqslant \underline{v}+\iota} - \mathbb{P}(U_t \leqslant \underline{u}+\iota, V_t \leqslant \underline{v}+\iota) \\ &+ \mathbb{P}(U_t \leqslant \underline{u}+\iota, V_t \leqslant \underline{v}+\iota) - \mathbb{P}(U_t \leqslant \underline{u}, V_t \leqslant \underline{v}).\end{aligned}$$

Note that $\mathbb{P}(U_t \leqslant \underline{u}+\iota, V_t \leqslant \underline{v}+\iota) - \mathbb{P}(U_t \leqslant \underline{u}, V_t \leqslant \underline{v}) \leqslant \mathbb{P}(U_t \in (\underline{u}, \underline{u}+\iota]) + \mathbb{P}(V_t \in (\underline{v}, \underline{v}+\iota]) \leqslant 2\iota$. We thus have $Y_n(\tilde{u}, \tilde{v}) - Y_n(\underline{u}, \underline{v}) \leqslant Y_n(\underline{u}+\iota, \underline{v}+\iota) - Y_n(\underline{u}, \underline{v}) + 2\iota\sqrt{n}$. We also have

$$Y_n(\tilde{u}, \tilde{v}) - Y_n(\underline{u}, \underline{v}) \geqslant -\mathbb{P}(U_t \leqslant \underline{u}+\iota, V_t \leqslant \underline{v}+\iota) + \mathbb{P}(U_t \leqslant \underline{u}, V_t \leqslant \underline{v}) \geqslant -2\iota\sqrt{n}.$$

Therefore

$$|Y_n(\tilde{u}, \tilde{v}) - Y_n(\underline{u}, \underline{v})| \leqslant |Y_n(\underline{u}+\iota, \underline{v}+\iota) - Y_n(\underline{u}, \underline{v})| + 2\iota\sqrt{n}.$$

For all $u \leqslant u^* \leqslant u+m\iota$ and $v \leqslant v^* \leqslant v+m\iota$, applying the previous inequality with $\underline{u} = u+(i-1)\iota$ and $\underline{v} = v+(j-1)\iota$ such that $\underline{u} \leqslant u^* < \underline{u}+\iota$ and $\underline{v} \leqslant v^* < \underline{v}+\iota$, we obtain

$$|Y_n(u,v) - Y_n(u^*,v^*)| \leqslant |Y_n(u,v) - Y_n(\underline{u}, \underline{v})| + |Y_n(u^*,v^*) - Y_n(\underline{u}, \underline{v})|$$

$$\leqslant |Y_n(u,v) - Y_n(\underline{u},\underline{v})| + |Y_n(\underline{u}+\iota,\underline{v}+\iota) - Y_n(\underline{u},\underline{v})| + 2\iota\sqrt{n}$$
$$\leqslant 3 \max_{0\leqslant i,j\leqslant m} |Y_n(u,v) - Y_n(u+i\iota, v+j\iota)| + 2\iota\sqrt{n}.$$

Taking $\varepsilon/n < \iota < \varepsilon/\sqrt{n}$, we obtain from (26) and the previous inequality

$$\mathbb{P}\left(\sup_{u\leqslant u^*\leqslant u+m\iota, v\leqslant v^*\leqslant v+m\iota} |Y_n(u,v) - Y_n(u^*,v^*)| \geqslant 5\varepsilon\right) \leqslant \frac{8}{\varepsilon^5} m^4\iota^2. \tag{27}$$

For any $\eta > 0$, for n large we can always chose a small ι such as $\frac{8}{\varepsilon^5}m^4\iota^2 \leqslant \eta$.

We thus have shown that for each $\varepsilon > 0$ and $\eta > 0$, there exists $\tau \in (0,1]$ such that

$$\mathbb{P}\left(\sup_{|u-u^*|<\tau, |v-v^*|<\tau} |Y_n(u,v) - Y_n(u^*,v^*)| \geqslant \varepsilon\right) \leqslant \eta \tag{28}$$

for large n. For any $\varepsilon > 0$ and $\delta > 0$, we thus have

$$\begin{aligned}
&\mathbb{P}\left(|e_n(x_n,y_n) - e_n(x_0,y_0)| \geqslant \varepsilon\right)\\
\leqslant &\mathbb{P}\left(\sup_{|x^*-x_0|\leqslant\delta, |y^*-y_0|\leqslant\delta} |e_n(x^*,y^*) - e_n(x_0,y_0)| \geqslant \varepsilon\right) + \mathbb{P}\left(|x_n - x_0| \geqslant \delta\right)\\
&+ \mathbb{P}\left(|y_n - y_0| \geqslant \delta\right).
\end{aligned}$$

The last two probabilities tend to zero as $n \to \infty$ because $x_n \to x_0$ and $y_n \to y_0$ in probability. Now note that the first term in the r.h.s. is bounded by

$$\mathbb{P}\left(\sup_{|u-u^*|\leqslant\delta\sup_x f_1(x), |v-v^*|\leqslant\delta\sup_x f_2(x)} |Y_n(u,v) - Y_n(u^*,v^*)| \geqslant \varepsilon\right) \leqslant \eta$$

when n is large enough and δ small enough to satisfy (28) with $\tau > \delta \sup_x \max\left\{g^{(1)}(x), g^{(2)}(x)\right\}$. Since η can be taken arbitrarily small, we have shown that $e_n(x_n,y_n) - e_n(x_0,y_0) = o_{\mathbb{P}}(1)$. $\square$

The convergence in (16) is deduced from the previous lemmas and **B6**. It remains to show (17) and (18).

Proof of (17). Follows from **B6** and the following lemma.

Lemma C.5. *Let $\mathcal{K}$ be a compact subset of $\mathbb{R}$. Then $\sup_{x,y\in\mathcal{K}} |\widehat{e}_2(x,y) - h_n^{(1)}(x,y) - h_n^{(2)}(x,y)| \to 0$ a.s.*

Proof. A Taylor expansion yields, for $x_t^* = x\sigma_{1t}(\boldsymbol{\theta}_t^{(1)^*})/\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})$ and $y_t^* = y\sigma_{2t}(\boldsymbol{\theta}_t^{(2)^*})/\sigma_{2t}(\boldsymbol{\theta}_0^{(2)})$ with $\boldsymbol{\theta}_t^{(i)^*}$ between $\widehat{\boldsymbol{\theta}}^{(i)}$ and $\boldsymbol{\theta}_0^{(i)}$,

$$\begin{aligned}
&|\widehat{e}_2(x,y) - h_n^{(1)}(x,y) - h_n^{(2)}(x,y)|\\
\leqslant\ & |x|G^{(2)}(y)\frac{1}{n}\sum_{t=1}^n \left| f_1(x_t^* \mid y)\frac{1}{\sigma_{1t}}\frac{\partial\sigma_{1t}(\boldsymbol{\theta}_t^{(1)^*})}{\partial\boldsymbol{\theta}^{(1)}} - f_1(x\mid y)\frac{1}{\sigma_{1t}}\frac{\partial\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}{\partial\boldsymbol{\theta}^{(1)}}\right| \left\|\sqrt{n}(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0^{(1)})\right\|\\
&+|y|G^{(1)}(x)\frac{1}{n}\sum_{t=1}^n \left| f_2(y_t^* \mid x)\frac{1}{\sigma_{2t}}\frac{\partial\sigma_{2t}(\boldsymbol{\theta}_t^{(2)^*})}{\partial\boldsymbol{\theta}^{(2)}} - f_2(y\mid x)\frac{1}{\sigma_{2t}}\frac{\partial\sigma_{2t}(\boldsymbol{\theta}_0^{(2)})}{\partial\boldsymbol{\theta}^{(2)}}\right| \left\|\sqrt{n}(\widehat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}_0^{(2)})\right\|.
\end{aligned}$$

The rest of the proof relies on the arguments given in FZ (Lemma 6.7). $\square$

Proof of (18). By

$$\widehat{e}_3(x,y) = \frac{1}{\sqrt{n}}\sum_{t=1}^n (\mathbb{1}_{\eta_{1t}\leqslant\widetilde{\chi}_{t,n}^{(1)}x} - \mathbb{1}_{\eta_{1t}\leqslant\chi_{t,n}^{(1)}x})\mathbb{1}_{\eta_{2t}\leqslant\widetilde{\chi}_{t,n}^{(2)}y} + (\mathbb{1}_{\eta_{2t}\leqslant\widetilde{\chi}_{t,n}^{(2)}y} - \mathbb{1}_{\eta_{2t}\leqslant\chi_{t,n}^{(2)}y})\mathbb{1}_{\eta_{1t}\leqslant\chi_{t,n}^{(1)}x},$$

we deduce that $\sup_{x,y\in\mathbb{R}}|\widehat{e}_3(x,y)|$ is less than

$$\sup_{x\in\mathbb{R}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|\mathbb{1}_{\eta_{1t}\leqslant\widetilde{\chi}^{(1)}_{t,n}x}-\mathbb{1}_{\eta_{1t}\leqslant\chi^{(1)}_{t,n}x}|+\sup_{y\in\mathbb{R}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}|\mathbb{1}_{\eta_{2t}\leqslant\widetilde{\chi}^{(2)}_{t,n}y}-\mathbb{1}_{\eta_{2t}\leqslant\chi^{(2)}_{t,n}y}|=o_{\mathbb{P}}(1),$$

where the last equality follows from the proof of Lemma 6.9 in FZ. □

Having shown (16)–(18), the first part of Theorem B.1 is established.

We now show that the asymptotic variance simplifies under **B7**. The equalities $\boldsymbol{\Omega}_i'\boldsymbol{J}_i^{-1}\boldsymbol{\Omega}_i=1$, for $i=1,2$ were established in Francq and Zakoïan (2013). Therefore $E(1-\boldsymbol{\Omega}_i'\boldsymbol{J}_i^{-1}\boldsymbol{D}_{it})^2=1-\boldsymbol{\Omega}_i'\boldsymbol{J}_i^{-1}\boldsymbol{\Omega}_i=0,\quad i=1,2$ and it follows that $\boldsymbol{\Omega}_i'\boldsymbol{J}_i^{-1}\boldsymbol{D}_{it}=1$ a.s. We thus have

$$E(1-\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\boldsymbol{D}_{1t})(1-\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\boldsymbol{D}_{2t})=1-\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\boldsymbol{\Omega}_1-\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\boldsymbol{\Omega}_2+\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\boldsymbol{J}_{12}J_2^{-1}\boldsymbol{\Omega}_2=0$$

where $\boldsymbol{J}_{12}=E(\boldsymbol{D}_{1t}\boldsymbol{D}_{2t}')$. Therefore $\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\boldsymbol{J}_{12}\boldsymbol{J}_2^{-1}\boldsymbol{\Omega}_2=1$. The expression of $\sigma^2_{x|y}$ straightforwardly follows. □

D Proofs of Theorem 3.1 and Corollary 3.1

Proof of Theorem 3.1. Note that $\widehat{F}(\cdot|y)$ is a step function with jumps of size $\{n\widehat{G}^{(2)}(y)\}^{-1}$. We thus have $\widehat{F}(\widehat{u}\mid\widehat{\xi}^{(2)}_{\alpha'})-\alpha\leqslant 1/n\alpha'$ and

$$\begin{aligned}\sqrt{n}\left\{\alpha-F\left(\widehat{u}\mid\xi^{(2)}_{\alpha'}\right)\right\}&=\sqrt{n}\left\{\widehat{F}\left(\widehat{u}\mid\widehat{\xi}^{(2)}_{\alpha'}\right)-F\left(\widehat{u}\mid\widehat{\xi}^{(2)}_{\alpha'}\right)\right\}\\&+\sqrt{n}\left\{F\left(\widehat{u}\mid\widehat{\xi}^{(2)}_{\alpha'}\right)-F\left(\widehat{u}\mid\xi^{(2)}_{\alpha'}\right)\right\}+o_{\mathbb{P}}(1).\end{aligned}\tag{29}$$

Now note that **B6** and **B8** entail **A5** and **A6**-**A7**. Proposition 2.3 thus entails that $\widehat{u}(\alpha,\alpha')$ strongly converges to $u(\alpha,\alpha')$. We have seen that $\widehat{\xi}^{(2)}_{\alpha'}$ strongly converges to $\xi^{(2)}_{\alpha'}$. Let

$$\begin{aligned}\nu_x=&\nu_x(\alpha,\alpha')=\frac{\partial}{\partial x}F(x\mid y)\Big|_{(x,y)=\left(u(\alpha,\alpha'),\xi^{(2)}_{\alpha'}\right)}=f_1\left(u(\alpha,\alpha')\mid\xi^{(2)}_{\alpha'}\right),\\\nu_y=&\frac{\partial}{\partial y}F(x\mid y)\Big|_{(x,y)=\left(u(\alpha,\alpha'),\xi^{(2)}_{\alpha'}\right)}=\frac{\Delta(u(\alpha,\alpha'),\xi^{(2)}_{\alpha'})}{G^{(2)}(\xi^{(2)}_{\alpha'})}.\end{aligned}$$

By the delta method, using **B8**, we thus have $\sqrt{n}\left(F(\widehat{u}\mid\widehat{\xi}^{(2)}_{\alpha'})-F(u(\alpha,\alpha')\mid\xi^{(2)}_{\alpha'})\right)=\nu_x\sqrt{n}\left\{\widehat{u}-u(\alpha,\alpha')\right\}+\nu_y\sqrt{n}\left(\widehat{\xi}^{(2)}_{\alpha'}-\xi^{(2)}_{\alpha'}\right)+o_{\mathbb{P}}(1)$ and

$$\sqrt{n}\left(F\left(u(\alpha,\alpha')\mid\xi^{(2)}_{\alpha'}\right)-F\left(\widehat{u}\mid\xi^{(2)}_{\alpha'}\right)\right)=\nu_x\sqrt{n}\left\{u(\alpha,\alpha')-\widehat{u}\right\}+o_{\mathbb{P}}(1).$$

Therefore we have

$$\sqrt{n}\left(F(\widehat{u}\mid\widehat{\xi}^{(2)}_{\alpha'})-F(\widehat{u}\mid\xi^{(2)}_{\alpha'})\right)=\nu_y\sqrt{n}\left(\widehat{\xi}^{(2)}_{\alpha'}-\xi^{(2)}_{\alpha'}\right)+o_{\mathbb{P}}(1).\tag{30}$$

By Theorem B.1, noting that $G^{(2)}(\xi^{(2)}_{\alpha'})=\alpha'$ and using (29)-(30), we have the Bahadur expansion

$$\sqrt{n}\left\{\alpha-F(\widehat{u}\mid\xi^{(2)}_{\alpha'})\right\}=\frac{1}{\sqrt{n}\alpha'}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}\leqslant u,\,\eta_{2t}\leqslant\xi^{(2)}_{\alpha'}}-H(u,\xi^{(2)}_{\alpha'})\}$$

$$
\begin{aligned}
&+\frac{uf_1(u\mid\xi^{(2)}_{\alpha'})}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^n(\eta_{1t}^2-1)\boldsymbol{D}_{1t}+\frac{\xi^{(2)}_{\alpha'}\Delta(u,\xi^{(2)}_{\alpha'})}{2\sqrt{n}\alpha'}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^n(\eta_{2t}^2-1)\boldsymbol{D}_{2t}\\
&-\frac{\alpha}{\alpha'}\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha'}}-\alpha'\right\}\\
&-\frac{\nu_y}{g^{(2)}(\xi^{(2)}_{\alpha'})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha'}}-\alpha')-\frac{\nu_y\xi^{(2)}_{\alpha'}}{2\sqrt{n}}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^n(\eta_{2t}^2-1)\boldsymbol{D}_{2t}+o_{\mathbb{P}}(1)\\
=&\frac{1}{\sqrt{n}\alpha'}\sum_{t=1}^n\{\mathbb{1}_{\eta_{1t}\leqslant u,\,\eta_{2t}\leqslant\xi^{(2)}_{\alpha'}}-H(u,\xi^{(2)}_{\alpha'})\}+\frac{uf_1(u\mid\xi^{(2)}_{\alpha'})}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^n(\eta_{1t}^2-1)\boldsymbol{D}_{1t}\\
&-\left(\frac{\alpha}{\alpha'}+\frac{\nu_y}{g^{(2)}(\xi^{(2)}_{\alpha'})}\right)\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\{\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha'}}-\alpha'\right\}+o_{\mathbb{P}}(1),
\end{aligned}
$$

noting that $F(u\mid\xi^{(2)}_{\alpha'})=\alpha$, $G^{(2)}(\xi^{(2)}_{\alpha'})=\alpha'$ and $\Delta(u,\xi^{(2)}_{\alpha'})=\alpha'\nu_y$. By the delta method applied with the function $F^{-1}\left(\cdot\mid\xi^{(2)}_{\alpha'}\right)$ (which exists in a neighborhood of $u(\alpha,\alpha')$ under **B8**), noting that $\partial F^{-1}(\alpha\mid\xi^{(2)}_{\alpha'})/\partial x=1/f_1(u\mid\xi^{(2)}_{\alpha'})=1/\nu_x$, we obtain the Bahadur expansion of the theorem. Noting that $\boldsymbol{\Sigma}_\Upsilon$ is the covariance matrix of the vector $\boldsymbol{\Upsilon}_t=\left(\mathbb{1}_{\eta_{1t}\leqslant u(\alpha,\alpha'),\,\eta_{2t}\leqslant\xi^{(2)}_{\alpha'}},\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha'}},\eta_{1t}^2\right)'$, the rest of the proof easily follows. □

Proof of Corollary 3.1. Using (4) and noting that $E(\eta_{1t}^2-1)\boldsymbol{\Upsilon}_t=\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_3$. □

E Proof of Theorem 4.1.

We start by the following lemma.

Lemma E.1. *Let the assumptions of Theorem 4.1 be satisfied. Let (x_n,y_n) be any sequence of random vectors converging almost surely to some $(x,y)\in\mathbb{R}^2$. We have For all $k\leqslant 4$, as $n\to\infty$*

$$
\frac{1}{n}\sum_{t=1}^n\widehat{\eta}_{1t}^k\mathbb{1}_{\widehat{\eta}_{1t}\leqslant x_n}\mathbb{1}_{\widehat{\eta}_{2t}\leqslant y_n}\to E\eta_{1t}^k\mathbb{1}_{\eta_{1t}\leqslant x}\mathbb{1}_{\eta_{2t}\leqslant y}\ \ a.s. \tag{31}
$$

Proof. Let $\eta_{1t}(\boldsymbol{\theta}^{(1)})=\epsilon_{1t}/\sigma_{1t}(\boldsymbol{\theta}^{(1)})$ and $\widetilde{\eta}_{1t}(\boldsymbol{\theta}^{(1)})=\epsilon_{1t}/\widetilde{\sigma}_{1t}(\boldsymbol{\theta}^{(1)})$, so that $\widehat{\eta}_{1t}=\widetilde{\eta}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)})$ and $\eta_{1t}=\eta_{1t}(\boldsymbol{\theta}_0^{(1)})$. By **A2*** and **A3*** we have $\sup_{\boldsymbol{\theta}^{(1)}\in\boldsymbol{\Theta}^{(1)}}\left|\eta_{1t}^k(\boldsymbol{\theta}^{(1)})-\widetilde{\eta}_{1t}^k(\boldsymbol{\theta}^{(1)})\right|\leqslant\frac{K}{\omega}\rho^t|\epsilon_{1t}|^k$. Under **A1**, there exists $s<1$ such that $E|\epsilon_{1t}|^{ks}<\infty$. Thus $\sum_{t=1}^n\rho^t|\epsilon_{1t}|^k$ is finite almost surely because $E\left|\sum_{t=1}^n\rho^t|\epsilon_{1t}|^k\right|^s\leqslant E|\epsilon_{1t}|^{ks}\sum_{t=1}^n\rho^{ts}<\infty$. It follows that $\frac{1}{n}\sum_{t=1}^n\widehat{\eta}_{1t}^k=\frac{1}{n}\sum_{t=1}^n\eta_{1t}^k(\widehat{\boldsymbol{\theta}}^{(1)})+O(n^{-1})$ a.s. By the mean value theorem $\frac{1}{n}\sum_{t=1}^n\eta_{1t}^k(\widehat{\boldsymbol{\theta}}^{(1)})=\frac{1}{n}\sum_{t=1}^n\eta_{1t}^k+\frac{1}{n}\sum_{t=1}^n\frac{\partial\eta_{1t}^k(\boldsymbol{\theta}_n^{(1)})}{\partial\boldsymbol{\theta}^{(1)'}}\left(\widehat{\boldsymbol{\theta}}^{(1)}-\boldsymbol{\theta}_0^{(1)}\right)$, with $\boldsymbol{\theta}_n^{(1)}$ between $\widehat{\boldsymbol{\theta}}^{(1)}$ and $\boldsymbol{\theta}_0^{(1)}$. By **B10** we have

$$
\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\left\|\frac{\partial\eta_{1t}^k(\boldsymbol{\theta}^{(1)})}{\partial\boldsymbol{\theta}^{(1)'}}\right\|=k\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\left\|\left(\frac{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})}\right)^k\boldsymbol{D}_{1t}(\boldsymbol{\theta}^{(1)})\eta_{1t}^k\right\|=u_t|\eta_{1t}^k|,
$$

where $u_t\in\mathcal{F}_{t-1}$, (u_t) is stationary and ergodic and $Eu_t<\infty$. We thus have, for $k\leqslant 4$, $\frac{1}{n}\sum_{t=1}^n\widehat{\eta}_{1t}^k\to E\eta_{1t}^k$, a.s.

Now, note that conditional on $(\boldsymbol{\epsilon}_t)$, we have $\frac{1}{n}\sum_{t=1}^n\widehat{\eta}_{1t}^k=E\eta_n^{*k}$ where $\eta_n^*\sim\widehat{G}^{(1)}$. Since $\widehat{G}^{(1)}$ converges to $G^{(1)}$, conditional on $(\boldsymbol{\epsilon}_t)$ and (x_n), as $n\to\infty$ the random variable $\eta_n^{*k}\mathbb{1}_{\eta_n^*\leqslant x_n}$

converges in distribution to $\eta^k \mathbb{1}_{\eta \leqslant x}$ where $\eta \sim G^{(1)}$ (noting that $\eta \mapsto \eta^k \mathbb{1}_{\eta \leqslant x}$ is continuous, except at $\eta = x$ which is such that $\mathbb{P}(\eta = x) = 0$). Theorem 3.6 in Billingsley (1999) shows that, conditional on $(\boldsymbol{\epsilon}_t)$, the random variables η_n^{*k} are uniformly integrable. It follows that, conditional on $(\boldsymbol{\epsilon}_t)$ and (x_n), the random variables $\eta_n^{*k} \mathbb{1}_{\eta_n^* \leqslant x_n}$ are also uniformly integrable. By Theorem 3.5 in Billingsley (1999), conditional on $(\boldsymbol{\epsilon}_t)$ and (x_n), we then have $\frac{1}{n}\sum_{t=1}^n \widehat{\eta}_{1t}^k \mathbb{1}_{\widehat{\eta}_{1t} \leqslant x_n} = E\eta_n^{*k} \mathbb{1}_{\eta_n^* \leqslant x_n} \to E\eta^k \mathbb{1}_{\eta \leqslant x}$ as $n \to \infty$. The rest of the proof follows by similar arguments. The proof of Lemma E.1 is complete. $\square$

Now, we turn to the proof of Theorem 4.1. In view of Remark 4, the consistency of $(\widehat{u}, \widehat{\xi})$ and **B9**, for almost all sequence $(\boldsymbol{\epsilon}_t)$ we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(1)*} - \widehat{\boldsymbol{\theta}}^{(1)}\right) = \frac{\boldsymbol{J}_1^{-1} + o(1)}{2\sqrt{n}} \sum_{t=1}^n \boldsymbol{x}_{t,n}^*, \quad \boldsymbol{x}_{t,n}^* = (\eta_{1t}^{*2} - \widehat{m}_2)\widehat{\boldsymbol{D}}_{1t},$$

and

$$\sqrt{n}\,(\widehat{u}^* - \widehat{u}) = \{\boldsymbol{\lambda} + o(1)\}' \frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{\Upsilon}_{t,n}^*,$$

where $\boldsymbol{\Upsilon}_{t,n}^* = \left(\mathbb{1}_{\eta_{1t}^* \leqslant \widehat{u},\, \eta_{2t}^* \leqslant \widehat{\xi}} - \widehat{\alpha}\widehat{\alpha}', \mathbb{1}_{\eta_{2t}^* \leqslant \widehat{\xi}} - \widehat{\alpha}', \eta_{1t}^{*2} - \widehat{m}_2\right)'$. Letting $\boldsymbol{y}_{t,n}^* = (\boldsymbol{x}_{t,n}^{*'}, \boldsymbol{\Upsilon}_{t,n}^{*'})'$, the convergence in distribution (6) follows by showing that, conditional on $(\boldsymbol{\epsilon}_t)$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{y}_{t,n}^* \xrightarrow{\mathcal{L}} \mathcal{N}\left\{\mathbf{0}, \boldsymbol{\Sigma} := \begin{pmatrix} (\kappa_1 - 1)\boldsymbol{J}_1 & \boldsymbol{\Omega}_1 \boldsymbol{e}_3' \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \boldsymbol{e}_3 \boldsymbol{\Omega}_1' & \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \end{pmatrix}\right\}. \tag{32}$$

Note that, conditional on $(\boldsymbol{\epsilon}_t)$, for each n the random vectors $\boldsymbol{y}_{1,n}^*, \boldsymbol{y}_{2,n}^*, \dots$ are independent and centered, with finite second-order moments. From Lindeberg's CLT for triangular arrays of square integrable martingale increments, and the Wold-Cramer device, it suffices to show that for any $\boldsymbol{c} \in \mathbb{R}^{d_1+3}$, $\boldsymbol{c} \neq \mathbf{0}$,

$$\frac{1}{n} \sum_{t=1}^n \operatorname{Var}\left(\boldsymbol{c}'\boldsymbol{y}_{t,n}^*\right) \to \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c} \quad \text{as } n \to \infty, \tag{33}$$

and for all $\varepsilon > 0$

$$\frac{1}{n} \sum_{t=1}^n E\left(\left\{\boldsymbol{c}'\boldsymbol{y}_{t,n}^*\right\}^2 \mathbb{1}_{\{|\boldsymbol{c}'\boldsymbol{y}_{t,n}^*| \geqslant \sqrt{n}\varepsilon\}}\right) \to 0 \quad \text{as } n \to \infty. \tag{34}$$

Conditional on $(\boldsymbol{\epsilon}_t)$, we have

$$\operatorname{Var}\left(\boldsymbol{y}_{t,n}^*\right) = \begin{pmatrix} (\widehat{m}_4 - \widehat{m}_2^2)\widehat{\boldsymbol{D}}_{1t}\widehat{\boldsymbol{D}}_{1t}' & \widehat{\alpha}'\widehat{\varrho}_{\widehat{\alpha},\widehat{\alpha}'}\widehat{\boldsymbol{D}}_{1t} & \widehat{\alpha}'\widehat{\nabla}_{\widehat{\alpha},\widehat{\alpha}'}\widehat{\boldsymbol{D}}_{1t} & (\widehat{m}_4 - \widehat{m}_2^2)\widehat{\boldsymbol{D}}_{1t} \\ \widehat{\alpha}'\widehat{\varrho}_{\widehat{\alpha},\widehat{\alpha}'}\widehat{\boldsymbol{D}}_{1t}' & \widehat{\alpha}\widehat{\alpha}'(1 - \widehat{\alpha}\widehat{\alpha}') & \widehat{\alpha}\widehat{\alpha}'(1 - \widehat{\alpha}') & \widehat{\alpha}'\widehat{\varrho}_{\widehat{\alpha},\widehat{\alpha}'} \\ \widehat{\alpha}'\widehat{\nabla}_{\widehat{\alpha},\widehat{\alpha}'}\widehat{\boldsymbol{D}}_{1t}' & \widehat{\alpha}\widehat{\alpha}'(1 - \widehat{\alpha}') & (1 - \widehat{\alpha}')\widehat{\alpha}' & \widehat{\alpha}'\widehat{\nabla}_{\widehat{\alpha},\widehat{\alpha}'} \\ (\widehat{m}_4 - \widehat{m}_2^2)\widehat{\boldsymbol{D}}_{1t}' & \widehat{\alpha}'\widehat{\varrho}_{\widehat{\alpha},\widehat{\alpha}'} & \widehat{\alpha}'\widehat{\nabla}_{\widehat{\alpha},\widehat{\alpha}'} & (\widehat{m}_4 - \widehat{m}_2^2) \end{pmatrix}$$

where $\widehat{\alpha}'\widehat{\varrho}_{\widehat{\alpha},\widehat{\alpha}'} = n^{-1}\sum_{t=1}^n \widehat{\eta}_{1t}^2 \mathbb{1}_{\widehat{\eta}_{1t} \leqslant \widehat{u}} \mathbb{1}_{\widehat{\eta}_{2t} \leqslant \widehat{\xi}} - \widehat{\alpha}\widehat{\alpha}'$ and $\widehat{\alpha}'\widehat{\nabla}_{\alpha,\alpha'} = n^{-1}\sum_{t=1}^n \widehat{\eta}_{1t}^2 \mathbb{1}_{\widehat{\eta}_{2t} \leqslant \widehat{\xi}} - \widehat{\alpha}'$. Lemma E.1 and the consistency of $\widehat{\boldsymbol{\theta}}^{(1)}$ and $(\widehat{u}, \widehat{\xi})$ show that, for t fixed and $n \to \infty$

$$\operatorname{Var}\left(\boldsymbol{y}_{t,n}^*\right) \to \begin{pmatrix} (\kappa_1 - 1)\widetilde{\boldsymbol{D}}_{1t}\widetilde{\boldsymbol{D}}_{1t}' & \widetilde{\boldsymbol{D}}_{1t}\boldsymbol{e}_3'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_3\widetilde{\boldsymbol{D}}_{1t}' & \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \end{pmatrix}$$

with $\widetilde{\boldsymbol{D}}_{1t} = \widetilde{\boldsymbol{D}}_{1t}(\boldsymbol{\theta}_0^{(1)})$. Now, the ergodic theorem implies that, for almost all sequence $(\boldsymbol{\epsilon}_t)$,

$$\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} (\kappa_1-1)\widetilde{\boldsymbol{D}}_{1t}\widetilde{\boldsymbol{D}}_{1t}' & \widetilde{\boldsymbol{D}}_{1t}\boldsymbol{e}_3'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_3\widetilde{\boldsymbol{D}}_{1t}' & \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \end{pmatrix} \to \boldsymbol{\Sigma} \quad \text{as } n\to\infty$$

which, in view of the form of $\mathrm{Var}\left(\boldsymbol{y}_{t,n}^*\right)$, entails (33).

Now we turn to the proof of (34). Let $\varepsilon > 0$ and $\boldsymbol{c} = (\boldsymbol{c}_1', \boldsymbol{c}_2')'$, with $\boldsymbol{c}_1 \in \mathbb{R}^{d_1}$ and $\boldsymbol{c}_2 \in \mathbb{R}^3$. We first show (34) when $\boldsymbol{c}_2 = \boldsymbol{0}_3$. Given $(\boldsymbol{\epsilon}_t)$, for some neighborhood $V(\boldsymbol{\theta}_0^{(1)})$ of $\boldsymbol{\theta}_0^{(1)}$ and n large enough we have

$$\begin{aligned} E\left\{\boldsymbol{c}_1'\boldsymbol{x}_{t,n}^*\right\}^2 \mathbb{1}_{\left\{\left|\boldsymbol{c}_1'\boldsymbol{x}_{t,n}^*\right|\geqslant\sqrt{n}\varepsilon\right\}} &\leqslant \mathbb{1}_{\left\{\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\left|\boldsymbol{c}_1'\widetilde{\boldsymbol{D}}_t(\boldsymbol{\theta}^{(1)})\right|>0\right\}} \sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\left\{\boldsymbol{c}_1'\widetilde{\boldsymbol{D}}_t(\boldsymbol{\theta}^{(1)})\right\}^2 \\ &\quad \times E\left|\eta_{1t}^{*2}-\widehat{m}_2\right|^2 \mathbb{1}_{\left\{\left|\eta_{1t}^{*2}-\widehat{m}_2\right|\geqslant\frac{\sqrt{n}\varepsilon}{\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\left|\boldsymbol{c}_1'\widetilde{\boldsymbol{D}}_t(\boldsymbol{\theta}^{(1)})\right|}\right\}}. \end{aligned} \tag{35}$$

For any $A > 0$ there exists n_A such that if $n > n_A$ then the expectation in the right-hand side of (35) is bounded by $E\left|\eta_{1t}^{*2}-\widehat{m}_2\right|^2 \mathbb{1}_{\left\{\left|\eta_{1t}^{*2}-\widehat{m}_2\right|\geqslant A\right\}}$. By the arguments of the proof of Lemma E.1, this term tends to $\int_{|x^2-1|\geqslant A}\left|x^2-1\right|^2 G^{(1)}(dx)$ which is arbitrarily small when A is sufficiently large. We then obtain (34) for $\boldsymbol{c} = (\boldsymbol{c}_1', \boldsymbol{0}_3')'$. A similar argument shows (34) for $\boldsymbol{c} = (\boldsymbol{0}_{d_1}', 0, 0, 1)'$. Now, note that

$$E\left\{\mathbb{1}_{\eta_{1t}^*\leqslant\hat{u},\,\eta_{2t}^*\leqslant\hat{\xi}} - \hat{\alpha}\hat{\alpha}'\right\}^2 \mathbb{1}_{\left\{\left|\mathbb{1}_{\eta_{1t}^*\leqslant\hat{u},\,\eta_{2t}^*\leqslant\hat{\xi}}-\hat{\alpha}\hat{\alpha}'\right|\geqslant\sqrt{n}\varepsilon\right\}} = 0$$

for n large enough, which shows (34) for $\boldsymbol{c} = (\boldsymbol{0}_{d_1}', 1, 0, 0)'$. By the same argument, the convergence holds for $\boldsymbol{c} = (\boldsymbol{0}_{d_1}', 0, 1, 0)'$. We thus have shown (34) and the proof of (6) is complete. It follows that the bootstrap distribution of $\sqrt{n}\left\{\mathrm{CoVaR}^* - \widehat{\mathrm{CoVaR}}_{n+1}\right\}$ is approximatively the distribution of $\sqrt{n}\left\{\widehat{\mathrm{CoVaR}}_{n+1} - \mathrm{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}(\alpha,\alpha')\right\}$ when n is large. The fact that the 3 CIs and SY have approximated $(1-\alpha_0)$ coverage probabilities $\square$

References

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

Adrian, T. and M. Brunnermeier (2011) CoVaR. Technical Report, NBER.

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

Amengual, D., Fiorentini, G. and E. Sentana (2013) Sequential estimation of shape parameters in multivariate dynamic models. *Journal of Econometrics* 177, 233-249.

Andrews, D.W. (2000) Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica* 68, 399–405.

Ang, A. and F.A. Longstaff (2013) Systemic Sovereign Credit Risk: Lessons from the U.S. and Europe. *Journal of Monetary Economics* 60, 493–510.

Banulescu, D., Hurlin, C., Leymarie, J. and O. Scaillet (2021) Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures. *Management Science* 67, 5730-5754.

Bartholomew, D.J. (1967) Hypothesis testing when the sample size is treated as a random variable. *Journal of the Royal Statistical Society: Series B* 29, 53–70.

Berkes, I. and L. Horváth (2003) Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and their Applications* 105, 271–298.

Beutner, E., Heinemann, A. and S. Smeekes (2024) A residual bootstrap for conditional Value-at-Risk. *Journal of Econometrics* 238, 105554.

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

Billingsley, P. (1968) *Convergence of probability measures.* Wiley, New York.

Billingsley, P. (1999) *Convergence of probability measures.* 2nd Edition, Wiley, New York.

Boldin, M.V. (1998) On residual empirical distribution functions in ARCH models with applications to testing and estimation. *Mitt. Math. Sem. Giessen* 235, 49–66.

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

Cavaliere, G., Nielsen, H.B., Pedersen R.S. and A. Rahbek (2022) Bootstrap inference on the boundary of the parameter space with application to conditional volatility models. *Journal of Econometrics* 227, 241–263.

Cavaliere, G. and I. Georgiev (2020) Inference under random limit bootstrap measures. *Econometrica* 88, 2547–2574.

Christoffersen, P.F. (2003) *Elements of Financial Risk Management.* Academic Press.

Davis, R.A. and T. Mikosch (2009) The extremogram: a correlogram for extreme event. *Bernoulli* 15, 977–1009.

Davis, R.A., Mikosch, T. and I. Cribben (2012) Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics* 170, 142–152.

Dimitriadis, T. and Y. Hoga (2022) Dynamic CoVaR modeling. arXiv preprint arXiv:2206.14275, revised feb 2024.

Embrechts, P. and M.A Hofert (2013) A note on generalized inverses. *Mathematical Methods of Operations Research* 77, 423–432.

Escanciano, J.C. and S.C. Goh (2014) Specification analysis of linear quantile models. *Journal of Econometrics* 178, 495–507.

Escanciano, J.C. (2008) Joint and marginal specification tests for conditional mean and variance models. *Journal of Econometrics* 143, 74–87.

Escanciano, J.C. and J. Olmo (2010) Backtesting parametric value-at-risk with estimation risk. *Journal of Business & Economic Statistics* 28, 36–51.

Escanciano, J.C. and P. Pei (2012) Pitfalls in backtesting historical simulation VaR models. *Journal of Banking & Finance* 36, 2233–2244.

Engle, R.F. and S. Manganelli (2004) CAViaR: Conditional Value at risk by Quantile Regression. *Journal of Business and Economic Statistics* 22, 367–381.

Falk, M., and E. Kaufmann (1991) Coverage probabilities of bootstrap-confidence intervals for quantiles. *The Annals of Statistics* 19, 485–495.

Francq, C. and J-M. Zakoïan (2013) Optimal predictions of powers of conditionally heteroscedastic processes. *Journal of the Royal Statistical Society: Series B* 75, 345–367.

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

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

FRANCQ, C. AND J-M. ZAKOÏAN (2019) *GARCH Models: Structure, Statistical Inference and Financial Applications.* 2nd Edition. John Wiley.

FRANCQ, C. AND J-M. ZAKOÏAN (2022) Adaptiveness of the empirical distribution of residuals in semi-parametric conditional location scale models. *Bernoulli* 28, 548–578.

GIESECKE, K. AND B. KIM (2011) Systemic risk: what defaults are telling us. *Management Science* 57, 1387–1405.

GIRARDI, G., AND A.T. ERGÜN (2013) Systemic risk measurement: Multivariate GARCH estimation of CoVaR. *Journal of Banking & Finance* 37, 3169–3180.

HALDANE, A.G. AND R.M. MAY (2011) Systemic risk in banking ecosystems. *Nature* 469, 351–355.

HALL, P. AND Q. YAO (2003) Inference in ARCH and GARCH models with heavyâĂŞtailed errors. *Econometrica* 71, 285–317.

HORVÁTH, L., KOKOSZKA, P. AND G. TEYSSIÈRE (2001) Empirical process of the squared residuals of an ARCH sequence. *The Annals of Statistics* 29, 445–469.

JORION, P. (2007). *Value-at-risk.* McGraw-Hill, Third edition.

KREISS, J.P., PAPARODITIS, E. AND D.N. POLITIS (2011) On the range of validity of the autoregressive sieve bootstrap. *Annals of Statistics* 39, 2103–2130.

KULPERGER, R. AND H. YU (2005) High moment partial sum processes of residuals in GARCH models and their applications. *The Annals of Statistics* 33, 2395–2422.

KUPIEC, P. (1995) Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* 3, 73–84.

LUCAS, A., SCHWAAB, B. AND X. ZHANG (2014) Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics* 32, 271–284.

MCNEIL, A.J., FREY, R., AND P. EMBRECHTS (2015) *Quantitative risk management: concepts, techniques and tools.* Revised edition. Princeton University Press.

MORELLI, J.M., OTTONELLO, P. AND D.J. PEREZ (2022) Global banks and systemic debt crises. *Econometrica* 90, 749–798.

OH, D.W. AND A.J. PATTON (2018) Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics* 36, 181–195.

PATTON, A.J., ZIEGEL, J.F. AND R. CHEN (2019) Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics* 211, 388–413.

SEGOVIANO, M.A. AND C. GOODHART (2009) Banking stability measures. IMF Working Paper WP/09/04, International Monetary Fund.

SHIMIZU, K. (2010) *Bootstrapping Stationary ARMA-GARCH Models.* Technische Universität Braunschweig.

SHIMIZU, K. (2013) The bootstrap does not always work for heteroscedastic models. *Statistics & Risk modeling* 30, 189–204.

SHORACK, G.R. AND J.A. WELLNER (1986) *Empirical processes with applications to statistics.* Wiley, New-York.

Supplementary Appendix to Inference on Dynamic Systemic Risk Measures

Christian Francq and Jean-Michel Zakoïan

This supplementary file provides additional proofs and complementary results.

D Assumptions A1-A4 for Model (13)

Consider the bivariate GARCH Model

$$\epsilon_{it} = \sigma_{it}\eta_{it}, \quad \sigma_{it}^2 = \omega_{0i} + \alpha_{0,ii}\epsilon_{i,t-1}^2 + \alpha_{0,ij}\epsilon_{j,t-1}^2 + \beta_{0i}\sigma_{i,t-1}^2, \quad i,j = 1,2. \tag{36}$$

where $E(\eta_{it}) = 0, E(\eta_{it}^2) = 1$. The joint distribution of the iid process $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t})'$ is not specified (i.e. not restricted to be a Student distribution as in Model (13) of Section 6), but we assume it is nondegenerate, in the sense that

$$a\eta_{1t}^2 + b\eta_{2t}^2 = c, \quad a.s. \Longrightarrow \quad a = b = c = 0. \tag{37}$$

We assume $\alpha_{0,ij}, \beta_{0i} \geqslant 0$ and $\omega_{0i} > 0$ for $i,j = 1,2$. The strict stationarity condition follows from the bivariate Markov representation

$$\boldsymbol{Z}_t := \begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} \omega_{01} \\ \omega_{02} \end{pmatrix} + \boldsymbol{A}(\boldsymbol{\eta}_{t-1})\boldsymbol{Z}_{t-1},$$

where, for $\boldsymbol{x} = (x_1, x_2)'$, $\boldsymbol{A}(x) = \begin{pmatrix} \alpha_{0,11}x_1^2 + \beta_{01} & \alpha_{0,12}x_2^2 \\ \alpha_{0,21}x_1^2 & \alpha_{0,22}x_2^2 + \beta_{02} \end{pmatrix}$. Let $\gamma_{\boldsymbol{A}}$ denote the top Lyapunov exponent of the sequence of random matrices $\boldsymbol{A}(\boldsymbol{\eta}_t)$. Brandt (1986) showed that $\gamma_{\boldsymbol{A}} < 0$ ensures the existence of a unique strictly stationary solution $(\boldsymbol{\epsilon}_t)$ to Model (36). Moreover, this solution is nonanticipative (i.e. $\boldsymbol{\eta}_t$ is independent from $\{\boldsymbol{\epsilon}_{t-u}, u > 0\}$). A consequence of the strict stationarity condition is that $\beta_i < 1$, for $i = 1,2$. The parameter space $\boldsymbol{\Theta}$ is such that, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, for $i,j = 1,2$, we have $\alpha_{ij} \geqslant 0, \beta_i \in (0,1)$ and $\omega_{0i} > \underline{\omega}$, for some $\underline{\omega} > 0$. The condition, $\beta_i \in (0,1)$ entails, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, an expansion of the form $\sigma_{it}^2(\boldsymbol{\theta}) = c_i(\boldsymbol{\theta}) + \sum_{j=1}^2 \sum_{k=1}^\infty d_{ij,k}(\boldsymbol{\theta})\epsilon_{j,t-k}^2$, where the coefficients $d_{ij,k}(\boldsymbol{\theta})$ decrease exponentially fast to 0 as $k \to \infty$. This, together with the existence of a small-order moment for the components of $\boldsymbol{\epsilon}_t$, allows to show that $|\sigma_{it}(\boldsymbol{\theta}^{(i)}) - \tilde{\sigma}_{it}(\boldsymbol{\theta}^{(i)})|) \leqslant K\rho^t$. Now assume that $\sigma_{it}(\boldsymbol{\theta}_0^{(i)}) = \sigma_{it}(\boldsymbol{\theta}^{(i)})$ $a.s.$ In view of (37), we have

$$(\omega_{0i} - \omega_i) + (\alpha_{0,i} - \alpha_i)\eta_{i,t-1}^2\sigma_{i,t-1}^2 + (\alpha_{0,ij} - \alpha_{0,ij})\eta_{j,t-1}^2\sigma_{j,t-1}^2 + (\beta_{0i} - \beta_i)\sigma_{i,t-1}^2 = 0,$$

thus $\alpha_{0,i} = \alpha_i$, $\alpha_{0,ij} = \alpha_{0,ij}$ and $(\omega_{0i} - \omega_i) + (\beta_{0i} - \beta_i)\sigma_{i,t-1}^2 = 0$. It is straightforward to see that (37) also entails that the variable $\sigma_{i,t-1}^2$ cannot be a.s. constant whenever $\alpha_{0,i1} + \alpha_{0,i2} \neq 0$. Thus $\omega_{0i} = \omega_i$ and $\beta_{0i} = \beta_i$, showing that identifiability holds (the last part of Assumption **A2**). Because the strictly stationary solution admits a small-order moment (see for instance Corollary 2.3 in Francq and Zakoian (2019)), the moment conditions in Assumptions **A3, A4** are satisfied.

We thus have shown that, for Model (36), Assumptions **A1-A4** reduce to
(i) the strict stationarity condition $\gamma_{\boldsymbol{A}} < 0$,
(ii) the identifiability condition (37), and
(iii) the invertibility and positivity conditions $\beta_i \in (0,1)$, $\alpha_{0,i1} + \alpha_{0,i2} \neq 0$ and $\omega_i > \underline{\omega}$ for $i = 1,2$.

E Proofs and complementary results for Section 2

Proof of Proposition 2.2. The first equivalence is a straightforward consequence of the positivity of σ_{1t} and the definitions of the conditional VaR and CoVaR. We have, using the continuity of the distributions, $u(\alpha,\alpha') \leqslant \xi_\alpha^{(1)} \iff \mathbb{P}[\eta_{1t} \leqslant u(\alpha,\alpha')] \leqslant \mathbb{P}\left[\eta_{1t} \leqslant \xi_\alpha^{(1)}\right] \iff \alpha\alpha' + \mathbb{P}\left[\eta_{1t} \leqslant u(\alpha,\alpha') | \eta_{1t} > \xi_{\alpha'}^{(2)}\right](1-\alpha') \leqslant \alpha \iff \mathbb{P}\left[\eta_{1t} \leqslant u(\alpha,\alpha') | \eta_{1t} > \xi_{\alpha'}^{(2)}\right] \leqslant \alpha = F[u(\alpha,\alpha') \mid \xi_{\alpha'}^{(2)}]$. We also have $u(\alpha,\alpha') \leqslant \xi_\alpha^{(1)} \iff \mathbb{P}\left[\eta_{1t} \leqslant u(\alpha,\alpha') \mid \eta_{2t} \leqslant \xi_{\alpha'}^{(2)}\right] \leqslant \mathbb{P}\left[\eta_{1t} \leqslant \xi_\alpha^{(1)} \mid \eta_{2t} \leqslant \xi_{\alpha'}^{(2)}\right] \iff \mathbb{P}\left[\eta_{1t} \leqslant \xi_\alpha^{(1)} \mid \eta_{2t} > \xi_{\alpha'}^{(2)}\right] \leqslant \alpha \leqslant P[\eta_{1t} \leqslant \xi_\alpha^{(1)} \mid \eta_{2t} \leqslant \xi_{\alpha'}^{(2)}]$, using the fact that $\mathbb{P}\left[\eta_{1t} \leqslant \xi_\alpha^{(1)}\right] = \alpha = \mathbb{P}\left[\eta_{1t} \leqslant \xi_\alpha^{(1)} \mid \eta_{2t} \leqslant \xi_{\alpha'}^{(2)}\right]\alpha' + \mathbb{P}\left[\eta_{1t} \leqslant \xi_\alpha^{(1)} \mid \eta_{2t} > \xi_{\alpha'}^{(2)}\right](1-\alpha')$.

Now suppose there exist a random variable U_t and two increasing right-continuous functions ϕ and ψ (with generalized inverses ϕ^- and ψ^-) such that $\eta_{1t} = \phi(U_t)$ and $\eta_{2t} = \psi(U_t)$ a.s. We have (see Embrechts and Hofert (2013) for the necessity of the right-continuity assumption),

$$P[\eta_{1t} \leqslant u(\alpha,\alpha') \mid \eta_{2t} \leqslant \xi_{\alpha'}^{(2)}] = P[U_t \leqslant \phi^-\{u(\alpha,\alpha')\} \mid U_t \leqslant \psi^-\{\xi_{\alpha'}^{(2)}\}].$$

The latter probability is equal to 1 when $\phi^-\{u(\alpha,\alpha')\} \geqslant \psi^-\{\xi_{\alpha'}^{(2)}\}$, so the third inequality of the proposition is satisfied in this case. When $\phi^-\{u(\alpha,\alpha')\} < \psi^-\{\xi_{\alpha'}^{(2)}\}$ we have

$$\begin{aligned}(1-\alpha')P[\eta_{1t} \leqslant u(\alpha,\alpha') \mid \eta_{2t} > \xi_{\alpha'}^{(2)}] &= P[\eta_{1t} \leqslant u(\alpha,\alpha'), \eta_{2t} \geqslant \xi_{\alpha'}^{(2)}] \\ &= P[U_t \leqslant \phi^-\{u(\alpha,\alpha')\}, U_t \geqslant \psi^-\{\xi_{\alpha'}^{(2)}\}] = 0\end{aligned}$$

thus the third inequality is again satisfied. □

Proof of Proposition 2.3. We will prove the extended version below, including the case where (ii) in $\mathbf{A6}(\alpha,\alpha')$ does not hold, which can be stated as follows. Recall that $u(\alpha,\alpha') = \inf\{x : F(x \mid \xi_{\alpha'}^{(2)}) \geqslant \alpha\}$. Let $u^+(\alpha,\alpha') = \inf\{x : F(x \mid \xi_{\alpha'}^{(2)}) > \alpha\}$.

Proposition E.1. *Under* **A1**-**A4**, $\mathbf{A5_1}$, $\mathbf{A5_2}(\xi_{\alpha'}^{(2)})$, $\mathbf{A6}(\alpha,\alpha')$, $\mathbf{A7}(\xi_{\alpha'}^{(2)})$, *we have* $\widehat{u}(\alpha,\alpha') \to u(\alpha,\alpha')$ *a.s. Without (ii)* $\mathbf{A6}(\alpha,\alpha')$, *we have*

$$[\liminf \widehat{u}(\alpha,\alpha'), \limsup \widehat{u}(\alpha,\alpha')] \subseteq [u(\alpha,\alpha'), u^+(\alpha,\alpha')], \quad a.s. \tag{38}$$

Proof. Let us prove (38). By definition of $u(\alpha,\alpha')$ and $u^+(\alpha,\alpha')$ we have, for any $\epsilon > 0$,

$$F(u(\alpha,\alpha') - \epsilon \mid \xi_{\alpha'}^{(2)}) < \alpha - \delta \quad \text{and} \quad F(u^+(\alpha,\alpha') + \epsilon \mid \xi_{\alpha'}^{(2)}) > \alpha + \delta, \tag{39}$$

for some $\delta > 0$. By Proposition B.1 *iii)*, assume n large enough so that $\sup_{x\in\mathbb{R}, y\in V(\xi_{\alpha'}^{(2)})} |\widehat{F}(x \mid y) - F(x \mid y)| < \delta/2$ a.s. By Corollary 4.1 in FZ, we know that under $\mathbf{A6}(\alpha')$ (i), $\widehat{\xi}_{\alpha'}^{(2)} \to \xi_{\alpha'}^{(2)}$ a.s. It thus follows that $\sup_{x\in\mathbb{R}} |F(x|\widehat{\xi}_{\alpha'}^{(2)}) - F(x|\xi_{\alpha'}^{(2)})| < \delta/2$ a.s. for n large enough, in view of $\mathbf{A7}(\xi_{\alpha'}^{(2)})$. Now we will show that (39) entails

$$u(\alpha,\alpha') - \epsilon \leqslant \widehat{u}(\alpha,\alpha') \leqslant u^+(\alpha,\alpha') + \epsilon \tag{40}$$

for n large enough. Indeed, if $u(\alpha,\alpha') - \epsilon > \widehat{u}(\alpha,\alpha')$ then, for n large enough,

$$\begin{aligned} & F(u(\alpha,\alpha') - \epsilon \mid \xi_{\alpha'}^{(2)}) \geqslant F(\widehat{u}(\alpha,\alpha') \mid \xi_{\alpha'}^{(2)}) \\ = \; & \underbrace{F(\widehat{u}(\alpha,\alpha') \mid \xi_{\alpha'}^{(2)}) - F(\widehat{u}(\alpha,\alpha') \mid \widehat{\xi}_{\alpha'}^{(2)})}_{>-\delta/2,\ a.s.\ \text{by } \mathbf{A6_1}(\alpha') \text{ and } \mathbf{A7}(\xi_{\alpha'}^{(2)})} + \underbrace{F(\widehat{u}(\alpha,\alpha') \mid \widehat{\xi}_{\alpha'}^{(2)}) - \widehat{F}(\widehat{u}(\alpha,\alpha') \mid \widehat{\xi}_{\alpha'}^{(2)})}_{>-\delta/2,\ a.s.\ \text{by } \mathbf{A6_1}(\alpha') \text{ and iii) of Proposition B.1}} \end{aligned}$$

$$+\underbrace{\widehat{F}(\widehat{u}(\alpha,\alpha') \mid \widehat{\xi}^{(2)}_{\alpha'})}_{\geqslant\alpha} \geqslant \alpha-\delta,$$

which contradicts the first inequality in (43). Moreover, if $u^+(\alpha,\alpha')+\epsilon < \widehat{u}(\alpha,\alpha')$ then by the same arguments $F(u^+(\alpha,\alpha')+\epsilon \mid \xi^{(2)}_{\alpha'}) \leqslant \widehat{F}(u^+(\alpha,\alpha')+\epsilon \mid \xi^{(2)}_{\alpha'})+\delta/2 \leqslant \alpha+\delta$, which contradicts the second inequality in (39). Hence (40) is shown. The strong convergence of $\widehat{u}(\alpha,\alpha')$ to the set $[u(\alpha,\alpha'),u^+(\alpha,\alpha')]$ follows from (40). Thus (38) is established. Now, if $\mathbf{A6_2}$ holds, the previous set reduces to the singleton $\{u(\alpha,\alpha')\}$. □

Proof of Proposition 2.4. As for Proposition 2.3, we will prove the following extended version, including the case where (ii) of $\mathbf{A6}(\alpha,0.5+\tau\alpha'')$ does not hold. Recalling that $\underline{u}(\alpha,\alpha'')=\inf\{x: F^{\Delta}(x \mid (\xi^{(2)}_{0.5-\alpha''},\xi^{(2)}_{\alpha''+0.5}]) \geqslant \alpha\}$, let $\underline{u}^+(\alpha,\alpha'')=\inf\{x: F^{\Delta}(x \mid (\xi^{(2)}_{0.5-\alpha''},\xi^{(2)}_{\alpha''+0.5}]) > \alpha\}$.

Proposition E.2. *Under* $\mathbf{A1}$-$\mathbf{A4}$, $\mathbf{A5_1}$ *and, for* $\tau\in\{-1,1\}$ *and* $\alpha''\in(0,1/2)$, $\mathbf{A5_2}(\xi^{(2)}_{0.5+\tau\alpha''})$, $\mathbf{A6}(\alpha,0.5+\tau\alpha'')$, $\mathbf{A7}(\xi^{(2)}_{0.5+\tau\alpha''})$, *we have* $\widehat{\underline{u}}(\alpha,\alpha'')\to\underline{u}(\alpha,\alpha'')$ *a.s. Without (ii) of* $\mathbf{A6}(\alpha,0.5+\tau\alpha'')$, *we have*

$$[\liminf \widehat{\underline{u}}(\alpha,\alpha''),\limsup \widehat{\underline{u}}(\alpha,\alpha'')] \subseteq [\underline{u}(\alpha,\alpha''),\underline{u}^+(\alpha,\alpha'')] \quad a.s. \tag{41}$$

Proof. Note that

$$F^{\Delta}(x|(y_1,y_2]) = \frac{F(x\mid y_2)G^{(2)}(y_2)-F(x\mid y_1)G^{(2)}(y_1)}{G^{(2)}(y_2)-G^{(2)}(y_1)}. \tag{42}$$

By definition of $\underline{u}(\alpha,\alpha'')$ and $\underline{u}^+(\alpha,\alpha'')$ we have, for any $\epsilon>0$,

$$F^{\Delta}(\underline{u}(\alpha,\alpha'')-\epsilon \mid A_{\alpha''}) < \alpha-\delta \quad \text{and} \quad F^{\Delta}(\underline{u}^+(\alpha,\alpha'')+\epsilon \mid A_{\alpha''}) > \alpha+\delta, \tag{43}$$

for some $\delta>0$. By arguments already used, we have under $\mathbf{A6}(0.5+\tau\alpha'')$ (i), $\widehat{\xi}^{(2)}_{0.5+\tau\alpha''}\to\xi^{(2)}_{0.5+\tau\alpha'}$ a.s. for $\tau\in\{-1,1\}$. It is clear that, in view of (42) and $\mathbf{A7}(\xi^{(2)}_{0.5+\tau\alpha''})$, $\sup_{x\in\mathbb{R}}|F^{\Delta}(x|\widehat{A}_{n,\alpha''})-F^{\Delta}(x|A_{\alpha''})|<\delta/2$ a.s. for n large enough.

Let, for A such that $\mathbb{P}(\eta_{2t}\in A)>0$ and n large enough, $\widehat{F}^{\Delta}_n(x|A)=\frac{\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{1t}\leqslant x,\widehat{\eta}_{2t}\in A}}{\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{2t}\in A}}$. It follows from (42), $\mathbf{A5_1}$ and $\mathbf{A5_2}(\xi^{(2)}_{0.5+\tau\alpha''})$ that, from Proposition B.1, $|F^{\Delta}(\widehat{\underline{u}}(\alpha,\alpha'')\mid\widehat{A}_{n,\alpha''})-\widehat{F}^{\Delta}(\widehat{\underline{u}}(\alpha,\alpha'')\mid\widehat{A}_{n,\alpha''})|\to 0$ *a.s.* We will show that (43) entails

$$\underline{u}(\alpha,\alpha'')-\epsilon \leqslant \widehat{\underline{u}}(\alpha,\alpha'') \leqslant \underline{u}^+(\alpha,\alpha'')+\epsilon \tag{44}$$

for n large enough. Indeed, if $\underline{u}(\alpha,\alpha'')-\epsilon > \widehat{\underline{u}}(\alpha,\alpha'')$ then, for n large enough,

$$\begin{aligned} & F^{\Delta}(\underline{u}(\alpha,\alpha'')-\epsilon \mid A_{\alpha''}) \geqslant F^{\Delta}(\widehat{\underline{u}}(\alpha,\alpha'')\mid A_{\alpha''}) \\ = \; & \underbrace{F^{\Delta}(\widehat{\underline{u}}(\alpha,\alpha'')\mid A_{\alpha''})-F^{\Delta}(\widehat{\underline{u}}(\alpha,\alpha'')\mid \widehat{A}_{n,\alpha''})}_{>-\delta/2,\ a.s.} \\ & +\underbrace{F^{\Delta}(\widehat{\underline{u}}(\alpha,\alpha'')\mid \widehat{A}_{n,\alpha''})-\widehat{F}^{\Delta}(\widehat{\underline{u}}(\alpha,\alpha'')\mid \widehat{A}_{n,\alpha''})}_{>-\delta/2,\ a.s.}+\underbrace{\widehat{F}^{\Delta}_n(\widehat{\underline{u}}(\alpha,\alpha'')\mid \widehat{A}_{n,\alpha''})}_{\geqslant\alpha-1/n} \geqslant \alpha-\delta, \end{aligned}$$

which contradicts the first inequality in (43). The rest of the proof is similar to that of Proposition 2.3. □

Proof of Proposition 2.5. Note that conditional on $(\boldsymbol{\epsilon}_t)$ and (x_n), we have $\frac{1}{n}\sum_{t=1}^n \widehat{\eta}_{1t}\mathbb{1}_{\widehat{\eta}_{2t}\leqslant x_n} = E\left(\eta^*_{1n}\mathbb{1}_{\eta^*_{2n}\leqslant x_n}\right)$ where $(\eta^*_{1n},\eta^*_{2n})\sim\widehat{H}$. Moreover,

$$\widehat{H}(x_n,y_n)-H(x,y)=\left\{\widehat{F}(x_n\mid y_n)-F(x_n\mid y_n)\right\}\widehat{G}^{(2)}(y_n)$$

$$+ F(x_n \mid y_n)\left\{\widehat{G}^{(2)}(y_n) - G^{(2)}(y_n)\right\} + H(x_n, y_n) - H(x, y).$$

Thus, for any sequence of random vectors (x_n, y_n) converging almost surely to some $(x, y) \in \mathbb{R}^2$, Proposition B.1, Theorem 2.1. in FZ and **B8** entail $|\widehat{H}(x_n, y_n) - H(x, y)| \to 0, \quad a.s.$ It follows that, conditional on $(\boldsymbol{\epsilon}_t)$ and (x_n), as $n \to \infty$ the random variable $\eta_{1n}^* \mathbb{1}_{\eta_{2n}^* \leqslant x_n}$ converges in distribution to $\eta_{10}\mathbb{1}_{\eta_{20} \leqslant x}$. To conclude by Theorem 3.5 in Billingsley (1999)[16], we need to show that the variables $\eta_{1n}^* \mathbb{1}_{\eta_{2n}^* \leqslant x_n}$ are uniformly integrable. It suffices to show that the variables η_{1n}^{*2} are uniformly integrable (see (3.18) in Billingsley (1999)). This will follow from Theorem 3.6 in Billingsley (1999)[17], if we show that

$$\frac{1}{n}\sum_{t=1}^{n} \widehat{\eta}_{1t}^2 = E\eta_{1n}^{*2} \to E\eta_{10}^2 \text{ as } n \to \infty. \tag{45}$$

Let $\eta_{1t}(\boldsymbol{\theta}^{(1)}) = \frac{\epsilon_{1t} - \mu_{1t}(\boldsymbol{\theta}^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})}$ and $\widetilde{\eta}_{1t}(\boldsymbol{\theta}^{(1)}) = \frac{\epsilon_{1t} - \widetilde{\mu}_{1t}(\boldsymbol{\theta}^{(1)})}{\widetilde{\sigma}_{1t}(\boldsymbol{\theta}^{(1)})}$, so that $\widehat{\eta}_{1t} = \widetilde{\eta}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)})$ and $\eta_{1t} = \eta_{1t}(\boldsymbol{\theta}_0^{(1)})$. By **A2** and **A3** we have $\sup_{\boldsymbol{\theta}^{(1)} \in \boldsymbol{\Theta}^{(1)}} \left|\eta_{1t}^2(\boldsymbol{\theta}^{(1)}) - \widetilde{\eta}_{1t}^2(\boldsymbol{\theta}^{(1)})\right| \leqslant \frac{K}{\underline{\omega}}\rho^t \epsilon_{1t}^2$. Under **A1**, there exists $s < 1$ such that $E|\epsilon_{1t}|^{2s} < \infty$. Thus $\sum_{t=1}^n \rho^t \epsilon_{1t}^2$ is finite almost surely because $E\left|\sum_{t=1}^n \rho^t \epsilon_{1t}^2\right|^s \leqslant E|\epsilon_{1t}|^{2s} \sum_{t=1}^n \rho^{ts} < \infty$. It follows that $\frac{1}{n}\sum_{t=1}^n \widehat{\eta}_{1t}^2 = \frac{1}{n}\sum_{t=1}^n \eta_{1t}^2(\widehat{\boldsymbol{\theta}}^{(1)}) + O(n^{-1})$ a.s. By the mean value theorem $\frac{1}{n}\sum_{t=1}^n \eta_{1t}^2(\widehat{\boldsymbol{\theta}}^{(1)}) = \frac{1}{n}\sum_{t=1}^n \eta_{1t}^2 + \frac{1}{n}\sum_{t=1}^n \frac{\partial \eta_{1t}^2(\boldsymbol{\theta}_n^{(1)})}{\partial \boldsymbol{\theta}^{(1)\prime}}\left(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0^{(1)}\right)$, with $\boldsymbol{\theta}_n^{(1)}$ between $\widehat{\boldsymbol{\theta}}^{(1)}$ and $\boldsymbol{\theta}_0^{(1)}$. By **A9** we have

$$\begin{aligned}
\sup_{\boldsymbol{\theta}^{(1)} \in V(\boldsymbol{\theta}_0^{(1)})} \left\|\frac{\partial \eta_{1t}^2(\boldsymbol{\theta}^{(1)})}{\partial \boldsymbol{\theta}^{(1)\prime}}\right\| &\leqslant 2\left\{\sup_{\boldsymbol{\theta}^{(1)} \in V(\boldsymbol{\theta}_0^{(1)})} \left\|\eta_{1t}^2(\boldsymbol{\theta}^{(1)})\boldsymbol{D}_{1t}(\boldsymbol{\theta}^{(1)})\right\| + \left\|\frac{\eta_{1t}(\boldsymbol{\theta}^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})}\frac{\partial \mu_{1t}(\boldsymbol{\theta}^{(1)})}{\partial \boldsymbol{\theta}}\right\|\right\}, \\
&\leqslant 2 \sup_{\boldsymbol{\theta}^{(1)} \in V(\boldsymbol{\theta}_0^{(1)})} \left\|\left(\frac{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}{\sigma_{1t}(\boldsymbol{\theta}^{(1)})}\right)^2 \boldsymbol{D}_{1t}(\boldsymbol{\theta}^{(1)})\eta_{1t}^2\right\| = u_t|\eta_{1t}^2|,
\end{aligned}$$

where $u_t \in \mathcal{F}_{t-1}$, (u_t) is stationary and ergodic and $Eu_t < \infty$. We thus have shown that (45) holds. The conclusion follows. □

F CoVaR in the Gaussian case

Let $(\eta_{1t}, \eta_{2t})' \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$.

F.1 Complement to Proposition 2.2.

Letting $F_0(x|y) = P[\eta_{1t} \leqslant x \mid \eta_{2t} \geqslant y]$, we will show that $F(x|y) > F_0(x|y)$ iff $\rho > 0$. Let ϕ and Φ denote, respectively, the density and the cdf of the standard Gaussian distribution. Let $\overline{\Phi}(x) = 1 - \Phi(x)$. We have $F_0(x|y) = \frac{\{1 - F(y|x)\}\Phi(x)}{\overline{\Phi}(y)}$. Hence $F(x|y) > F_0(x|y) \iff H(\rho) > 0$, where $H(\rho) = F(x|y)\overline{\Phi}(y) - \{1 - F(y|x)\}\Phi(x)$. We have $\frac{\partial H(\rho)}{\partial \rho} = \frac{1}{\sqrt{1-\rho^2}}\frac{\phi(x)}{\Phi(y)}\phi\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right) > 0$. Since $H(0) = 0$, the conclusion follows. □

[16] If X_n are uniformly integrable and $X_n \Rightarrow X$, then X is integrable and $EX_n \to EX$.

[17] If X and X_n are nonnegative and integrable, and if $X_n \Rightarrow X$ and $EX_n \to EX$, then the X_n are uniformly integrable.

F.2 Computation of the conditional density f_2

We have

$$\begin{aligned} F(y \mid x) &= \frac{H(x,y)}{\Phi(x)} = \frac{1}{\Phi(x)} \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_{\eta_{2t}\mid\eta_{1t}=u}(v)dv \right) \phi(u)du \\ &= \frac{1}{\Phi(x)} \int_{-\infty}^{x} \left(\int_{-\infty}^{y} \frac{1}{\sqrt{1-\rho^2}} \phi\left(\frac{v-\rho u}{\sqrt{1-\rho^2}} \right) dv \right) \phi(u)du \\ &= \frac{1}{\Phi(x)} \int_{-\infty}^{x} \Phi\left(\frac{y-\rho u}{\sqrt{1-\rho^2}} \right) \phi(u)du. \end{aligned}$$

It follows that

$$\begin{aligned} f_2(y \mid x) &= \frac{\partial}{\partial y} F(y \mid x) = \frac{1}{\Phi(x)} \int_{-\infty}^{x} \frac{1}{\sqrt{1-\rho^2}} \phi\left(\frac{y-\rho u}{\sqrt{1-\rho^2}} \right) \phi(u)du \\ &= \frac{1}{\Phi(x)} \int_{-\infty}^{x} \frac{1}{\sqrt{1-\rho^2}} \phi\left(\frac{u-\rho y}{\sqrt{1-\rho^2}} \right) \phi(y)du = \frac{\phi(y)}{\Phi(x)} \Phi\left(\frac{x-\rho y}{\sqrt{1-\rho^2}} \right). \end{aligned}$$

F.3 Confidence interval for the CoVaR

If the joint distribution of the innovations is known to be Gaussian, more accurate estimation can be expected from using this information. In this case, the coefficient u is entirely determined by the unknown correlation ρ and will be written $u(\rho)$. Let $\widehat{\rho}$ denote the sample autocorrelation obtained from the residuals, namely $\widehat{\rho} = \frac{1}{n}\sum_{t=1}^{n} \widehat{\eta}_{1t}\widehat{\eta}_{2t}$. An estimator of $u(\rho)$ is thus $u(\widehat{\rho})$ and the following asymptotic distribution holds.

Proposition F.1. *Under the assumptions of Theorem 3.1 and* **B7**, *in the Gaussian case we have*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}^{(1)'} - \boldsymbol{\theta}_0^{(1)'}, u(\widehat{\rho}) - u(\rho) \right)' \xrightarrow{\mathcal{L}} \mathcal{N}\left\{ \mathbf{0}, \boldsymbol{\Sigma}^{\mathcal{N}}(\alpha,\alpha') \right\},$$

where

$$\boldsymbol{\Sigma}^{\mathcal{N}}(\alpha,\alpha') = \begin{pmatrix} \frac{1}{2}\boldsymbol{J}_1^{-1} & \frac{-\rho}{2}\sqrt{1-\rho^2}K_{\alpha,\alpha'}(\rho)\boldsymbol{J}_1^{-1}\boldsymbol{\Omega}_1 \\ \frac{-\rho}{2}\sqrt{1-\rho^2}K_{\alpha,\alpha'}(\rho)\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1} & K^2_{\alpha,\alpha'}(\rho)(1-\rho^2) \end{pmatrix}$$

with $K_{\alpha,\alpha'}(\rho) = \frac{\phi\{z_{\alpha,\alpha'}(\rho)\}}{\Phi\{z_{\alpha,\alpha'}(\rho)\}}$ *and* $z_{\alpha,\alpha'}(\rho) = \frac{\xi_{\alpha'}^{(2)} - \rho u(\rho)}{\sqrt{1-\rho^2}}$.

Proof. We have

$$\begin{aligned} \sqrt{n}(\widehat{\rho}-\rho) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{\eta}_{1t}\widehat{\eta}_{2t}-\rho) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_{1t}\eta_{2t}-\rho) + \frac{1}{n}\sum_{t=1}^{n}\left[\eta_{2t}\frac{\partial \eta_{1t}(\boldsymbol{\theta}_0^{(1)})}{\partial \boldsymbol{\theta}^{(1)'}} \quad \eta_{1t}\frac{\partial \eta_{2t}(\boldsymbol{\theta}_0^{(2)})}{\partial \boldsymbol{\theta}^{(2)'}} \right] \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_{\mathbb{P}}(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_{1t}\eta_{2t}-\rho) - \rho\left[\boldsymbol{\Omega}_1' \ \ \boldsymbol{\Omega}_2' \right]\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_{\mathbb{P}}(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\eta_{1t}\eta_{2t}-\rho) - \frac{\rho}{2\sqrt{n}}\sum_{t=1}^{n}(\eta_{1t}^2+\eta_{2t}^2-2) + o_{\mathbb{P}}(1), \end{aligned}$$

using (4) and **B7**. We also have $\mathrm{Var}(\eta_{1t}\eta_{2t}) = 1 + \rho^2$, $\mathrm{Cov}(\eta_{1t}^2, \eta_{2t}^2) = 2\rho^2$ and $\mathrm{Cov}(\eta_{1t}\eta_{2t}, \eta_{it}^2) = 2\rho$, for $i = 1, 2$. It follows that $\mathrm{Var}_{as}\{\sqrt{n}(\widehat{\rho} - \rho)\} = (1 - \rho^2)^2$. We also have $\mathrm{Cov}_{as}\{\sqrt{n}(\widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}_0^{(1)}), \sqrt{n}(\widehat{\rho} - \rho)\} = \frac{1}{2}\rho(1 - \rho^2)\boldsymbol{J}_1^{-1}\boldsymbol{\Omega}_1$. In view of $\sqrt{n}\{u(\widehat{\rho}) - u(\rho)\} = \frac{\partial u(\rho)}{\partial \rho}\sqrt{n}(\widehat{\rho} - \rho) + o_{\mathbb{P}}(1)$, we thus have $\mathrm{Var}_{as}\sqrt{n}\{u(\widehat{\rho}) - u(\rho)\} = \left\{\frac{\partial u(\rho)}{\partial \rho}\right\}^2 (1 - \rho^2)^2$.Now, by differentiating with respect to ρ the equality

$$\int_{x<u(\rho)} \Phi\left(\frac{\xi_{\alpha'} - \rho x}{\sqrt{1-\rho^2}}\right)\phi(x)dx = \alpha\alpha',$$

where $\xi_{\alpha'}$ denotes the α'-quantile of the $\mathcal{N}(0,1)$ distribution, we get

$$\begin{aligned}
0 &= \int_{x<u(\rho)} \phi\left(\frac{\xi_{\alpha'} - \rho x}{\sqrt{1-\rho^2}}\right)\phi(x)\frac{\rho\xi_{\alpha'} - x}{(1-\rho^2)^{3/2}}dx + \frac{\phi(u(\rho))}{\sqrt{1-\rho^2}}\Phi\left(\frac{\xi_{\alpha'} - \rho u(\rho)}{\sqrt{1-\rho^2}}\right)\frac{\partial u(\rho)}{\partial \rho} \\
&= \frac{\phi(\xi_{\alpha'})}{\sqrt{1-\rho^2}}\phi\left(\frac{u(\rho) - \rho\xi_{\alpha'}}{\sqrt{1-\rho^2}}\right) + \phi(u(\rho))\Phi\left(\frac{\xi_{\alpha'} - \rho u(\rho)}{\sqrt{1-\rho^2}}\right)\frac{\partial u(\rho)}{\partial \rho} \\
&= \frac{\phi(u(\rho))}{\sqrt{1-\rho^2}}\phi\left(\frac{\xi_{\alpha'} - \rho u(\rho)}{\sqrt{1-\rho^2}}\right) + \phi(u(\rho))\Phi\left(\frac{\xi_{\alpha'} - \rho u(\rho)}{\sqrt{1-\rho^2}}\right)\frac{\partial u(\rho)}{\partial \rho}.
\end{aligned}$$

Thus

$$\frac{\partial u(\rho)}{\partial \rho} = \frac{-1}{\sqrt{1-\rho^2}}\frac{\phi\{z_{\alpha,\alpha'}(\rho)\}}{\Phi\{z_{\alpha,\alpha'}(\rho)\}}.$$

The asymptotic distribution of Theorem F.1 follows. □

By the delta method, an approximate $(1-\alpha_0)$ CI for $\mathrm{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}$, where $\alpha_0 \in (0,1)$, has bounds given by

$$-\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}^{(1)})u(\widehat{\rho}) \pm \frac{1}{\sqrt{n}}\Phi^{-1}(1 - \alpha_0/2)\left\{\boldsymbol{\delta}_{n+1}'\widehat{\boldsymbol{\Sigma}}^{\mathcal{N}}(\alpha,\alpha')\boldsymbol{\delta}_{n+1}\right\}^{1/2},$$

where $\Phi^{-1}(u)$ denotes the u-quantile of the standard Gaussian distribution, $u \in (0,1)$, $\widehat{\boldsymbol{\Sigma}}^{\mathcal{N}}(\alpha,\alpha')$ is a consistent estimator of $\boldsymbol{\Sigma}^{\mathcal{N}}(\alpha,\alpha')$, and $\boldsymbol{\delta}_{n+1}' = \left[\frac{\partial\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}^{(1)})}{\partial\boldsymbol{\theta}'}u(\widehat{\rho}) \quad \widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}^{(1)})\right]$.

G Asymptotic distribution of $\widehat{\underline{u}}(\alpha,\alpha'')$

The next result provides the asymptotic distribution of the estimator $\widehat{\underline{u}}(\alpha,\alpha'')$ of $\underline{u}(\alpha,\alpha'')$, simply denoted $\widehat{\underline{u}}$ and $\underline{u}$ in the sequel.

Theorem G.1. *Under the assumptions of Theorem 3.1 we have,*

$$\begin{aligned}
\sqrt{n}\left(\widehat{\underline{u}} - \underline{u}\right) &= \frac{1}{\Delta f_1(\underline{u},\alpha'')}\left\{\frac{-1}{\sqrt{n}}\sum_{t=1}^n\{\mathbb{1}_{\eta_{1t}\leqslant\underline{u}}\mathbb{1}_{\eta_{2t}\in A_{\alpha''}^{(2)}} - 2\alpha\alpha''\}\right. \\
&\left.+ G^{(1)}(\underline{u})\{S_n(0.5+\alpha'') - S_n(0.5-\alpha'')\}\right\} - \frac{\underline{u}}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^n(\eta_{1t}^2 - 1)\boldsymbol{D}_{1t} + o_{\mathbb{P}}(1),
\end{aligned}$$

where $\delta_{\alpha''} = \xi_{0.5+\alpha''}^{(2)}g^{(2)}(\xi_{0.5+\alpha''}^{(2)}) - \xi_{0.5-\alpha''}^{(2)}g^{(2)}(\xi_{0.5-\alpha''}^{(2)})$, $\Delta f_1(\underline{u},\alpha'') = (0.5+\alpha'')f_1(\underline{u} \mid \xi_{0.5+\alpha''}^{(2)}) - (0.5-\alpha'')f_1(\underline{u} \mid \xi_{0.5-\alpha''}^{(2)})$ *and, for any* $\alpha^* \in (0,1)$,

$$S_n(\alpha^*) = \frac{f_2(\xi_{\alpha^*}^{(2)} \mid \underline{u})}{g^{(2)}(\xi_{\alpha^*}^{(2)})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi_{\alpha^*}^{(2)}} - \alpha^*).$$

Under $\mathbf{B7_1}$ *we have,* $\sqrt{n}\{\hat{\underline{u}}-\underline{u}\} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \underline{\sigma}^2(\alpha,\alpha'')=\boldsymbol{\nu}'\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}\boldsymbol{\nu}\right)$, *where*

$$\boldsymbol{\nu}' = \left(-\frac{\underline{u}}{2}, \frac{-1}{\Delta f_1(\underline{u},\alpha'')}, \frac{G^{(1)}(\underline{u})}{\Delta f_1(\underline{u},\alpha'')}\frac{f_2(\xi^{(2)}_{0.5+\alpha''}\mid \underline{u})}{g^{(2)}(\xi^{(2)}_{0.5+\alpha''})}, \frac{-G^{(1)}(\underline{u})}{\Delta f_1(\underline{u},\alpha'')}\frac{f_2(\xi^{(2)}_{0.5-\alpha''}\mid \underline{u})}{g^{(2)}(\xi^{(2)}_{0.5-\alpha''})}\right).$$

and (the upper triangular part of) $\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}$ *is*

$$\begin{pmatrix} \kappa_1-1 & 2\alpha''\varrho_{\alpha,A^{(2)}_{\alpha''}}(\underline{u}) & (0.5+\alpha'')\nabla_{\alpha,0.5+\alpha''} & (0.5-\alpha'')\nabla_{\alpha,0.5-\alpha''} \\ & 2\alpha\alpha''(1-2\alpha\alpha'') & 2\alpha\alpha''(0.5-\alpha'') & -2\alpha\alpha''(0.5-\alpha'') \\ & & 0.5^2-(\alpha'')^2 & 0.5^2-\alpha''(1-\alpha'') \\ & & & 0.5^2-(\alpha'')^2 \end{pmatrix}$$

where $\varrho_{\alpha,A}(\underline{u}) = E(\eta_{1t}^2\mathbb{1}_{\eta_{1t}\leqslant\underline{u}}|\eta_{2t}\in A)-\alpha$.

Proof. Let

$$\widehat{F}^{\Delta}(x|[y_1,y_2)) = \frac{\widehat{F}(x\mid y_2)\widehat{G}^{(2)}(y_2)-\widehat{F}(x\mid y_1)\widehat{G}^{(2)}(y_1)}{\widehat{G}^{(2)}(y_2)-\widehat{G}^{(2)}(y_1)}. \tag{46}$$

Letting $\Delta G^{(2)}(y_1,y_2) = G^{(2)}(y_2)-G^{(2)}(y_1)$, $\quad \Delta F(x|y_1,y_2) = F(x|y_2)-F(x|y_1)$, we have

$$\begin{aligned}
&\sqrt{n}\{\widehat{F}^{\Delta}(x_n|[y_{1n},y_{2n}))-F^{\Delta}(x_n|[y_{1n},y_{2n}))\} \\
=&\frac{1}{\Delta G^{(2)}(y_1,y_2)}\sum_{i=1}^{2}(-1)^i\sqrt{n}\{\widehat{F}(x_n\mid y_{in})-F(x_n\mid y_{in})\}G^{(2)}(y_i)+ \\
&\frac{\Delta F(x|y_1,y_2)}{\{\Delta G^{(2)}(y_1,y_2)\}^2}\left[G^{(2)}(y_2)\sqrt{n}\{\widehat{G}^{(2)}(y_{1n})-G^{(2)}(y_{1n})\}-G^{(2)}(y_1)\sqrt{n}\{\widehat{G}^{(2)}(y_{2n})-G^{(2)}(y_{2n})\}\right]
\end{aligned}$$

up to an $o_{\mathbb{P}}(1)$ term. We also have, for $y_n\to y$,

$$\sqrt{n}\{\widehat{G}^{(2)}(y_n)-G^{(2)}(y_n)\} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{2t}\leqslant y}-G^{(2)}(y)\}+\frac{yg^{(2)}(y)}{2\sqrt{n}}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_t^2-1)\boldsymbol{D}_{2t}$$

up to an $o_{\mathbb{P}}(1)$ term. It follows that

$$\begin{aligned}
&\sqrt{n}\{\widehat{F}^{\Delta}(x_n|[y_{1n},y_{2n}))-F^{\Delta}(x_n|[y_{1n},y_{2n}))\} \\
=&\frac{1}{\sqrt{n}\Delta G^{(2)}(y_1,y_2)}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}\leqslant x,\,\eta_{2t}\in(y_1,y_2]}-\Delta H(x|y_1,y_2)\} \\
&+x\frac{G^{(2)}(y_2)f_1(x\mid y_2)-G^{(2)}(y_1)f_1(x\mid y_1)}{2\sqrt{n}\Delta G^{(2)}(y_1,y_2)}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^{n}(\eta_{1t}^2-1)\boldsymbol{D}_{1t} \\
&+\frac{a(x,y_1,y_2)}{2\sqrt{n}}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t} \\
&-\frac{F^{\Delta}(x|(y_1,y_2])}{\sqrt{n}\Delta G^{(2)}(y_1,y_2)}\sum_{t=1}^{n}\left\{\mathbb{1}_{\eta_{2t}\in(y_1,y_2]}-\Delta G^{(2)}(y_1,y_2)\right\}+o_{\mathbb{P}}(1),
\end{aligned}$$

where $\Delta H(x,y_1,y_2) = H(x,y_2)-H(x,y_1)$, and

$$\begin{aligned}
a(x,y_1,y_2) =&\frac{1}{\{\Delta G^{(2)}(y_1,y_2)\}^2}\left[\{y_1g^{(2)}(y_1)-y_2g^{(2)}(y_2)\}\Delta H(x,y_1,y_2)\right. \\
&\left.+\{y_2f_2(y_2|x)-y_1f_2(y_1|x)\}G^{(1)}(x)\Delta G^{(2)}(y_1,y_2)\right].
\end{aligned}$$

Note that $a(x, y_1, y_2) = 0$ when η_{1t} and η_{2t} are independent. It follows that

$$\begin{aligned}
&\sqrt{n}\{\widehat{F}^{\Delta}(x_n|[y_{1n}, y_{2n})) - F^{\Delta}(x_n|[y_{1n}, y_{2n}))\}\\
=&\frac{1}{\sqrt{n}\Delta G^{(2)}(y_1, y_2)} \sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}\leqslant x} - F^{\Delta}(x|[y_1, y_2))\}\mathbb{1}_{\eta_{2t}\in(y_1,y_2]}\\
&+x\frac{G^{(2)}(y_2)f_1(x \mid y_2) - G^{(2)}(y_1)f_1(x \mid y_1)}{2\sqrt{n}\Delta G^{(2)}(y_1, y_2)}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^{n}(\eta_{1t}^2 - 1)\boldsymbol{D}_{1t}\\
&+\frac{a(x, y_1, y_2)}{2\sqrt{n}}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2 - 1)\boldsymbol{D}_{2t} + o_{\mathbb{P}}(1).
\end{aligned}$$

Proceeding as in the proof of Theorem 3.1, we note that $\widehat{F}^{\Delta}(\underline{\widehat{u}} \mid \widehat{A}^{(2)}_{n,\alpha''}) - \alpha \leqslant 1/n\alpha''$ and

$$\begin{aligned}
\sqrt{n}\left\{\alpha - F^{\Delta}\left(\underline{\widehat{u}} \mid A^{(2)}_{\alpha''}\right)\right\} &= \sqrt{n}\left\{\widehat{F}^{\Delta}\left(\underline{\widehat{u}} \mid \widehat{A}^{(2)}_{n,\alpha''}\right) - F^{\Delta}\left(\underline{\widehat{u}} \mid \widehat{A}^{(2)}_{n,\alpha''}\right)\right\}\\
&+ \sqrt{n}\left\{F^{\Delta}\left(\underline{\widehat{u}} \mid \widehat{A}^{(2)}_{n,\alpha''}\right) - F^{\Delta}\left(\underline{\widehat{u}} \mid A^{(2)}_{\alpha''}\right)\right\} + o_{\mathbb{P}}(1). \qquad (47)
\end{aligned}$$

With a slight abuse of notation, denote by $f_1(\cdot \mid A)$ the density of η_{1t} conditional on $\eta_{2t} \in A$ for any measurable set A. Let

$$\begin{aligned}
\lambda_x =&\lambda_x(\alpha, \alpha'') = \frac{\partial}{\partial x}F^{\Delta}(x \mid (y_1, y_2])\bigg|_{(x,y_1,y_2)=\left(\underline{u},\xi^{(2)}_{0.5-\alpha''},\xi^{(2)}_{\alpha''+0.5}\right)} = f_1\left(\underline{u} \mid A^{(2)}_{\alpha''}\right),\\
\lambda_{y_1} =& \frac{\partial F^{\Delta}(x \mid (y_1, y_2])}{\partial y_1}\bigg|_{(x,y_1,y_2)=\left(\underline{u},\xi^{(2)}_{0.5-\alpha''},\xi^{(2)}_{\alpha''+0.5}\right)} = \frac{-f_2(\xi^{(2)}_{0.5-\alpha''} \mid \underline{u})G^{(1)}(\underline{u}) + g^{(2)}(\xi^{(2)}_{0.5-\alpha''})\alpha}{2\alpha''},\\
\lambda_{y_2} =& \frac{\partial F^{\Delta}(x \mid (y_1, y_2])}{\partial y_2}\bigg|_{(x,y_1,y_2)=\left(\underline{u},\xi^{(2)}_{0.5-\alpha''},\xi^{(2)}_{\alpha''+0.5}\right)} = \frac{f_2(\xi^{(2)}_{0.5+\alpha''} \mid \underline{u})G^{(1)}(\underline{u}) - g^{(2)}(\xi^{(2)}_{0.5+\alpha''})\alpha}{2\alpha''}.
\end{aligned}$$

By arguments already given, we thus have

$$\begin{aligned}
&\sqrt{n}\left(F^{\Delta}(\underline{\widehat{u}} \mid \widehat{A}^{(2)}_{n,\alpha''}) - F^{\Delta}(\underline{u}(\alpha, \alpha'') \mid A^{(2)}_{\alpha''})\right) = \lambda_x\sqrt{n}\left\{\underline{\widehat{u}} - \underline{u}(\alpha, \alpha'')\right\}\\
&+ \lambda_{y_1}\sqrt{n}\left(\widehat{\xi}^{(2)}_{0.5-\alpha''} - \xi^{(2)}_{0.5-\alpha''}\right) + \lambda_{y_2}\sqrt{n}\left(\widehat{\xi}^{(2)}_{\alpha''+0.5} - \xi^{(2)}_{\alpha''+0.5}\right) + o_{\mathbb{P}}(1)
\end{aligned}$$

and

$$\sqrt{n}\left(F^{\Delta}\left(\underline{u}(\alpha, \alpha'') \mid A^{(2)}_{\alpha''}\right) - F^{\Delta}\left(\underline{\widehat{u}} \mid A^{(2)}_{\alpha''}\right)\right) =\lambda_x\sqrt{n}\left\{\underline{u}(\alpha, \alpha'') - \underline{\widehat{u}}\right\} + o_{\mathbb{P}}(1),$$

thus $\sqrt{n}\left(F^{\Delta}(\underline{\widehat{u}} \mid \widehat{A}^{(2)}_{n,\alpha''}) - F^{\Delta}(\underline{\widehat{u}} \mid A^{(2)}_{\alpha''})\right) = \lambda_{y_1}\sqrt{n}\left(\widehat{\xi}^{(2)}_{0.5-\alpha''} - \xi^{(2)}_{0.5-\alpha''}\right) + \lambda_{y_2}\sqrt{n}\left(\widehat{\xi}^{(2)}_{\alpha''+0.5} - \xi^{(2)}_{\alpha''+0.5}\right) + o_{\mathbb{P}}(1)$.

Noting that

$$\frac{(0.5 + \alpha'')f_1(\underline{u} \mid \xi^{(2)}_{0.5+\alpha''}) - (0.5 - \alpha'')f_1(\underline{u} \mid \xi^{(2)}_{0.5-\alpha''})}{2\alpha''} = \Delta f_1(\underline{u} \mid A^{(2)}_{\alpha''}),$$

and that $a\left(\underline{u}, \xi^{(2)}_{0.5-\alpha''}, \xi^{(2)}_{\alpha''+0.5}\right) - \lambda_{y_1}\xi^{(2)}_{0.5-\alpha''} - \lambda_{y_2}\xi^{(2)}_{\alpha''+0.5} = 0$, we deduce, using Corollary 4.2 in FZ, the Bahadur expansion

$$\sqrt{n}\left\{\alpha - F^{\Delta}\left(\underline{\widehat{u}} \mid A^{(2)}_{\alpha''}\right)\right\} = \frac{1}{2\sqrt{n}\alpha''}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}\leqslant \underline{u}} - \alpha\}\mathbb{1}_{\eta_{2t}\in A^{(2)}_{\alpha''}}$$

$$
\begin{aligned}
&+\frac{\underline{u}f_1(\underline{u}\mid A^{(2)}_{\alpha''})}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^n(\eta_{1t}^2-1)\boldsymbol{D}_{1t}-\frac{\lambda_{y_1}}{g^{(2)}(\xi^{(2)}_{0.5-\alpha''})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{0.5-\alpha''}}-\alpha''+0.5)\\
&-\frac{\lambda_{y_2}}{g^{(2)}(\xi^{(2)}_{\alpha''+0.5})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha''+0.5}}-0.5-\alpha'')+o_{\mathbb{P}}(1)\\
=&\frac{1}{2\sqrt{n}\alpha''}\sum_{t=1}^n\{\mathbb{1}_{\eta_{1t}\leqslant\underline{u}}-\alpha\}\mathbb{1}_{\eta_{2t}\in A^{(2)}_{\alpha''}}+\frac{\underline{u}f_1(\underline{u}\mid A^{(2)}_{\alpha''})}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^n(\eta_{1t}^2-1)\boldsymbol{D}_{1t}\\
&+\frac{\alpha}{2\alpha''}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\in A^{(2)}_{\alpha''}}-2\alpha'')+\frac{f_2(\xi^{(2)}_{0.5-\alpha''}\mid\underline{u})G^{(1)}(\underline{u})}{2\alpha''g^{(2)}(\xi^{(2)}_{0.5-\alpha''})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{0.5-\alpha''}}-\alpha''+0.5)\\
&-\frac{f_2(\xi^{(2)}_{0.5+\alpha''}\mid\underline{u})G^{(1)}(\underline{u})}{2\alpha''g^{(2)}(\xi^{(2)}_{\alpha''+0.5})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha''+0.5}}-0.5-\alpha'')+o_{\mathbb{P}}(1)\\
=&\frac{1}{2\sqrt{n}\alpha''}\sum_{t=1}^n\{\mathbb{1}_{\eta_{1t}\leqslant\underline{u}}\mathbb{1}_{\eta_{2t}\in A^{(2)}_{\alpha''}}-2\alpha\alpha''\}+\frac{\underline{u}f_1(\underline{u}\mid A^{(2)}_{\alpha''})}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^n(\eta_{1t}^2-1)\boldsymbol{D}_{1t}\\
&+\frac{f_2(\xi^{(2)}_{0.5-\alpha''}\mid\underline{u})G^{(1)}(\underline{u})}{2\alpha''g^{(2)}(\xi^{(2)}_{0.5-\alpha''})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{0.5-\alpha''}}-\alpha''+0.5)\\
&-\frac{f_2(\xi^{(2)}_{0.5+\alpha''}\mid\underline{u})G^{(1)}(\underline{u})}{2\alpha''g^{(2)}(\xi^{(2)}_{\alpha''+0.5})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha''+0.5}}-0.5-\alpha'')+o_{\mathbb{P}}(1).
\end{aligned}
$$

Similarly to the proof of Theorem 3.1, we conclude by applying the delta method to the latter expansion, using the inverse of the function $F^{\Delta}\left(\cdot\mid A^{(2)}_{\alpha''}\right)$. We have

$$
\begin{aligned}
&\sqrt{n}\,(\hat{\underline{u}}-\underline{u})\\
=&-\frac{1}{2\sqrt{n}\alpha''f_1(\underline{u}\mid A^{(2)}_{\alpha''})}\sum_{t=1}^n\{\mathbb{1}_{\eta_{1t}\leqslant\underline{u}}\mathbb{1}_{\eta_{2t}\in A^{(2)}_{\alpha''}}-2\alpha\alpha''\}-\frac{\underline{u}}{2\sqrt{n}}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\sum_{t=1}^n(\eta_{1t}^2-1)\boldsymbol{D}_{1t}\\
&-\frac{f_2(\xi^{(2)}_{0.5-\alpha''}\mid\underline{u})G^{(1)}(\underline{u})}{2\alpha''g^{(2)}(\xi^{(2)}_{0.5-\alpha''})f_1(\underline{u}\mid A^{(2)}_{\alpha''})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{0.5-\alpha''}}-\alpha''+0.5)\\
&+\frac{f_2(\xi^{(2)}_{0.5+\alpha''}\mid\underline{u})G^{(1)}(\underline{u})}{2\alpha''g^{(2)}(\xi^{(2)}_{\alpha''+0.5})f_1(\underline{u}\mid A^{(2)}_{\alpha''})}\frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{\alpha''+0.5}}-0.5-\alpha'')+o_{\mathbb{P}}(1).
\end{aligned}
$$

The asymptotic distribution follows, noting that $\Delta f_1(\underline{u},\alpha'')=2\alpha''f_1(\underline{u}\mid A^{(2)}_{\alpha''})$ and that $\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}$ is the covariance matrix of the vector $\boldsymbol{\Psi}_t=\left(\eta_{1t}^2,\ \mathbb{1}_{\eta_{1t}\leqslant\underline{u},\,\eta_{2t}\in A^{(2)}_{\alpha''}},\ \mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{0.5+\alpha''}},\ \mathbb{1}_{\eta_{2t}\leqslant\xi^{(2)}_{0.5-\alpha''}}\right)'$. □

We have $\underline{u}(\alpha,1/2)=\xi^{(1)}_{\alpha}$ and $\hat{\underline{u}}(\alpha,1/2)=\hat{\xi}^{(1)}_{\alpha}$, so it can be expected that the asymptotic variance $\underline{\sigma}^2(\alpha,1/2)$ coincides with the asymptotic variance in (5). This is indeed true noting that, when $\alpha''=1/2$, $\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}$ has the bloc-diagonal form

$$
\boldsymbol{\Sigma}_{\boldsymbol{\Psi}}=\begin{pmatrix}\boldsymbol{\Sigma}^{11}_{\boldsymbol{\Psi}}&\mathbf{0}\\\mathbf{0}&\mathbf{0}\end{pmatrix}\quad\text{where}\quad\boldsymbol{\Sigma}^{11}_{\boldsymbol{\Psi}}=\begin{pmatrix}\kappa_1-1&\varrho(\xi^{(1)}_{\alpha})\\\varrho(\xi^{(1)}_{\alpha})&\alpha(1-\alpha)\end{pmatrix}.
$$

The joint asymptotic distribution of the estimator of $\boldsymbol{\theta}^{(1)}_0$ and the difference appearing in (3) is straightforwardly deduced from Theorems 3.1 and G.1.

Corollary G.1. *Under the assumptions of Theorem 3.1, including* $\mathbf{B7_1}$*, we have*

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}^{(1)}-\boldsymbol{\theta}_0^{(1)} \\ \widehat{u}-\widehat{\underline{u}}-u+\underline{u} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{\mathbf{0},\underline{\boldsymbol{\Sigma}}(\alpha,\alpha',\alpha''):=\begin{pmatrix} \frac{\kappa_1-1}{4}\boldsymbol{J}_1^{-1} & \frac{\boldsymbol{J}_1^{-1}}{2}\boldsymbol{\Omega}_1\boldsymbol{\mu}'\boldsymbol{\Sigma_\Phi}\boldsymbol{e}_3 \\ \frac{1}{2}\boldsymbol{e}_3'\boldsymbol{\Sigma_\Phi}\boldsymbol{\mu}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1} & \boldsymbol{\mu}'\boldsymbol{\Sigma_\Phi}\boldsymbol{\mu} \end{pmatrix}\right\}.$$

where $\boldsymbol{\Sigma_\Phi}$ *is the variance of the vector* $\boldsymbol{\Phi}_t=\left(\mathbb{1}_{\eta_{1t}\leqslant u,\,\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}},\,\mathbb{1}_{\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}},\boldsymbol{\Psi}_t'\right)'$, $\boldsymbol{e}_3=(0,0,1,0,0,0)'$ *and* $\boldsymbol{\mu}=(\lambda_1,\lambda_2,\lambda_3-\nu_1,-\nu_2,-\nu_3,-\nu_4)'$.

H Proofs for Section 5

Proof of Proposition 5.1. We start by deriving the joint asymptotic distribution of $(\widehat{\xi}_\alpha^{(1)},\widehat{u}(\alpha,\alpha'))$. Using Theorem 3.1 with $\alpha'=1$, we get

$$\sqrt{n}(\widehat{\xi}_\alpha^{(1)}-\xi_\alpha^{(1)})=\frac{-1}{g^{(1)}(\xi_\alpha^{(1)})\sqrt{n}}\sum_{t=1}^n\{\mathbb{1}_{\eta_{1t}\leqslant\xi_\alpha^{(1)}}-\alpha\}-\frac{\xi_\alpha^{(1)}}{2\sqrt{n}}\sum_{t=1}^n(\eta_{1t}^2-1)+o_{\mathbb{P}}(1).$$

Let $\boldsymbol{\Sigma_Z}$ denote the covariance matrix of the vector $\boldsymbol{Z}_t=\left(\mathbb{1}_{\eta_{1t}\leqslant u(\alpha,\alpha'),\,\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}},\,\mathbb{1}_{\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}},\,\eta_{1t}^2,\,\mathbb{1}_{\eta_{1t}\leqslant\xi_\alpha^{(1)}}\right)'$. In view of Theorem 3.1 we thus have

$$\sqrt{n}\begin{pmatrix} \widehat{u}(\alpha,\alpha')-u(\alpha,\alpha') \\ \widehat{\xi}_\alpha^{(1)}-\xi_\alpha^{(1)} \end{pmatrix}\xrightarrow{\mathcal{L}}\mathcal{N}\left\{\mathbf{0},\boldsymbol{A}'\boldsymbol{\Sigma_Z}\boldsymbol{A}\right\},\qquad \boldsymbol{A}'=\begin{pmatrix} & \boldsymbol{\lambda}' & & 0 \\ 0 & 0 & \frac{-\xi_\alpha^{(1)}}{2} & \frac{-1}{g^{(1)}(\xi_\alpha^{(1)})}\end{pmatrix}.$$

The critical region is obtained at the boundary of the null hypothesis, noting that if $u(\alpha,\alpha')=\xi_\alpha^{(1)}$, we have $\mathrm{Cov}\left(\mathbb{1}_{\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}},\,\mathbb{1}_{\eta_{1t}\leqslant\xi_\alpha^{(1)}}\right)=0$ and thus $\sqrt{n}\{\widehat{u}(\alpha,\alpha')-\widehat{\xi}_\alpha^{(1)}\}=\daleth'\{\beth_t-E(\beth_t)\}+o_{\mathbb{P}}(1)$, where $\beth_t=\left(\mathbb{1}_{\eta_{1t}\leqslant\xi_\alpha^{(1)},\,\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}},\,\mathbb{1}_{\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}},\,\mathbb{1}_{\eta_{1t}\leqslant\xi_\alpha^{(1)}}\right)'$ and

$$\daleth=\left(\frac{-1}{\alpha' f_1(\xi_\alpha^{(1)}|\xi_{\alpha'}^{(2)})},\frac{\alpha}{\alpha' f_1(\xi_\alpha^{(1)}|\xi_{\alpha'}^{(2)})}\frac{f_2(\xi_{\alpha'}^{(2)}|\xi_\alpha^{(1)})}{g^{(2)}(\xi_{\alpha'}^{(2)})},\frac{1}{g^{(1)}(\xi_\alpha^{(1)})}\right)'.$$

Define the covariance matrix $\boldsymbol{\Sigma}_\beth$ of the vector $\beth_t$. We have $s_{\alpha,\alpha'}=(\daleth'\boldsymbol{\Sigma}_\beth\daleth)^{1/2}$ with

$$\boldsymbol{\Sigma}_\beth=\begin{pmatrix} \alpha\alpha'(1-\alpha\alpha') & \alpha\alpha'(1-\alpha') & \alpha\alpha'(1-\alpha) \\ \alpha\alpha'(1-\alpha') & (1-\alpha')\alpha' & 0 \\ \alpha\alpha'(1-\alpha) & 0 & \alpha(1-\alpha)\end{pmatrix}$$

and the conclusion follows. Note that when the two components of $\boldsymbol{\eta}_t$ are independent, the asymptotic variance reduces to $s_{\alpha,\alpha'}^2=\frac{\alpha(1-\alpha)(1-\alpha')}{\alpha'[g^{(1)}(\xi_\alpha^{(1)})]^2}$. The behavior of the test under H_0^{sys} follows. Under H_1^{sys}, as $n\to\infty$ we have

$$\begin{aligned}&\mathbb{P}\left\{\sqrt{n}\left(\widehat{u}(\alpha,\alpha')-\widehat{\xi}_\alpha^{(1)}\right)<\hat{s}_{\alpha,\alpha'}\Phi^{-1}(1-\alpha_0)\right\}\\ =&\mathbb{P}\left\{\sqrt{n}\left(\widehat{u}(\alpha,\alpha')-u(\alpha,\alpha')-\widehat{\xi}_\alpha^{(1)}+\xi_\alpha^{(1)}\right)<\hat{s}_{\alpha,\alpha'}\Phi^{-1}(1-\alpha_0)-\sqrt{n}\left(u(\alpha,\alpha')-\xi_\alpha^{(1)}\right)\right\}\to 1.\end{aligned}$$

□

Proof of Proposition 5.2. First note that, under Proposition 2.3, H_0 holds with

$$\boldsymbol{\theta}_0=\left(\boldsymbol{\theta}_0^{(1)'},\boldsymbol{\theta}_0^{(2)'}\right)',\ \boldsymbol{u}_0=\left(u(\alpha,\alpha'),\xi_{\alpha'}^{(2)}\right)',\ h_t=\mathbb{1}_{\eta_{1t}\leqslant u(\alpha,\alpha'),\,\eta_{2t}\leqslant\xi_{\alpha'}^{(2)}}.$$

As $R \to \infty$, $\widehat{h}_t \to h_t$ for all $t = R+1, \ldots, n$ and $\sqrt{P} K_P \sim \mathcal{B}(P, \alpha\alpha') - P\alpha\alpha'$. The result concerning the test (8) is obtained by noting that

$$\sum_{i: \left|\frac{i-n\alpha}{\sqrt{n}}\right| > c} \binom{n}{i} \alpha^i (1-\alpha)^{n-i} = \mathbb{P}\left(\frac{|\mathcal{B}(n,\alpha) - n\alpha|}{\sqrt{n}} > c \right).$$

The result concerning the tests in (9) follows from Corrolary 1 and Theorem 2 of Escanciano and Olmo (2010) in the fixed scheme case. □

Proof of Proposition 5.3. It suffices to note that $P' \sim \mathcal{B}(P, \alpha')$ when $\mathrm{VaR}_t^{\epsilon_2}(\alpha')$ is correctly specified. □

References

Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18, 221-254.