Preprint

Estimating conditional systemic risk measures in semi-parametric volatility models

LOÏC CANTIN[1], CHRISTIAN FRANCQ[1] and JEAN-MICHEL ZAKOÏAN[1]

[1] *CREST, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France*
E-mail: loic.cantin@polytechnique.edu *E-mail:* christian.francq@ensae.fr
E-mail: zakoian@ensae.fr

We propose a two-step semi-parametric estimation approach for dynamic Conditional VaR (CoVaR), from which other important systemic risk measures such as the Delta-CoVaR can be derived. The CoVaR allows to define reserves for a given financial entity, in order to limit exceeding losses when a system is in distress. We assume that all financial returns in the system follow semi-parametric GARCH-type models. Our estimation method relies on the fact that the dynamic CoVaR is the product of the volatility of the financial entity's return and a conditional quantile term involving the innovations of the different returns. We show that the latter quantity can be easily estimated from residuals of the GARCH-type models estimated by Quasi-Maximum Likelihood (QML). The study of the asymptotic behaviour of the corresponding estimator and the derivation of asymptotic confidence intervals for the dymanic CoVaR are the main purposes of the paper. Our theoretical results are illustrated via Monte-Carlo experiments and real financial time series.

Keywords: conditional CoVaR and Delta-Covar, empirical distribution of bivariate residuals, model-free estimation risk, multivariate risks.

1. Introduction

The study of risk has long been neglected by practitioners in the financial world, who prefer the search for profit through the maximization of returns. The topic only 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. However, it has long been known in the investment world that ”there's no such thing as a free lunch”, translating the idea that any return is in fact only the remuneration of a proportional risk from which it is not possible to escape. The idea of seeking to optimize the risk/return ratio is not new and has been embodied in the extensive use of the Sharpe ratio in the financial industry, introduced by Sharpe (see Sharpe, 1994). However, the massive losses experienced during the successive crises have refocused attention on risks whose occurrence is rarer but whose consequences are more severe. These risks are called ”tail risks”. This preoccupation with measuring tail risks is common to various practitioners in the financial world, from portfolio managers to company executives in the banking or insurance sector, and even to regulators. Therefore, much of the econometric tools used to measure these risks are similar. The best-known measure is the VaR (Value at Risk) which has gradually gained popularity since the 90s. This period saw the first attempts to standardize the measurement of extreme risks faced by financial institutions. The so-called "financial" risk characterizes a loss of value (of a security, a portfolio, a currency, etc.) and can be decomposed - for modeling purposes - into two parts: i) an "innovation" which is an event that affects the value positively or negatively and that cannot be predicted (this quantity is therefore considered random) and ii) the so-called "conditional" volatility which can be estimated by statistical and econometric models. All risk forecasting is based on the choice of a suitable volatility model (a GARCH-type model, for example) and whether or not to make parametric assumptions about the distribution of innovations.

This univariate approach to risk - which only focuses on the asset under consideration and which omits its relationship with other 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. There are mainly two approaches to study these risks: i) the network approach summarized - as an example - in an OECD article by Poledna (2020) and ii) the econometric approach of co-movement analysis - which we will adopt - via (conditional) systemic risk measures such as CoVaR in Adrian and Brunnermeier (2011), Delta-CoVaR in Adrian and Brunnermeier (2016), SES in Acharya et al. (2017), SRISK in Brownlees and Engle (2017). See also Benoit et al. (2019), Banulescu et al. (2020). Just as with VaR, it is necessary to choose a model to estimate conditional volatility and to deal with the randomness of the innovations. However, until now, most of the literature has used fully parametric models, assuming that the distribution of innovations is known. However, the choice of a distribution that does not fit the reality of the data can have dramatic consequences in practice. The popular and well-known article by journalist Felix Salmon "The Formula that Killed Wall Street" in Salmon (2012) holds the use of Gaussian copula as an important cause of the crisis mentioned above. Indeed, the use of Gaussian distributions underestimated tail-risks and encouraged banks to under-provision their capital reserves required to face periods of stress. It is in particular to overcome this shortcoming that we propose a semi-parametric approach to estimate conditional systemic risk measures.

In this article we focus on the estimation of the *dynamic CoVaR* and $\Delta CoVaR$ between several assets or financial entities. CoVaR stands for *Conditional VaR* and the "conditional" here refers to the fact that we are interested in the VaR of a series under the condition that the loss of one or several other series exceed their VaRs. Because we consider dynamic VaRs (i.e. conditional on past returns) we refer to dynamic CoVaR. The ΔCoVaR is defined as the differrence between CoVars, one computed when the variables in the conditionning set are "in distress" and the other one when such variables are in a "median" state. We propose a novel approach for estimating such dynamic CoVaRs and ΔCoVaRs in a semi-parametric framework, allowing us to obtain asymptotic results using QML (Quasi-Maximum Likelihood) estimation.

The rest of the paper is organized as follows. Starting by considering the case of two assets, Section 2 introduces the semi-parametric CoVaR and ΔCoVaR estimators. Its consistency is studied in Section 3 and its asymptotic distribution is derived in Section 4. In Section 6 we extend the setting to handle multi-assets conditioning events. Numerical illustrations are proposed in Section 7. Proofs and complementary results are displayed in an Appendix. Section 8 concludes.

2. Semi-parametric CoVaR and ΔCoVaR estimators

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

$$\mathrm{VaR}_t^X(\alpha) = -\inf\{z_t : P_{t-1}[X_t \leq z_t] \geq \alpha\}$$

where P_{t-1} denotes the historical distribution conditional on $\{X_u, u < t\}$.

When (ϵ_t) is a non-anticipative solution of a GARCH-type model of the form $\epsilon_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \boldsymbol{\theta}_0)\eta_t$ where (η_t) is an iid(0,1) process, the conditional VaR at level α is given by

$$\mathrm{VaR}_t^{\epsilon}(\alpha) = -\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \boldsymbol{\theta}_0)\xi_\alpha,$$

where $\xi_\alpha = \inf\{x : F(x) \geq \alpha\}$ is the α-quantile of the cumulative distribution function (cdf) F of η_t.

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 $\mathrm{CoVaR}_t^{X|Y}(\alpha, \alpha')$, can be defined as

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

The *conditional* 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

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

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

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

where $A_t(\alpha'') = (-\mathrm{VaR}_t^Y(50\% - \alpha''), -\mathrm{VaR}_t^Y(50\% + \alpha'')]$.

Now let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{mt})'$ a vector of non-anticipative solutions of a GARCH-type model. Assume the first two components satisfy

$$\epsilon_{it} = \sigma_i(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots; \boldsymbol{\theta}_0^{(i)})\eta_{it} := \sigma_{it}\eta_{it}, \tag{2.1}$$

where (η_{1t}, η_{2t}) is an iid process with $E\eta_{it}^2 = 1, i = 1, 2$, and $\boldsymbol{\theta}_0^{(i)}, i = 1, 2$, are vectors of unknown parameters which belong to compact parameter sets $\boldsymbol{\Theta}^{(i)} \subset \mathbb{R}^{d_i}$ for some positive integers d_i. Let $d = d_1 + d_2$. Note that the volatility of each component may depend on the past of both components of $\boldsymbol{\epsilon}_t$. Assume that the variables η_{it} are independent from $\{\boldsymbol{\epsilon}_{t-u}, u > 0\}$.

Our first result shows that the conditional CoVaR and ΔCoVaR of ϵ_{1t} relative to ϵ_{2t} are proportional to the volatility of the first component. Let

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

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

***Proposition* 2.1.** The conditional CoVaR and ΔCoVaR at levels $\alpha, \alpha', \alpha''$ of the first relative to the second component are given by

$$\mathrm{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha') = -\sigma_{1t}u(\alpha,\alpha'), \quad \Delta\mathrm{CoVaR}_t^{\epsilon_1|\epsilon_2}(\alpha,\alpha',\alpha'') = -\sigma_{1t}\{u(\alpha,\alpha') - \underline{u}(\alpha,\alpha'')\},$$

where

$$u(\alpha,\alpha') = \inf\left\{x : F\left(x \mid \xi_{\alpha'}^{(2)}\right) \ge \alpha\right\}, \quad \underline{u}(\alpha,\alpha'') = \inf\left\{x : F^{\Delta}\left(x \mid A_{\alpha''}^{(2)}\right) \ge \alpha\right\},$$

$A_{\alpha''}^{(2)} = \left(\xi_{0.5-\alpha''}^{(2)}, \xi_{\alpha''+0.5}^{(2)}\right]$ and $\xi_{\alpha}^{(2)}$ is the α-quantile of η_{2t}.

The multiplicative forms of the conditional CoVaR and ΔCoVaR are crucial for our study. They enable to decompose the risk as a product of a fixed characteristic of the joint distribution of the innovations and the time-varying volatility of the first component, therefore suggesting a two-step estimation method. As in Amengual et al. (2013) we will propose a sequential estimation method but, contrary to this reference, the second step will be nonparametric.

Given observations $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$, and using arbitrary initial values $\tilde{\boldsymbol{\epsilon}}_j$ for $j \le 0$, we define for any $\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}$,

$$\tilde{\sigma}_{it}(\boldsymbol{\theta}^{(i)}) = \sigma_i(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \boldsymbol{\epsilon}_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \boldsymbol{\theta}^{(i)}),$$

which will be used as a proxy of $\sigma_{it}(\boldsymbol{\theta}^{(i)}) = \sigma_i(\boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \dots, \boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_{-1}, \dots; \boldsymbol{\theta}^{(i)})$. The parameters $\boldsymbol{\theta}_0^{(i)}$ in (2.1) can be estimated equation-by-equation using the Gaussian QML approach (see Francq and Zakoïan (2016)):

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

Note that this approach does not require full specification of the dynamics of $\boldsymbol{\epsilon}_t$, since we made no assumption concerning the joint distribution of (η_{1t}, η_{2t}).

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

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

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

where $\widehat{A}^{(2)}_{n,\alpha''} = \left(\widehat{\xi}^{(2)}_{n,0.5-\alpha''}, \widehat{\xi}^{(2)}_{n,\alpha''+0.5}\right]$, $\rho_\alpha(z) = z(\alpha - \mathbb{1}_{z<0})$ is the usual "check" function and $\widehat{\xi}^{(2)}_{n,\alpha'}$ 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 $\mathrm{CoVaR}^{\epsilon_1|\epsilon_2}_t(\alpha, \alpha')$ and $\Delta\mathrm{CoVaR}^{\epsilon_1|\epsilon_2}_t(\alpha, \alpha', \alpha'')$ are thus

$$\widehat{\mathrm{CoVaR}}^{\epsilon_1|\epsilon_2}_t(\alpha, \alpha') = -\widetilde{\sigma}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)}_n)\widehat{u}_n(\alpha, \alpha'),$$
$$\Delta\widehat{\mathrm{CoVaR}}^{\epsilon_1|\epsilon_2}_t(\alpha, \alpha', \alpha'') = -\widetilde{\sigma}_{1t}(\widehat{\boldsymbol{\theta}}^{(1)}_n)\{\widehat{u}_n(\alpha, \alpha') - \underline{\widehat{u}}_n(\alpha, \alpha'')\}$$

with obvious notations. To our knowledge, no asymptotic statistical results have been established so far for any estimator of the dynamic CoVaR and ΔCoVaR in a semi-parametric framework.

Obtaining the joint asymptotic distribution of $\widehat{u}_n(\alpha, \alpha')$ and $\underline{\widehat{u}}_n(\alpha, \alpha'')$ is a complex task. Our strategy is to first derive the asymptotic distribution of an empirical cdf and then use the delta method to get the asymptotic distribution of the inverse. The same approach was developped in Francq and Zakoïan (2022) (hereafter FZ) for the empirical distribution of the residuals. However, the introduction of conditioning variables in the present paper induces additional technical difficulties.

In the next section, we start by showing the consistency of an estimator of $F(x|y)$ defined by

$$\widehat{F}_n(x|y) = \frac{\frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{1t} \le x, \widehat{\eta}_{2t} \le y}}{\frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{2t} \le y}} := \frac{\widehat{H}_n(x, y)}{\widehat{G}^{(2)}_n(y)},$$

provided the denominator is not equal to zero.

3. Consistency

We start by showing the uniform consistency of $\widehat{F}_n(\cdot|y)$. Let $K > 0$ be a generic constant or random variable measurable with respect to $\mathcal{F}_0$, where $\mathcal{F}_t$ denotes the σ-algebra generated by $\{\eta_{is}, s \le t, i = 1, 2\}$. Let $\rho \in (0, 1)$. The next first 3 conditions ensure the strong consistency of the QML estimator of $\boldsymbol{\theta}^{(i)}_0$, for $i = 1, 2$, while the 4th one will be used to control the difference between the innovations and residuals.

A1$_\mathbf{i}$: $(\boldsymbol{\epsilon}_t)$ is a strictly stationary and ergodic process. The variables η_{it} are independent from $\{\boldsymbol{\epsilon}_{t-u}, u > 0\}$. Moreover, $E|\sigma_{it}|^r < \infty$ for some $r > 0$.

A2$_\mathbf{i}$: For any real sequence (x_j), the function $\boldsymbol{\theta}^{(i)} \mapsto \sigma(x_1, x_2, \dots; \boldsymbol{\theta}^{(i)})$ is continuously differentiable. 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, $\sigma_{it}(\boldsymbol{\theta}^{(i)}_0)/\sigma_{it}(\boldsymbol{\theta}^{(i)}) = 1$ *a.s.* iff $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i)}_0$.

A3$_\mathbf{i}$: $\sup_{\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}^{(i)}} |\sigma_{it}(\boldsymbol{\theta}^{(i)}) - \widetilde{\sigma}_{it}(\boldsymbol{\theta}^{(i)})| \le K\rho^t$.

A4$_\mathbf{i}$: For $r > 0$, there exists a neighborhood $V(\boldsymbol{\theta}_0^{(i)})$ of $\boldsymbol{\theta}_0^{(i)}$ such that $E\left(\sup_{\boldsymbol{\theta}^{(i)}\in V(\boldsymbol{\theta}_0^{(i)})} \frac{\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}{\sigma_{it}(\boldsymbol{\theta}^{(i)})}\right)^r < \infty$ and $E\sup_{\boldsymbol{\theta}^{(i)}\in V(\boldsymbol{\theta}_0^{(i)})} \left\|\boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)})\right\|^r < \infty$, where $\boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)}) = \sigma_{it}^{-1}(\boldsymbol{\theta}^{(i)})\partial\sigma_{it}(\boldsymbol{\theta}^{(i)})/\partial\boldsymbol{\theta}^{(i)}$.

When an assumption holds for $i = 1, 2$ we omit the index for ease of notation. We also introduce the following assumption, for $i = 1, 2$, which is simply denoted **A5$_\mathbf{i}$** when it holds for all $x \in \mathbb{R}$.

A5$_\mathbf{i}$ (x): For $x \in \mathbb{R}$, the cdf $G^{(i)}$ of η_{it} is Lipschitz continuous in a neighborhood of x.

Theorem 3.1. *Under* **A1**-**A4**,

i) If **A5$_\mathbf{1}$**(x) *and* **A5$_\mathbf{2}$**(y) *hold for* $x, y \in \mathbb{R}$, *and if* $G^{(2)}(y) > 0$, *we have* $|\widehat{F}_n(x|y) - F(x|y)| \to 0$ *a.s.*

ii) If **A5$_\mathbf{1}$** *and* **A5$_\mathbf{2}$**(y) *hold for* $y \in \mathbb{R}$ *such that* $G^{(2)}(y) > 0$, *we have* $\sup_{x\in\mathbb{R}} |\widehat{F}_n(x|y) - F(x|y)| \to 0$ *a.s.*

iii) If **A5$_\mathbf{1}$** *and* **A5$_\mathbf{2}$**(y_0) *hold for* $y_0 \in \mathbb{R}$ *such that* $G^{(2)}(y_0) > 0$, *for any small enough neighborhood* $V(y_0)$ *of* y_0 *we have* $\sup_{x\in\mathbb{R}, y\in V(y_0)} |\widehat{F}_n(x|y) - F(x|y)| \to 0$ *a.s.*

Next, we turn to the consistency of the empirical conditional quantile $\widehat{u}_n(\alpha, \alpha')$. Recall that $u(\alpha, \alpha') = \inf\{x : F(x \mid \xi_{\alpha'}^{(2)}) \geq \alpha\}$. Let $u^+(\alpha, \alpha') = \inf\{x : F(x \mid \xi_{\alpha'}^{(2)}) > \alpha\}$. In the following assumptions (which will be simply denoted **A6**(α, α') when both of them hold), we impose that the quantile function of η_{2t} and the conditional quantile function $F^-(\cdot \mid \xi_{\alpha'}^{(2)})$ of η_{1t} be right-continuous at α' and α, respectively.

A6$_\mathbf{1}$ (α'): The cdf of η_{2t} satisfies: $G^{(2)}(y) > \alpha' \quad \text{whenever} \quad y > \xi_{\alpha'}^{(2)}$.

A6$_\mathbf{2}$ (α, α'): For $\alpha, \alpha' \in (0, 1)$ such that $G^{(2)}(\xi_{\alpha'}^{(2)}) > 0$, the conditional cdf of η_{1t} satisfies: $F(x \mid \xi_{\alpha'}^{(2)}) > \alpha \quad \text{whenever} \quad x > u(\alpha, \alpha')$.

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

A7(y_0): The conditional cdf of η_1 satisfies: $\sup_{x\in\mathbb{R}} |F(x|y) - F(x|y_0)| \to 0$ when $y \to y_0$.

Theorem 3.2. *Under* **A1**-**A4**, **A5$_\mathbf{1}$**, **A5$_\mathbf{2}$**$(\xi_{\alpha'}^{(2)})$, **A6**(α, α'), **A**7$(\xi_{\alpha'}^{(2)})$, *if* $G^{(2)}(\xi_{\alpha'}^{(2)}) > 0$ *we have the strong convergence*

$$\widehat{u}_n(\alpha, \alpha') \to u(\alpha, \alpha') \quad a.s.$$

Without Assumption **A6$_\mathbf{2}$**(α, α'), *we have*

$$[\liminf \widehat{u}_n(\alpha, \alpha'), \limsup \widehat{u}_n(\alpha, \alpha')] \subseteq [u(\alpha, \alpha'), u^+(\alpha, \alpha')] \quad a.s. \tag{3.1}$$

Now we turn to the consistency of $\underline{\widehat{u}}_n(\alpha, \alpha'')$. Recalling that $\underline{u}(\alpha, \alpha'') = \inf\{x : F^{\Delta}(x \mid [\xi_{0.5-\alpha''}^{(2)}, \xi_{\alpha''+0.5}^{(2)})) \geq \alpha\}$, let $\underline{u}^+(\alpha, \alpha'') = \inf\{x : F^{\Delta}(x \mid [\xi_{0.5-\alpha''}^{(2)}, \xi_{\alpha''+0.5}^{(2)})) > \alpha\}$.

Theorem 3.3. *Under* **A1**-**A4**, $\mathbf{A5_1}$ *and, for* $\tau \in \{-1,1\}$, $\mathbf{A5_2}(\xi^{(2)}_{0.5+\tau\alpha''})$, $\mathbf{A6}(\alpha, 0.5+\tau\alpha'')$, $\mathbf{A7}(\xi^{(2)}_{0.5+\tau\alpha''})$, *if* $G^{(2)}(\xi^{(2)}_{0.5+\tau\alpha''}) > G^{(2)}(\xi^{(2)}_{0.5-\tau\alpha''})$ *we have the strong convergence*

$$\underline{\widehat{u}}_n(\alpha,\alpha'') \to \underline{u}(\alpha,\alpha'') \quad a.s.$$

Without Assumptions $\mathbf{A6_2}(\alpha, 0.5+\tau\alpha'')$, *we have*

$$[\liminf \underline{\widehat{u}}_n(\alpha,\alpha''), \limsup \underline{\widehat{u}}_n(\alpha,\alpha'')] \subseteq [\underline{u}(\alpha,\alpha''), \underline{u}^+(\alpha,\alpha'')] \quad a.s. \tag{3.2}$$

4. Asymptotic distributions

In this section, we first derive the asymptotic distribution of the empirical conditional quantile $\widehat{u}_n(\alpha,\alpha')$. We first need to establish the asymptotic distribution of the empirical conditional cdf $\widehat{F}_n(x \mid y)$.

The assumptions below ensure the asymptotic normality of the QMLE of $\boldsymbol{\theta}_0^{(i)}$, as a consequence of the Bahadur expansion

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(i)} - \boldsymbol{\theta}_0^{(i)}) = \frac{\boldsymbol{J}_i^{-1}}{2\sqrt{n}} \sum_{t=1}^{n} (\eta_{it}^2 - 1)\boldsymbol{D}_{it} + o_P(1), \tag{4.1}$$

where $\boldsymbol{J}_i = E(\boldsymbol{D}_{it}\boldsymbol{D}_{it}')$ and $\boldsymbol{D}_{it} = \boldsymbol{D}_{it}(\boldsymbol{\theta}_0^{(i)})$. Finally, let $\boldsymbol{\Omega}_i = E(\boldsymbol{D}_{it})$ and $\boldsymbol{J}_{12} = E\boldsymbol{D}_{1t}\boldsymbol{D}_{2t}'$.

$\mathbf{B1_i}$: $\boldsymbol{\theta}_0^{(i)}$ belongs to the interior of $\boldsymbol{\Theta}^{(i)}$.

$\mathbf{B2_i}$: There exist no non-zero $\boldsymbol{x} \in \mathbb{R}^{d_i}$ such that $\boldsymbol{x}' \frac{\partial \sigma_{it}(\boldsymbol{\theta}_0^{(i)})}{\partial \boldsymbol{\theta}^{(i)}} = 0, \quad a.s.$

$\mathbf{B3_i}$: The function $\boldsymbol{\theta}^{(i)} \mapsto \sigma(x_1, x_2, \ldots; \boldsymbol{\theta}^{(i)})$ has continuous second-order derivatives, and

$$\sup_{\boldsymbol{\theta}^{(i)} \in \boldsymbol{\Theta}} \left\| \frac{\partial \sigma_{it}(\boldsymbol{\theta}^{(i)})}{\partial \boldsymbol{\theta}^{(i)}} - \frac{\partial \widetilde{\sigma}_t(\boldsymbol{\theta}^{(i)})}{\partial \boldsymbol{\theta}^{(i)}} \right\| \leq K\rho^t.$$

$\mathbf{B4_i}$: There exists a neighborhood $V(\boldsymbol{\theta}_0^{(i)})$ of $\boldsymbol{\theta}_0^{(i)}$ such that

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

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

$\mathbf{B5_i}$: 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 $\mathbb{R}^2$.

It is clear that under **B6**, the variables η_{it} admit a density, denoted $g^{(i)}$, so that Assumptions **A5** − **A7** are satisfied. In addition, we make the next assumption on the volatility functions which entails formidable simplifications in the upcoming asymptotic results. This assumption is satisfied by all commonly used GARCH-type models.

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

4.1. For the empirical conditional cdf

We first derive the asymptotic distribution of $\widehat{F}_n(x \mid y)$. To deduce the law of the empirical conditional quantiles, it will not be 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).

Let $H(x, y) = P\{\eta_{1t} \le x, \eta_{2t} \le y\}$. 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 $\eta_{2t} \le y$ (resp. $\eta_{1t} \le x$).

Theorem 4.1. *Assume* **A1**-**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 $G^{(2)}(y) \neq 0$, we have*

$$\sqrt{n}\left(\widehat{F}_n(x_n \mid y_n) - F(x_n \mid y_n)\right) = \frac{1}{\sqrt{n}G^{(2)}(y)} \sum_{t=1}^{n} \{\mathbb{1}_{\eta_{1t} \le x,\, \eta_{2t} \le y} - H(x, y)\}$$
$$+ \frac{x f_1(x \mid y)}{2\sqrt{n}} \boldsymbol{\Omega}_1' \boldsymbol{J}_1^{-1} \sum_{t=1}^{n} (\eta_{1t}^2 - 1)\boldsymbol{D}_{1t} + \frac{y\Delta(x, y)}{2\sqrt{n}G^{(2)}(y)} \boldsymbol{\Omega}_2' \boldsymbol{J}_2^{-1} \sum_{t=1}^{n} (\eta_{2t}^2 - 1)\boldsymbol{D}_{2t}$$
$$- \frac{F(x \mid y)}{G^{(2)}(y)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{\mathbb{1}_{\eta_{2t} \le y} - G^{(2)}(y)\right\} + o_P(1),$$

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

If in addition **B7$_\mathbf{i}$** *holds for $i = 1, 2$, we have $\boldsymbol{\Omega}_i' \boldsymbol{J}_i^{-1} \boldsymbol{D}_{it} = 1$ a.s. and*

$$\sqrt{n}\left(\widehat{F}_n(x_n \mid y_n) - F(x_n \mid y_n)\right) \overset{\mathcal{L}}{\to} \mathcal{N}\left(0, \sigma_{x|y}^2\right)$$

where

$$\sigma_{x|y}^2 = \frac{F(x \mid y)\{1 - F(x \mid y)\}}{G^{(2)}(y)} + \frac{\{x f_1(x \mid y)\}^2}{4}(\kappa_1 - 1) + \frac{y^2 \Delta^2(x, y)}{4\{G^{(2)}(y)\}^2}(\kappa_2 - 1)$$
$$+ \frac{x f_1(x \mid y)}{G^{(2)}(y)} \varrho_1(x, y) + \frac{y\Delta(x, y)}{\{G^{(2)}(y)\}^2} \varrho_2(x, y) + \frac{xy f_1(x \mid y)\Delta(x, y)}{2G^{(2)}(y)} \left\{E(\eta_{1t}^2 \eta_{2t}^2) - 1\right\}$$

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

***Remark* 4.1.** It is worth noting that under the (mild) assumptions $\mathbf{B7_i}$ on the volatility functions, the asymptotic variance of the empirical conditional distribution function of the residuals is model-free (i.e. independent of the volatility parameters). However, estimation matters: if the residuals were replaced by the (supposedly observed) innovations, the asymptotic variance would reduce to the first term in the formula of $\sigma^2_{x|y}$.

***Remark* 4.2.** In the case where η_{1t} and η_{2t} are independent we have $\Delta(x,y) = 0$ and the asymptotic variance reduces to

$$\sigma^2_{x|y} = \frac{G^{(1)}(x)\{1 - G^{(1)}(x)\}}{G^{(2)}(y)} + \frac{\{xg^{(1)}(x)\}^2}{4}(\kappa_1 - 1) + xg^{(1)}(x)E\{\eta^2_{it}\mathbb{1}_{\eta_{1t}\le x} - G^{(1)}(x)\}.$$

For the sake of comparison, note that the asymptotic variance of $\sqrt{n}\left(\widehat{G}^{(1)}_n(x_n) - G^{(1)}(x_n)\right)$ is (see FZ, Theorem 2.2)

$$\sigma^2_x := G^{(1)}(x)\{1 - G^{(1)}(x)\} + \frac{\{xg^{(1)}(x)\}^2}{4}(\kappa_1 - 1) + xg^{(1)}(x)E\{\eta^2_{it}\mathbb{1}_{\eta_{1t}\le x} - G^{(1)}(x)\}.$$

The two asymptotic variances, $\sigma^2_{x|y}$ and σ^2_x, differ only by the presence of $G^{(2)}(y)$ in the denominator of the first summand of $\sigma^2_{x|y}$, coming from the fact that the estimator of $G^{(1)}(x_n)$ is based on the proportion of the observations satisfying the constraint $\widehat{\eta}_{2t} < y_n$.

***Remark* 4.3.** Under **B7**, estimating the asymptotic variance $\sigma^2_{x|y}$ 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. We provide in Appendix A a closed form formula in the Gaussian case.

4.2. For the empirical conditional quantile $\widehat{u}_n(\alpha, \alpha')$

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

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

Define the covariance matrix $\mathbf{\Sigma}_\Upsilon$ of the vector

$$\mathbf{\Upsilon}_t = \left(\mathbb{1}_{\eta_{1t}\le u(\alpha,\alpha'),\, \eta_{2t}\le \xi^{(2)}_{\alpha'}},\ \eta^2_{1t},\ \mathbb{1}_{\eta_{2t}\le\xi^{(2)}_{\alpha'}}\right).$$

To alleviate the notations, write $\widehat{u}_n = \widehat{u}_n(\alpha, \alpha')$, $u = u(\alpha, \alpha')$.

Theorem 4.2. *Let* **A1**-**A4**, **B1**-**B6** *and* **B8** *hold. We have*

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

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

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

where

$$\boldsymbol{\lambda}'=\left(\frac{1}{\alpha' f_1(u|\xi^{(2)}_{\alpha'})},\frac{u}{2},-\frac{1}{\alpha'}\frac{G^{(1)}(u)}{f_1(u|\xi^{(2)}_{\alpha'})}\frac{f_2(\xi^{(2)}_{\alpha'}|u)}{g^{(2)}(\xi^{(2)}_{\alpha'})}\right).$$

***Remark* 4.4.** It is worth noticing that the sole assumption $\mathbf{B7_1}$ (without $\mathbf{B7_2}$) is sufficient to ensure that the asymptotic distribution of $\widehat{u}_n$ is model-free, that is, independent of the volatility specifications.

***Remark* 4.5.** Comparing the asymptotic expansion of $\sqrt{n}\left(\widehat{F}_n(x_n\mid y_n)-F(x_n\mid y_n)\right)$ in Theorem 4.1 with that of $\sqrt{n}\left\{\widehat{u}_n-u\right\}$ we note that the sum $\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}$ no longer appears in the latter expansion. Thus, the effect of estimation on the asymptotic distribution of of $\sqrt{n}\left\{\widehat{u}_n-u\right\}$ is only due to the estimation of the volatility of the first component.

***Remark* 4.6.** 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

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

where f denotes the density of η_{1t} and $\varrho(\xi^{(1)}_{\alpha})=E(\eta_{1t}^2\mathbb{1}_{\eta_{1t}\le\xi^{(1)}_{\alpha}})-\alpha$, whereas the asymptotic distribution of the empirical quantile of the residuals $\widehat{\eta}_{1t}$ is (see FZ)

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

Unsurprisingly, the estimator $\widehat{u}_n$ is asymptotically less accurate (particularly when α' is small), as a price paid for the unnecessary inclusion of the residuals of the second volatlity model in the estimation of $\xi^{(1)}_{\alpha}$. 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).

***Remark* 4.7.** An explicit expression for $\mathbf{\Sigma}_{\mathbf{\Upsilon}}$ is

$$\mathbf{\Sigma}_{\mathbf{\Upsilon}} = \begin{pmatrix} \alpha\alpha'(1-\alpha\alpha') & \alpha'\varrho_{\alpha,\alpha'} & \alpha\alpha'(1-\alpha') \\ \alpha'\varrho_{\alpha,\alpha'} & \kappa_1 - 1 & \alpha'\nabla_{\alpha,\alpha'} \\ \alpha\alpha'(1-\alpha') & \alpha'\nabla_{\alpha,\alpha'} & (1-\alpha')\alpha' \end{pmatrix}$$

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

4.3. For the empirical conditional quantile $\widehat{\underline{u}}_n(\alpha, \alpha'')$

Now we turn to the asymptotic distribution of $\widehat{\underline{u}}_n(\alpha, \alpha'')$, simply denoted $\widehat{\underline{u}}_n$ in the sequel. 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.

Theorem 4.3. *Let* **A1-A4**, **B1-B6** *and* **B8** *hold. We have*

$$\begin{aligned}
&\sqrt{n}\,(\widehat{\underline{u}}_n - \underline{u}) \\
&= -\frac{1}{2\sqrt{n}\alpha'' f_1(\underline{u} \mid A^{(2)}_{\alpha''})} \sum_{t=1}^n \{\mathbb{1}_{\eta_{1t}\le \underline{u}} \mathbb{1}_{\eta_{2t}\in A^{(2)}_{\alpha''}} - 2\alpha\alpha''\} - \frac{\underline{u}}{2\sqrt{n}} \mathbf{\Omega}_1' \boldsymbol{J}_1^{-1} \sum_{t=1}^n (\eta_{1t}^2 - 1)\boldsymbol{D}_{1t} \\
&- \frac{(1-2\alpha'')\delta_{\alpha''}}{2 f_1(\underline{u} \mid A^{(2)}_{\alpha''})(2\alpha'')^2} \frac{1}{\sqrt{n}} \mathbf{\Omega}_2' \boldsymbol{J}_2^{-1} \sum_{t=1}^n (\eta_{2t}^2 - 1)\boldsymbol{D}_{2t} \\
&- \frac{f_2(\xi^{(2)}_{0.5-\alpha''} \mid \underline{u}) G^{(1)}(\underline{u})}{2\alpha'' g^{(2)}(\xi^{(2)}_{0.5-\alpha''}) f_1(\underline{u} \mid A^{(2)}_{\alpha''})} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\eta_{2t}\le \xi^{(2)}_{0.5-\alpha''}} - \alpha'' + 0.5) \\
&+ \frac{f_2(\xi^{(2)}_{0.5+\alpha''} \mid \underline{u}) G^{(1)}(\underline{u})}{2\alpha'' g^{(2)}(\xi^{(2)}_{\alpha''+0.5}) f_1(\underline{u} \mid A^{(2)}_{\alpha''})} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\eta_{2t}\le \xi^{(2)}_{\alpha''+0.5}} - 0.5 - \alpha'')
\end{aligned}$$

up to some $o_P(1)$, *where* $\delta_{\alpha''} = \xi^{(2)}_{0.5-\alpha''} g^{(2)}(\xi^{(2)}_{0.5-\alpha''}) - \xi^{(2)}_{0.5+\alpha''} g^{(2)}(\xi^{(2)}_{0.5+\alpha''})$. *With the additional assumption* $\mathbf{B7_1}$ *we have,*

5. Confidence intervals for the CoVaR

5.1. Using the asymptotic distribution

***Remark* 5.1.** Let $\widetilde{\boldsymbol{D}}_{it}(\boldsymbol{\theta}^{(i)}) = \widetilde{\sigma}_{it}^{-1}(\boldsymbol{\theta}^{(i)}) \partial\widetilde{\sigma}_{it}(\boldsymbol{\theta}^{(i)})/\partial\boldsymbol{\theta}^{(i)}$ and $\widehat{\boldsymbol{D}}_{it} = \widetilde{\boldsymbol{D}}_{it}\left(\widehat{\boldsymbol{\theta}}_n^{(i)}\right)$. Under **A1–A4** and **B1–B6**, it can be shown that

$$\sup_{\boldsymbol{\theta}^{(i)}\in\mathbf{\Theta}} \left\| \widetilde{\boldsymbol{D}}_{it}(\boldsymbol{\theta}^{(i)}) - \boldsymbol{D}_{it}(\boldsymbol{\theta}^{(i)}) \right\| \le K\rho^t u_t, \qquad \widehat{\boldsymbol{J}}_{in} := \frac{1}{n}\sum_{t=1}^n \widehat{\boldsymbol{D}}_{it} \widehat{\boldsymbol{D}}_{it}' \to \boldsymbol{J}_i \text{ a.s.},$$

with (u_t) a positive stationary process such that $Eu_t^4 < \infty$.

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 $g^{(2)}$. Similarly, for y such that $P(\eta_{2t} \le y) > 0$, one can estimate $f_1(\mid y)$ by the kernel density estimator

$$\widehat{f}_1(x \mid y) = \frac{1}{n_2(y) b_{n_2(y)}} \sum_{t=1}^{n} K\left(\frac{x - \widehat{\eta}_{1t}}{b_{n_2(y)}}\right) \mathbb{1}_{\widehat{\eta}_{2t} \le y} \mathbb{1}_{n_2(y) > 0},$$

where $n_2(y)$ is the number of $\widehat{\eta}_{2t}$'s for $t = 1, \dots, n$ such as $\widehat{\eta}_{2t} \le y$, the kernel K is a standardized probability density and (b_n) is a sequence of positive numbers satisfying some regularity conditions (for our numerical illustrations we kept the default values of the `R` function `density()`).

We do not specify particular estimators for the previous densities, but assume the following.

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 $n \to \infty$.

To obtain approximate confidence intervals for $\mathrm{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}$ based on $\widehat{\mathrm{CoVaR}}_{n+1} = -\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}_n^{(1)})\widehat{u}$, it is necessary to estimate the joint distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(1)'} - \boldsymbol{\theta}_0^{(1)'}, \widehat{u} - u)'$ whose asymptotic distribution is[1]

$$\mathcal{N}\left\{\mathbf{0}, \boldsymbol{\Sigma}(\alpha, \alpha')\right\}, \quad \boldsymbol{\Sigma}(\alpha, \alpha') = \begin{pmatrix} \frac{\kappa_1 - 1}{4}\boldsymbol{J}_1^{-1} & \frac{-1}{2}\boldsymbol{J}_1^{-1}\boldsymbol{\Omega}_1\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_2 \\ \frac{-1}{2}\boldsymbol{e}_2'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{\lambda}\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1} & \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{\lambda} \end{pmatrix}. \tag{5.1}$$

5.2. Bootstrapped CoVaR

Monte Carlo experiments revealed that the finite sample behaviours 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), Pascual et al. (2006), Hidalgo and Zaffaroni (2007), Corradi and Iglesias (2008), Shimizu (2010), Spierdijk (2016), Cavaliere et al. (2018), Beutner et al. (2020) 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 Shimizu (2013), Cavaliere et al. (2017) 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.

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

[1] Using (4.1) and noting that $E(\eta_{1t}^2 - 1)\boldsymbol{\Upsilon}_t = \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_2$ with $\boldsymbol{e}_2 = (0, 1, 0)'$

[2] As noted by Beutner et al. (2020) it is useless to standardized the residuals, but it can be done.

2. Independently of the observations, generate independent and $\widehat{H}_n$-distributed vectors $(\eta_{1t}^*, \eta_{2t}^*)'$, $t = 1, \dots, n$. Generate the Newton-Raphson bootstrap estimates

$$\widehat{\boldsymbol{\theta}}_n^{(1)*} = \widehat{\boldsymbol{\theta}}_n^{(1)} + \frac{\widehat{\boldsymbol{J}}_{1n}^{-1}}{2n} \sum_{t=1}^{n} (\eta_{1t}^{*2} - \widehat{m}_2) \widehat{\boldsymbol{D}}_{1t}$$

and

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

where $\widehat{\xi} = \widehat{\xi}_{n,\alpha'}^{(2)}$, $\hat{\alpha}' = \widehat{G}_n^{(2)}(\widehat{\xi})$, $\hat{\alpha} = \widehat{H}_n(\widehat{u}, \widehat{\xi})/\hat{\alpha}'$ and $\widehat{m}_k = n^{-1} \sum_{t=1}^{n} \widehat{\eta}_{1t}^k$.[3] The bootstrap estimator of $\text{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}(\alpha, \alpha')$ is $\text{CoVaR}^* = -\widetilde{\sigma}_{1,n+1}(\widehat{\boldsymbol{\theta}}_n^{(1)*})\widehat{u}^*$.

The next result establish the validity of the resampling algorithm.

Theorem 5.1. *Let* **A1**-**A4** *and* **B1**-**B9** *hold. For almost all realization* $(\boldsymbol{\epsilon}_t)$*, as* $n \to \infty$ *we have, given* $(\boldsymbol{\epsilon}_t)$*,*

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n^{(1)*} - \widehat{\boldsymbol{\theta}}_n^{(1)} \\ \widehat{u}^* - \widehat{u} \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{ \mathbf{0}, \boldsymbol{\Sigma}(\alpha, \alpha') \right\}. \tag{5.2}$$

The resampling algorithm thus provides a way to approximate the distribution (5.1), and any statistics depending of this distribution, without having to estimate $\boldsymbol{\Sigma}(\alpha, \alpha')$ directly. For instance an approximate confidence interval for the CoVaR can be obtained by adding the following step to the algorithm.

3. Repeat B times Step 2, and denote by $\text{CoVaR}_1^*, \dots, \text{CoVaR}_B^*$ the bootstrap estimates of $\text{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}(\alpha, \alpha')$. An approximate $100(1 - \underline{\alpha})\%$ confidence interval for $\text{CoVaR}_{n+1}^{\epsilon_1|\epsilon_2}(\alpha, \alpha')$ is $\left[\text{CoVaR}_{(\underline{\alpha}/2)}^*,\ \text{CoVaR}_{(1-\underline{\alpha}/2)}^*\right]$ where $\text{CoVaR}_{(\underline{\alpha})}^*$ denotes the empirical $\underline{\alpha}$-quantile of the B bootstrap CoVaR estimates

Lemma 5.1. *Suppose that the assumptions of Theorem 5.1 are 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*

$$|\widehat{H}_n(x_n, y_n) - H(x, y)| \to 0 \quad \textit{and} \quad \left|\widehat{G}_n^{(2)}(y_n) - G^{(2)}(y)\right| \to 0 \textit{ a.s.} \tag{5.3}$$

For all $k \leq 4$*, as* $n \to \infty$

$$\frac{1}{n} \sum_{t=1}^{n} \widehat{\eta}_{1t}^k \mathbb{1}_{\widehat{\eta}_{1t} \leq x_n} \mathbb{1}_{\widehat{\eta}_{2t} \leq y_n} \to E\eta_{1t}^k \mathbb{1}_{\eta_{1t} \leq x} \mathbb{1}_{\eta_{2t} \leq y} \textit{ a.s.} \tag{5.4}$$

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

Proof of Lemma 5.1. The second convergence in (5.3) is already known (see Theorem 2.1. in Francq and Zakoian, 2022). We have

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

Theorem 3.1 and the continuity of H (under **B6**) then entail the first convergence of (5.3).

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 $\widehat{\eta}_{1t} = \widetilde{\eta}_{1t}(\widehat{\boldsymbol{\theta}}_n^{(1)})$ and $\eta_{1t} = \eta_{1t}(\boldsymbol{\theta}_0^{(1)})$. By $\mathbf{A2_1}$ and $\mathbf{A3_1}$ we have

$$\sup_{\boldsymbol{\theta}^{(1)}\in\boldsymbol{\Theta}^{(1)}} \left|\eta_{1t}^k(\boldsymbol{\theta}^{(1)}) - \widetilde{\eta}_{1t}^k(\boldsymbol{\theta}^{(1)})\right| \leq \frac{K}{\underline{\omega}}\rho^t|\epsilon_{1t}|^k.$$

Under $\mathbf{A1_1}$, 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 $E|\sum_{t=1}^n \rho^t|\epsilon_{1t}|^k|^s \leq E|\epsilon_{1t}|^{ks}\sum_{t=1}^n \rho^{ts}\infty$. It follows that

$$\frac{1}{n}\sum_{t=1}^n \widehat{\eta}_{1t}^k = \frac{1}{n}\sum_{t=1}^n \eta_{1t}^k(\widehat{\boldsymbol{\theta}}_n^{(1)}) + O(n^{-1}) \text{ a.s.}$$

By the mean value theorem

$$\frac{1}{n}\sum_{t=1}^n \eta_{1t}^k(\widehat{\boldsymbol{\theta}}_n^{(1)}) = \frac{1}{n}\sum_{t=1}^n \eta_{1t}^k + \frac{1}{n}\sum_{t=1}^n \frac{\partial \eta_{1t}^k(\boldsymbol{\theta}_n^{(1)})}{\partial \boldsymbol{\theta}^{(1)'}}\left(\widehat{\boldsymbol{\theta}}_n^{(1)} - \boldsymbol{\theta}_0^{(1)}\right),$$

with $\boldsymbol{\theta}_n^{(1)}$ between $\widehat{\boldsymbol{\theta}}_n^{(1)}$ and $\boldsymbol{\theta}_0^{(1)}$. By $\mathbf{B4_1}$ 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) stationary ergodic and $Eu_t<\infty$. We thus have

$$\frac{1}{n}\sum_{t=1}^n \widehat{\eta}_{1t}^k \to E\eta_{1t}^k, \qquad \text{a.s.}$$

for $k \leq 4$.

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 $\eta_n^* \sim \widehat{G}_n^{(1)}$. Since $\widehat{G}_n^{(1)}$ converges to $G^{(1)}$, conditional on $(\boldsymbol{\epsilon}_t)$ and (x_n), as $n\to\infty$ the random variable $\eta_n^{*k}\mathbb{1}_{\eta_n^*\leq x_n}$ converges in distribution to $\eta^k\mathbb{1}_{\eta\leq x}$ where $\eta\sim G^{(1)}$. 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 $\eta_n^{*k}\mathbb{1}_{\eta_n^*\leq x_n}$

are also uniformly integrable. By Theorem 3.5 in Billingsley (1999), conditional on $(\boldsymbol{\epsilon}_t)$ and (x_n), we then have

$$\frac{1}{n}\sum_{t=1}^{n}\widehat{\eta}_{1t}^{k}\mathbb{1}_{\widehat{\eta}_{1t}\leq x_n} = E\eta_n^{*k}\mathbb{1}_{\eta_n^*\leq x_n} \to E\eta^k\mathbb{1}_{\eta\leq x} \text{ as } n\to\infty.$$

The rest of the proof follows by similar arguments. The proof of Lemma 5.1 is complete.
Proof of Theorem 5.1. In view of Remark 5.1, the consistency of $(\widehat{u},\widehat{\xi})$ and **A9**, for almost all sequence $(\boldsymbol{\epsilon}_t)$ we have

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

and

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

where

$$\boldsymbol{\Upsilon}_{t,n}^* = \left(\mathbb{1}_{\eta_{1t}^*\leq\widehat{u},\,\eta_{2t}^*\leq\widehat{\xi}}-\hat{\alpha}\hat{\alpha}', \eta_{1t}^{*2}-\widehat{m}_2, \mathbb{1}_{\eta_{2t}^*\leq\widehat{\xi}}-\hat{\alpha}'\right)'.$$

Letting $\boldsymbol{y}_{t,n}^* = (\boldsymbol{x}_{t,n}^{*'}, \boldsymbol{\Upsilon}_{t,n}^{*'})'$, the result follows by showing that, conditional on $(\boldsymbol{\epsilon}_t)$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{y}_{t,n}^* \xrightarrow{\mathcal{L}} \mathcal{N}\left\{\boldsymbol{0}, \boldsymbol{\Sigma} := \begin{pmatrix} (\kappa_1-1)\boldsymbol{J}_1 & \boldsymbol{\Omega}_1\boldsymbol{e}_2'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_2\boldsymbol{\Omega}_1' & \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \end{pmatrix}\right\}. \tag{5.5}$$

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 $\boldsymbol{c}\in\mathbb{R}^{d_1+3}$, $\boldsymbol{c}\neq\boldsymbol{0}$,

$$\frac{1}{n}\sum_{t=1}^{n}\mathrm{Var}\left(\boldsymbol{c}'\boldsymbol{y}_{t,n}^*\right) \to \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c} \quad \text{as } n\to\infty, \tag{5.6}$$

and for all $\varepsilon>0$

$$\frac{1}{n}\sum_{t=1}^{n}E\left(\left\{\boldsymbol{c}'\boldsymbol{y}_{t,n}^*\right\}^2\mathbb{1}_{\{|\boldsymbol{c}'\boldsymbol{y}_{t,n}^*|\geq\sqrt{n}\varepsilon\}}\right) \to 0 \quad \text{as } n\to\infty. \tag{5.7}$$

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

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

where $\widehat{\alpha}'\hat{\varrho}_{\widehat{\alpha},\widehat{\alpha}'} = n^{-1}\sum_{t=1}^n \widehat{\eta}_{1t}^2 \mathbb{1}_{\widehat{\eta}_{1t}<\widehat{u}}\mathbb{1}_{\widehat{\eta}_{2t}<\widehat{\xi}} - \widehat{\alpha}\widehat{\alpha}'$ and $\widehat{\alpha}'\widehat{\nabla}_{\alpha,\alpha'} = n^{-1}\sum_{t=1}^n \widehat{\eta}_{1t}^2 \mathbb{1}_{\widehat{\eta}_{2t}<\widehat{\xi}} - \widehat{\alpha}'$. Lemma 5.1 and the consistency of $\widehat{\boldsymbol{\theta}}_n^{(1)*}$ and $(\widehat{u}, \widehat{\xi})$ show that, for t fixed and $n \to \infty$

$$\mathrm{Var}\left(\boldsymbol{y}_{t,n}^*\right) \to \begin{pmatrix} (\kappa_1 - 1)\widetilde{\boldsymbol{D}}_{1t}\widetilde{\boldsymbol{D}}_{1t}' & \widetilde{\boldsymbol{D}}_{1t}\boldsymbol{e}_2'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_2\widetilde{\boldsymbol{D}}_{1t}' & \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \end{pmatrix}$$

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

$$\frac{1}{n}\sum_{t=1}^n \begin{pmatrix} (\kappa_1 - 1)\widetilde{\boldsymbol{D}}_{1t}\widetilde{\boldsymbol{D}}_{1t}' & \widetilde{\boldsymbol{D}}_{1t}\boldsymbol{e}_2'\boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \\ \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}}\boldsymbol{e}_2\widetilde{\boldsymbol{D}}_{1t}' & \boldsymbol{\Sigma}_{\boldsymbol{\Upsilon}} \end{pmatrix} \to \boldsymbol{\Sigma} \text{ as } n \to \infty$$

which entails (5.6).

Now we turn to the proof of (5.7). 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 (5.7) 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

$$\begin{aligned}
&E\left\{\boldsymbol{c}_1'\boldsymbol{x}_{t,n}^*\right\}^2 \mathbb{1}_{\left\{\left|\boldsymbol{c}_1'\boldsymbol{x}_{t,n}^*\right| \geq \sqrt{n}\varepsilon\right\}} \\
&\leq \mathbb{1}_{\left\{\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\sup_{t\geq 1}\left|\boldsymbol{c}_1'\widetilde{\boldsymbol{D}}_t(\boldsymbol{\theta}^{(1)})\right|>0\right\}} \sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0)}\sup_{t\geq 1}\left\{\boldsymbol{c}_1'\widetilde{\boldsymbol{D}}_t(\boldsymbol{\theta}^{(1)})\right\}^2 \\
&\times E\left|\eta_{1t}^{*2} - \widehat{m}_2\right|^2 \mathbb{1}_{\left\{\left|\eta_{1t}^{*2}-\widehat{m}_2\right| \geq \frac{\sqrt{n}\varepsilon}{\sup_{\boldsymbol{\theta}^{(1)}\in V(\boldsymbol{\theta}_0^{(1)})}\sup_{t\geq 1}\left|\boldsymbol{c}_1'\widetilde{\boldsymbol{D}}_t(\boldsymbol{\theta}^{(1)})\right|}\right\}}.
\end{aligned} \tag{5.8}$$

For any $A > 0$ there exists n_A such that if $n > n_A$ then the expectation in the right-hand side of (5.8) is bounded by

$$E\left|\eta_{1t}^{*2} - \widehat{m}_2\right|^2 \mathbb{1}_{\left\{\left|\eta_{1t}^{*2}-\widehat{m}_2\right| \geq A\right\}}.$$

By the arguments of the proof of Lemma 5.1, this term tends to

$$\int_{|x^2-1|\geq A} \left|x^2 - 1\right|^2 G^{(1)}(dx)$$

which is arbitrarily small when A is sufficiently large. We then obtain (5.7) for $\boldsymbol{c} = (\boldsymbol{c}_1', \boldsymbol{0}_3')'$. A similar argument shows (5.7) for $\boldsymbol{c} = (\boldsymbol{0}_{d_1}', 0, 1, 0)'$. Now, note that

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

for n large enough, which shows (5.7) for $\boldsymbol{c} = (\boldsymbol{0}_{d_1}', 1, 0, 0)'$. By the same argument, the convergence holds for $\boldsymbol{c} = (\boldsymbol{0}_{d_1}', 0, 0, 1)'$. We thus have shown (5.7) and the proof is complete. □

6. Multi-asset extensions

In this section, we extend the notion of CoVaR and its estimation to handle situations where the risks of more than two assets are considered. The *dynamic* multi-CoVaR (MCoVaR) of a process (X_t) relative to a sequence of process (Y_{jt}), $j = 1, \dots, m$ at risk levels $\alpha, \alpha'_j \in (0, 1)$, denoted by $\mathrm{MCoVaR}_t^{X|(Y_j)}(\alpha, (\alpha'_j))$, can be defined as

$$P_{t-1}\left[X_t \leq -\mathrm{MCoVaR}_t^{X|(Y_j)}(\alpha, (\alpha'_j)_{1\leq j\leq m}) \,\middle|\, \bigcap_{j=1}^{m} \left\{Y_{jt} \leq -\mathrm{VaR}_{jt}^{Y}(\alpha'_j)\right\}\right] = \alpha.$$

This definition measures the systemic risk of a firm by (the opposite of) a quantile, conditional on the fact that *all* firms of the system are in distress. The levels of distress are allowed to be firm-dependent through the introduction of different risk levels α'_j. Similarly, we could define a quantile conditional on the fact that *at least one* firm is in distress. More generally, denoting $(\alpha'_\ell)_{1\leq \ell\leq m}$ by $\boldsymbol{\alpha}'$,

$$P_{t-1}\left[X_t \leq -\mathrm{MCoVaR}_t^{X|(Y_j)_{[i]}}(\alpha, \boldsymbol{\alpha}') \,\middle|\, \bigcup_{\{j_1,\dots,j_i\}\subset\{1,\dots m\}} \bigcap_{\ell=1}^{i} \left\{Y_{j_\ell t} \leq -\mathrm{VaR}_{j_\ell t}^{Y}(\alpha'_{j_\ell})\right\}\right] = \alpha$$

defines a CoVaR at level α for X_t, conditional on a proportion i/m of firms in distress in the system.

Now let the $m+1$-dimensional vector $\boldsymbol{\epsilon}_t = (\epsilon_{0t}, \epsilon_{1t}, \dots, \epsilon_{m,t})'$, whose entries satisfy models of the form (2.1). The following is a straightforward extension of Proposition 2.1. To simplify notation, we simply denote $(\epsilon_j)_{[i]}$ by $\boldsymbol{\epsilon}_{[i]}$.

***Proposition* 6.1.** The conditional CoVaR at levels $\alpha, \boldsymbol{\alpha}'$ of the first relative to a proportion i/m of the other components is given by

$$\mathrm{MCoVaR}_t^{\epsilon_0|\boldsymbol{\epsilon}_{[i]}}(\alpha, \boldsymbol{\alpha}') = -\sigma_{0t} u^{[i]}(\alpha, \boldsymbol{\alpha}'),$$

where $u^{[i]}(\alpha, \boldsymbol{\alpha}')$ is such that

$$P\left[\eta_{0t} \leq u^{[i]}(\alpha, \boldsymbol{\alpha}') \,\middle|\, \bigcup_{\{j_1,\dots,j_i\}\subset\{1,\dots m\}} \bigcap_{\ell=1}^{i} \left(\eta_{j_\ell t} \leq \xi_{\alpha'_{j_\ell}}^{(j_\ell)}\right)\right] = \alpha,$$

and $\xi_\alpha^{(j)}$ is the α-quantile of η_{jt}.

Let, for $i = 1, \dots, m$,

$$F^{[i]}(x|(y_j)_{j=1,\dots,m}) = F^{[i]}(x|\boldsymbol{y}) = P\left[\eta_{0t} \leq x \,\middle|\, \bigcup_{\{j_1,\dots,j_i\}\subset\{1,\dots m\}} \bigcap_{\ell=1}^{i} (\eta_{j_\ell t} \leq y_{j_\ell})\right]$$

for $x \in \mathbb{R}$, and $y_j \in \mathbb{R}$ such that the probability of the conditioning set is non zero. An estimator of $F^{[i]}(x|\boldsymbol{y})$ is

$$\widehat{F}_n^{[i]}(x|\boldsymbol{y}) = \frac{\frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{0t}\leq x}\mathbb{1}_{\bigcup_{\{j_1,\dots,j_i\}\subset\{1,\dots m\}}\bigcap_{\ell=1}^i(\widehat{\eta}_{j_\ell t}\leq y_{j_\ell})}}{\frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\bigcup_{\{j_1,\dots,j_i\}\subset\{1,\dots m\}}\bigcap_{\ell=1}^i(\widehat{\eta}_{j_\ell t}\leq y_{j_\ell})}} := \frac{\widehat{H}_n^{[i]}(x,\boldsymbol{y})}{\widehat{G}_n^{[i]}(\boldsymbol{y})},$$

provided the denominator is not equal to zero. Let $F_n^{[i]}(x|\boldsymbol{y}) = \frac{H_n^{[i]}(x,\boldsymbol{y})}{G_n^{[i]}(\boldsymbol{y})}$ be similarly defined, with residuals replaced by innovations.

Let $\widehat{u}_n^{[i]}(\alpha,\boldsymbol{\alpha}')$ the estimator of $u(\alpha,\alpha')$ such that

$$\widehat{u}_n^{[i]}(\alpha,\boldsymbol{\alpha}') = \inf\arg\min_{z\in\mathbb{R}}\sum_{t=1}^n \rho_\alpha(\widehat{\eta}_{0t}-z)\mathbb{1}_{\bigcup_{\{j_1,\dots,j_i\}\subset\{1,\dots m\}}\bigcap_{\ell=1}^i\left(\widehat{\eta}_{j_\ell t}<\widehat{\xi}^{(j_\ell)}_{n,\alpha'_{j_\ell}}\right)}$$

where $\widehat{\xi}^{(j)}_{n,\alpha'_j}$ is the α'_j-quantile of $\widehat{\eta}_{j1},\dots,\widehat{\eta}_{jn}$, that is the $\lceil n\alpha'_j\rceil$-th order statistics of these residuals. An estimator of $\mathrm{MCoVaR}_t^{\epsilon_0|\boldsymbol{\epsilon}_{[i]}}(\alpha,\boldsymbol{\alpha}')$ is thus

$$\widehat{\mathrm{MCoVaR}}_t^{\epsilon_0|\boldsymbol{\epsilon}_{[i]}}(\alpha,\boldsymbol{\alpha}') = -\widetilde{\sigma}_{1t}(\widehat{\boldsymbol{\theta}}_n^{(1)})\widehat{u}_n^{[i]}(\alpha,\boldsymbol{\alpha}'),$$

with obvious notations.

We show the following extension of Theorem 3.1.

Theorem 6.1. *Assume that the variables η_{it} admit densities. Then*

i) for $x,\boldsymbol{y}\in\mathbb{R}\times\mathbb{R}^m$, and if $G^{[i]}(\boldsymbol{y})>0$ we have $\sup_{x\in\mathbb{R}}|\widehat{F}_n^{[i]}(x|\boldsymbol{y})-F^{[i]}(x|\boldsymbol{y})|\to 0$ a.s.

ii) For any small enough neighborhood $V(\boldsymbol{y}_0)$ of $\boldsymbol{y}_0$ with $G^{[i]}(\boldsymbol{y}_0)>0$ we have $\sup_{x\in\mathbb{R},\boldsymbol{y}\in V(\boldsymbol{y}_0)}|\widehat{F}_n^{[i]}(x|\boldsymbol{y})-F^{[i]}(x|\boldsymbol{y})|\to 0$ a.s.

The strong consistency of $\widehat{u}_n^{[i]}(\alpha,\boldsymbol{\alpha}')$ is established in the next result.

Theorem 6.2. *Under* **A1**-**A4**, $\mathbf{A5_1}$, $\mathbf{A5_2}(\xi^{(2)}_{\alpha'})$, **A6**-**A7**, *if $G^{(2)}(\xi^{(2)}_{\alpha'})>0$ we have the strong convergence*

$$\widehat{u}_n^{[i]}(\alpha,\boldsymbol{\alpha}')\to u_n^{[i]}(\alpha,\boldsymbol{\alpha}')\quad a.s.$$

7. Numerical illustrations

7.1. Numerical computations of VaR and CoVaR

We first consider a bivariate GARCH-type model with

$$\sigma_{it}^2 = \omega_i + \alpha_{ii}\epsilon_{i,t-1}^2 + \alpha_{ij}\epsilon_{j,t-1}^2 + \beta_i\sigma_{i,t-1}^2,\quad i,j=1,2 \tag{7.1}$$

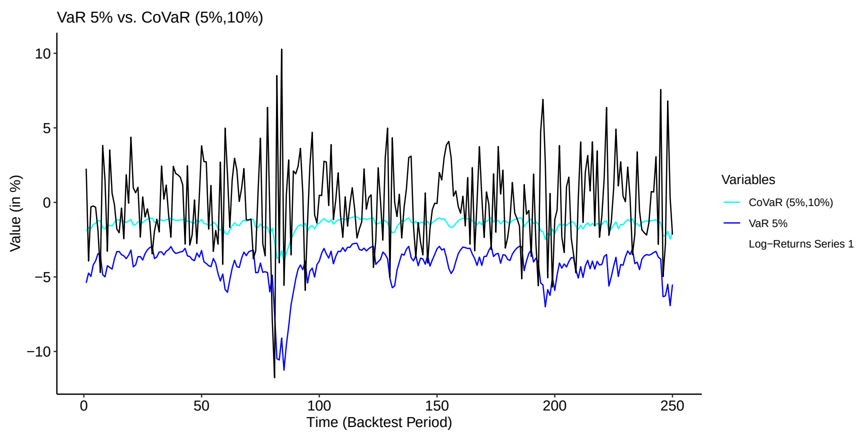

Figure 1. CoVaR versus VaR for a bivariate GARCH with Gaussian innovations with correlation $\rho = -0.5$

where the innovation vector follows a Gaussian distribution with $\rho = \text{cor}(\eta_{1t}, \eta_{2t})$. The volatility of each asset is thus impacted by the past values of the other asset. Figures 1 and 2 compare the theoretical conditional VaR and CoVaR. As expected the CoVaR is much larger than the VaR when ρ is positive. When ρ is negative, large negative returns for ϵ_{2t} are in average associated with positive values for ϵ_{2t}, leading to a CoVaR smaller than the VaR. This is a favorable situation where a long position in asset 1 can serve to hedge against a fall in the price of the asset 2. Figure 3 provides another illustration that the risks and co-risks of the two assets vary in the same way or not, depending on the sign of the correlation coefficient, which is intuitively clear for Gaussian innovations.

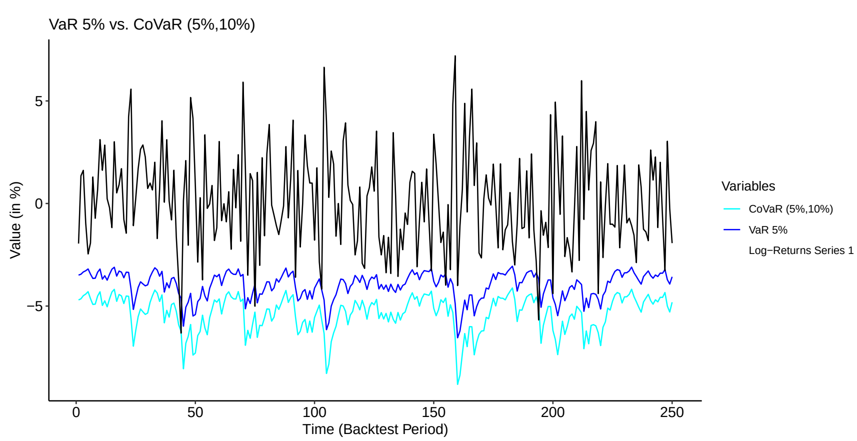

Figure 2. As Figure 1 but for $\rho = 0.5$

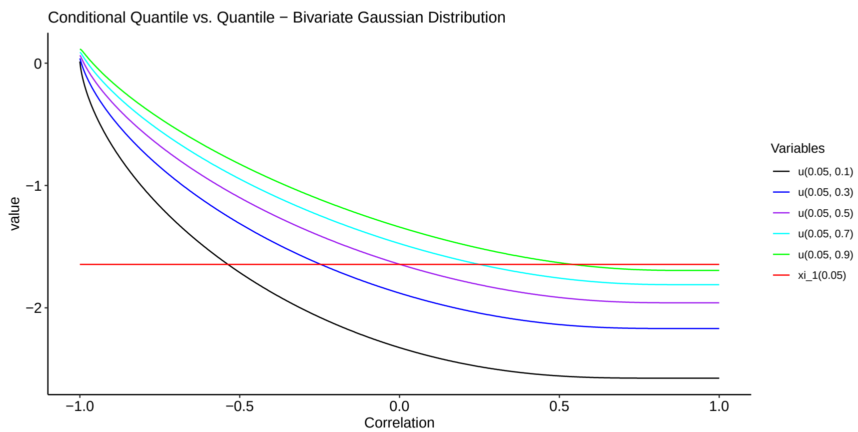

Figure 3. Quantile $\xi^{(2)}_{\alpha'}$ and CoVaR $u(\alpha,\alpha')$ as a function of ρ for a bivariate Gaussian distribution with correlation ρ

7.2. Monte Carlo experiments

In these experiments we aim at illustrating the asymptotic results in Theorems 4.1 and 4.2. In a first set of Monte Carlo experiments, we generated a GARCH model with volatilities

$$\begin{aligned}\sigma^2_{1t} &= 1+0.05\epsilon^2_{1,t-1}+0.01\epsilon^2_{2,t-1}+0.9\sigma^2_{1,t-1},\\ \sigma^2_{2t} &= 1+0.01\epsilon^2_{1,t-1}+0.1\epsilon^2_{2,t-1}+0.85\sigma^2_{1,t-1},\end{aligned}$$

where $\boldsymbol{\eta}_t$ follows a Gaussian distribution with $\rho = \text{cor}(\eta_{1t},\eta_{2t}) = 0.5$. The GARCH model is estimated equation by equation by QML. The empirical distribution of $\sqrt{n}\left\{\widehat{F}_n\left(u(\alpha,\alpha')\mid \xi^{(2)}_{\alpha'}\right)-F\left(u(\alpha,\alpha')\mid \xi^{(2)}_{\alpha'}\right)\right\}$ over 1000 independent replications is displayed in Figure 4. Figure 5 displays the empirical distribution of $\sqrt{n}\left\{\widehat{u}_n(\alpha,\alpha')-u(\alpha,\alpha')\right\}$. The empirical results are in accordance with the asymptotic Gaussian distributions of Theorems 4.1 and 4.2. Table 1 compares the empirical variance of $\sqrt{n}\left\{\widehat{F}_n\left(u(0.05,0.5)\mid \xi^{(2)}_{0.5}\right)-F\left(u(0.05,0.5)\mid \xi^{(2)}_{0.5}\right)\right\}$ computed through 1,000 simulations of a Monte Carlo experiment with the asymptotic variance of Theorem 4.1 (evaluated by numerical integration). The latter is seen to provide a good approximation for n large of the empirical variance.

7.3. A real data example

Figures 6 and 7 illustrate the joint behavior of the opposite of VaR and the opposite of CoVaR on a real data example estimated using a bivariate GARCH model satisfying the

Table 1. Empirical vs Asymptotic variance of Theorem 4.1

	n=1000	n=5000	n=15000
Empirical variance	0.071	0.075	0.073
Asymptotic variance	0.074	0.074	0.074

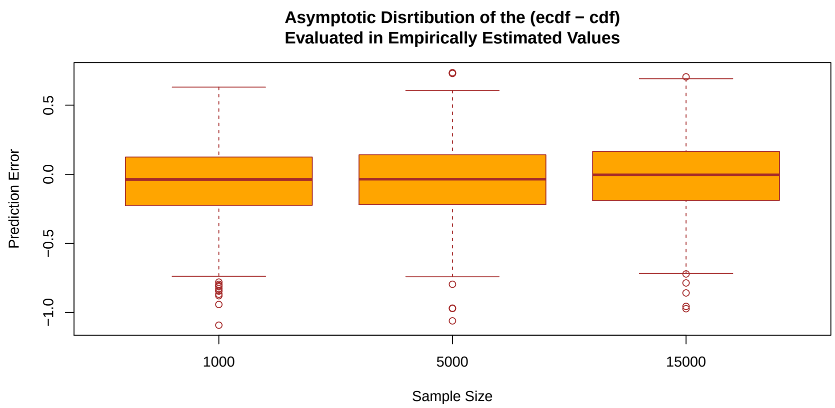

Figure 4. Empirical distribution of $\sqrt{n}\left\{\widehat{F}_n\left(\widehat{u}(\alpha,\alpha'),\widehat{\xi}^{(2)}_{\alpha'}\right) - F\left(\widehat{u}(\alpha,\alpha'),\widehat{\xi}^{(2)}_{\alpha'}\right)\right\}$ over 1000 independent replications

previously mentioned assumptions. The log-returns of series 1 are the log-returns of the Goldman Sachs stock over the period 12/02/2015 to 11/24/2017 (the model has been calibrated on the 01/02/2004 to 11/30/2015 period) and series 2 (the "conditioning" series) consists of the log-returns of the Dow Jones over the same period. Two different risk levels have been chosen here: $\alpha = 5\%$ and $\alpha' = 10\%$. The red dotted lines represent the violations of the VaR(10%) of the series 2 (Dow Jones), the green dotted lines represent the violations of the CoVaR(5%, 10%) of GS returns and the purple dotted lines represent the joint violations of both VaR for series 2 and CoVaR for series 1. Figure 6 illustrates the -VaR and -CoVaR of series 1 conditional on series 2 which is represented by Figure 7 - which shows the log-returns of the series 2 vs. its -VaR(10%) (the same as the one used to calculate the CoVaR in Figure 6). We do not observe systematic correlations between CoVaR violations for series 1 and VaR violations for series 2 which is consistent with the design of the model. Finally, the CoVaR is well below the VaR in Figure 6 which implies positive co-movements between the two series as we have seen in Figure 3. Indeed the empirical correlation correlation between the two series is 0.78.

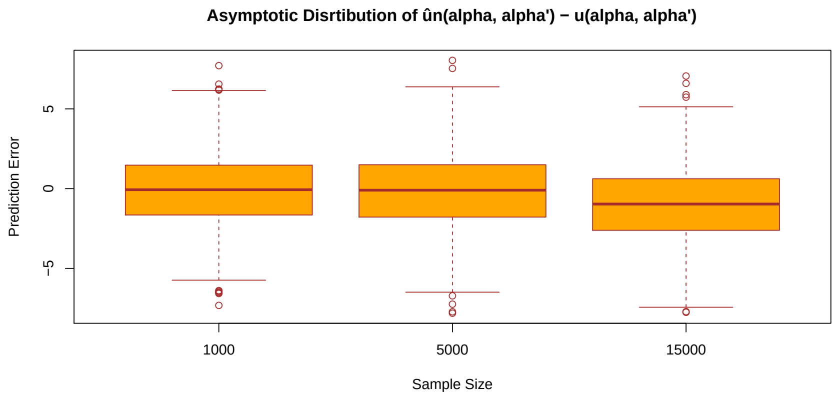

Figure 5. Empirical distribution of $\sqrt{n}\,\{\widehat{u}_n(\alpha,\alpha') - u(\alpha,\alpha')\}$ over 1000 independent replications.

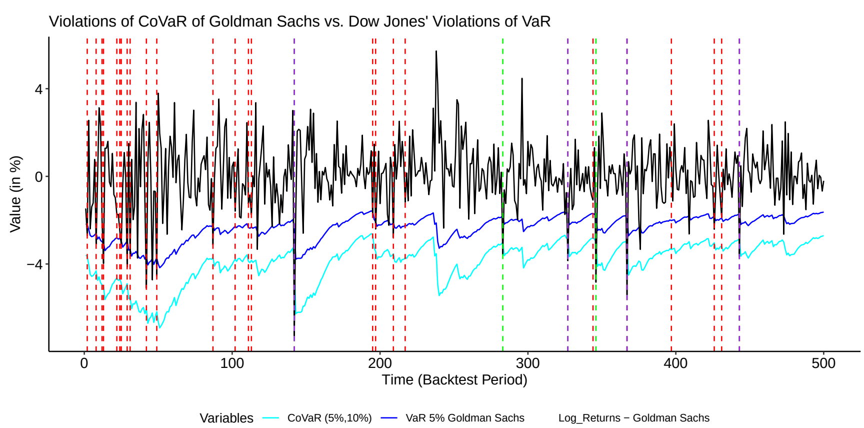

Figure 6.

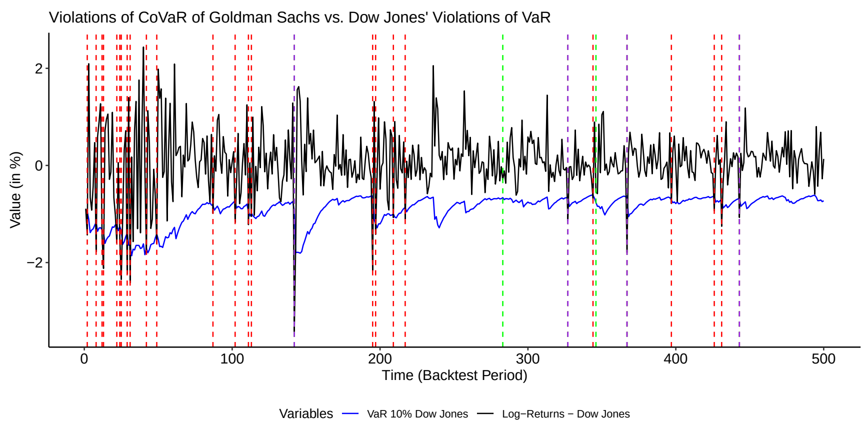

Figure 7.

8. Concluding remarks

In this paper we proposed an econometric approach for estimating the dynamic CoVaR and ΔCoVaR between financial entities in a semi-parametric GARCH-type framework. In this setting, the dynamic CoVaR and ΔCoVaR take the form of a product of the volatility of one asset times a characteristic of the joint distribution of the innovations. The derivation of the asymptotic distribution of the QML estimator of the latter quantity was achieved under general assumptions on the volatility processes. It allows for the construction of asymptotic confidence intervals for the dynamic CoVaR, characterizing the *estimation risk* often neglected in empirical studies. The validity of the approach could be assessed by introducing backtests, generalizing the tests introduced for the VaR and other risk measures (see for instance Jorion (2007)). This is left for further work.

Appendix A: Computation of f_2 in the Gaussian case

Let $(\eta_{1t}, \eta_{2t})' \sim \mathcal{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right)\right)$. Let ϕ and Φ denote, respectively, the density and the cdf of the standard Gaussian distribution. We have

$$\begin{aligned} F(y \mid x) &= \frac{H(x,y)}{\Phi(x)} = \frac{1}{\Phi(x)} \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_{\eta_2 \mid \eta_1 = u}(v) dv \right) \phi(u) du \\ &= \frac{1}{\Phi(x)} \int_{-\infty}^{x} \left(\int_{-\infty}^{y} \frac{1}{\sqrt{1-\rho^2}} \phi\left(\frac{v - \rho u}{\sqrt{1-\rho^2}} \right) dv \right) \phi(u) du. \end{aligned}$$

It follows that

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

Appendix B: Proofs for Sections 2 and 3

Proof of Proposition 2.1. This 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 $\boldsymbol{\epsilon}_t$. □

Proof of Theorem 3.1. Recall that $\widehat{H}_n(x,y) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{1t} \le x, \widehat{\eta}_{2t} \le y}$ and $\widehat{G}_n^{(2)}(x) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\widehat{\eta}_{2t} \le x}$. Let $H_n(x,y) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\eta_{1t} \le x, \eta_{2t} \le y}$, $G_n^{(2)}(x) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{\eta_{2t} \le x}$ and $F_n(x|y) = H_n(x,y)/G_n^{(2)}(y)$.

We have

$$|\widehat{F}_n(x|y) - F_n(x|y)| \le \frac{G_n^{(2)}(y)|\widehat{H}_n(x,y) - H_n(x,y)| + H_n(x,y)|G_n^{(2)}(y) - \widehat{G}_n^{(2)}(y)|}{G_n^{(2)}(y)\widehat{G}_n^{(2)}(y)}. \tag{B.1}$$

Given that $|G_n^{(2)}(y) - \widehat{G}_n^{(2)}(y)| \to 0$ 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 $|\widehat{H}_n(x,y) - H_n(x,y)| \to 0$ a.s. for all x, y. The result is straightforward because

$$|\mathbb{1}_{\widehat{\eta}_{1t} \le x}\mathbb{1}_{\widehat{\eta}_{2t} \le y} - \mathbb{1}_{\eta_{1t} \le x}\mathbb{1}_{\eta_{2t} \le y}| \quad \le \quad |\mathbb{1}_{\widehat{\eta}_{1t} \le x} - \mathbb{1}_{\eta_{1t} \le x}| + |\mathbb{1}_{\widehat{\eta}_{2t} \le y} - \mathbb{1}_{\eta_{2t} \le y}|$$

entails

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

which goes to 0 a.s. by the proof of Theorem 2.1 in FZ. Result *ii)* is also a consequence of this theorem. Result *iii)* follows from the fact that the denominator in (B.1) is bounded away from 0 on $V(y_0)$. □

Proof of Theorem 3.2. Let us prove (3.1). By definition of $u(\alpha, \alpha')$ and $u^+(\alpha, \alpha')$ we have, for any $\epsilon > 0$,

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

for some $\delta > 0$. By Theorem 3.1 *iii)*, assume n large enough so that $\sup_{x\in\mathbb{R},y\in V(\xi^{(2)}_{\alpha'})}|\widehat{F}_n(x\mid y)-F(x\mid y)|<\delta/2$ a.s. By Corollary 4.1 in FZ, we know that under $\mathbf{A6_1}(\alpha')$, $\widehat{\xi}^{(2)}_{n,\alpha'}\to\xi^{(2)}_{\alpha'}$ a.s. It thus follows that $\sup_{x\in\mathbb{R}}|F(x|\widehat{\xi}^{(2)}_{n,\alpha'})-F(x|\xi^{(2)}_{\alpha'})|<\delta/2$ a.s. for n large enough, in view of $\mathbf{A7}(\xi^{(2)}_{\alpha'})$. Now we will show that (B.2) entails

$$u(\alpha,\alpha')-\epsilon\leq\widehat{u}_n(\alpha,\alpha')\leq u^+(\alpha,\alpha')+\epsilon \tag{B.3}$$

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

$$\begin{aligned}F(u(\alpha,\alpha')-\epsilon\mid\xi^{(2)}_{\alpha'}) &\geq F(\widehat{u}_n(\alpha,\alpha')\mid\xi^{(2)}_{\alpha'})\\ &= \underbrace{F(\widehat{u}_n(\alpha,\alpha')\mid\xi^{(2)}_{\alpha'})-F(\widehat{u}_n(\alpha,\alpha')\mid\widehat{\xi}^{(2)}_{n,\alpha'})}_{>-\delta/2,\ a.s.\ \text{by } \mathbf{A6_1}(\alpha') \text{ and } \mathbf{A7}(\xi^{(2)}_{\alpha'})}\\ &\quad+\underbrace{F(\widehat{u}_n(\alpha,\alpha')\mid\widehat{\xi}^{(2)}_{n,\alpha'})-\widehat{F}_n(\widehat{u}_n(\alpha,\alpha')\mid\widehat{\xi}^{(2)}_{n,\alpha'})}_{>-\delta/2,\ a.s.\ \text{by } \mathbf{A6_1}(\alpha') \text{ and iii) of Th. 3.1}}\\ &\quad+\underbrace{\widehat{F}_n(\widehat{u}_n(\alpha,\alpha')\mid\widehat{\xi}^{(2)}_{n,\alpha'})}_{\geq\alpha}\geq\alpha-\delta,\end{aligned}$$

which contradicts the first inequality in (B.5). Moreover, if $u^+(\alpha,\alpha')+\epsilon<\widehat{u}_n(\alpha,\alpha')$ then by the same arguments

$$F(u^+(\alpha,\alpha')+\epsilon\mid\xi^{(2)}_{\alpha'})\leq\widehat{F}_n(u^+(\alpha,\alpha')+\epsilon\mid\xi^{(2)}_{\alpha'})+\delta/2\leq\alpha+\delta$$

which contradicts the second inequality in (B.2). Hence (B.3) is shown. The strong convergence of $\widehat{u}_n(\alpha,\alpha')$ to the set $[u(\alpha,\alpha'),u^+(\alpha,\alpha')]$ follows from (B.3). Thus (3.1) is established. Now, if $\mathbf{A6_2}$ holds, the previous set reduces to the singleton $\{u(\alpha,\alpha')\}$. The conclusion follows. □

Proof of Theorem 3.3. Note that

$$F^{\Delta}(x|(y_1,y_2])=\frac{F(x\mid y_2)G^{(2)}(y_2)-F(x\mid y_1)G^{(2)}(y_1)}{G^{(2)}(y_2)-G^{(2)}(y_1)}. \tag{B.4}$$

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

$$F^{\Delta}(\underline{u}(\alpha,\alpha'')-\epsilon\mid A_{\alpha''})<\alpha-\delta\quad\text{and}\quad F^{\Delta}(\underline{u}^+(\alpha,\alpha'')+\epsilon\mid A_{\alpha''})>\alpha+\delta, \tag{B.5}$$

for some $\delta>0$. By arguments already used, we have under $\mathbf{A6_1}(0.5+\tau\alpha'')$, $\widehat{\xi}^{(2)}_{n,0.5+\tau\alpha''}\to\xi^{(2)}_{0.5+\tau\alpha'}$ a.s. for $\tau\in\{-1,1\}$. It is clear that, in view of (B.4) 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 $P(\eta_2\in A)>0$ and n large enough,

$$\widehat{F}^{\Delta}_n(x|A)=\frac{\frac{1}{n}\sum_{t=1}^n\mathbb{1}_{\widehat{\eta}_{1t}\leq x,\widehat{\eta}_{2t}\in A}}{\frac{1}{n}\sum_{t=1}^n\mathbb{1}_{\widehat{\eta}_{2t}\in A}}.$$

It follows from (B.4), $\mathbf{A5_1}$ and $\mathbf{A5_2}(\xi^{(2)}_{0.5+\tau\alpha''})$ that, from Theorem 3.1,

$$|F^{\Delta}(\widehat{\underline{u}}_n(\alpha,\alpha'') \mid \widehat{A}_{n,\alpha''}) - \widehat{F}^{\Delta}_n(\widehat{\underline{u}}_n(\alpha,\alpha'') \mid \widehat{A}_{n,\alpha''})| \to 0 \quad a.s.$$

We will show that (B.5) entails

$$\underline{u}(\alpha,\alpha'') - \epsilon \leq \widehat{\underline{u}}_n(\alpha,\alpha'') \leq \underline{u}^{+}(\alpha,\alpha'') + \epsilon \tag{B.6}$$

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

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

which contradicts the first inequality in (B.5). The rest of the proof is similar to that of Theorem 3.2. □

Appendix C: Proof of Theorem 4.1

First note that

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

Write

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

Let $\chi^{(i)}_{t,n} = \sigma_t(\widehat{\boldsymbol{\theta}}^{(i)}_n)/\sigma_t(\boldsymbol{\theta}^{(i)}_0)$,

$$\begin{aligned}
&\widehat{e}_n(x,y) = \sqrt{n}\{\widehat{H}_n(x,y) - H(x,y)\}, \quad e_n(x,y) = \sqrt{n}\{H_n(x,y) - H(x,y)\}, \\
&h^{(1)}_n(x,y) = x f_1(x \mid y) G^{(2)}(y) \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{D}'_{1t}\right) \sqrt{n}(\widehat{\boldsymbol{\theta}}^{(1)}_n - \boldsymbol{\theta}^{(1)}_0),
\end{aligned}$$

$$h_n^{(2)}(x,y) = yf_2(y \mid x)G^{(1)}(x)\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{D}'_{2t}\right)\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{(2)} - \boldsymbol{\theta}_0^{(2)}).$$

We have

$$\begin{aligned}\widehat{e}_n(x,y) &= \underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{1}_{\eta_{1t}\leq x\chi_{t,n}^{(1)},\,\eta_{2t}\leq y\chi_{t,n}^{(2)}} - H\left(x\chi_{t,n}^{(1)},\, y\chi_{t,n}^{(2)}\right)}_{\widehat{e}_{n,1}(x,y)}\\ &+ \underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}H\left(x\chi_{t,n}^{(1)},\, y\chi_{t,n}^{(2)}\right) - H(x,y)}_{\widehat{e}_{n,2}(x,y)}\\ &+ \underbrace{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{1}_{\eta_{1t}\leq\widetilde{\chi}_{t,n}^{(1)}x,\,\eta_{2t}\leq\widetilde{\chi}_{t,n}^{(2)}y} - \mathbb{1}_{\eta_{1t}\leq x\chi_{t,n}^{(1)},\,\eta_{2t}\leq y\chi_{t,n}^{(2)}}}_{\widehat{e}_{n,3}(x,y)}.\end{aligned}$$

We will first show that

$$\widehat{e}_{n,1}(x_n,y_n) = e_n(x,y) + o_P(1), \qquad \text{(C.1)}$$

$$\widehat{e}_{n,2}(x_n,y_n) = h_n^{(1)}(x,y) + h_n^{(2)}(x,y) + o_P(1), \qquad \text{(C.2)}$$

$$\widehat{e}_{n,3}(x_n,y_n) = o_P(1). \qquad \text{(C.3)}$$

The last three results, (4.1), **B6** and the Glivenko-Cantelli result in Theorem 3.1 entail

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

Using Theorem 2.2 in Francq and Zakoian (2022)

$$b_n(x_n,y_n) = \frac{-F(x \mid y)}{G^{(2)}(y)}\left\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{2t}<y} - G^{(2)}(y)\} + \frac{yf^{(2)}(y)}{2\sqrt{n}}\boldsymbol{\Omega}'_2\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}\right\} + o_P(1).$$

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

C.1. Proof of (C.1)

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 $\boldsymbol{a} = (\boldsymbol{a}^{(1)'}, \boldsymbol{a}^{(2)'})'$, let

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

$$\gamma_{t,n}^{(i)}(\boldsymbol{a}^{(i)}) = \frac{\sigma_{it}(\boldsymbol{\theta}_0^{(i)} + \frac{\boldsymbol{a}^{(i)}}{\sqrt{n}})}{\sigma_{it}(\boldsymbol{\theta}_0^{(i)})}.$$

Note that $\widehat{e}_{n,1}(x, y) = e_{n,1}(x, y, \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))$ and $e_{n,1}(x, y, \mathbf{0}) = e_n(x, y)$. Write

$$e_{n,1}(x, y, \boldsymbol{a}) - e_n(x, y) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_{t,n}(x, y, \boldsymbol{a}), \tag{C.4}$$

where

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

We will establish a number of auxiliary lemmas.

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

$$P\left(\left|\frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_{t,n}(x, y, \boldsymbol{a})\right| > u\right) \le \frac{K}{nu^4}\{(x^2 + y^2)\|\boldsymbol{a}\|^2 + 1\}.$$

Proof. By the proof of Lemma 6.2 in FZ, it suffices to show that

$$E\left(\sum_{t=1}^{n} E(z_{t,n}^2(x, \boldsymbol{a})|\mathcal{F}_{t-1})\right)^2 \le nK(x^2 + y^2)\|\boldsymbol{a}\|^2. \tag{C.5}$$

Noting that the second-order conditional moment of $z_{t,n}(x, y, \boldsymbol{a})$ is the variance of a Bernoulli distribution, we have, using **B6**,

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

where $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_t^{(1)'}, \boldsymbol{\theta}_t^{(2)'})'$ is between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0 + \boldsymbol{a}/\sqrt{n}$. It follows that

$$E\left\{\sum_{t=1}^n E[z_{t,n}^2(x,y,\boldsymbol{a})|\mathcal{F}_{t-1}]\right\}^2 \leq \frac{Kx^2}{n}\sum_{s,t=1}^n E\left\|\frac{1}{\sigma_{1s}(\boldsymbol{\theta}_0^{(1)})}\frac{\partial\sigma_{1s}(\boldsymbol{\theta}_s^*)}{\partial\boldsymbol{\theta}^{(1)}}\right\|\left\|\frac{1}{\sigma_{1t}(\boldsymbol{\theta}_0^{(1)})}\frac{\partial\sigma_{1t}(\boldsymbol{\theta}_t^*)}{\partial\boldsymbol{\theta}^{(1)}}\right\|\|\boldsymbol{a}^{(1)}\|^2 + \frac{Ky^2}{n}\sum_{s,t=1}^n E\left\|\frac{1}{\sigma_{2s}(\boldsymbol{\theta}_0^{(2)})}\frac{\partial\sigma_{2s}(\boldsymbol{\theta}_s^*)}{\partial\boldsymbol{\theta}^{(2)}}\right\|\left\|\frac{1}{\sigma_{2t}(\boldsymbol{\theta}_0^{(2)})}\frac{\partial\sigma_{2t}(\boldsymbol{\theta}_t^*)}{\partial\boldsymbol{\theta}^{(2)}}\right\|\|\boldsymbol{a}^{(2)}\|^2$$

and thus (C.5) holds. □

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

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

$$P\left(\max_{-N\leq i,j\leq N}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n z_{t,n}(x_i,y_j,\boldsymbol{a})\right| > u\right) \leq \frac{K}{\sqrt{n}}. \tag{C.6}$$

As in FZ, it therefore suffices to show that

$$\limsup_{n\to\infty} P\left\{\max_{-N\leq i,j\leq N-1}\delta(i,j,\boldsymbol{a}) > u\right\} = 0, \tag{C.7}$$

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 $i,j = 0,\dots,N-1$,

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

$$- \left\{ \mathbb{1}_{\eta_{1t} \le x_i \gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), \eta_{2t} \le y_j \gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})} - H(x_i \gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)}), y_j \gamma_{t,n}^{(2)}(\boldsymbol{a}^{(2)})) \right\} \Big|$$
$$+ 2W_n(i,j,\boldsymbol{a}),$$

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,

$$\delta(i,j,\boldsymbol{a}) \quad \le \quad \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{n} z_{t,n}(x_{i+1}, y_{j+1}, \boldsymbol{a}) \right| + \frac{1}{\sqrt{n}} \left| \sum_{t=1}^{n} z_{t,n}(x_i, y_j, \boldsymbol{a}) \right| + V_n(i,j) + 2W_n(i,j,\boldsymbol{a}), \tag{C.8}$$

where

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

By Assumption **B6** and the mean-value theorem, $H(x_{i+1}, y_{j+1}) - H(x_i, y_j) \le M\varepsilon/\sqrt{n}$ where $M = \sup_{x \in \mathbb{R}} f(x)$. Thus $W_n(i,j,\mathbf{0}) \le M\epsilon$ and

$$\delta(i,j,\mathbf{0}) \le V_n(i,j) + 2M\epsilon. \tag{C.9}$$

Therefore, from (C.8)-(C.9),

$$\max_{0 \le i,j \le N-1} \sup_{x \in [x_i, x_{i+1}], y \in [y_j, y_{j+1}]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_{t,n}(x,y,\boldsymbol{a}) - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_{t,n}(x_i, y_j, \boldsymbol{a}) \right|$$
$$\le \max_{0 \le i,j \le N} \frac{2}{\sqrt{n}} \left| \sum_{t=1}^{n} z_{t,n}(x_i, y_j, \boldsymbol{a}) \right| + 2 \max_{0 \le i,j \le N-1} W_n(i,j,\boldsymbol{a}) + 2 \max_{0 \le i,j \le N-1} V_n(i,j) + 2M\varepsilon. \tag{C.10}$$

By the properties of the modulus of continuity of the empirical process (see Shorack and Wellner (1986), p. 542), under Assumption **B6** we have

$$\max_{0 \le i,j \le N-1} V_n(i,j) = o_P(1). \tag{C.11}$$

By the arguments used in FZ to complete the proof of Lemma 6.3, we also have

$$\max_{0 \le i,j \le N-1} W_n(i,j,\boldsymbol{a}) \quad = \quad \varepsilon \times O_P(1). \tag{C.12}$$

Thus (C.7) is established and the conclusion follows. □

Lemma C.3. *Let $\mathcal{K}$ be a compact subset of $\mathbb{R}$. For any $A>0$ and $\boldsymbol{A}=[-A,A]^d$,*

$$\sup_{x,y\in\mathcal{K}}\sup_{\boldsymbol{a}\in\boldsymbol{A}}|e_{n,1}(x,y,\boldsymbol{a})-e_n(x,y)|=o_P(1).$$

Proof. In view of (C.4) will show that

$$\sup_{\boldsymbol{a}\in\boldsymbol{A}}X_n(\boldsymbol{a})=o_P(1),\quad\text{where}\quad X_n(\boldsymbol{a})=\sup_{x,y\in\mathcal{K}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n z_{t,n}(x,y,\boldsymbol{a})\right|.\tag{C.13}$$

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

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

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 $E|\eta_t|<\infty$, we have $\sup_x|x|f(x)<\infty$. 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 $H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^1),y\gamma_{t,n}^{(2)}(a_{d_1+1},\dots,a_{j-1},a(k_j),a_{j+1},\dots,a_d)\right)-H\left(x\gamma_{t,n}^{(1)}(\boldsymbol{a}^{(1)})\right)$ where $j=d_{1+1},\dots,d$. Therefore, for n large enough,

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

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

$$\sup_{\boldsymbol{a}\in\boldsymbol{A}(\boldsymbol{k})}\sup_{x,y\in\mathcal{K}}\left|\sum_{t=1}^n z_{t,n}(x,y,\boldsymbol{a})-z_{t,n}(x,y,\boldsymbol{a}(\boldsymbol{k}))\right|$$

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

Note that

$$\begin{aligned}
&\sup_{\boldsymbol{a}\in\boldsymbol{A}}X_n(\boldsymbol{a}) \\
&\leq \max_{\boldsymbol{k}\in\{1,\dots,N\}^d}\sup_{\boldsymbol{a}\in\boldsymbol{A}(\boldsymbol{k})}\sup_{x\in\mathcal{K}}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[z_{t,n}(x,y,\boldsymbol{a})-z_{t,n}(x,y,\boldsymbol{a}(\boldsymbol{k}))]\right|+\max_{\boldsymbol{k}\in\{1,\dots,N\}^d}X_n(\boldsymbol{a}(\boldsymbol{k})) \\
&\leq \frac{2K\varepsilon}{n}\sum_{t=1}^{n}\max_{i=1,2}\sup_{\boldsymbol{\theta}^{(i)}\in V(\boldsymbol{\theta}_0^{(i)})}\left\|\frac{1}{\sigma_{it}(\boldsymbol{\theta}^{(i)})}\frac{\partial\sigma_{it}(\boldsymbol{\theta}^{(i)})}{\partial\boldsymbol{\theta}^{(i)}}\right\| \\
&\quad+2\max_{\boldsymbol{k}\in\{1,\dots,N\}^d}X_n(\boldsymbol{a}(\boldsymbol{k}))+\max_{\boldsymbol{k}\in\{1,\dots,N\}^d}X_n(\boldsymbol{a}^*(\boldsymbol{k})).
\end{aligned}$$

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 $X_n(\boldsymbol{a})=o_P(1)$ by Lemma C.2 and the maximas are over a finite number of points. Therefore (C.13) is established. □

Lemma C.4. *Let (x_n,y_n) be a sequence of real random vectors 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 $e_n(x_n,y_n)-e_n(x_0,y_0)=o_P(1)$.*

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

$$\begin{aligned}
e_n(x,y) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{1}_{U_t\leq G^{(1)}(x),V_t\leq G^{(2)}(y)}-P(U_t\leq G^{(1)}(x),V_t\leq G^{(2)}(y)) \\
&= Y_n(G^{(1)}(x),G^{(2)}(x)),
\end{aligned}$$

where

$$Y_n(u,v)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{1}_{U_t\leq u,V_t\leq v}-P(U_t\leq u,V_t\leq v).$$

Billingsley (1968) studied the modulus of continuity of $\{Z_n(u), u \in [0,1]\}$ where $Z_n(u) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{1}_{U_t \leq u} - u$ 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,

$$P\left(\sup_{|u-v|<\tau} |Z_n(u) - Z_n(v)| \geq \varepsilon\right) \leq \eta$$

We will extend this inequality in (C.17) below to the sequence $\{Y_n(u,v), u, v \in [0,1]\}$.

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

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

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

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

If, for $\varepsilon \in (0,1)$,

$$\frac{\varepsilon}{n} \leq |u-u^*| \wedge |v-v^*|$$

we then have

$$E\left|Y_n(u,v) - Y_n(u^*,v^*)\right|^4 \leq \frac{4}{\varepsilon}|u-u^*|^2 \wedge |v-v^*|^2 \leq \frac{2}{\varepsilon}\left(|u-u^*|^2 + |v-v^*|^2\right). \quad \text{(C.14)}$$

Let ι a number such that $\varepsilon/n \leq \iota$ and m a positive integer such that $u + m\iota \leq 1$ and $v + m\iota \leq 1$. 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

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

In view of (C.14), for $0 \leq i \leq i^* \leq m$ and $1 \leq j, j^* \leq m+1$ such that $i(m+1) + j \leq i^*(m+1) + j^*$, we have

$$E\left|S_{i(m+1)+j} - S_{i^*(m+1)+j^*}\right|^4 \leq \frac{2\iota^2}{\varepsilon}\left\{(i^*-i)^2 + |j^*-j|^2\right\}.$$

This is of the form of inequality (12.42) in Billingsley (1968), with $\gamma = 4$, $\alpha = 2$,

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

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

$$P\left(\max_{0\le i,j\le m} |Y_n(u,v) - Y_n(u+i\iota, v+j\iota)| \ge \tau\right) \le \frac{8}{\varepsilon\tau^4} m^4 \iota^2. \tag{C.15}$$

Now, for $\underline{u} \le \widetilde{u} \le \underline{u} + \iota$ and $\underline{v} \le \widetilde{v} \le \underline{v} + \iota$, we have

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

Note that $P(U_t \le \underline{u}+\iota, V_t \le \underline{v}+\iota) - P(U_t \le \underline{u}, V_t \le \underline{v}) \le P(U_t \in (\underline{u}, \underline{u}+\iota]) + P(V_t \in (\underline{v}, \underline{v}+\iota]) \le 2\iota$. We thus have

$$Y_n(\widetilde{u}, \widetilde{v}) - Y_n(\underline{u}, \underline{v}) \le Y_n(\underline{u}+\iota, \underline{v}+\iota) - Y_n(\underline{u}, \underline{v}) + 2\iota\sqrt{n}.$$

We also have

$$Y_n(\widetilde{u}, \widetilde{v}) - Y_n(\underline{u}, \underline{v}) \ge -P(U_t \le \underline{u}+\iota, V_t \le \underline{v}+\iota) + P(U_t \le \underline{u}, V_t \le \underline{v}) \ge -2\iota\sqrt{n}.$$

Therefore

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

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

$$\begin{aligned} |Y_n(u,v) - Y_n(u^*,v^*)| &\le |Y_n(u,v) - Y_n(\underline{u}, \underline{v})| + |Y_n(u^*,v^*) - Y_n(\underline{u}, \underline{v})| \\ &\le |Y_n(u,v) - Y_n(\underline{u}, \underline{v})| + |Y_n(\underline{u}+\iota, \underline{v}+\iota) - Y_n(\underline{u}, \underline{v})| + 2\iota\sqrt{n} \\ &\le 3 \max_{0\le i,j\le m} |Y_n(u,v) - Y_n(u+i\iota, v+j\iota)| + 2\iota\sqrt{n}. \end{aligned}$$

Taking $\varepsilon/n < \iota < \varepsilon/\sqrt{n}$, we obtain from (C.15) and the previous inequality

$$P\left(\sup_{u\le u^*\le u+m\iota, v\le v^*\le v+m\iota} |Y_n(u,v) - Y_n(u^*,v^*)| \ge 5\varepsilon\right) \le \frac{8}{\varepsilon^5} m^4 \iota^2. \tag{C.16}$$

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

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

$$P\left(\sup_{|u-u^*|<\tau, |v-v^*|<\tau} |Y_n(u,v) - Y_n(u^*,v^*)| \ge \varepsilon\right) \le \eta \tag{C.17}$$

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

$$P\left(|e_n(x_n, y_n) - e_n(x_0, y_0)| \ge \varepsilon\right)$$

$$\leq P\left(\sup_{|x^*-x_0|\leq\delta,|y^*-y_0|\leq\delta}|e_n(x^*,y^*)-e_n(x_0,y_0)|\geq\varepsilon\right)+P\left(|x_n-x_0|\geq\delta\right)$$
$$+P\left(|y_n-y_0|\geq\delta\right).$$

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

$$P\left(\sup_{|x^*-x_0|\leq\delta,|y^*-y_0|\leq\delta}|e_n(x^*,y^*)-e_n(x_0,y_0)|\geq\varepsilon\right)$$
$$\leq P\left(\sup_{|u-u^*|\leq\delta\sup_x f_1(x),|v-v^*|\leq\delta\sup_x f_2(x)}|Y_n(u,v)-Y_n(u^*,v^*)|\geq\varepsilon\right)\leq\eta$$

when n is large enough and δ small enough to satisfy (C.17) 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 $e_n(x_n,y_n)-e_n(x_0,y_0)=o_P(1)$. □

The proof of (C.1) is straightforwardly deduced from the previous lemmas and **B6**.

C.2. Proof of (C.2)

Lemma C.5. *Let $\mathcal{K}$ be a compact subset of $\mathbb{R}$. Then $\sup_{x,y\in\mathcal{K}}|\widehat{e}_{n,2}(x,y)-h_n^{(1)}(x,y)-h_n^{(2)}(x,y)|\to 0$ 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 $\widehat{\boldsymbol{\theta}}_n^{(i)}$ and $\boldsymbol{\theta}_0^{(i)}$,

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

The rest of the proof relies on the arguments given in FZ. □

We obtain (C.2) from the previous lemma and **B6**.

C.3. Proof of (C.3)

By

$$\widehat{e}_{n,3}(x,y)\ =\ \frac{1}{\sqrt{n}}\sum_{t=1}^n(\mathbb{1}_{\eta_{1t}\leq\widetilde{\chi}_{t,n}^{(1)}x}-\mathbb{1}_{\eta_{1t}\leq\chi_{t,n}^{(1)}x})\mathbb{1}_{\eta_{2t}\leq\widetilde{\chi}_{t,n}^{(2)}y}+(\mathbb{1}_{\eta_{2t}\leq\widetilde{\chi}_{t,n}^{(2)}y}-\mathbb{1}_{\eta_{2t}\leq\chi_{t,n}^{(2)}y})\mathbb{1}_{\eta_{1t}\leq\chi_{t,n}^{(1)}x},$$

we deduce

$$\begin{aligned}
&\sup_{x,y\in\mathbb{R}} |\widehat{e}_{n,3}(x,y)| \\
\leq\ & \sup_{x\in\mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |\mathbb{1}_{\eta_{1t}\leq\widetilde{\chi}^{(1)}_{t,n}x} - \mathbb{1}_{\eta_{1t}\leq\chi^{(1)}_{t,n}x}| + \sup_{y\in\mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |\mathbb{1}_{\eta_{2t}\leq\widetilde{\chi}^{(2)}_{t,n}y} - \mathbb{1}_{\eta_{2t}\leq\chi^{(2)}_{t,n}y}| = o_P(1),
\end{aligned}$$

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

C.4. Simplification of the asymptotic variance 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

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

Therefore $\boldsymbol{\Omega}_1'\boldsymbol{J}_1^{-1}\boldsymbol{J}_{12}\boldsymbol{J}_2^{-1}\boldsymbol{\Omega}_2 = 1$. □

Appendix D: Proof of Theorem 4.2.

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

$$\begin{aligned}
\sqrt{n}\left\{\alpha - F\left(\widehat{u}_n \mid \xi^{(2)}_{\alpha'}\right)\right\} &= \sqrt{n}\left\{\widehat{F}_n\left(\widehat{u}_n \mid \widehat{\xi}^{(2)}_{n,\alpha'}\right) - F\left(\widehat{u}_n \mid \widehat{\xi}^{(2)}_{n,\alpha'}\right)\right\} \\
&+ \sqrt{n}\left\{F\left(\widehat{u}_n \mid \widehat{\xi}^{(2)}_{n,\alpha'}\right) - F\left(\widehat{u}_n \mid \xi^{(2)}_{\alpha'}\right)\right\} + o_P(1). \qquad \text{(D.1)}
\end{aligned}$$

Now note that **B6** and **B8** entail **A5** and **A6-A7**. Theorem 3.2 thus entails that $\widehat{u}_n(\alpha, \alpha')$ strongly converges to $u(\alpha, \alpha')$, and Corollary 4.1 in FZ entails that $\widehat{\xi}^{(2)}_{n,\alpha'}$ 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

$$\sqrt{n}\left(F(\widehat{u}_n \mid \widehat{\xi}^{(2)}_{n,\alpha'}) - F(u(\alpha, \alpha') \mid \xi^{(2)}_{\alpha'})\right)$$

$$=\nu_x\sqrt{n}\left\{\widehat{u}_n-u(\alpha,\alpha')\right\}+\nu_y\sqrt{n}\left(\widehat{\xi}^{(2)}_{n,\alpha'}-\xi^{(2)}_{\alpha'}\right)+o_P(1)$$

and

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

Therefore we have

$$\sqrt{n}\left(F(\widehat{u}_n\mid\widehat{\xi}^{(2)}_{n,\alpha'})-F(\widehat{u}_n\mid\xi^{(2)}_{\alpha'})\right)=\nu_y\sqrt{n}\left(\widehat{\xi}^{(2)}_{n,\alpha'}-\xi^{(2)}_{\alpha'}\right)+o_P(1). \tag{D.2}$$

By Theorem 4.1 and Theorem 2.2 in FZ, noting that $G^{(2)}(\xi^{(2)}_{\alpha'})=\alpha'$ and using (D.1)-(D.2), we have the Bahadur expansion

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

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. The rest of the proof easily follows. □

Appendix E: Proof of Theorem 4.3.

Let

$$\widehat{F}_n^{\Delta}(x|[y_1,y_2))=\frac{\widehat{F}_n(x\mid y_2)\widehat{G}_n^{(2)}(y_2)-\widehat{F}_n(x\mid y_1)\widehat{G}_n^{(2)}(y_1)}{\widehat{G}_n^{(2)}(y_2)-\widehat{G}_n^{(2)}(y_1)}. \tag{E.1}$$

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

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

We also have, for $y_n\to y$,

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

It follows that

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

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

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

Note that $a(x,y_1,y_2)=0$ when η_1 and η_2 are independent. It follows that

$$\begin{aligned}
&\sqrt{n}\{\widehat{F}_n^{\Delta}(x_n|[y_{1n},y_{2n}))-F^{\Delta}(x_n|[y_{1n},y_{2n}))\}\\
=&\frac{1}{\sqrt{n}\Delta G^{(2)}(y_1,y_2)}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}<x}-F^{\Delta}(x|[y_1,y_2))\}\mathbb{1}_{\eta_{2t}\in(y_1,y_2]}
\end{aligned}$$

$$+x\frac{G^{(2)}(y_2)f_1(x \mid y_2) - G^{(2)}(y_1)f_1(x \mid y_1)}{2\sqrt{n}\Delta G^{(2)}(y_1, y_2)}\boldsymbol{\Omega}_1' \boldsymbol{J}_1^{-1}\sum_{t=1}^{n}(\eta_{1t}^2 - 1)\boldsymbol{D}_{1t}$$
$$+\frac{a(x, y_1, y_2)}{2\sqrt{n}}\boldsymbol{\Omega}_2' \boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2 - 1)\boldsymbol{D}_{2t} + o_P(1).$$

Note that we retrieve the expansion of Theorem 4.2 when $y_1 \to -\infty$ (provided $y_1 g^{(2)}(y_1) \to 0$).

Proceeding as in the proof of Theorem 4.2, we note that $\widehat{F}_n^{\Delta}(\underline{\widehat{u}}_n \mid \widehat{A}_{n,\alpha''}^{(2)}) - \alpha \leq 1/n\alpha''$ and

$$\sqrt{n}\left\{\alpha - F^{\Delta}\left(\underline{\widehat{u}}_n \mid A_{\alpha''}^{(2)}\right)\right\} = \sqrt{n}\left\{\widehat{F}_n^{\Delta}\left(\underline{\widehat{u}}_n \mid \widehat{A}_{n,\alpha''}^{(2)}\right) - F^{\Delta}\left(\underline{\widehat{u}}_n \mid \widehat{A}_{n,\alpha''}^{(2)}\right)\right\}$$
$$+ \sqrt{n}\left\{F^{\Delta}\left(\underline{\widehat{u}}_n \mid \widehat{A}_{n,\alpha''}^{(2)}\right) - F^{\Delta}\left(\underline{\widehat{u}}_n \mid A_{\alpha''}^{(2)}\right)\right\} + o_P(1). \tag{E.2}$$

Let

$$\lambda_x = \lambda_x(\alpha, \alpha'') = \frac{\partial}{\partial x}F^{\Delta}(x \mid (y_1, y_2])\Big|_{(x,y_1,y_2)=\left(\underline{u},\xi_{0.5-\alpha''}^{(2)},\xi_{\alpha''+0.5}^{(2)}\right)} = f_1\left(\underline{u} \mid A_{\alpha''}^{(2)}\right),$$
$$\lambda_{y_1} = \frac{\partial}{\partial y_1}F^{\Delta}(x \mid (y_1, y_2])\Big|_{(x,y_1,y_2)=\left(\underline{u},\xi_{0.5-\alpha''}^{(2)},\xi_{\alpha''+0.5}^{(2)}\right)}$$
$$= \frac{-f_2(\xi_{0.5-\alpha''}^{(2)} \mid \underline{u})G^{(1)}(\underline{u}) + g^{(2)}(\xi_{0.5-\alpha''}^{(2)})\alpha}{2\alpha''},$$
$$\lambda_{y_2} = \frac{\partial}{\partial y_2}F^{\Delta}(x \mid (y_1, y_2])\Big|_{(x,y_1,y_2)=\left(\underline{u},\xi_{0.5-\alpha''}^{(2)},\xi_{\alpha''+0.5}^{(2)}\right)}$$
$$= \frac{f_2(\xi_{0.5+\alpha''}^{(2)} \mid \underline{u})G^{(1)}(\underline{u}) - g^{(2)}(\xi_{0.5+\alpha''}^{(2)})\alpha}{2\alpha''}.$$

By arguments already given, we thus have

$$\sqrt{n}\left(F^{\Delta}(\underline{\widehat{u}}_n \mid \widehat{A}_{n,\alpha''}^{(2)}) - F^{\Delta}(\underline{u}(\alpha, \alpha'') \mid A_{\alpha''}^{(2)})\right) = \lambda_x\sqrt{n}\left\{\underline{\widehat{u}}_n - \underline{u}(\alpha, \alpha'')\right\}$$
$$+ \lambda_{y_1}\sqrt{n}\left(\widehat{\xi}_{n,0.5-\alpha''}^{(2)} - \xi_{0.5-\alpha''}^{(2)}\right) + \lambda_{y_2}\sqrt{n}\left(\widehat{\xi}_{n,\alpha''+0.5}^{(2)} - \xi_{\alpha''+0.5}^{(2)}\right) + o_P(1)$$

and

$$\sqrt{n}\left(F^{\Delta}\left(\underline{u}(\alpha, \alpha'') \mid A_{\alpha''}^{(2)}\right) - F^{\Delta}\left(\underline{\widehat{u}}_n \mid A_{\alpha''}^{(2)}\right)\right) = \lambda_x\sqrt{n}\left\{\underline{u}(\alpha, \alpha'') - \underline{\widehat{u}}_n\right\} + o_P(1),$$

thus

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

Noting that

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

and that

$$a\left(\underline{u},\xi^{(2)}_{0.5-\alpha''},\xi^{(2)}_{\alpha''+0.5}\right)-\lambda_{y_1}\xi^{(2)}_{0.5-\alpha''}-\lambda_{y_2}\xi^{(2)}_{\alpha''+0.5}=\frac{1-2\alpha''}{(2\alpha'')^2}\{\xi^{(2)}_{0.5-\alpha''}g^{(2)}(\xi^{(2)}_{0.5-\alpha''})-\xi^{(2)}_{0.5+\alpha''}g^{(2)}(\xi^{(2)}_{0.5+\alpha''})\}$$

we deduce, using Corollary 4.2 in FZ, the Bahadur expansion

$$\begin{aligned}
&\sqrt{n}\left\{\alpha-F^{\Delta}\left(\widehat{\underline{u}}_n\mid A^{(2)}_{\alpha''}\right)\right\}=\frac{1}{2\sqrt{n}\alpha''}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}<\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{1-2\alpha''}{2\sqrt{n}(2\alpha'')^2}\{\xi^{(2)}_{0.5-\alpha''}g^{(2)}(\xi^{(2)}_{0.5-\alpha''})-\xi^{(2)}_{0.5+\alpha''}g^{(2)}(\xi^{(2)}_{0.5+\alpha''})\}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}\\
&-\frac{\lambda_{y_1}}{g^{(2)}(\xi^{(2)}_{0.5-\alpha''})}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\mathbb{1}_{\eta_{2t}<\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}<\xi^{(2)}_{\alpha''+0.5}}-0.5-\alpha'')+o_P(1)\\
&=\frac{1}{2\sqrt{n}\alpha''}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}<\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{1-2\alpha''}{2\sqrt{n}(2\alpha'')^2}\{\xi^{(2)}_{0.5-\alpha''}g^{(2)}(\xi^{(2)}_{0.5-\alpha''})-\xi^{(2)}_{0.5+\alpha''}g^{(2)}(\xi^{(2)}_{0.5+\alpha''})\}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}\\
&+\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}<\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}<\xi^{(2)}_{\alpha''+0.5}}-0.5-\alpha'')+o_P(1)\\
&=\frac{1}{2\sqrt{n}\alpha''}\sum_{t=1}^{n}\{\mathbb{1}_{\eta_{1t}<\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{1-2\alpha''}{2\sqrt{n}(2\alpha'')^2}\{\xi^{(2)}_{0.5-\alpha''}g^{(2)}(\xi^{(2)}_{0.5-\alpha''})-\xi^{(2)}_{0.5+\alpha''}g^{(2)}(\xi^{(2)}_{0.5+\alpha''})\}\boldsymbol{\Omega}_2'\boldsymbol{J}_2^{-1}\sum_{t=1}^{n}(\eta_{2t}^2-1)\boldsymbol{D}_{2t}\\
&+\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}<\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}<\xi^{(2)}_{\alpha''+0.5}}-0.5-\alpha'')+o_P(1).
\end{aligned}$$

Similarly to the proof of Theorem 4.2, we conclude by applying the delta method to the latter expansion, using the inverse of the function $F^{\Delta}\left(\cdot \mid A^{(2)}_{\alpha''}\right)$.

$$
\begin{aligned}
&\sqrt{n}\,(\widehat{\underline{u}}_n - \underline{u}) \\
&= -\frac{1}{2\sqrt{n}\alpha'' f_1(\underline{u} \mid A^{(2)}_{\alpha''})} \sum_{t=1}^{n} \{\mathbb{1}_{\eta_{1t}<\underline{u}} \mathbb{1}_{\eta_{2t}\in A^{(2)}_{\alpha''}} - 2\alpha\alpha''\} - \frac{\underline{u}}{2\sqrt{n}} \boldsymbol{\Omega}_1' \boldsymbol{J}_1^{-1} \sum_{t=1}^{n} (\eta_{1t}^2 - 1)\boldsymbol{D}_{1t} \\
&- \frac{1-2\alpha''}{2 f_1(\underline{u} \mid A^{(2)}_{\alpha''})(2\alpha'')^2} \{\xi^{(2)}_{0.5-\alpha''} g^{(2)}(\xi^{(2)}_{0.5-\alpha''}) - \xi^{(2)}_{0.5+\alpha''} g^{(2)}(\xi^{(2)}_{0.5+\alpha''})\} \frac{1}{\sqrt{n}} \boldsymbol{\Omega}_2' \boldsymbol{J}_2^{-1} \sum_{t=1}^{n} (\eta_{2t}^2 - 1)\boldsymbol{D}_{2t} \\
&- \frac{f_2(\xi^{(2)}_{0.5-\alpha''} \mid \underline{u}) G^{(1)}(\underline{u})}{2\alpha'' g^{(2)}(\xi^{(2)}_{0.5-\alpha''}) f_1(\underline{u} \mid A^{(2)}_{\alpha''})} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbb{1}_{\eta_{2t}<\xi^{(2)}_{0.5-\alpha''}} - \alpha'' + 0.5) \\
&+ \frac{f_2(\xi^{(2)}_{0.5+\alpha''} \mid \underline{u}) G^{(1)}(\underline{u})}{2\alpha'' g^{(2)}(\xi^{(2)}_{\alpha''+0.5}) f_1(\underline{u} \mid A^{(2)}_{\alpha''})} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbb{1}_{\eta_{2t}<\xi^{(2)}_{\alpha''+0.5}} - 0.5 - \alpha'') + o_P(1).
\end{aligned}
$$

Appendix F: Proofs for Section 6

Proof of Theorem 6.1. We have

$$|\widehat{G}^{[i]}_n(\boldsymbol{y}) - G^{[i]}_n(\boldsymbol{y})| \leq \frac{1}{n} \sum_{t=1}^{n} \sum_{j=1}^{m} |\mathbb{1}_{\widehat{\eta}_{jt}<y_j} - \mathbb{1}_{\eta_{jt}<y_j}| \to 0, \quad a.s.$$

for all $\boldsymbol{y}$. Similarly,

$$|\widehat{H}^{[i]}_n(x, \boldsymbol{y}) - H^{[i]}_n(x, \boldsymbol{y})| \leq \frac{1}{n} \sum_{t=1}^{n} \left(|\mathbb{1}_{\widehat{\eta}_{0t}<x} - \mathbb{1}_{\eta_{0t}<x}| + \sum_{j=1}^{m} |\mathbb{1}_{\widehat{\eta}_{jt}<y_j} - \mathbb{1}_{\eta_{jt}<y_j}| \right) \to 0, \quad a.s.$$

from which it follows, using arguments of the proof of Theorem 3.1, that

$$|\widehat{F}^{[i]}_n(x|\boldsymbol{y}) - F^{[i]}_n(x|\boldsymbol{y})| \to 0, \quad a.s.$$

The proofs of i) and ii) can be deduced using the arguments given in the proof of Theorem 3.1. □

Proof of Theorem 6.2. The result follows by a straightforward adaptation of the proof of Theorem 3.2. □

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.

Banulescu, D., Hurlin, C., Leymarie, J. and O. Scaillet (2020) Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures. *Management Science*, 67.

Benoît, S., Colletaz, G., Hurlin, C. and C. Pérignon (2019) A Theoretical and Empirical Comparison of Systemic Risk Measures. Working Papers hal-02292323, HAL.

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

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

Engle, R.F. and G. Gonzalez-Rivera (1991) Semiparametric ARCH models. *Journal of Business & Economic Statistics* 9, 345–359.

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

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

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

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

Sharpe, W. F. (1994) The sharpe ratio. *Journal of portfolio management* 21, 49–58.