


How do empirical estimators of popular risk measures impact pro-cyclicality?

Marcel Bräutigam^{1,2} and Marie Kratz^{1*} 

¹ESSEC Business School Paris, CREAR, Paris, France; and ²Sorbonne University, LPSM, Paris, France

*Corresponding author. E-mail: kratz@essec.edu

(Received 13 January 2022; revised 03 February 2023; accepted 08 February 2023; first published online 29 March 2023)

Abstract

Risk measurements are clearly central to risk management, in particular for banks, (re)insurance companies, and investment funds. The question of the appropriateness of risk measures for evaluating the risk of financial institutions has been heavily debated, especially after the financial crisis of 2008/2009. Another concern for financial institutions is the pro-cyclicality of risk measurements. In this paper, we extend existing work on the pro-cyclicality of the Value-at-Risk to its main competitors, Expected Shortfall, and Expectile: We compare the pro-cyclicality of historical quantile-based risk estimation, taking into account the market state. To characterise the latter, we propose various estimators of the realised volatility. Considering the family of augmented GARCH(p, q) processes (containing well-known GARCH models and iid models, as special cases), we prove that the strength of pro-cyclicality depends on the three factors: the choice of risk measure and its estimators, the realised volatility estimator and the model considered, but, no matter the choices, the pro-cyclicality is always present. We complement this theoretical analysis by performing simulation studies in the iid case and developing a case study on real data.

Keywords: Asymptotic normality; Expected shortfall; Expectile; Estimators; (Filtered) Historical simulation; Measure of dispersion; Pro-cyclicality; Risk measure; Sample quantile; Value-at-risk

2020 AMS classification: Primary: 60F05; 62H20; 62P05; 62P20; 91B30

JEL classification: C13; C14; C30; C58; C69; G32

1. Introduction

Risk measurements are clearly central to risk management, in particular for financial institutions such as banks, insurance companies, and investment funds. Risk can be measured in terms of probability distributions. However, for practical purpose, it is often summarised into one real number, interpreted as a capital amount. Risk measures denote the tools that map loss distributions or random variables to capital amounts required as a buffer against insolvency. Proposing risk measures and studying their properties and impact in terms of risk management and model validation have been the topic of much research in actuarial science (including risk analysis and management), economics, finance, and mathematics; see e.g. Föllmer & Schied (2016), Jarrow (2017), McNeil *et al.* (2015), Miller (2018), Rüschemdorf (2013), for a non-exhaustive list of books on the topic. Here, our focus is *not* on risk measures but on the impact of historical risk estimation on pro-cyclicality, when considering popular or regulatory quantile-based risk measures as Value-at-Risk (VaR), Expected Shortfall (ES; named also Average VaR, Conditional VaR or Tail-VaR), and expectile. While the case of the VaR has been considered in Bräutigam *et al.* (2022) to lay down a methodology for quantifying pro-cyclicality, here, we investigate the pro-cyclicality of historical risk measurements in bigger generality by also considering other quantile-based risk measures, in particular ES, widely used by practitioners and more recently by regulators (as the

Basel Committee on Banking Supervision has replaced VaR with the coherent risk measure ES in internal market risk models).

The pro-cyclicality of risk measurements is an important concern for financial institutions; there is a consensus that, in times of crisis, risk measurements overestimate the future risk, while they underestimate it in quiet times, leaving the financial institutions unprepared. Quoting Gilles Moec, Chief Economist of AXA, 'The major mistake made in 2010 was imposing austerity at the worst time' (*Le Monde*, 2020/01/21. Translated). Pro-cyclicality has been a topic investigated mainly in economics, modelling agent behaviour (see e.g. Adrian & Shin, 2013; Catarineu-Rabell *et al.*, 2005). In macro-economics, the general issue of pro-cyclicality has been analysed with respect to the implications for banking regulation. In particular, there have been many discussions on the fact that, in times of crisis, banks reduce lending to firms, accentuating then the liquidity crisis. It is thus not surprising that discussions have mostly focused on credit risk measurement. For a more general review on bank pro-cyclicality, we refer to Athanasoglou *et al.* (2014), Quagliariello (2008), and the references therein. Note also the paper by Behn *et al.* (2016), where the authors use a quasi-experimental setup to show that model-based regulation in credit risk, which implies an increase in capital requirements, can increase the lack of credit. Regulatory authorities in different sectors (as BIS¹ for Basel III (where a "counter-cyclical" capital buffer has been created), EIOPA² for Solvency 2, and ESMA³ for derivative exchanges), have then proposed solutions to reduce pro-cyclicality (see Basel Committee on Banking Supervision, 2019, RBC30 & RBC40, Solvency 2 Directive, 2009 art.77b-d, 106, RTS (European Union), 2013 art.28). Studies such as Repullo & Saurina (2011), Rubio & Carrasco-Gallego (2016) and Llacay & Peffer (2019) deal with these new proposals in regulation and review its consequences on the phenomenon. In Bräutigam (2020), we tackled the topic differently, taking a statistical point of view to investigate, empirically and theoretically, the pro-cyclicality of the VaR historical estimation (i.e. a VaR estimation based purely on the empirical distribution of the data), showing that pro-cyclicality exists even beyond business cycles. Our approach, presented in Bräutigam *et al.* (2022), allowed us to quantify the pro-cyclicality for the first time. To evaluate the quality of the VaR prediction 1 year ahead, we introduced a ratio, referred to as "look-forward ratio", that compares the VaR prediction with its realisation. Using the realised volatility to characterise the market state, we observed an opposite behaviour over time between the estimated look-forward ratio and the realised volatility (see e.g. Bräutigam *et al.*, 2022, Figure 2). By how much this ratio is over- or underestimated for high or low volatility can also be directly represented through volatility binning, as in Bräutigam *et al.* (2022, Figure 4), providing the amount of over- and under-estimated capital. This might be a useful representation for risk management purposes. This led us to assess the pro-cyclicality via the degree of negative dependence between the two estimators. Subsequently, we identified two main factors characterising the observed pro-cyclical effect for the VaR. The clustering and return-to-the-mean of volatility, as expected, and, more surprisingly, the very way risk is measured on historical data, independently of the business cycles.

To investigate further this pro-cyclicality issue, we explore it more generally, considering quantile-based risk measures beyond the VaR, namely ES and expectile. In doing so, we question the appropriateness of measuring risk based on the empirical distribution of the data when considering those quantile-based risk measures for both models, static (iid) and dynamic (GARCH) ones. Moreover, as the realised volatility is used as a proxy for the market state, as in Bräutigam (2020) and Bräutigam *et al.* (2022), we also take into account the impact of the choice of its estimator when quantifying the pro-cyclicality. In the financial literature, the standard deviation is the usual realised volatility estimator; how does the pro-cyclicality behaviour change when

¹Bank for International Settlements.

²European Insurance and Occupational Pensions Authority.

³European Securities and Markets Authority.

choosing other realised volatility estimators? To answer these two questions, we build on the methodology set in Bräutigam (2020) and Bräutigam *et al.* (2022) when working with the VaR. We prove asymptotic normality of the joint distribution of the logarithm of the look-forward ratio and the dispersion measures estimators, whatever the risk measure, VaR, ES or expectile, and whatever the choice of dispersion measure to estimate the volatility. We thereby assess theoretically the pro-cyclicality of risk measurements made on historical data in a general framework through the degree of negative correlation between the components of this asymptotic distribution. This contribution complements the economic studies on the topic of pro-cyclicality and sets the results in Bräutigam *et al.* (2022) in a broader perspective. To our knowledge, this was an unexplored subject, and timely in its exploration given that this historical method is still used predominantly in practice, whatever the approach, unconditionally or conditionally; see e.g. Pérignon & Smith (2010) or European Banking Authority (2019) for quantitative surveys on this matter.

Resorting to historical types of measurement can be explained not only by an easy implementation but also by the reduction to a one-dimensional risk-measure estimation problem, without having to estimate the multivariate distribution of the risk vector nor taking into account the dependence between risk-factor changes. By construction, an unconditional approach neglects the dynamics of the data, leading generally to a poor estimation. Considering conditional approaches based on historical simulation from a univariate time series loss model certainly represents an improvement. For instance, a well-known method named *filtered historical simulation* applies empirical quantile estimation to the residuals of GARCH models. Although the conditional information will be more restricted than if considering the full multivariate setting, this method, of current use in practice, often works well. See e.g. McNeil *et al.* (2015) for further discussions. Our results cover both approaches of historical estimation of risk measures, conditional and unconditional, and conclude to the existence of pro-cyclicality in either case, and whatever the quantile-based risk measure considered. They also confirm the two main factors characterising pro-cyclicality.

To examine the relevance of our asymptotic theoretical results in a finite sample setting, we perform a simulation study in the iid case considering light and heavy tails of distributions. This study validates the appropriateness of the theoretical results to practically relevant situations, ensuring that one can expect the same conclusions in practice. Therefore, we can meaningfully compare the impact of choices of risk measures and dispersion measures on the degree of pro-cyclicality, when taking into account the fatness of the tail of the underlying distribution. This step being done, we turn to real data to verify that historical risk measurements lead to pro-cyclicality, irrespectively of the choice of quantile-based risk measures and dispersion measures, but also irrespectively of the choice of conditional versus unconditional approaches for risk estimation. Indeed, using Theorem 1 in conjunction with the good finite sample performance, we are able to confirm both main factors of pro-cyclicality, the clustering and return-to-the-mean of volatility and the way risk is measured on historical data independently of the business cycles.

The paper is organised as follows. In Section 2, we introduce the notation, the definitions of risk measures and measures of dispersion, and the setup of the statistical framework for our study. Section 3 derives the asymptotic joint distribution of the involved estimators proving the pro-cyclicality of risk measurements made on historical data. In Section 4, a simulation study is undertaken in the iid case to compare the effect of the choice of risk measure and dispersion measure and their estimators on pro-cyclicality. Then, in Section 5, we check about the two main factors explaining pro-cyclicality when considering real data. With this, we conclude to the presence of the pro-cyclical effect of traditional risk measurements (i.e. purely based on historical data), whatever the choice of risk measure and of volatility estimators. Proofs of theoretical results are deferred to the Appendix, while complementary material to Sections 4 and 5 are deferred to the Online Supplementary Appendix hosted by the journal.

2. Framework

Notation

Let X be a random variable (rv) with cumulative distribution function (cdf) F_X , and, given they exist, probability density function (pdf) f_X , mean μ , variance σ^2 , as well as, for any integer $r \geq 1$, the r -th absolute centred moment, $\mu(X, r) := \mathbb{E}[|X - \mu|^r]$ and quantile of order p defined as $q_X(p) := \inf\{x \in \mathbb{R}: F_X(x) \geq p\}$. For a n -sample (X_1, \dots, X_n) , we denote the associated order statistics by $X_{(1)} \leq \dots \leq X_{(n)}$.

Recall the sample estimator of the quantile for any order $p \in (0, 1]$, defined as $q_n(p) = X_{(\lceil np \rceil)}$, where $\lceil x \rceil = \min\{m \in \mathbb{Z}: m \geq x\}$, $\lfloor x \rfloor = \max\{m \in \mathbb{Z}: m \leq x\}$ and $[x]$, are the rounded-up, rounded-off integer parts and the nearest integer of a real number $x \in \mathbb{R}$, respectively.

The r -th absolute centred sample moment is defined, for $r \in \mathbb{N}$, by

$$\hat{m}(X, n, r) := \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}_n|^r, \tag{1}$$

\bar{X}_n denoting the empirical mean. Special cases of this latter estimator include the sample variance ($r = 2$) and the sample Mean Absolute Deviation (denoted as MAD) around the sample mean ($r = 1$).

Some standard notations: u^T for the transpose of a vector u and, for the signum function, $\text{sgn}(x) := -\mathbb{1}_{(x < 0)} + \mathbb{1}_{(x > 0)}$. Moreover the notations \xrightarrow{d} , $\xrightarrow{a.s.}$, \xrightarrow{P} , and $\xrightarrow{D_d[0,1]}$ correspond to the convergence in distribution, almost surely, in probability and in distribution of a random vector in the d -dimensional Skorohod space $D_d[0, 1]$. Further, for real-valued functions f, g , we write $f(x) = O(g(x))$ (as $x \rightarrow \infty$) if and only if there exists a positive constant M and a real number x_0 s.t. $|f(x)| \leq Mg(x)$ for all $x \geq x_0$, and $f(x) = o(g(x))$ (as $x \rightarrow \infty$) if for all $\epsilon > 0$ there exists a real number x_0 s.t. $|f(x)| \leq \epsilon g(x)$ for all $x \geq x_0$. Analogously, for a sequence of rv's X_n and constants a_n , we denote by $X_n = o_P(a_n)$ the convergence in probability to 0 of X_n/a_n .

Risk Measures

Let us recall the definitions of the risk measures we consider in this paper. One of the most used risk measures is Value-at-Risk (VaR), popularised by JP Morgan in 1996 (see Morgan & Reuters, 1996), and defined as follows: If we assume a loss random variable L having a continuous, strictly increasing distribution function F_L , the VaR at level p of L is simply the quantile $q_L(p)$ of order p of L :

$$\text{VaR}_p = \inf \left\{ x : P[L \leq x] \geq p \right\} = q_L(p). \tag{2}$$

Despite the availability of other approaches, in practice the VaR is usually estimated on historical data (see e.g. Pérignon & Smith, 2010 or European Banking Authority, 2019) for quantitative surveys on this matter), using the empirical quantile $\widehat{\text{VaR}}_n(p) = q_n(p)$ associated to a n -loss sample (L_1, \dots, L_n) with $p \in (0, 1)$.

VaR has been shown not to be a coherent risk measure, Artzner *et al.* (1999), contrary to Expected Shortfall (ES), introduced in slightly different formulations in Artzner *et al.* (1997, 1999), Acerbi & Tasche (2002), Rockafellar & Uryasev (2002). ES is defined as follows (e.g. Acerbi & Tasche, 2002) for a loss random variable L and a level $p \in (0, 1)$:

$$\text{ES}_p = \frac{1}{1-p} \int_p^1 q_L(u) du = \mathbb{E}[L | L \geq q_L(p)]. \tag{3}$$

While the first equality in (3) is the definition of ES, the second one holds only if L is continuous. There are different ways of estimating ES, we focus on the two most direct ones when using historical estimation.

First, simply approximating the conditional expectation in (3) by averaging over k sample quantiles, i.e.

$$\tilde{ES}_{n,k}(p) := \frac{1}{k} \sum_{i=1}^k q_n(p_i), \tag{4}$$

for a specific choice of $p = p_1 < p_2 < \dots < p_k < 1$. This was e.g. proposed in Emmer *et al.* (2015) in the context of backtesting expected shortfall (using $p_i = 0.25 p(5 - i) + 0.25(i - 1)$, $i = 1, \dots, 4$).

Another way was proposed in Chen (2008) and can be seen as a special case of $\tilde{ES}_{n,k}(p)$ choosing $k = n - [np] + 1$ and the p_i accordingly, estimating it as a conditional mean of the sample:

$$\widehat{ES}_n(p) := \frac{1}{n - [np] + 1} \sum_{i=1}^n L_i \mathbb{1}_{(L_i \geq q_n(p))}. \tag{5}$$

The discussions about which risk measure would be most appropriate to use for evaluating the risk of financial institutions have often included a third risk measure, the expectile. It was introduced, in the context of least-squares estimation in Newey & Powell (1987) and later used as a risk measure in finance and actuarial science (see e.g. Kuan *et al.*, 2009; Bellini *et al.*, 2021 and references therein, to refer to the pioneering paper and a very recent one, respectively). This risk measure satisfies many favourable properties (in particular for backtesting), making it appealing from a theoretical point of view (see e.g. Bellini & Bignozzi, 2015; Ziegel, 2016 and references therein). It is defined, for a square-integrable loss random variable L and level $p \in (0, 1)$, by the following minimiser

$$e_p = \operatorname{argmin}_{x \in \mathbb{R}} (p \mathbb{E}[\max(L - x, 0)^2] + (1 - p) \mathbb{E}[\max(x - L, 0)^2]). \tag{6}$$

There are various ways to estimate the expectile, the most natural one being the argmin of the empirical version of (6). Here, for simplicity, we choose to estimate it through a sample quantile using the following relation given in Yao & Tong (1996): Let $q_L(p)$ be the quantile at level $p \in (0, 1)$, then there exists a bijection $\kappa: (0, 1) \mapsto (0, 1)$ such that $e_{\kappa(p)}(L) = q_L(p)$ with

$$\kappa(p) = \frac{p \int_{-\infty}^{q_L(p)} x dF_L(x)}{\mathbb{E}[L] - 2 \int_{-\infty}^{q_L(p)} x dF_L(x) - (1 - 2p)q_L(p)}. \tag{7}$$

Thus, we consider the sample estimator for the expectile at level p denoted as

$$\hat{e}_n(p) := q_n(\kappa^{-1}(p)). \tag{8}$$

As unified notation representing these risk measures and their estimators defined above, we introduce, for $j = 1, \dots, 4$:

$$\zeta^{(j)}(p) = \begin{cases} \operatorname{VaR}_p & \text{for } j = 1, \\ \operatorname{ES}_p & \text{for } j = 2, \\ \operatorname{ES}_p & \text{for } j = 3, \\ e_p & \text{for } j = 4, \end{cases} \quad \text{with estimators } \zeta_n^{(j)}(p) = \begin{cases} \widehat{\operatorname{VaR}}_n(p) & \text{for } j = 1, \\ \widehat{ES}_n(p) & \text{for } j = 2, \\ \tilde{ES}_{n,k}(p) & \text{for } j = 3, \\ \hat{e}_n(p) & \text{for } j = 4. \end{cases} \tag{9}$$

Since we will work in a dynamic setting, time needs to be introduced. Therefore, we introduce a time-series notation of our estimated quantities. By

$$\widehat{\operatorname{VaR}}_{n,t}(p), \widehat{ES}_{n,t}(p), \tilde{ES}_{n,k,t}(p), \hat{e}_{n,t}(p), \zeta_{n,t}^{(j)}(p), \tag{10}$$

we denote the corresponding estimators estimated at time t over the last n observations before time t .

Framework setup

The samples considered in this study will be either realisations from iid rv’s (white noise process) or from augmented GARCH(p, q) processes (with the latter naturally including the former as a special case), since they include two families that allow one to isolate the pro-cyclicality effects, as shown in Bräutigam (2020) (see also Bräutigam *et al.*, 2022): One is the iid model that exemplifies the inherent part of pro-cyclicality (due to the use of historical estimation), and the other is the GARCH(1,1) model that shows that pro-cyclicality is caused by GARCH effects, i.e. return-to-the-mean and clustering of volatility (thus further amplifying the effect of pro-cyclicality).

Recall that an augmented GARCH(p, q) process $X = (X_t)_{t \in \mathbb{Z}}$, due to Duan (1997), satisfies, for integers $p \geq 1$ and $q \geq 0$,

$$X_t = \sigma_t \epsilon_t \tag{11}$$

$$\text{with } \Lambda(\sigma_t^2) = \sum_{i=1}^p g_i(\epsilon_{t-i}) + \sum_{j=1}^q c_j(\epsilon_{t-j})\Lambda(\sigma_{t-j}^2), \tag{12}$$

where (ϵ_t) is a series of iid rv’s with mean 0 and variance 1, $\sigma_t^2 = \text{Var}(X_t)$ and $\Lambda, g_i, c_j, i = 1, \dots, p, j = 1, \dots, q$, are real-valued measurable functions. Also, as in Lee (2014), we restrict the choice of Λ to the so-called group of either polynomial GARCH(p, q) or exponential GARCH(p, q) processes:

(Lee)
$$\Lambda(x) = x^\delta, \text{ for some } \delta > 0, \quad \text{or} \quad \Lambda(x) = \log(x)$$

Using this family of processes, let us now explain how we proceed to assess the pro-cyclicality of the proposed risk measure estimators (defined in (9) and (10)), based on the methodology developed in Bräutigam (2020) and Bräutigam *et al.* (2022). Namely, we use an indicator, the look-forward ratio, that quantifies the difference between the historically predicted risk and the estimated realised future risk (measured *ex post*) by considering the ratio of those two quantities: $\zeta_{n,t+n}^{(j)} / \zeta_{n,t}^{(j)}$ (with the time-series notation, i.e. estimation at any time $t, t + n$ over the last n observations before time $t, t + n$). We study this ratio as a function of the realised volatility (a proxy for market states). Then, we investigate theoretically the dependence between the look-forward ratio and the realised volatility estimated with the r -th absolute central sample moment (1), denoted $\hat{m}(X, n, r, t)$ (or $\hat{m}(n, r, t)$ when no confusion possible). A negative dependence will characterise the degree of pro-cyclicality of the considered risk measures estimators. As their dependence is nonlinear, we study the joint asymptotic distribution between *the logarithm* of the look-forward ratio and the r -th absolute central sample moment.

Note that the two quantities defining the look-forward ratio are estimators defined on disjoint, finite samples. Thus, some care has to be taken to translate the setting from a finite sample of overall size n , into an asymptotic one, where $n \rightarrow \infty$. In order to simplify computations, we use a little trick to obtain the disjointness of the estimators. Namely, we consider $\zeta_{n/2,t+n/2}^{(j)}, \zeta_{n/2,t}^{(j)}$ and $\hat{m}(n/2, r, t)$, where we assume w.l.o.g. that $n/2$ is an integer. It means that the estimators are evaluated on a sample of size $n/2$ each.

Denoting the correlation of the *asymptotic distribution* for the two quantities of interest, to ease and by abuse of notation, as

$$\lim_{n \rightarrow \infty} \text{Cor} \left(\log \left| \frac{\zeta_{n/2,t+n/2}^{(j)}}{\zeta_{n/2,t}^{(j)}} \right|, \hat{m}(X, n/2, r, t) \right), \quad j = 1, \dots, 4, r > 0, \tag{13}$$

the measure of the pro-cyclicality of risk measure estimators amounts then to the degree of negative correlation of (13).

3. An Asymptotic Theorem for Assessing Pro-Cyclicality

Let us now present the main result on pro-cyclicality. For the ease of presentation, the main theorem is presented in Section 3.1 while the discussion and explanation of the specific conditions under which this result holds are deferred to Section 3.2.

3.1 Main result

Extending the study Bräutigam *et al.* (2022) (see also Bräutigam, 2020) performed for the VaR, we provide here a general asymptotic theorem to assess theoretically the negative dependence for the historical estimators of any of the considered risk measures and any of the dispersion measures. With such a generality, it highlights even more the impact of historical risk measurements on pro-cyclicality. Moreover, we state this main theorem for the large family of augmented GARCH processes (defined in (11) and (12)), including of course the two models, iid and GARCH(1,1), which we used to isolate and prove the effect of pro-cyclicality. This generalisation is made for modelers using more sophisticated GARCH models as e.g. HAR, EGARCH, to model the volatility dynamics. To allow for such an extension, we need a set of conditions for the asymptotic theorem to hold. The simpler the model, the weaker the conditions. Those conditions will be made specific after the statement of the theorem (in Section 3.2) to discuss them, from the simplest iid case, which is also more intuitive, to the general augmented GARCH case that requires more work.

Let us present the result in its full generality. We refer to Appendix A for its proof.

Theorem 1 *Let $X = (X_t)_{t \in \mathbb{Z}}$ be either a white noise process or from the family of augmented GARCH(p,q) processes and have stationary distribution with its cdf, pdf, and quantile denoted as F_X, f_X, q_X , respectively. Consider a risk measure estimator $\zeta_n^{(j)}(p)$, $j \in \{1, \dots, 4\}$, and the r-th absolute central sample moment $\hat{m}(n, r)$, for a chosen integer $r > 0$. Given the choice of process, risk and dispersion measures, assume its corresponding set of conditions (given explicitly in Section 3.2), i.e. either one of $(S_1), (S_2), (S_3), (S_4)$ in the iid case, or one of $(S_1^*), (S_2^*), (S_3^*), (S_4^*)$ when X belongs to the family of augmented GARCH(p,q) processes.*

Then, the asymptotic distribution of the logarithm of the look-forward ratio of the risk measure estimator with the r-th absolute central sample moment, at any given fixed time t, is bivariate normal, i.e.

$$\sqrt{n} \begin{pmatrix} \log \left| \frac{\zeta_{n/2, t+n/2}^{(j)}(p)}{\zeta_{n/2, t}^{(j)}(p)} \right| \\ \hat{m}(n/2, r, t) - \mu(X, r) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \tilde{\Gamma}),$$

with $\tilde{\Gamma} = (\tilde{\Gamma}_{ik})_{1 \leq i, k \leq 2}$ and $\tilde{\Gamma}_{ik} = \begin{cases} 4\Gamma_{ik} / (\zeta^{(j)}(p))^2 & \text{for } i = k = 1, \\ 2\Gamma_{ik} & \text{for } i = k = 2, \\ -2\Gamma_{ik} / \zeta^{(j)}(p) & \text{otherwise,} \end{cases}$

Γ being the covariance matrix of the asymptotic bivariate distribution between $\zeta_n^{(j)}(p)$ and $\hat{m}(n, r)$ (see Appendix A.1 for the explicit expressions in the different cases). In particular, the correlation of this asymptotic bivariate distribution equals

$$\frac{\tilde{\Gamma}_{12}}{\sqrt{\tilde{\Gamma}_{11}}\sqrt{\tilde{\Gamma}_{22}}} = \frac{-1}{\sqrt{2}} \text{sgn}(\zeta^{(j)}(p)) \frac{\Gamma_{12}}{\sqrt{\Gamma_{11}}\sqrt{\Gamma_{22}}} = \frac{-1}{\sqrt{2}} \frac{|\Gamma_{12}|}{\sqrt{\Gamma_{11}}\sqrt{\Gamma_{22}}}, \tag{14}$$

thus, it is always nonpositive.

3.2 Discussion of the conditions sets

Let us introduce and explain the different types of conditions we need to consider.

First, we impose four conditions on the distribution function F_X . They are needed for the asymptotic representation of the different risk measure (and measure of dispersion) estimators: The continuity of F_X (or its l -th derivative), the l -fold differentiability of F_X for any integer $l > 0$, and the positivity of its density f_X . These conditions hold at a given point or neighbourhood and are named as

- (C_0) F_X is continuous,
- (C_l) the l -th derivative of F_X is continuous,
- (C'_l) F_X is l -times differentiable,
- (P) f_X is positive.

Equally, to establish any Central Limit Theorem (CLT), we need a condition on the finiteness of the moments of the innovation process of the augmented GARCH process, (M_k), for $k \in \mathbb{N}$,

$$(M_k) \quad \mathbb{E}[|\epsilon_0|^{2k}] < \infty.$$

To motivate which of these conditions are exactly needed for asymptotic theorem, let us informally explain how to establish (13) considering the iid case. In a later step, we then point out the additional conditions needed for the class of augmented GARCH(p, q) processes.

IID case – In this case, recall that the sample used for any risk measure estimator at time $t + n/2$, $\zeta_{n/2, t+n/2}^{(j)}(p)$, is, by construction, disjoint from the sample used at time t . Thus, the estimator $\zeta_{n/2, t+n/2}^{(j)}(p)$ will be uncorrelated with the r -th absolute centred sample moment $\hat{m}(n/2, r, t)$, at time t , as well as with the risk measure estimator $\zeta_{n/2, t}^{(j)}(p)$ at time t .

Translating this for the correlation of the asymptotic distribution (again abusing the notation), i.e. (13), it should hold, for $j = 1, \dots, 4$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Cor} \left(\log \left| \frac{\zeta_{n/2, t+n/2}^{(j)}(p)}{\zeta_{n/2, t}^{(j)}(p)} \right|, \hat{m}(n/2, r, t) \right) &= \lim_{n \rightarrow \infty} \frac{\text{Cov}(-\log |\zeta_{n/2, t}^{(j)}(p)|, \hat{m}(n/2, r, t))}{\sqrt{2\text{Var}(\log |\zeta_{n/2, t}^{(j)}(p)|)}\sqrt{\text{Var}(\hat{m}(n/2, r, t))}} \\ &= \frac{-1}{\sqrt{2}} \lim_{n \rightarrow \infty} \text{Cor}(\log |\zeta_{n, t}^{(j)}(p)|, \hat{m}(n, r, t)) = \frac{-1}{\sqrt{2}} \left| \lim_{n \rightarrow \infty} \text{Cor}(\zeta_{n, t}^{(j)}(p), \hat{m}(n, r, t)) \right|, \end{aligned} \quad (15)$$

where the first equality follows by the uncorrelatedness (which also holds for finite n), the second by the scale invariance of the correlation, and the third as a consequence of the Delta-method with the logarithm.

Thus, in the iid case, establishing (13) reduces to proving a joint asymptotic normality between $\zeta_{n, t}^{(j)}(p)$ and $\hat{m}(n, r, t)$. To do so, we need conditions to obtain an appropriate representation of the risk measure estimator $\zeta_{n, t}^{(j)}(p)$ and moment conditions depending on the chosen dispersion measure $\hat{m}(n, r, t)$. We group the set of conditions (S_j) by the choice of risk measure $\zeta_{n, t}^{(j)}(p)$, $j = 1, \dots, 4$, for a given measure of dispersion $\hat{m}(n, r, t)$, $r \in \mathbb{N}$, where

$$(S_1) = S(p) \quad \text{and} \quad (S_4) = S(\kappa^-(p)) \quad \text{with} \quad S(\cdot): \begin{cases} (C'_1)(\text{see } (C'_1)) \text{ at } q_X(\cdot) \\ (P) \text{ at } \mu \text{ for } r = 1, \text{ and at } q_X(\cdot) \\ (M_r)(\text{see } (M_k)) \end{cases}$$

$$(S_2): \begin{cases} (C_3)(\text{see } (C_l)) \text{ in a neighbourhood of } q_X(p) \\ F_X \text{ absolutely continuous} \\ (M_r) \text{ and } (M_{1+\gamma})(\text{see } (M_k)) \text{ for some } \gamma > 0 \\ (P) \text{ at } \mu \text{ for } r = 1 \end{cases} \quad \text{and} \quad (S_3): \begin{cases} \text{For } l = 2, \dots, k: \\ S(p_l) \\ (C'_1)(\text{see } (C'_1)) \text{ at } q_X(p_l) \\ (P) \text{ at } q_X(p_l) \end{cases}$$

Augmented GARCH(p, q) – Clearly, in the case of the augmented GARCH(p, q) processes family, we will need additional conditions for the asymptotic normality to hold. There are two reasons for this. First, the conditions to establish such a limit theorem are stronger. Second, any two estimators, even if computed over disjoint samples, might be correlated (in contrast to the iid case), so the argumentation in (13) does not hold without further requirements.

Let us first discuss conditions for establishing a bivariate asymptotic theorem between $\zeta_{n,t}^{(j)}(p)$ and $\hat{m}(n, r, t)$ for the family of augmented GARCH(p, q) processes.

Note that already for a strictly stationary solution to (11) and (12) to exist, the functions Λ, g_i, c_j as well as the innovation process $(\epsilon_t)_{t \in \mathbb{Z}}$ have to fulfil some regularity conditions (see e.g. Lee, 2014, Lemma 1), namely the positivity of the functions used, (A), and the boundedness in L_r -norm for either the polynomial GARCH, (P_ν), or exponential/logarithmic GARCH, (L_ν), respectively, for a given integer $\nu > 0$:

$$(A) \quad g_i \geq 0, c_j \geq 0, i = 1, \dots, p, j = 1, \dots, q,$$

$$(P_\nu) \quad \sum_{i=1}^p \|g_i(\epsilon_0)\|_\nu < \infty, \quad \sum_{j=1}^q \|c_j(\epsilon_0)\|_\nu < 1,$$

$$(L_\nu) \quad \mathbb{E} \left[\exp \left(4\nu \sum_{i=1}^p |g_i(\epsilon_0)|^2 \right) \right] < \infty, \quad \sum_{j=1}^q |c_j(\epsilon_0)| < 1.$$

Note that Condition (L_ν) requires the c_j to be bounded functions.

It was shown by Lee that these conditions are sufficient conditions for establishing a CLT for the volatility process (see Lee, 2014, Corollary 1) and, together with the finiteness of the innovations process, i.e. (M_r) (see (M_k)), also for the CLT of the augmented GARCH(p, q) process itself.

For this, Lee exploited the known fact that the L_2 -near-epoch dependence (L_2 -NED) paired with corresponding finite moments is a sufficient condition for establishing the CLT.

Let us recall the concept of L_p -near-epoch dependence (L_p -NED), using a definition due to Andrews (1988) but restricted to stationary processes. Let $(Z_n)_{n \in \mathbb{Z}}$, be a sequence of rv's and $\mathcal{F}_s^t = \sigma(Z_s, \dots, Z_t)$, for $s \leq t$, the corresponding sigma-algebra. By $|\cdot|$, we denote the Euclidean norm and the usual L_p -norm is denoted by $\|\cdot\|_p := \mathbb{E}^{1/p}[|\cdot|^p]$.

Definition 2 (L_p -NED, Andrews, 1988) For $p > 0$, a stationary sequence $(X_n)_{n \in \mathbb{Z}}$ is called L_p -NED on $(Z_n)_{n \in \mathbb{Z}}$ if, for $k \geq 0$,

$$\|X_1 - \mathbb{E}[X_1 | \mathcal{F}_{n-k}^{n+k}]\|_p \leq v(k),$$

for non-negative constants $v(k)$ such that $v(k) \rightarrow 0$ as $k \rightarrow \infty$.

If $v(k) = O(k^{-\tau-\epsilon})$ for some $\epsilon > 0$, we say that X_n is L_p -NED of size $(-\tau)$.

If $v(k) = O(e^{-\delta k})$ for some $\delta > 0$, we say that X_n is geometrically L_p -NED.

Now that we have discussed the conditions for establishing a CLT, let us go back to (13) and the fact that two estimators, even if computed over disjoint samples, might be correlated as those processes exhibit dependence. In this case, we show that the condition of strong mixing with geometric rate makes the estimators on disjoint samples asymptotically uncorrelated, thus, recovering structurally the pro-cyclicality behaviour as in the iid case. It means that, besides L_2 -NED dependence, also strong mixing with geometric rate is an additional condition needed for the class of augmented GARCH(p, q) processes.

Let us recall for completeness the notion of strong mixing, denoting for a sequence of random variables $(Z_n)_{n \in \mathbb{Z}}$ the corresponding σ -algebra as $\mathcal{F}_s^t = \sigma(Z_s, \dots, Z_t)$ for $s \leq t$:

Definition 3 (Strong mixing) *Define, as measure of dependence, for any integer $n \geq 1$,*

$$\alpha(n) := \sup_{j \in \mathbb{Z}} \sup_{C \in \mathcal{F}_{-\infty}^j, D \in \mathcal{F}_{j+n}^{\infty}} |P(C \cap D) - P(C)P(D)|. \tag{16}$$

The sequence of rv's $(Z_n)_{n \in \mathbb{Z}}$ is called strongly mixing if $\alpha(n) \rightarrow 0$ as $n \rightarrow \infty$. It is called strongly mixing with geometric rate if there exist constants $\lambda \in (0, 1)$ and c such that $\alpha(n) \leq c\lambda^n$ for every n .

If we denote the set of conditions for the family of augmented GARCH(p, q) processes X by (S_j^*) (as extension to the set of conditions (S_j) in the iid case), again grouped by the choice of risk measure $\zeta_{n,t}^{(j)}(p), j = 1, \dots, 4$, for a given measure of dispersion $\hat{m}(n, r, t), r \in \mathbb{N}$, we get:

$$(S_1^*) = S^*(p) \text{ and } (S_4^*) = S^*(\kappa^-(p)) \text{ with } S^*(\cdot): \begin{cases} (C'_2)(\text{see } (C'_1)) \text{ at } q_X(\cdot) \\ (P) \text{ at } \mu \text{ for } r = 1, \text{ and at } q_X(\cdot) \\ (M_{r+\tau})(\text{see } (M_k)) \text{ for some } \tau > 0 \\ X \text{ strongly mixing with geometric rate} \\ (Lee) \text{ and } (A) \\ (P_{\max(1, \frac{r}{\delta})})(\text{see } (P_V)) \text{ if } X \text{ is polynomial GARCH} \\ (L_r)(\text{see } (L_V)) \text{ if } X \text{ is exponential GARCH} \end{cases}$$

$$(S_2^*): \begin{cases} (S_1^*) \\ F_X \text{ absolutely continuous} \\ (C_3)(\text{see } (C_l)) \text{ in a neighbourhood of } q_X(p) \\ (M_{1+\gamma})(\text{see } (M_k)) \text{ for some } \gamma > 0 \text{ for } r = 1 \\ \text{All the 2nd partial derivatives of the bivariate} \\ \text{distribution of } (X_1, X_{k+1}) \text{ for } k \geq 1, \text{ are} \\ \text{bounded in a neighbourhood of } q_X(p) \end{cases} \quad \text{and} \quad (S_3^*): \begin{cases} \text{For } l = 2, \dots, k: \\ S^*(p_l) \\ (C'_1)(\text{see } (C'_l)) \text{ at } q_X(p_l) \\ (P) \text{ at } q_X(p_l) \end{cases}$$

4. Comparing the Pro-Cyclicity of Risk Measures

We consider two different applications in this section. Both aim at further understanding the theoretically proven pro-cyclicity behaviour of the different risk measures in conjunction with the corresponding volatility estimators.

First, in Section 4.1, we evaluate empirically in a simulation study for light (Gaussian) and heavy-tailed distributions (Student-t distribution with 5 and 3 degrees of freedom) how well the finite sample results, as one encounters in practice, approximate the theoretical asymptotics.

Subsequently, in Section 4.2, we compare, for the same distributions and estimators as in Section 4.1, the strength of the theoretical correlation of the asymptotic distribution – thus comparing the existing degree of pro-cyclicity between the different risk measures.

4.1 Simulation study: finite sample performance

As in practice the estimation of risk measures occurs on a finite sample (usually a sample of 1 year of data which corresponds to 252 data points), we want to assess the finite sample performance, in view of the asymptotic results obtained in Theorem 1. When working with data, we estimate the risk measure and volatility estimators on finite samples of size n . To subsequently evaluate the

corresponding covariance and correlation empirically we need, say l , independent realisations of these risk measure and realised volatility estimators.

To assess the finite sample performance, we conduct a simulation study following a similar but more general setup than in Section 4.1 of Bräutigam *et al.* (2022). We simulate an iid sample with, e.g. mean $\mu = 0$ and variance $\sigma^2 = 1$, from three different distributions each: Either a Gaussian distribution or Student t distributions with 3 and 5 degrees of freedom, respectively. The sample is of varying size $n \times l$. It is determined by the fact that we use different sample sizes n for the estimation of either the risk measure or volatility, with $n = 126, 252, 504, 1,008$ (being multiples or fractions of 1 year of data, i.e. 252 working days/data points), and different lengths l corresponding to the independent realisations of the risk measure or realised volatility estimator to estimate the sample correlation of interest.

Taking the example of risk levels relevant in practice (this means $p = 0.99$ for the VaR and $p = 0.975$ for the ES in the standard Basel regulation; for the expectile, it has been suggested in Bellini & Di Bernardino (2017) to consider $p = 0.99855$), we consider $p = 0.95, 0.975, 0.99, 0.99855$. For each risk level, we compute l independent realisations of each risk measure estimator $(\zeta_{n,k}^{(j)}(p))_{k=1,\dots,l}, j = 1, 2, 3, 4$ on disjoint samples and, accordingly, l realisations of the realised volatility estimator $(\hat{m}(X, n, r))_{k=1,\dots,l}, r = 1, 2$, i.e. focussing on the standard deviation and MAD for the latter.

We then estimate $\text{Cor} \left(\log \left| \frac{\zeta_{n/2, t+n/2}^{(j)}(p)}{\zeta_{n/2, t}^{(j)}(p)} \right|, \hat{m}(X, n/2, r, t) \right), j = 1, 2, 3, 4, r = 1, 2$ by its sample correlation (denoted by $\widehat{\text{Cor}}$), using these l pairs of independent realisations of the estimators. This procedure is repeated 1,000-fold in each case. We report the averages of the 1,000-fold repetition with, into brackets, the corresponding empirical 95% confidence interval. Further, we provide as benchmark the theoretical value of the correlation in its asymptotic distribution, denoted as “ $(n \rightarrow \infty)$ ” in the last column. The explicit expressions in the case of a Gaussian or Student distribution of the correlation of the asymptotic distribution used to calculate the theoretical values in the tables can be derived from Theorem 1 (and can be found in Appendix B.1).

In this section, we study exemplarily the finite sample performance of the historical Expected Shortfall estimator as risk measure estimator and the MAD as realised volatility estimator. We focus on the approximation of the correlation of the asymptotic distribution by its sample correlation as a function of the sample size n and the three different distributions considered. As risk levels, we consider $p = 0.95$ and $p = 0.99$. Thus, we fix the length of the sample correlation time series to $l = 50$ (from simulations performed with different sample sizes, we saw that this is long enough for a good estimation of the correlation; see Appendix D).

The other cases, VaR, expectile as risk measures, standard deviation as realised volatility estimator and other risk levels, show a similar behaviour and can be found in the extensive appendix for the simulation studies (see Appendix D).

Thus, let us look at the results displayed in Table 1. First, we consider the risk level $p = 0.95$. For the three distributions, we see that a sample size of $n = 126$ suffices to estimate on average the correlation of the asymptotic distribution well enough (with slightly less precision for heavier tailed distributions). For the higher risk level $p = 0.99$, one can observe that the convergence to the theoretical value is slower and a sample size of $n = 126$ does not yield as accurate results as for $p = 0.95$. Nevertheless, considering a sample size frequently used in practice, $n = 252$, i.e. 1 year of data, gives a sufficiently accurate picture. Note also that the size of the empirical confidence intervals depend on the size of the sample correlation, here chosen to be $l = 50$ (recall that the results for other values of l can be found in Appendix D). They are in line with what is to be expected for $l = 50$: To show this, we can build confidence intervals for the sample Pearson correlation coefficient around the theoretical value (using the classical variance-stabilising Fisher transform of the correlation coefficient for a bivariate normal distribution to compute the confidence intervals – see the original paper Fisher, 1921 or e.g. a standard encyclopedia entry Rodriguez, 1982).

Table 1. Average values from a 1,000-fold repetition. Comparing the sample Pearson correlations of the log-ratio of the historical Expected Shortfall with the sample MAD, as a function of the sample size n on which the quantile is estimated (fixed length $l = 50$ of the bivariate sample used to estimate the correlation). We consider the thresholds $p = 0.95, 0.99$. Underlying samples are simulated from a Gaussian, Student(5) and Student(3) distributions. Average empirical values are written first (with empirical 95% confidence interval in brackets). The theoretical correlation value in the asymptotic distribution “ $(n \rightarrow \infty)$ ” are provided as benchmark in the last column.

ES and MAD	$n = 126$	$n = 252$	$n = 504$	$n = 1,008$	Theoretical value ($n \rightarrow \infty$)
Gaussian distr.					
$p = 0.95$	-32(- 56, -7)	-32(- 56, -5)	-32(- 55, -7)	-32(- 55, -7)	-32
$p = 0.99$	-22(- 48, 5)	-21(- 46, 8)	-21(- 46, 6)	-20(- 45, 8)	-19
Student(5) distr.					
$p = 0.95$	-38(- 60, -13)	-37(- 58, -12)	-38(- 59, -14)	-38(- 57, -14)	-37
$p = 0.99$	-30(- 55, -4)	-27(- 53, -2)	-28(- 51, -2)	-27(- 50, -2)	-26
Student(3) distr.					
$p = 0.95$	-42(- 63, -15)	-42(- 64, -17)	-43(- 63, -17)	-43(- 63, -21)	-43
$p = 0.99$	-37(- 59, -9)	-36(- 59, -7)	-36(- 59, -9)	-36(- 58, -11)	-35

4.2 Theoretical comparison

Having confirmed in the previous section the validity of the theoretical results in a sample size setting used in practice, we are interested in the following in comparing the degree of (theoretical) pro-cyclicality depending on the choice of volatility estimator, the risk measure (estimator), and the heavy-tailedness of the distribution. As for the simulation study, we consider underlying iid models and evaluate the degree of negative correlation in the asymptotic distribution of the log-ratio of risk measure estimators with measure of dispersion estimators, given in (13).

As risk measure estimators, we consider the VaR estimator $\widehat{\text{VaR}}_n(p)$, the expectile estimator $\hat{e}_n(p)$, and three ES estimators, $\widetilde{\text{ES}}_{n,4}$, $\widetilde{\text{ES}}_{n,50}$ and $\widetilde{\text{ES}}_{n,\infty}$ ($=\widehat{\text{ES}}_n$). As in the simulation study, we focus on the sample MAD $\hat{m}(X, n, 1)$ and the sample variance $\hat{m}(X, n, 2)$ as they are the two most common realised volatility estimators.

As, from a risk management perspective, only large values for the risk level $p \in (0, 1)$ are relevant, we focus on those cases. When considering the tail of the distribution, we choose the Gaussian distribution, $\mathcal{N}(0, 1)$, for its light tail, and, for heavy tailed distributions, the Student- t ones with varying degrees of freedom ν ($\nu = 3, 4, 5, 10, 50$) but always normalised to have mean 0 and variance 1.

The closed form solutions for the degree of pro-cyclicality in the aforementioned cases follow from Theorem 1 and its corresponding bivariate CLT’s, and can be found in Appendix B.1. Here, we focus on visualising the solutions and comparing them.

Gaussian Distribution. In Figure 1, we plot the correlations in the asymptotic distribution of the different risk measure estimators with the sample standard deviation (left) and the sample MAD (right), respectively.

Observing the different plots, we can make the following claims on the pro-cyclicality behaviour for the Gaussian distribution (as an example of a light-tailed distribution).

We observe a switch of behaviour in the tail. The degree of pro-cyclicality for tail values for different risk measures has a turning point in the tail in which the ordering is reversed. For quantile values below the turning point, the ES exhibits the highest pro-cyclicality, then the VaR and then the expectile. After the turning point, this exact ordering is reversed. The choice of dispersion measure has no real impact in this case, as the behaviour looks similar with MAD and variance. The ordering of the different risk measures with respect to pro-cyclicality is the same for MAD and variance – also roughly their magnitude of pro-cyclicality. What differs is the location of the turning point: It is further in the tail with the variance (around 0.97) than with MAD (around 0.92).

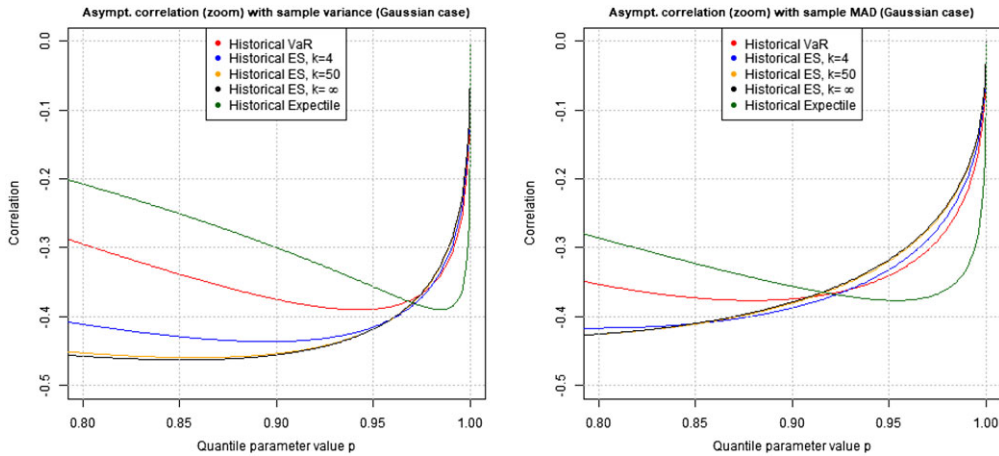


Figure 1. *Gaussian case.* Pro-cyclicality as defined in (13), considering on each plot three different risk measures (VaR, ES, evaluated in three possible ways, and expectile) for a risk level higher than 80%. On the left with the standard deviation, on the right with the sample MAD.

Considering VaR and ES, the values of the correlation are very similar after the turning point, while before, the pro-cyclicality for ES can be markedly (depending on the estimation method) higher than for VaR. From the plots it is also clear that the expectile, with the estimation method used, has the highest degree of pro-cyclicality in the far tail. It could be tested if this observation remains true when taking another method to estimate the expectile. We do not do it here, since the expectile is not yet a risk measure used in practice.

Student t Distribution. By considering the Student t distribution, we are interested in understanding how the observed behaviour may change with heavy-tailed distributions. Exemplary, we consider the case $\nu = 5$ in Figure 2 since we need $\nu > 4$ for (M_2) (see (M_k)) to hold. However, in Appendix B (see Figure 1), we include an analysis when looking at different degrees of freedom ($\nu = 3, 4, 5, 10, 40$), investigating how the correlations change as a function of ν , also comparing them with the Gaussian limiting case.

In Figure 2, we observe that, for this heavy-tailed distribution, the correlation behaviour depends on the dispersion measure considered. Indeed, in contrast to the Gaussian case, the existence of the turning points depends on whether we use the sample variance or the MAD: It does not exist with the sample variance, but appears with the MAD (as for the Gaussian distribution) – with the location of the turning point having shifted further into the tail. Moreover, the degree of pro-cyclicality for the VaR and ES is very different when choosing the variance, whereas for the MAD (where the turning point still exists as in the Gaussian case), the tail behaviour is similar for both risk measures. For the expectile, the correlation behaviour is similar as in the Gaussian case, except in the extreme case when the quantile level p tends to 1: in such a case, the correlation does not tend to 0 but to a non-zero value, for both measures of dispersion. Moreover, the pro-cyclicality is the highest in the far tail for the sample MAD (right plots), as in the Gaussian case, but not anymore for the sample variance.

Implications of the pro-cyclicality for the choice of risk measure. Let us end the comparison of the pro-cyclicality in Gaussian and Student iid models for the different risk measures by commenting on its implications for the choice of risk measure.

From the observations made on the figures, we have seen that the pro-cyclicality behaviour depends on the choice of the underlying risk measure, the dispersion measure, and also on the type of tail distribution. But, as already proven through Theorem 1, pro-cyclicality is inherent to and present (to a significant degree) with all risk measures due to the method of estimation, namely the historical estimation.

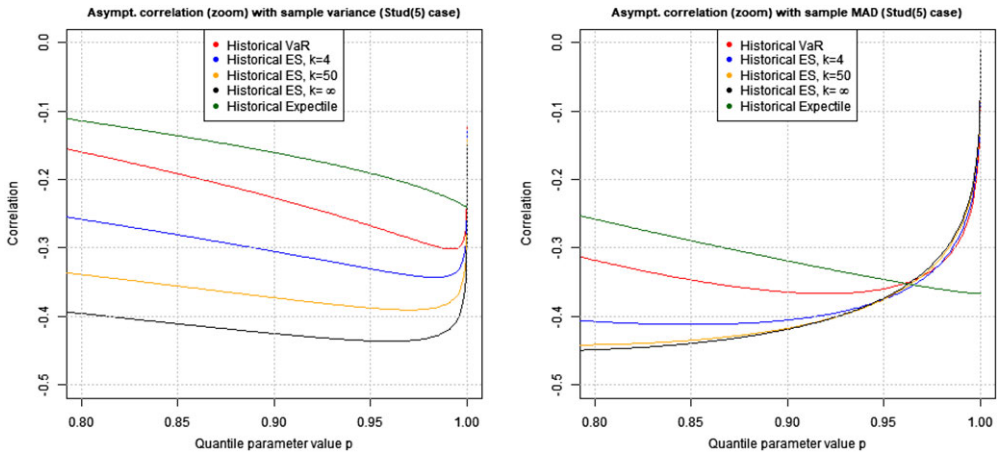


Figure 2. Case of a Student-*t* distribution with 5 degrees of freedom. Pro-cyclicality as defined in (13), (VaR, ES, evaluated in three possible ways, and expectile) quantified for a risk level higher than 80%. On the left with the standard deviation, on the right with the sample MAD.

5. A Case Study on Real Data

In this part, we want to go one step beyond the illustration of the pro-cyclicality in the iid case and check about the effect of pro-cyclicality when considering real data. We want to use the results of Theorem 1 to statistically verify the empirical claims on the causes of pro-cyclicality. Namely, that the pro-cyclicality observed is partly from an intrinsic effect of using the method of historical estimation and partly due to the clustering and return-to-the-mean behaviour of volatility, as modeled with a GARCH(1, 1), for any risk and dispersion measures.

We could compute the theoretical pro-cyclicality value for a GARCH(1, 1) process given in Theorem 1 and compare it with the value obtained for the real data. But we do not have closed form solutions of the correlation of the asymptotic distribution for this family of models. Further, it is known that, for GARCH processes, the convergence to its asymptotic distribution is slow (as argued e.g. in Mikosch & Stărică (2000) for the autocovariance/autocorrelation process). Hence, we consider the alternative and easier way suggested in Bräutigam *et al.* (2022) for the VaR case: Instead of analysing the theoretical correlation for a GARCH model, we analyse the residuals of a GARCH(1, 1) fitted to the data, then the pro-cyclicality for this residual process.

Pro-cyclicality Analysis of Residuals. To start with, recall the GARCH(1,1) model:

$$X_{t+1} = \epsilon_t \sigma_t, \quad \text{with } \sigma_t^2 = \omega + \alpha X_t^2 + \beta \sigma_{t-1}^2 \text{ and } \omega > 0, \alpha \geq 0, \beta \geq 0,$$

where $(\epsilon_t, t \in \mathbb{Z})$ is an iid series with mean 0 and variance 1.

As in Bräutigam *et al.* (2022) and Bräutigam (2020), we consider 11 stock indices. The data used are the daily closing prices from Friday, January 2, 1987 to Friday, September 28, 2018 (detailed information about the countries and indices used can be found in Appendix C.1, Table 3). As measure of dispersion we choose the MAD. An important motivation for choosing the MAD over the variance as measure of dispersion is that it implies weaker conditions on the moments of the underlying distribution (recall Theorem 1).

For each of the 11 indices, we initialise $\hat{\sigma}_t$ by using 1 year of data (as “burn-in” sample) using the fitted GARCH parameters (see Table 4 in Appendix C.1 for details). This enables us to consider the time series of empirical residuals $\hat{\epsilon}_t := X_{t+1}/\hat{\sigma}_t$.

Then, to assess the pro-cyclicality for these residuals, we compute the sample correlation between the logarithm of the look-forward ratio and the sample MAD, recall (13) – but here, on the time series of residuals $\hat{\epsilon}_t$ (and not the real data itself!). In theory, this time series of residuals

should be iid with mean 0 and variance 1. Hence, we can exactly assess the pro-cyclicality (i.e. the degree of negative correlation in the asymptotic distribution of the logarithm of the look-forward ratio and the MAD) for iid models, by applying Theorem 1 (for the iid case, i.e. under milder conditions than what would be needed for the GARCH).

To compare the sample correlation (based on a finite sample of about 300) with the theoretical asymptotic value of the correlation, we provide the corresponding confidence intervals for the sample Pearson linear correlation coefficient (for details, see Appendix C.2). Note that they are computed assuming a bivariate normal sample, although the bivariate normality of the logarithm of the look-forward ratio with the sample MAD holds only asymptotically. But, from preliminary simulation results, we observed that, for such a sample size, the empirical and theoretical confidence intervals for underlying Gaussian and Student samples are similar. Thus, we feel confident in providing those theoretical confidence intervals as approximate guidance. We then verify if the sample correlation based on the residuals falls in these iid confidence intervals and how the sample correlation based on the real data behaves in comparison (the corresponding raw values, i.e. the pro-cyclicality values for the 11 stock indices, can be found in Appendix C.1).

In the theoretical results, we have considered three different risk measures (VaR, ES, and expectile) and in the application for the iid case (Section 4) have looked at five different risk measure estimators. Here, dealing with an empirical setup, we restrict ourselves to the two risk measures that are effectively used in practice, VaR and ES, and their simplest form of estimation $\widehat{\text{VaR}}_n(p)$ and $\widehat{\text{ES}}_{n,\infty}(= \widehat{\text{ES}}_n)$. Note that, for completeness and to allow for comparison with ES, we include the case of the VaR already studied in Bräutigam *et al.* (2022).

For each risk measure, we consider the results on each of the 11 indices, using one plot per index. In those plots, we compare for four thresholds $p = 0.95, 0.975, 0.99, 0.995$, the measured pro-cyclicality (i.e. the sample correlation between the log-ratio of the VaR or ES estimators, respectively, and the sample MAD) on the real data versus the one on the residuals. Further, 95%-confidence intervals for a sample correlation assuming an underlying white noise process are given – considering as alternatives a Gaussian or Student distribution, the latter with varying degrees of freedom, $\nu = 4, \dots, 7$. In Figure 3, we consider, exemplarily, the S&P 500 and the FTSE index for the VaR and ES, respectively. The totality of the 22 plots are to be found in Appendix C.3 in Figures 2 and 3. Considering the VaR (in the first row of Figure 3), we observe that for the FTSE the pro-cyclicality value for the residuals is always in the confidence interval, whereas for the S&P 500 only two out of four times. For the ES in the second row, it is in all cases for both indices in the iid confidence interval.

Considering all indices in the case of the ES (Figure 3 in Appendix C.3), in 37 out of 44 cases (84%), the sample correlation of the real data falls in the 95% confidence interval of the sample correlation of iid rv's, while it is in 38 out of 44 cases (86%) for the VaR (see Figure 2 in Appendix C.3). Moreover, for exactly 1 out of 44 cases for ES and none for VaR, the sample correlation of the real data falls in these confidence intervals.

Thus, the same conclusion holds for ES as for VaR: We are left with a pro-cyclicality behaviour for the residuals like for iid data.

6. Conclusion

In this study, we investigated the pro-cyclicality of historical risk estimation considering popular quantile-based risk measures, using the methodology set in Bräutigam *et al.* (2022), based on the joint behaviour of the look-forward ratio and the realised volatility. We extended the results obtained for the VaR in Bräutigam *et al.* (2022) by considering ES and expectile and different realised volatility estimators for characterising the market states. The main theoretical result consists in establishing the joint asymptotic normality between the logarithm of the look-forward ratio (defined in terms of risk measures estimators) and dispersion measures estimators. By this,

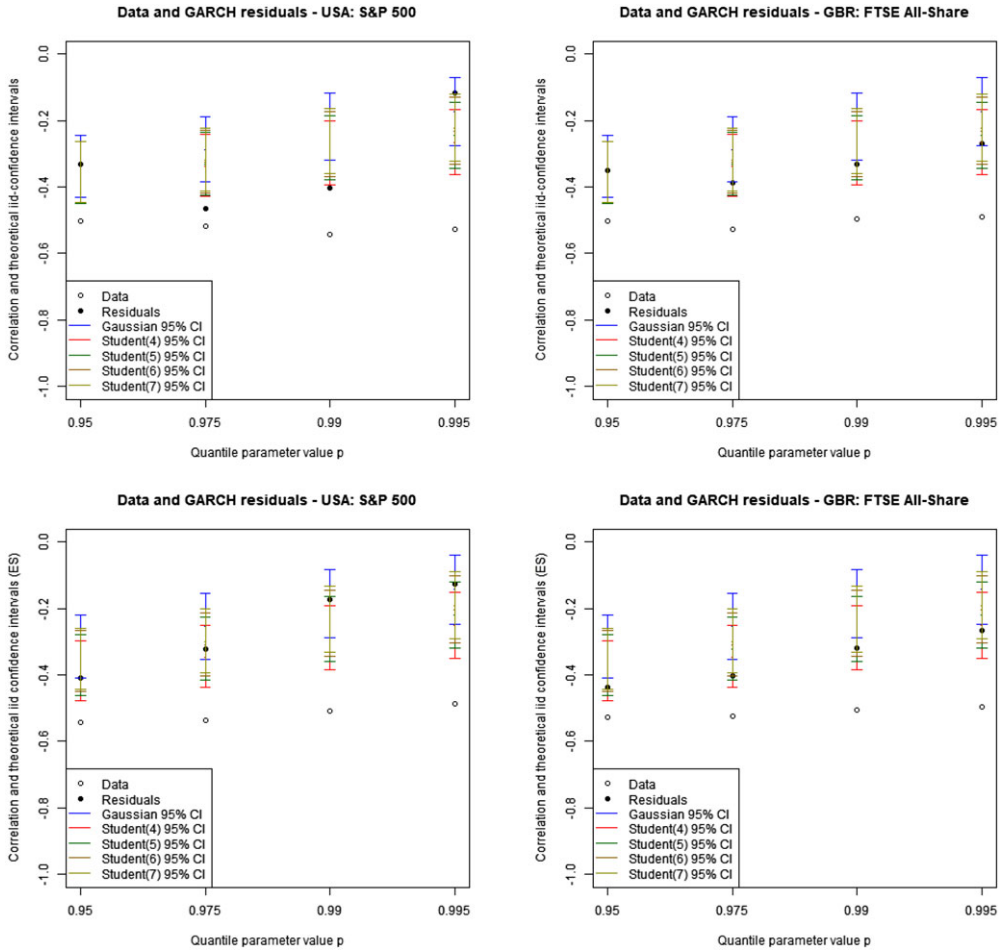


Figure 3. Comparison of pro-cyclicality for the real data (blank circle) with the pro-cyclicality for the GARCH(1, 1)-residuals (filled circle), considering the S&P500 (on the left) and the FTSE (on the right). The first row considers the VaR as risk measure (estimator $\widehat{\text{VaR}}_n(p)$), the second row the ES (estimator $\widetilde{\text{ES}}_{n,\infty}$). Each plot contains the correlation for the four different quantile values p . For each of them, corresponding theoretical confidence intervals (for the sample correlation) assuming a specific underlying distribution (Gaussian or Student with different degrees of freedom) are plotted.

we are able to quantify the pro-cyclicality by the degree of negative dependence between the components of this distribution.

As the result reveals a negative correlation, it shows that pro-cyclicality exists, whatever the quantile-based risk measure, VaR, ES, or expectile, whether we consider a conditional or unconditional approach, and whatever the choice of dispersion measure to estimate the volatility.

To test the relevance of the theoretical results on finite samples, we undertake extensive simulation studies in the iid case. The results show in all cases a good approximation of the asymptotic negative dependence by the finite sample counterpart, highlighting the applicability of such asymptotic theorem. Using the corresponding closed-form solutions based on the theorem, we then compare the impact of choices of quantile-based risk measures and dispersion measures on the degree of pro-cyclicality, also taking into account the fatness of the tail of the underlying distribution.

Finally, in a case study on real data, we verify that historical risk measurements indeed lead to pro-cyclicality, irrespectively of the choice of quantile-based risk measures and dispersion measures, but also irrespectively of the choice of conditional versus unconditional approaches for historical risk estimation.

While we focused on the impact of historical estimation on pro-cyclicality, further research could be done when considering other risk estimation methods, with different (parametric or not) models, as e.g. (GARCH-)EVT one. Other popular stochastic processes for modelling financial returns (like e.g. ARMA-GARCH) could be included in the framework by extending the class of processes currently considered; related theoretical work on the topic developed for a broader class of augmented GARCH processes (Bräutigam & Kratz, 2021) could be used as a starting point.

This work lays the ground for tackling pro-cyclicality at its root, namely the way risk is measured, rather than resorting to operational means as e.g. the transitional measures of Solvency 2 that temper with economic valuation. Our study opens the door for finding new direct ways to mitigate pro-cyclicality. Such a mitigation is sought after by risk managers and regulators, and our conclusions should help building a counter-cyclical risk measure, or at least a risk measure that limits the pro-cyclicality. This is what we are currently investigating.

Conflicts of interest. None.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S1748499523000039>

Acknowledgement. The authors would like to acknowledge the financial support of LabEx MME-DII (ANR-11-LBX-0023-01) during the PhD thesis of Marcel Bräutigam, who was awarded one of the two special mentions at the *Prix des Sciences du Risque* 2020.

References

- Acerbi, C. & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking & Finance*, **26**(7), 1487–1503.
- Adrian, T. & Shin, H. (2013). Pro-cyclical leverage and value-at-risk. *The Review of Financial Studies*, **27**(2), 373–403.
- Andrews, D. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory*, **4**(3), 458–467.
- Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1997). Thinking coherently. *Risk*, **10**, 68–71.
- Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999). Coherent measures of risks. *Mathematical Finance*, **9**, 203–228.
- Athanasoglou, P., Daniilidis, I. & Delis, M. (2014). Bank procyclicality and output: issues and policies. *Journal of Economics and Business*, **72**, 58–83.
- Aue, A., Hörmann, S., Horváth, L. & Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, **37**(6B), 4046–4087.
- Basel Committee on Banking Supervision. (2019). The basel framework. Available online at the address https://www.bis.org/basel_framework/.
- Behn, M., Haselmann, R. & Wachtel, P. (2016). Pro-cyclical capital regulation and lending. *Journal of Finance*, **71**(2), 919–955.
- Bellini, F. & Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, **15**(5), 725–733.
- Bellini, F., Cesarone, F., Colombo, C. & Tardella, F. (2021). Risk parity with expectiles. *European Journal of Operational Research*, **291**, 1149–1163.
- Bellini, F. & Di Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance*, **23**(6), 487–506.
- Boussama, F. (1998). *Ergodicité, mélange et estimation dans les modeles GARCH*. PhD thesis, Université 7 Paris.
- Bräutigam, M. (2020). *Pro-cyclicality of Risk Measurements - Empirical Quantification and Theoretical Confirmation*. Theses, Sorbonne Université. Available online at the address <https://www.theses.fr/2020SORUS100>.
- Bräutigam, M., Dacorogna, M. & Kratz, M. (2023). Pro-cyclicality beyond business cycles. *Mathematical Finance*, **33**(2), pp. 308–341.

- Bräutigam, M. & Kratz, M.** (2021). Joint asymptotics for the sample quantile and measures of dispersion for functionals of mixing processes. arXiv:2111.07650.
- Carrasco, M. & Chen, X.** (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, **18**(1), 17–39.
- Catarineu-Rabell, E., Jackson, P. & Tsomocos, D.P.** (2005). Procyclicality and the new basel accord-banks' choice of loan rating system. *Economic Theory*, **26**(3), 537–557.
- Chen, S.X.** (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, **6**(1), 87–107.
- Duan, J.** (1997). Augmented garch (p, q) process and its diffusion limit. *Journal of Econometrics*, **79**(1), 97–127.
- Ekström, M.** (2014). A general central limit theorem for strong mixing sequences. *Statistics & Probability Letters*, **94**, 236–238.
- Emmer, S., Kratz, M. & Tasche, D.** (2015). What is the best risk measure in practice? a comparison of standard risk measures. *Journal of Risk*, **18**(2), 31–60.
- European Banking Authority. (2019). Results from the 2018 market risk benchmarking exercise. EBA report.
- Fisher, R.** (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.
- Föllmer, H. & Schied, A.** (2016). Stochastic finance. In *Stochastic Finance*. de Gruyter.
- Jarrow, R.** (2017). *The Economic Foundations of Risk Management. Theory, Practice, and Applications*. World Scientific.
- Kuan, C.-M., Yeh, J.-H. & Hsu, Y.-C.** (2009). Assessing value at risk with care, the conditional autoregressive expectile models. *Journal of Econometrics*, **150**(2), 261–270.
- Lee, O.** (2014). Functional central limit theorems for augmented garch (p, q) and figarch processes. *Journal of the Korean Statistical Society*, **43**(3), 393–401.
- Llacay, B. & Pfeffer, G.** (2019). Impact of Basel III countercyclical measures on financial stability: an agent-based model. *Journal of Artificial Societies & Social Simulation*, **22**(1), 6. <https://doi.org/10.18564/jasss.3927>.
- McNeil, A.J., Frey, R. & Embrechts, P.** (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition*. Princeton University Press, Princeton, NJ.
- Mikosch, T. & Stărică, C.** (2000). Limit theory for the sample autocorrelations and extremes of a garch (1,1) process. *The Annals of Statistics*, **28**(5), 1427–1451.
- Miller, M.** (2018). *Quantitative Financial Risk Management*. Wiley.
- Morgan, J. & Reuters.** (1996). Riskmetrics - technical document. Available online at the address <https://www.msci.com/documents/10199/5915b101-4206-4ba0-ae2-3449d5c7e95a>.
- Newey, W. & Powell, J.** (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, **55**, 819–847.
- Pérignon, C. & Smith, D.R.** (2010). The level and quality of value-at-risk disclosure by commercial banks. *Journal of Banking & Finance*, **34**(2), 362–377.
- Politis, D., Romano, J. & Wolf, M.** (1997). Subsampling for heteroskedastic time series. *Journal of Econometrics*, **81**(2), 281–317.
- Quagliariello, M.** (2008). Does macroeconomy affect bank stability? a review of the empirical evidence. *Journal of Banking Regulation*, **9**(2), 102–115.
- Repetto, R. & Saurina, J.** (2011). The countercyclical capital buffer of Basel III: a critical assessment. CEPR Discussion Paper 8304, Available online at the address https://cepr.org/active/publications/discussion_papers/dp.php?dpno=8304.
- Rockafellar, R. & Uryasev, S.** (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, **26**(7), 1443–1471.
- Rodriguez, R.** (1982). Correlation. In **S. Kotz, N. Balakrishnan, C. Read, B. Vidakovic & N. Johnson** (Eds.), *Encyclopedia of Statistical Sciences* (pp. 1375–1385), 2nd edition. Wiley, New York.
- Roussas, G.G. & Ioannides, D.** (1987). Moment inequalities for mixing sequences of random variables. *Stochastic Analysis and Applications*, **5**(1), 60–120.
- RTS (European Union). (2013). Commission delegated regulation (eu) no 153/2013. *Official Journal of the European Union L*, **52**, 41–70.
- Rubio, M. & Carrasco-Gallego, J.A.** (2016). The new financial regulation in Basel III and monetary policy: a macroprudential approach. *Journal of Financial Stability*, **26**, 294–305.
- Rüschendorf, L.** (2013). *Mathematical Risk Analysis*. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-33590-7>.
- Solvency 2 Directive. (2009). European union directive 2009/138/ec (Solvency 2). *Official Journal of European Union L*, 335.
- Wendler, M.** (2011). Bahadur representation for u-quantiles of dependent data. *Journal of Multivariate Analysis*, **102**(6), 1064–1079.
- Yao, Q. & Tong, H.** (1996). Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, **6**(2–3), 273–292.
- Ziegel, J.F.** (2016). Coherence and elicibility. *Mathematical Finance*, **26**(4), 901–918.

APPENDIX

The Appendix consists of four parts. The first one, Appendix A, given here, contains the proof of Theorem 1 and all the necessary auxiliary results (as for example the (F)CLTs between risk measure estimators and the r -th absolute centred sample moment in Appendix A.1). To ease its understanding, an outline of the structure of the proof is given at the beginning. The three other appendices are available online, with parts taken from Bräutigam *et al.* (2022) and included only for self-containedness. We explicitly point this out throughout the online appendix where applicable. Appendix B provides additional material related to Section 4; first, the explicit formulas for the examples computed in Section 4, then, additional plots illustrating the pro-cyclicality behaviour in the iid case for a Student t distribution with different degrees of freedom. Appendix C presents additional plots related to Section 5. We display the plots of the pro-cyclicality of the residuals for all 11 indices considered – in the case of VaR and ES; the results for the VaR, which can be found in Bräutigam *et al.* (2022), are included here for comparative completeness. Finally, all tables corresponding to the simulation study developed in Section 4 are given in Appendix D.

Note that for brevity and notational convenience, we refer to the asymptotic theorem, Theorem 1, throughout the appendix as CLT, although in the main body of the paper we avoided this terminology to point out that the given theorem is not simply a limit theorem based on an average of random variables, but needs more work than that.

Appendix A. Proof of Theorem 1

Let us start by briefly laying out the structure of the proof to make the partition into the given subsections comprehensible. Recalling Equation (13), we claimed that, in the iid case, we can reduce the CLT of Theorem 1 (i.e. between the log-ratio $\log \left| \frac{\zeta_{n/2, t+n/2}^{(j)}(p)}{\zeta_{n/2, t}^{(j)}(p)} \right|$ and the r -th absolute central sample moment $\hat{m}(X, n/2, r, t)$) to a CLT between the risk measure estimator itself $\zeta_{n/2, t}^{(j)}(p)$ and the dispersion measure estimator $\hat{m}(X, n/2, r, t)$. For the family of augmented GARCH(p, q) processes, some additional assumptions on the dependence and existence of moments are needed, but this reduction is equally possible.

Thus, we start in Section A.1 collecting the FCLTs between risk measure and measure of dispersion estimators. While these results are known for the VaR and expectile estimators we consider (recalled in Section A.1.1), we need to establish them in the case of the ES estimator, $\widehat{ES}_n(p)$, as a novel result in the literature (Section A.1.2). Given these FCLT's, we present in Section A.2 the reduction theorem, which formally proves why the bivariate asymptotics between $\zeta_{n/2, t}^{(j)}(p)$ and $\hat{m}(X, n/2, r, t)$ are enough to deduce the trivariate asymptotics of $\log \left| \zeta_{n/2, t}^{(j)}(p) \right|$, $\log \left| \zeta_{n/2, t+n/2}^{(j)}(p) \right|$ and $\hat{m}(X, n/2, r, t)$. Finally, in Section A.3, the proof of Theorem 1 is given, which, after all this preparation, is a simple application of the reduction theorem.

Note that, as mentioned, the results in the iid case hold with weaker conditions, as a special case of the family of augmented GARCH(p, q) processes. Here, we present and prove more general results for augmented GARCH(p, q) processes. For direct proofs of all results in the iid case, we refer to Bräutigam *et al.* (2022) (for the VaR) or Bräutigam (2020).

A.1 FCLT's between risk and dispersion measure estimators

In this section, we provide the FCLT's between $\zeta_n^{(j)}(p)$, $j = 1, \dots, 4$, and $\hat{m}(X, n, r)$ when considering augmented GARCH(p, q) processes. We separately consider risk measure estimators based on the sample quantile (which we call "VaR-based") in Section A.1.1 and the Expected Shortfall estimator in Section A.1.2.

A.1.1 VaR-based risk measure estimators

The bivariate FCLT for the estimator $\widehat{\text{VaR}}_n(p)$, already proven in Bräutigam (2020, Theorem 4.3), is stated here for completeness only.

To ease its presentation, we introduce a trivariate normal random vector (functionals of X), $(U, V, W)^T$, with mean zero and the following covariance matrix:

$$(D) \left\{ \begin{aligned} \text{Var}(U) &= \text{Var}(X_0) + 2 \sum_{i=1}^{\infty} \text{Cov}(X_i, X_0) \\ \text{Var}(V) &= \text{Var}(|X_0|^r) + 2 \sum_{i=1}^{\infty} \text{Cov}(|X_i|^r, |X_0|^r) \\ \text{Var}(W) &= \frac{p(1-p)}{f_X^2(q_X(p))} + \frac{2}{f_X^2(q_X(p))} \sum_{i=1}^{\infty} (\mathbb{E}[\mathbb{1}_{(X_0 \leq q_X(p))} \mathbb{1}_{(X_i \leq q_X(p))}] - p^2) \\ \text{Cov}(U, V) &= \sum_{i \in \mathbb{Z}} \text{Cov}(|X_i|^r, X_0) = \sum_{i \in \mathbb{Z}} \text{Cov}(|X_0|^r, X_i) \\ \text{Cov}(U, W) &= \frac{-1}{f_X(q_X(p))} \sum_{i \in \mathbb{Z}} \text{Cov}(\mathbb{1}_{(X_i \leq q_X(p))}, X_0) = \frac{-1}{f_X(q_X(p))} \sum_{i \in \mathbb{Z}} \text{Cov}(\mathbb{1}_{(X_0 \leq q_X(p))}, X_i) \\ \text{Cov}(V, W) &= \frac{-1}{f_X(q_X(p))} \sum_{i \in \mathbb{Z}} \text{Cov}(|X_0|^r, \mathbb{1}_{(X_i \leq q_X(p))}) = \frac{-1}{f_X(q_X(p))} \sum_{i \in \mathbb{Z}} \text{Cov}(|X_i|^r, \mathbb{1}_{(X_0 \leq q_X(p))}). \end{aligned} \right.$$

Theorem 4 (Theorem 4.3 in Bräutigam, 2020.) *For an integer $r > 0$, consider an augmented GARCH(p, q) process X as defined in (11) and (12) satisfying condition (Lee), (C_0) at 0 for $r = 1$, and both conditions (C'_2) (see (C'_1)), (P) at $q_X(p)$. Assume also conditions (M_r) (see (M_k)), (A) , and either $(P_{\max(1,r/\delta)})$ (see (P_v)) for X belonging to the group of polynomial GARCH, or (L_r) (see (L_v)) for the group of exponential GARCH. Introducing the random vector $T_{n,r}(X) = \begin{pmatrix} q_n(p) - q_X(p) \\ \hat{m}(X, n, r) - m(X, r) \end{pmatrix}$, we have the following FCLT: For $t \in [0, 1]$, as $n \rightarrow \infty$,*

$$\sqrt{n} t T_{[nt],r}(X) \xrightarrow{D_2[0,1]} \mathbf{W}_{\Gamma^{(r)}}(t),$$

where $(\mathbf{W}_{\Gamma^{(r)}}(t))_{t \in [0,1]}$ is the 2-dimensional Brownian motion with covariance matrix $\Gamma^{(r)} \in \mathbb{R}^{2 \times 2}$ defined, for any $(s, t) \in [0, 1]^2$, by $\text{Cov}(\mathbf{W}_{\Gamma^{(r)}}(t), \mathbf{W}_{\Gamma^{(r)}}(s)) = \min(s, t)\Gamma^{(r)}$, where

$$\begin{aligned} \Gamma_{11}^{(r)} &= \text{Var}(W), \\ \Gamma_{22}^{(r)} &= r^2 \mathbb{E}[X_0^{r-1} \text{sgn}(X_0)^r]^2 \text{Var}(U) + \text{Var}(V) - 2r \mathbb{E}[X_0^{r-1} \text{sgn}(X_0)^r] \text{Cov}(U, V), \\ \Gamma_{12}^{(r)} &= \Gamma_{21}^{(r)} = -r \mathbb{E}[X_0^{r-1} \text{sgn}(X_0)^r] \text{Cov}(U, W) + \text{Cov}(V, W), \end{aligned}$$

$(U, V, W)^T$ being the trivariate normal vector (functionals of X) with mean zero and covariance given in (D), all series being absolute convergent.

Note that the conditions for establishing such a CLT are weaker than (S_1^*) , since we are not considering the log-ratio, but only the asymptotics between the risk measure and the measure of dispersion estimators. Notably we do not require the process to be strongly mixing with geometric rate and the moment condition (M_r) (see (M_k)) is sufficient (instead of $(M_{r+\tau})$, for $\tau > 0$).

Extension to the estimators $\zeta_n^{(j)}(p)$ for $j = 3, 4$, which are expressed as a VaR estimator. Theorem 4 can also be applied to establish a FCLT for $\zeta_n^{(3)}(p) = \hat{e}_n(p) = \widehat{\text{VaR}}_n(\kappa^{-1}(p))$ for κ given. Also, it can

be directly extended to a FCLT for a k -vector of estimators $\widehat{\text{VaR}}_n(p_i), i = 1, \dots, k$. Applying then the continuous mapping theorem yields the case of $\zeta_n^{(4)}(p) = \widehat{\text{ES}}_{n,k}(p)$.

A.1.2 ES-based risk measure estimators

Contrary to the other presented risk measure estimators, the asymptotics of $\widehat{\text{ES}}_n(p)$ with $\hat{m}(X, n, r)$ have not yet been proven in the literature. To do so and establish the bivariate FCLT for $\widehat{\text{ES}}_n(p)$, we proceed in a similar way to the case of $\widehat{\text{VaR}}_n(p)$. We introduce, to ease the presentation of the FCLT, a 4-dimensional normal random vector (functionals of X), $(U, V, \tilde{W}, R)^T$, with mean zero and the following covariance matrix:

$$(\tilde{D}) \left\{ \begin{aligned}
 \text{Var}(U) &= \text{Var}(X_0) + 2 \sum_{i=1}^{\infty} \text{Cov}(X_i, X_0), \\
 \text{Var}(V) &= \text{Var}(|X_0|^r) + 2 \sum_{i=1}^{\infty} \text{Cov}(|X_i|^r, |X_0|^r), \\
 \text{Var}(\tilde{W}) &= q_X^2(p) \left(\text{Var}(\mathbb{1}_{(X_0 \geq q_X(p))}) + 2 \sum_{i=1}^{\infty} \text{Cov}(\mathbb{1}_{(X_i \geq q_X(p))}, \mathbb{1}_{(X_0 \geq q_X(p))}) \right), \\
 \text{Var}(R) &= \text{Var}(X_0 \mathbb{1}_{(X_0 \geq q_X(p))}) + 2 \sum_{i=1}^{\infty} \text{Cov}(X_i \mathbb{1}_{(X_i \geq q_X(p))}, X_0 \mathbb{1}_{(X_0 \geq q_X(p))}), \\
 \text{Cov}(U, V) &= \sum_{i \in \mathbb{Z}} \text{Cov}(|X_i|^r, X_0) = \sum_{i \in \mathbb{Z}} \text{Cov}(|X_0|^r, X_i), \\
 \text{Cov}(U, \tilde{W}) &= q_X(p) \sum_{i \in \mathbb{Z}} \text{Cov}(\mathbb{1}_{(X_i \geq q_X(p))}, X_0) = q_X(p) \sum_{i \in \mathbb{Z}} \text{Cov}(\mathbb{1}_{(X_0 \geq q_X(p))}, X_i), \\
 \text{Cov}(V, \tilde{W}) &= q_X(p) \sum_{i \in \mathbb{Z}} \text{Cov}(|X_0|^r, \mathbb{1}_{(X_i \geq q_X(p))}) = q_X(p) \sum_{i \in \mathbb{Z}} \text{Cov}(|X_i|^r, \mathbb{1}_{(X_0 \geq q_X(p))}), \\
 \text{Cov}(\tilde{W}, R) &= q_X(p) \sum_{i \in \mathbb{Z}} \text{Cov}(X_i \mathbb{1}_{(X_i \geq q_X(p))}, \mathbb{1}_{(X_0 \geq q_X(p))}) \\
 &= q_X(p) \sum_{i \in \mathbb{Z}} \text{Cov}(X_0 \mathbb{1}_{(X_0 \geq q_X(p))}, \mathbb{1}_{(X_i \geq q_X(p))}), \\
 \text{Cov}(U, R) &= \sum_{i \in \mathbb{Z}} \text{Cov}(X_i \mathbb{1}_{(X_i \geq q_X(p))}, X_0) = \sum_{i \in \mathbb{Z}} \text{Cov}(X_0 \mathbb{1}_{(X_0 \geq q_X(p))}, X_i), \\
 \text{Cov}(V, R) &= \sum_{i \in \mathbb{Z}} \text{Cov}(|X_0|^r, X_i \mathbb{1}_{(X_i \geq q_X(p))}) = \sum_{i \in \mathbb{Z}} \text{Cov}(|X_i|^r, X_0 \mathbb{1}_{(X_0 \geq q_X(p))}).
 \end{aligned} \right.$$

Using this 4-dimensional vector, we can now describe the joint asymptotic distribution of $\widehat{\text{ES}}_n(p)$ and $\hat{m}(X, n, r)$. Here also, the conditions given in Proposition 5 are less restrictive than S_2^* . Nevertheless, in contrast to the VaR case, strong mixing with geometric rate is necessary (and, for $r = 1$, also a slightly stronger moment condition than (M_r) – see (M_k)). A more detailed comparison on the difference between the conditions for VaR and ES is given in Remark 7.

Proposition 5 Consider an augmented GARCH(p, q) process X as defined in (11) and (12), strongly mixing with geometric rate and satisfying the (Lee) condition. For any integer $r > 0$, assume that: (M_r) (see (M_k)) and (A) hold, F_X is absolutely continuous, (C_3) (see (C_l)) holds in a neighbourhood of $q_X(p)$, and all 2nd partial derivatives of the joint distribution of (X_1, X_{k+1}) , for $k \geq 1$, are bounded in a neighbourhood of $q_X(p)$. Assume also either $(P_{\max(1, r/\delta)})$ (see (P_v)) for polynomial GARCH, or (L_r) (see (L_v)) for exponential GARCH and, if $r = 1$, (C_0) at the mean μ and $(M_{1+\delta})$ (see (M_k)) for some $\delta > 0$.

Introducing the random vector $T_{n,r}(X) = \left(\begin{matrix} \widehat{\text{ES}}_n(p) - \text{ES}_p \\ \widehat{m}(X, n, r) - m(X, r) \end{matrix} \right)$, for $r \in \mathbb{Z}$, we have the following FCLT. For $t \in [0, 1]$, as $n \rightarrow \infty$,

$$\sqrt{n} t T_{[nt],r}(X) \xrightarrow{D_2[0,1]} \mathbf{W}_{\Gamma^{(r)}}(t),$$

where $(\mathbf{W}_{\Gamma^{(r)}}(t))_{t \in [0,1]}$ is the 2-dimensional Brownian motion with covariance matrix $\Gamma^{(r)} \in \mathbb{R}^{2 \times 2}$ defined for any $(s, t) \in [0, 1]^2$ by $\text{Cov}(\mathbf{W}_{\Gamma^{(r)}}(t), \mathbf{W}_{\Gamma^{(r)}}(s)) = \min(s, t)\Gamma^{(r)}$, where

$$\begin{aligned} \Gamma_{11}^{(r)} &= \text{Var}(\tilde{W}) + \text{Var}(R) - 2\text{Cov}(\tilde{W}, R), \\ \Gamma_{22}^{(r)} &= r^2 \mathbb{E}[X_0^{r-1} \text{sgn}(X_0)^r]^2 \text{Var}(U) + \text{Var}(V) - 2r \mathbb{E}[X_0^{r-1} \text{sgn}(X_0)^r] \text{Cov}(U, V), \\ \Gamma_{12}^{(r)} &= \Gamma_{21}^{(r)} = \text{Cov}(R, V) - \text{Cov}(\tilde{W}, V) - r \mathbb{E}[X_0^{r-1} \text{sgn}(X_0)^r] \text{Cov}(R, U) \\ &\quad + r \mathbb{E}[X_0^{r-1} \text{sgn}(X_0)^r] \text{Cov}(\tilde{W}, U), \end{aligned}$$

$(U, V, \tilde{W}, R)^T$ being the 4-dimensional normal vector (functionals of X) with mean zero and covariance given in (\tilde{D}) , all series being absolutely convergent.

Remark 6 How restrictive is the condition of strong mixing with geometric rate for the augmented GARCH(p, q) processes? While we cannot make a general statement covering all cases, there exist different results in the literature linking GARCH processes and strong mixing: Boussama proves in Boussama (1998), Theorem 3.4.2, the strong mixing with geometric rate for GARCH(p,q) processes. Carrasco & Chen (2002) prove in Proposition 5(i), that a big class of augmented GARCH(1,1) processes are strongly mixing with geometric rate. Therein, in Proposition 12, they also prove that power GARCH(p,q) (PGARCH) processes are strongly mixing with geometric rate.

Remark 7 Comparing the conditions in Proposition 5 with those for $\widehat{\text{VaR}}_n(p)$ in Theorem 4, we see that we need here the absolute continuity of F_X and the continuity of the second derivative of f_X in a neighbourhood of $q_X(p)$ (instead of (C'_2) (see (C'_1)) and (P) at $q_X(p)$). Also, for $r = 1$, we need $(M_{1+\delta})$ instead of (M_1) (see (M_k)). Note that those conditions are sufficient to obtain the CLT in the iid case. Here, we also require the process X to be strongly mixing with geometric rate, as well as all second partial derivatives of the joint distribution of (X_1, X_{k+1}) , for $k \geq 1$, to be bounded (in a neighbourhood of $q_X(p)$). These conditions come from the use of the Bahadur representation of the ES, see Chen (2008).

Proof of Proposition 5 The proof follows the lines of the corresponding FCLT between the sample quantile and the r -th absolute centred sample moment (Theorem 4.3 in Bräutigam, 2020), also keeping the same structure of the proof in four steps: In Step 1, we check the conditions of the Bahadur representation of the ES. Obviously, this is different than Step 1 in the case of the VaR. Checking the conditions of the representation of the r -th absolute sample moment in Step 2 is identical to the VaR-case and only recalled for self-containedness. In Step 3, we check the conditions to apply the FCLT. Naturally, it is what requires most of the work. But as the Bahadur representation for the ES contains elements of the Bahadur representation for the VaR, we can partially use results from the VaR case. Finally, Step 4 has exactly the same reasoning as in the case of the VaR, even if using the ES.

Step 1: Bahadur representation of the ES – conditions.

We want to use the Bahadur representation of the ES. Such a representation holds under the necessary conditions (i) and (ii) as given in Chen (2008), which here are fulfilled by assumption:

- (i) The process X is strongly mixing with geometric rate.
- (ii) The stationarity of the process follows from Assumption $(P_{\max(1,r/\delta)})$ (see (P_v)) or (L_r) (see (L_v)), respectively, with Lemma 1 of Lee (2014). The conditions on continuity and moments imposed by Chen (2008) are fulfilled by assumption, namely, the absolute continuity of F_X , continuous second derivative of f_X in a neighbourhood of $q_X(p)$, the boundedness in a neighbourhood of $q_X(p)$ of all 2nd partial derivatives of the joint distribution of (Y_1, Y_{k+1}) for $k \geq 1$.

Thus, we can apply the Bahadur representation of the ES:

$$\widehat{ES}_n(p) - ES_p = \frac{1}{(1-p)n} \sum_{i=1}^n (X_i - q_X(p)) \mathbb{1}_{(X_i \geq q_X(p))} - (ES_p - q_X(p)) + o_p(n^{-3/4+\kappa}), \quad (A.1)$$

for an arbitrary $\kappa > 0$.

Step 2: Representation of the r -th absolute centred sample moment – Conditions.

This step is exactly the same as for the FCLT with the VaR and can be found in the proof of Theorem 4.3 in Bräutigam (2020). We recall it for self-containedness:

As shown in Proposition 4.8 in Bräutigam (2020), we have a representation of the r -th absolute centred sample moment under the following conditions:

- A stationary and ergodic time-series $(X_n, n \geq 1)$ with “short-memory”, i.e. $\sum_{i=0}^{\infty} |\text{Cov}(X_0, X_i)| < \infty$
- An existing r -th moment of X_0 and (C_0) at μ for $r = 1$.

Under that conditions, it holds, as $n \rightarrow \infty$, that

$$\begin{aligned} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}_n|^r \right) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n |X_i - \mu|^r \right) \\ &\quad - r\sqrt{n}(\bar{X}_n - \mu) \mathbb{E}[(X_0 - \mu)^{r-1} \text{sgn}(X_0 - \mu)^r] + o_p(1). \end{aligned} \quad (A.2)$$

We recall here, why these conditions are satisfied:

1. As mentioned in Step 1, the stationarity follows from assumption $(P_{\max(1,r/\delta)})$ (see (P_v)) or (L_r) (see (L_v)), respectively, and Lemma 1 of Lee (2014).
2. For the moment condition, short-memory property and ergodicity, we simply verify that the conditions for a CLT of X_t^r (or $|X_t|^r$) are fulfilled, distinguishing between the polynomial and exponential case. Conditions (M_r) (see (M_k)), (A) , $(P_{\max(1,r/\delta)})$ (see (P_v)) in the polynomial case, and (M_r) , (A) , (L_r) (see (L_v)) in the exponential case respectively, imply the CLT, using Corollary 2 and 3 in Lee (2014), respectively.
3. Finally, (C_0) at μ for $r = 1$ holds by assumption.

Step 3: Conditions for applying the FCLT

We adapt Step 3 in the proof of Theorem 4.3 in Bräutigam (2020) to the ES instead of the VaR. Here, we are using a four-dimensional version of the FCLT (Lemma 4.9 in Bräutigam, 2020, a slight modification of Theorem A.1 in Aue et al., 2009), which we cite here for self-containedness and to better understand the proof.

Lemma 8 (Theorem A.1 in Aue et al., 2009) Consider a d -dimensional random process $(u_j, j \in \mathbb{Z})$, which is centred and has finite variance, i.e.

$$\mathbb{E}[u_j] = 0, \quad \|u_j\|_2^2 < \infty \quad \forall j \in \mathbb{Z}, \quad (A.3)$$

and has a causal (possibly non-linear) representation in terms of an iid process, i.e.

$$u_j = f(\epsilon_j, \epsilon_{j-1}, \dots), \quad (A.4)$$

where $f : \mathbb{R}^{1 \times \infty} \rightarrow \mathbb{R}^d$ is a measurable function and $(\epsilon_j, j \in \mathbb{Z})$ is a sequence of real valued iid rv's with mean 0 and variance 1.

Suppose further, there exists a Δ -dependent approximation of u_j , i.e. a sequence of d -dimensional random vectors $(u_j^{(\Delta)}, j \in \mathbb{Z})$ such that, for any $\Delta \geq 1$, we have

$$u_j^{(\Delta)} = f^{(\Delta)}(\epsilon_{j-\Delta}, \dots, \epsilon_j, \dots, \epsilon_{j+\Delta}) \tag{A.5}$$

$$\text{and } \sum_{\Delta \geq 1} \|u_0 - u_0^{(\Delta)}\|_2 < \infty, \tag{A.6}$$

where $f^{(\Delta)} : \mathbb{R}^{1 \times (2\Delta+1)} \rightarrow \mathbb{R}^d$ is a measurable function.

Then, the series $\Gamma = \sum_{j \in \mathbb{Z}} \text{Cov}(u_0, u_j)$ converges (coordinatewise) absolutely and a FCLT holds for $U_n := \frac{1}{n} \sum_{j=1}^n u_j$

$$\sqrt{nt}U_{[nt]} \xrightarrow{D_d[0,1]} W_\Gamma(t),$$

where the convergence takes place in the d -dimensional Skorohod space $D_d[0, 1]$ and $(W_\Gamma(t), t \in [0, 1])$ is a d -dimensional Brownian motion with covariance matrix Γ , i.e. it has mean 0 and $\text{Cov}(W_\Gamma(s), W_\Gamma(t)) = \min(s, t)\Gamma$.

Anticipating the use of this Lemma in Step 4 to establish the FCLT for $U_n(X) := \frac{1}{n} \sum_{j=1}^n u_j$, where

$$u_j = \begin{pmatrix} X_j \\ |X_j|^r - m(X, r) \\ q_X(p)\mathbb{1}_{(X_j \geq q_X(p))} - (1-p)q_X(p) \\ X_j\mathbb{1}_{(X_j \geq q_X(p))} - \mathbb{E}[X_j\mathbb{1}_{(X_j \geq q_X(p))}] \end{pmatrix},$$

we verify that its conditions (Equations (A.3)–(A.6)) hold. We have that u_j fulfils (A.3) as $\mathbb{E}[u_j] = 0$ holds by construction, and $\mathbb{E}[|X_j|^{2r}] < \infty$ is guaranteed since $|X_t|^r$ satisfies a CLT (see Step 2), thus also $\mathbb{E}[u_j^2] < \infty$. As we assume (A), it follows from Lemma 1 in Lee (2014) that $X_j = f(\epsilon_j, \epsilon_{j-1}, \dots)$. This latter relation also holds for functionals of X_j , i.e. u_j , thus (A.4) holds.

Then, we define a Δ -dependent approximation $u_0^{(\Delta)}$ satisfying (A.5) and (A.6).

Denote, for the ease of notation, $X_{0\Delta} := \mathbb{E}[X_0 | \mathcal{F}_{-\Delta}^+]$, and set

$$u_0^{(\Delta)} = \begin{pmatrix} X_{0\Delta} \\ \mathbb{E}[|X_0|^r | \mathcal{F}_{-\Delta}^+] - m(X, r) \\ q_X(p)\mathbb{1}_{(X_{0\Delta} \geq q_X(p))} - (1-p)q_X(p) \\ X_{0\Delta}\mathbb{1}_{(X_{0\Delta} \geq q_X(p))} - \mathbb{E}[X_j\mathbb{1}_{(X_j \geq q_X(p))}] \end{pmatrix}$$

with $\mathcal{F}_s^t = \sigma(\epsilon_s, \dots, \epsilon_t)$ for $s \leq t$. Thus, (A.5) is fulfilled by construction. Let us verify (A.6). We can write

$$\begin{aligned} \sum_{\Delta \geq 1} \|u_0 - u_0^{(\Delta)}\|_2 &\leq \sum_{\Delta \geq 1} \left(\|X_0 - X_{0\Delta}\|_2 + \||X_0|^r - \mathbb{E}[|X_0|^r | \mathcal{F}_{-\Delta}^+]\|_2 \right. \\ &\quad \left. + q_X^2(p) \|\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_2 + \|X_0\mathbb{1}_{(X_0 \geq q_X(p))} - X_{0\Delta}\mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_2 \right). \end{aligned} \tag{A.7}$$

Obviously, a sufficient condition for (A.7) is the finiteness of its summands. Note that we have already shown the finiteness for the first three parts of the sum in (A.7) (in Step 3 of the proof

of Theorem 4.3 in Bräutigam, 2020) which we recall here for self-containedness: If it holds that each summand is geometrically L_2 -NED, then its sum will be finite. E.g. assuming that X_0 is geometrically L_2 -NED, i.e. $\|X_0 - X_{0\Delta}\|_2 = O(e^{-\kappa\Delta})$ for some $\kappa > 0$, it follows that $\sum_{\Delta \geq 1} \|X_0 - X_{0\Delta}\|_2 < \infty$.

The condition of geometric L_2 -NED of X_0 and $|X_0^*|$ is satisfied, on the one hand in the polynomial case under (M_r) (see (M_k)), (A) and $(P_{\max(1,r/\delta)})$ (see (P_v)) via Corollary 2 in Lee (2014), on the other hand in the exponential case under (M_r) , (A) and (L_r) (see (L_v)) via Corollary 3 in Lee (2014). Thus, as X_0 is geometric L_2 -NED, this follows also for its bounded functional $\mathbb{1}_{(X_0 \leq q_X(p))}$ using Lemma 3.5 in Wendler (2011).

Finally, we only need to consider the fourth sum. This follows directly by an algebraic manipulation. Using first the triangle inequality, then the Hölder inequality (with $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$), we have

$$\begin{aligned} & \|X_0 \mathbb{1}_{(X_0 \geq q_X(p))} - X_{0\Delta} \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_2 \\ &= \|X_0 (\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}) + \mathbb{1}_{(X_{0\Delta} \geq q_X(p))} (X_0 - X_{0\Delta})\|_2 \\ &\leq \|X_0 (\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))})\|_2 + \|\mathbb{1}_{(X_{0\Delta} \geq q_X(p))} (X_0 - X_{0\Delta})\|_2 \\ &\leq \|X_0\|_{2p} \|\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_{2q} + \|X_0 - X_{0\Delta}\|_2. \end{aligned}$$

Choosing $p = 1 + \delta$, for δ as in Proposition 5, $\|X_0\|_{2+2\delta}$ is finite by assumption. Further, note that we can write, for any q ,

$$\|\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_{2q} = \|\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_2^{1/q}.$$

As we have recalled above that $\sum_{\Delta \geq 1} \|X_0 - X_{0\Delta}\|_2 < \infty$ and $\|\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_2 = O(e^{-\kappa\Delta})$ for some $\kappa > 0$, then $\sum_{\Delta \geq 1} \|\mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_2^{1/q} < +\infty$. Hence, we can conclude that

$$\sum_{\Delta \geq 1} \|X_0 \mathbb{1}_{(X_0 \geq q_X(p))} - X_{0\Delta} \mathbb{1}_{(X_{0\Delta} \geq q_X(p))}\|_2 < \infty,$$

which means that (A.6) is fulfilled.

Step 4: Multivariate FCLT

Having checked the conditions for the FCLT of Lemma 8 in Step 3, we can apply a 4-dimensional FCLT for u_j

$$\sqrt{n} \frac{1}{n} \sum_{j=1}^{[nt]} u_j = \sqrt{n} t \left(\begin{array}{c} \bar{X}_{[nt]} \\ \frac{1}{[nt]} \sum_{j=1}^{[nt]} |X_j|^r - m(X, r) \\ \frac{q_X(p)}{[nt]} \sum_{j=1}^{[nt]} (\mathbb{1}_{X_j \geq q_X(p)} - (1-p)) \\ \frac{1}{[nt]} \sum_{j=1}^{[nt]} (X_j \mathbb{1}_{X_j \geq q_X(p)} - \mathbb{E}[X_j \mathbb{1}_{X_j \geq q_X(p)}]) \end{array} \right) \xrightarrow{D_4[0,1]} \mathbf{W}_{\tilde{\Gamma}^{(r)}}(t) \quad \text{as } n \rightarrow \infty, \tag{A.8}$$

where $\mathbf{W}_{\tilde{\Gamma}^{(r)}}(t), t \in [0, 1]$ is the 4-dimensional Brownian motion with covariance matrix $\tilde{\Gamma}^{(r)} \in \mathbb{R}^{4 \times 4}$, i.e. the components $\tilde{\Gamma}_{ij}^{(r)}, 1 \leq i, j \leq 4$, satisfy the dependence structure (\tilde{D}) , with all series being absolutely convergent.

Recalling the representation of $\hat{m}(X, n, r)$, (A.2), and the Bahadur representation (A.1) of the sample ES (ignoring the remainder terms for the moment), we apply to (A.8) the multivariate continuous mapping theorem using the function $f(w, x, y, z) \mapsto (aw + x, b(z - y))$ with $a = -r\mathbb{E}[(X - \mu)^{r-1} \text{sgn}(X - \mu)^r], b = 1/(1 - p)$, and obtain

$$\sqrt{n} t \left(\frac{a(\bar{X}_{[nt]}) + \frac{1}{[nt]} \sum_{j=1}^{[nt]} |X_j|^r - m(X, r)}{\frac{1}{1-p} \left(\frac{1}{[nt]} \sum_{j=1}^{[nt]} \mathbb{1}_{(X_j \geq q_X(p))} (X_j - q_X(p)) - (1-p)(ES_p - q_X(p)) \right)} \right) \xrightarrow{D_2[0,1]} \mathbf{W}_{\Gamma(r)}(t). \tag{A.9}$$

As by Slutsky’s theorem, a remainder term that converges in probability to 0, does not change the limiting distribution, we get from (A.9),

$$\sqrt{n} t \left(\frac{\hat{m}(X, [nt], r) - m(X, r)}{\widehat{ES}_{[nt]}(p) - ES_p} \right) \xrightarrow{D_2[0,1]} \mathbf{W}_{\Gamma(r)}(t),$$

where $\Gamma^{(r)}$ follows from the specifications of $\tilde{\Gamma}^{(r)}$ above and the continuous mapping theorem. □

A.2 Reduction Theorem

For the proof of Theorem 1, we establish a slightly more general result in Theorem 9, named reduction theorem, which form appears less related with the focus of our paper (hence the choice of putting it in the Appendix). To apply this reduction theorem to our setting of pro-cyclicality, we need to make a specific choice of functions f, g to recover Equation (13), then to prove the FCLT (A.10) for the class of augmented GARCH(p, q) processes.

Theorem 9 Consider a univariate, stationary stochastic process $(X_j, j \in \mathbb{Z})$. Assume that, for given real functions f and g , the bivariate rv $u_j := \begin{pmatrix} f(X_j) - \mathbb{E}[f(X_j)] \\ g(X_j) - \mathbb{E}[g(X_j)] \end{pmatrix}$ satisfies the FCLT, i.e.

$$\sqrt{nt} \begin{pmatrix} \sum_{j=1}^{[nt]} (f(X_j) - \mathbb{E}[f(X_j)]) / [nt] \\ \sum_{j=1}^{[nt]} (g(X_j) - \mathbb{E}[g(X_j)]) / [nt] \end{pmatrix} \xrightarrow{D_2[0,1]} \mathbf{W}_{\Gamma}(t), \quad \text{as } n \rightarrow \infty, \tag{A.10}$$

where $(\mathbf{W}_{\Gamma}(t))_{t \in [0,1]}$ is the 2-dimensional Brownian motion with covariance matrix $\Gamma(s, t) \in \mathbb{R}^{2 \times 2}$, defined for any $(s, t) \in [0, 1]^2$ by $\text{Cov}(\mathbf{W}_{\Gamma}(t), \mathbf{W}_{\Gamma}(s)) = \min(s, t)\Gamma$ (Γ being the covariance matrix of u_0). Define

$$Q_j = \begin{cases} 0 & \text{for } j \leq \lfloor n/2 \rfloor \\ f(X_j) & \text{for } j > \lfloor n/2 \rfloor \end{cases}, \quad Y_j = \begin{cases} f(X_j) & \text{for } j \leq \lfloor n/2 \rfloor \\ 0 & \text{for } j > \lfloor n/2 \rfloor \end{cases}, \quad Z_j = \begin{cases} g(X_j) & \text{for } j \leq \lfloor n/2 \rfloor \\ 0 & \text{for } j > \lfloor n/2 \rfloor \end{cases}. \tag{A.11}$$

Denote their sample averages (normalised to mean 0) as

$$\bar{Q}_n = \frac{1}{n} \sum_{j=1}^n (Q_j - \mathbb{E}[Q_j]), \quad \bar{Y}_n = \frac{1}{n} \sum_{j=1}^n (Y_j - \mathbb{E}[Y_j]), \quad \bar{Z}_n = \frac{1}{n} \sum_{j=1}^n (Z_j - \mathbb{E}[Z_j]). \tag{A.12}$$

Then, if the process X_j is strongly mixing with geometric rate and there exists a $\delta > 0$ s.t.

$$\mathbb{E}[|Q_j - \mathbb{E}[Q_j]|^{2+2\delta}] < \infty, \quad \mathbb{E}[|Y_j - \mathbb{E}[Y_j]|^{2+2\delta}] < \infty, \quad \mathbb{E}[|Z_j - \mathbb{E}[Z_j]|^{2+2\delta}] < \infty, \quad \forall j, \tag{A.13}$$

it holds that

$$\sqrt{n} \begin{pmatrix} \bar{Q}_n \\ \bar{Y}_n \\ \bar{Z}_n \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma), \text{ where } \Sigma = (\Sigma_{ij})_{1 \leq i, j \leq 3} \text{ satisfies } \Sigma_{ij} = \begin{cases} \frac{1}{2}\Gamma_{11} & \text{for } i = j \in \{1, 2\}, \\ \frac{1}{2}\Gamma_{22} & \text{for } i = j = 3, \\ \frac{1}{2}\Gamma_{12} & \text{for } i, j \in \{2, 3\} \text{ with } i \neq j, \\ 0 & \text{otherwise.} \end{cases} \tag{A.14}$$

Proof The proof consists of two steps. First, we need to establish univariate CLT’s for each of the components of the vector in (A.14), using a CLT for non-stationary strongly mixing sequences. Secondly, we argue why we can deduce the trivariate asymptotics directly via the Cramér-Wold device. To do so, we need to show that the covariances between estimators over disjoint samples vanish asymptotically. For this, we will use covariance bounds for strongly mixing processes.

Step 1: Univariate CLT’s

To establish the univariate CLT’s, we use a CLT for non-stationary sequences by Politis *et al.* (1997), Ekström (2014), which we simplify to our purposes, as follows:

Consider a stochastic process, denoted by $(W_j, j \in \mathbb{Z})$, which is strongly mixing with coefficient $\alpha(k)$. Denote $\bar{W}_n = \frac{1}{n} \sum_{j=1}^n W_j$ and $\sigma_n^2 = \text{Var}(\sqrt{n} \bar{W}_n)$. If the following three conditions hold,

$$\mathbb{E}[|W_j - E[W_j]|^{2+2\delta}] \leq c, \quad \forall j \tag{A.15}$$

$$\sigma^2 := \lim_n \sigma_n \in (0, \infty) \tag{A.16}$$

$$\sum_{k=0}^{\infty} (k+1)^2 \alpha(k)^{\delta/(4+\delta)} \leq d, \text{ for a finite constant } d \text{ independent of } k, \tag{A.17}$$

then $\sqrt{n} (\bar{W}_n - E[\bar{W}_n]) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ as $n \rightarrow \infty$.

Note that a stronger condition than (A.16), is introduced in Politis *et al.* (1997), namely

$$\forall (d_n) \text{ s.t. } d_n \rightarrow \infty: \sup_t |\text{Var}(\sqrt{d_n} \frac{1}{d_n} \sum_{j=t}^{t+d_n-1} W_j) - \sigma^2| \rightarrow 0, \text{ as } n \rightarrow \infty, \tag{A.18}$$

under which the authors conclude that $\frac{1}{\sqrt{d_n}} \sum_{i=1}^{d_n} X_i \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ holds (with $\sigma^2 := \lim_{n \rightarrow \infty} \sigma_n$) for any sequence $d_n \leq n$ such that $d_n \rightarrow \infty$ as $n \rightarrow \infty$. To ensure this asymptotic Gaussian behaviour, (A.18) is reasonable, i.e. the CLT should hold for any d_n with always the same variance σ^2 . In our case, we only need the CLT to hold for $d_n = n$ (and we do not care what would happen for other choices of d_n). This is why we consider Ekström (2014), where the author shows that (A.18) is actually superfluous, but at the price of accepting potentially degenerate limiting distributions. As a compromise between the two, we demand (A.16), which ensures that we do not have a degenerate limiting distribution for the case $d_n = n$.

The proof for each of the three univariate CLT’s is analogous. Thus, we prove it for Q_j and only state the results for the two other cases.

Let us verify the conditions (A.15)–(A.17) so that we can apply the CLT. First, we note that (A.15) corresponds, in our case, to our assumption (A.13), hence is satisfied. Direct computations lead to (A.16):

$$\begin{aligned} \sigma_Q^2 &= \lim_n \text{Var}(\sqrt{n} \bar{Q}_n) = \lim_n \frac{1}{n} \left(\sum_{j=n/2+1}^n \text{Var}(Q_j) + 2 \sum_{n/2+1 \leq i < j \leq n} \text{Cov}(Q_i, Q_j) \right) \\ &= \frac{1}{2} \text{Var}(f(X_0)) + \lim_n \frac{2}{n} \sum_{i=1}^{n/2-1} (n/2 - i) \text{Cov}(f(X_0), f(X_i)) \\ &= \frac{1}{2} \text{Var}(f(X_0)) + \sum_{i=1}^{\infty} \text{Cov}(f(X_0), f(X_i)), \end{aligned}$$

which is non-degenerate by (A.10).

As Q_j is a functional of X_j , we can bound from above the mixing coefficient of Q_j , denoted by $\alpha_Q(k)$, by the one of X_j , i.e. $\alpha_Q(k) \leq \alpha(k)$. As we know that X_j is strongly mixing with geometric rate, we have that $\alpha_Q(k) \leq C\lambda^k$ for some constants $C > 0$ and $\lambda \in (0, 1)$, which implies

$$\sum_{k=0}^{\infty} (k+1)^2 \alpha_Q(k)^{\delta/(4+\delta)} \leq \sum_{k=0}^{\infty} (k+1)^2 (C\lambda^k)^{\delta/(4+\delta)} = C^{\delta/(4+\delta)} \sum_{k=1}^{\infty} k^2 \lambda^{(k-1)\delta/(4+\delta)}.$$

We perform a ratio test to confirm the convergence of this series

$$L = \lim_{k \rightarrow \infty} \left| \frac{(k+1)^2 \lambda^{k\delta/(4+\delta)}}{k^2 \lambda^{(k-1)\delta/(4+\delta)}} \right| = \lim_{k \rightarrow \infty} \left| \left(1 + \frac{2}{k} + \frac{1}{k^2}\right) \lambda^{\delta/(4+\delta)} \right| = \lambda^{\delta/(4+\delta)} < 1.$$

Thus, the series is convergent, from which we deduce (A.17). We conclude to the CLT, as $n \rightarrow \infty$,

$$\sqrt{n}(\bar{Q}_n - \mathbb{E}[\bar{Q}_n]) \xrightarrow{d} \mathcal{N}(0, \sigma_Q^2).$$

In the same manner, we obtain, as $n \rightarrow \infty$,

$$\sqrt{n}(\bar{Y}_n - \mathbb{E}[\bar{Y}_n]) \xrightarrow{d} \mathcal{N}(0, \sigma_Y^2) \quad \text{and} \quad \sqrt{n}(\bar{Z}_n - \mathbb{E}[\bar{Z}_n]) \xrightarrow{d} \mathcal{N}(0, \sigma_Z^2),$$

where

$$\sigma_Q^2 = \sigma_Y^2 \quad \text{and} \quad \sigma_Z^2 = \frac{1}{2} \text{Var}(g(X_0)) + \sum_{i=1}^{\infty} \text{Cov}(g(X_0), g(X_i)).$$

Step 2: Trivariate CLT

By the Cramér-Wold Device, it suffices to show that all linear combinations of the components of $(\bar{Q}_n, \bar{Y}_n, \bar{Z}_n)^T$ are normally distributed, to conclude their trivariate normality.

For any $a, b, c \in \mathbb{R}$, we establish the CLT for

$$U_j := a(Q_j - \mathbb{E}[Q_j]) + b(Y_j - \mathbb{E}[Y_j]) + c(Z_j - \mathbb{E}[Z_j]),$$

i.e.

$$\sqrt{n} \sum_{j=1}^n U_j/n \xrightarrow{d} \mathcal{N}(0, \sigma^2), \text{ as } n \rightarrow \infty,$$

with σ^2 to be determined - in a similar way as in Step 1. Note that, by construction, $\mathbb{E}[U_j] = 0$. We need to verify the strong mixing of U_j and the three conditions (A.15)–(A.17). By the Minkowski inequality, we have that

$$\mathbb{E}[|U_j|^{2+2\delta}] = \|U_j\|_{2+2\delta}^{2+2\delta} \leq (a\|Q_j - \mathbb{E}[Q_j]\|_{2+2\delta} + b\|Y_j - \mathbb{E}[Y_j]\|_{2+2\delta} + c\|Z_j - \mathbb{E}[Z_j]\|_{2+2\delta})^{2+2\delta}.$$

Thus, (A.15) is fulfilled by the assumption (A.13).

By construction, each U_j is a functional of X_j (which is strongly mixing with geometric rate, by assumption). We can bound from above the mixing coefficient of U_j , denoted by $\alpha_U(k)$, by the one of X_j , i.e. $\alpha_U(k) \leq \alpha(k)$. Therefore, (A.12) holds by the same argumentation as in the univariate case.

So, we are left with computing $\sigma^2 = \lim_{n \rightarrow \infty} \sigma_n^2$. We write it as

$$\begin{aligned} \sigma_n^2 &= \text{Var}(\sqrt{n} \sum_{j=1}^n U_j/n) = \frac{1}{n} \text{Var}(\sqrt{n} \sum_{j=1}^n (aQ_j + bY_j + cZ_j)/n) \\ &= a^2 \text{Var}(\sqrt{n} \sum_{j=1}^n Q_j/n) + b^2 \text{Var}(\sqrt{n} \sum_{j=1}^n Y_j/n) + c^2 \text{Var}(\sqrt{n} \sum_{j=1}^n Z_j/n) \\ &\quad + 2ab \text{Cov}(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Y_i/n) + 2ac \text{Cov}(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n) \\ &\quad + 2bc \text{Cov}(\sqrt{n} \sum_{j=1}^n Y_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n). \end{aligned} \tag{A.19}$$

As this expression for σ_n^2 will involve some computations, we split it into different parts. First, note that the respective variances in (A.19) are known from the univariate asymptotics:

$$\lim_{n \rightarrow \infty} \text{Var}(\sqrt{n} \sum_{j=1}^n Q_j/n) = \sigma_Q^2, \quad \lim_{n \rightarrow \infty} \text{Var}(\sqrt{n} \sum_{j=1}^n Y_j/n) = \sigma_Y^2, \quad \lim_{n \rightarrow \infty} \text{Var}(\sqrt{n} \sum_{j=1}^n Z_j/n) = \sigma_Z^2. \tag{A.20}$$

Thus, we are left with the covariances that we assess one after the other.

- Computation of the first covariance of (A.19):

$$\begin{aligned} \text{Cov}(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Y_i/n) &= \frac{1}{n} \sum_{j=n/2+1}^n \sum_{i=1}^{n/2} \text{Cov}(f(X_j), f(X_i)) \\ &= \frac{1}{n} \sum_{j=n/2+1}^n \sum_{i=1}^{n/2} \text{Cov}(f(X_{j-i}), f(X_0)) \\ &= \frac{1}{n} \left(\sum_{k=1}^{n/2} k \text{Cov}(f(X_k), f(X_0)) + \sum_{k=n/2+1}^{n-1} (n-k) \text{Cov}(f(X_k), f(X_0)) \right) \\ &= \frac{1}{n} \left(\sum_{k=1}^{n/2} k \text{Cov}(f(X_k), f(X_0)) + \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) \text{Cov}(f(X_{k+n/2}), f(X_0)) \right), \end{aligned} \tag{A.21}$$

where we used the stationarity of the underlying process X .

To bound the two sums in (A.21), we use covariance bounds provided in Roussas & Ioannides (1987), Theorem 7.3. For convenience, we recall them here for a process $(X_j, j \in \mathbb{Z})$: For chosen positive integers $l, k > 0$,

1. if $f(X_k)$ is \mathcal{F}_{l+k}^∞ measurable and $f(X_0)$ is $\mathcal{F}_{-\infty}^l$ measurable
2. if $\mathbb{E}[|f(X_0)|^p] < \infty$ and $\mathbb{E}[|f(X_k)|^q] < \infty$ for some $p, q > 1$ s.t. $\frac{1}{p} + \frac{1}{q} < 1$,
3. if the process $(X_j, j \in \mathbb{Z})$, is strongly mixing, with mixing coefficient $\alpha(k)$,

then we have $|\text{Cov}(f(X_0), f(X_k))| \leq 10 \alpha(k)^{1-\frac{1}{p}-\frac{1}{q}} \|f(X_0)\|_p \|f(X_k)\|_q$.

Choosing $q = 2$ and $p = 2 + 2\delta$ (as those moments will exist under Assumption (A.13)), and $l = 0$, we can write the inequality above as

$$|\text{Cov}(f(X_0), f(X_k))| \leq M \alpha(k)^{1-\frac{1}{p}-\frac{1}{q}},$$

where $M := 10\|f(X_0)\|_p\|f(X_k)\|_q$.

The process being strong mixing with geometric rate, recall that there exist constants $C > 0$ and $\lambda \in (0, 1)$ s.t. $\alpha(k) \leq C\lambda^k$. We use this geometric rate and the covariance bound to show the finiteness of the first covariance sum of (A.21):

$$\sum_{k=1}^{n/2} k \text{Cov}(f(X_k), f(X_0)) \leq \sum_{k=1}^{n/2} k |\text{Cov}(f(X_k), f(X_0))| \leq \sum_{k=1}^{n/2} k M \alpha(k)^{1-\frac{1}{p}-\frac{1}{q}} \leq MC^{1-\frac{1}{p}-\frac{1}{q}} \sum_{k=1}^{n/2} k \lambda^{k(1-\frac{1}{p}-\frac{1}{q})}.$$

Using once again the ratio test for the finiteness of the latter series (as $n \rightarrow \infty$)

$$L = \lim_{k \rightarrow \infty} \left| \frac{(k+1)\lambda^{(k+1)(1-\frac{1}{p}-\frac{1}{q})}}{k\lambda^{k(1-\frac{1}{p}-\frac{1}{q})}} \right| = \lim_{k \rightarrow \infty} (1 + 1/k)\lambda^{(1-\frac{1}{p}-\frac{1}{q})} = \lambda^{(1-\frac{1}{p}-\frac{1}{q})} < 1,$$

we deduce that

$$\lim_n \frac{1}{n} \sum_{k=1}^{n/2} k \text{Cov}(f(X_k), f(X_0)) = 0. \tag{A.22}$$

Now we need to look at the second sum of (A.21). We proceed in the same way using the strong mixing rate as well as the covariance bounds:

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) \text{Cov}(f(X_{k+n/2}), f(X_0)) &\leq \frac{1}{n} \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) |\text{Cov}(f(X_{k+n/2}), f(X_0))| \\ &\leq \frac{1}{n} \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) M \alpha(k+n/2)^{1-\frac{1}{p}-\frac{1}{q}} \leq \frac{1}{n} \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) M (C\lambda^{k+n/2})^{1-\frac{1}{p}-\frac{1}{q}}. \end{aligned} \tag{A.23}$$

For the ease of notation, define $\tilde{\lambda} = \lambda^{1-\frac{1}{p}-\frac{1}{q}}$ and $\tilde{M} = MC^{1-\frac{1}{p}-\frac{1}{q}}$, such that we have from (A.23)

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) \text{Cov}(f(X_{k+n/2}), f(X_0)) &\leq \tilde{M} \tilde{\lambda}^{n/2} \sum_{k=1}^{n/2-1} \left(\frac{1}{2} - \frac{k}{n}\right) \tilde{\lambda}^k \leq \tilde{M} \tilde{\lambda}^{n/2} \\ &\sum_{k=1}^{n/2-1} \left(\frac{1}{2} - \frac{k}{n}\right) = \tilde{M} \tilde{\lambda}^{n/2} \frac{n-2}{8}, \end{aligned}$$

which tends to 0, as $n \rightarrow \infty$, as $\tilde{\lambda} < 1$. Thus, we can conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) \text{Cov}(f(X_{k+n/2}), f(X_0)) = 0. \tag{A.24}$$

Combining (A.21) with (A.22) and (A.24), we conclude for the first covariance sum of (A.19) that:

$$\lim_n \text{Cov}\left(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Y_i/n\right) = 0. \tag{A.25}$$

- Computation of the second covariance of (A.19):

The computation of the limit of the second covariance of (A.19) is analogous to the first one, simply replacing Y_i by Z_i and thus $f(X_i)$ by $g(X_i)$. I.e., from (A.21), we deduce that

$$\begin{aligned} \text{Cov}(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n) &= \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n \text{Cov}(Q_j, Z_i) = \dots \\ &= \frac{1}{n} \left(\sum_{k=1}^{n/2} k \text{Cov}(f(X_k), g(X_0)) + \sum_{k=1}^{n/2-1} \left(\frac{n}{2} - k\right) \text{Cov}(f(X_{k+n/2}), g(X_0)) \right). \end{aligned} \tag{A.26}$$

The covariance bounds are again applicable. Choosing $p = 2$ and $q = 2 + 2\delta$, those moments exist by (A.13). Thus, we obtain analogous results to (A.22) and (A.24) and can conclude, as for the first covariance of (A.19), that

$$\lim_{n \rightarrow \infty} \text{Cov}(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n) = 0. \tag{A.27}$$

• Computation of the third covariance of (A.19): We are left with

$$\begin{aligned} \text{Cov}(\sqrt{n} \sum_{j=1}^n Y_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n) &= \frac{1}{n} \sum_{j=1}^{n/2} \sum_{i=1}^{n/2} \text{Cov}(f(X_j), f(X_i)) \\ &= \frac{1}{n} \left(\frac{n}{2} \text{Cov}(f(X_0), f(X_0)) + 2 \sum_{i=1}^{n/2-1} \left(\frac{n}{2} - i\right) \text{Cov}(f(X_i), f(X_0)) \right). \end{aligned} \tag{A.28}$$

Thus, we have

$$\lim_{n \rightarrow \infty} \text{Cov}(\sqrt{n} \sum_{j=1}^n Y_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n) = \text{Var}(f(X_0))/2 + \sum_{i=1}^{\infty} \text{Cov}(f(X_i), f(X_0)). \tag{A.29}$$

Therefore, we can finally compute σ_n^2 . We get, recalling the expressions for the variances in (A.20) and for the covariances in (A.25), (A.27), and (A.29), that

$$\begin{aligned} \sigma_n^2 &= \text{Var}(\sqrt{n} \sum_{j=1}^n U_j/n) = \frac{1}{n} \text{Var}(\sqrt{n} \sum_{j=1}^n (aQ_j + bY_j + cZ_j)/n) \\ &= a^2 \text{Var}(\sqrt{n} \sum_{j=1}^n Q_j/n) + b^2 \text{Var}(\sqrt{n} \sum_{j=1}^n Y_j/n) + c^2 \text{Var}(\sqrt{n} \sum_{j=1}^n Z_j/n) \\ &\quad + 2ab \text{Cov}(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Y_i/n) + 2ac \text{Cov}(\sqrt{n} \sum_{j=1}^n Q_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n) \\ &\quad + 2bc \text{Cov}(\sqrt{n} \sum_{j=1}^n Y_j/n, \sqrt{n} \sum_{i=1}^n Z_i/n). \end{aligned}$$

Hence, we have in the limit

$$\lim_{n \rightarrow \infty} \sigma_n^2 = a^2 \sigma_Q^2 + b^2 \sigma_Y^2 + c^2 \sigma_Z^2 + 2bc \left(\frac{1}{2} \text{Var}(f(X_0)) + \sum_{i=1}^{\infty} \text{Cov}(f(X_i), f(X_0)) \right). \tag{A.30}$$

Recalling the univariate asymptotics of $\bar{Q}_n, \bar{Y}_n,$ and $\bar{Z}_n,$ respectively, $\Sigma_{11} = \sigma_Q^2, \Sigma_{22} = \sigma_Y^2, \Sigma_{33} = \sigma_Z^2,$ we can deduce from (A.30) that it must hold $\Sigma_{12} = \Sigma_{13} = 0$ and

$\Sigma_{23} = \text{Var}(f(X_0))/2 + \sum_{i=1}^{\infty} \text{Cov}(f(X_i), f(X_0))$ to have the trivariate normality of the asymptotic distribution of $\sqrt{n} \begin{pmatrix} \bar{Q}_n \\ \bar{Y}_n \\ \bar{Z}_n \end{pmatrix}$ with covariance matrix Σ .

The claims on the relation of Σ and Γ follow directly by comparing. □

A.3 Main proof

After having proved Theorem 9, which was the biggest part of the work, we can proceed with the proof of Theorem 1. It consists of two parts. In the first part, we check all conditions to apply Theorem 9 to establish trivariate asymptotics. The second part uses Slutsky’s theorem, the Delta method, and the continuous mapping theorem to deduce from the trivariate asymptotics the claimed bivariate asymptotics of Theorem 1.

Step 1: Applicability of Theorem 9

The first assumption in Theorem 9, needs the bivariate FCLT, (A.13), to hold. For this, we need to identify the functions f and g with the risk measure and measure of dispersion estimators, respectively, and then explain why the bivariate FCLT between them holds.

Recall that we already know that, for $j = 1, \dots, 4$,

$$\zeta_n^{(j)}(p) = \sum_{i=1}^n (f_j(X_i) - \mathbb{E}[f_j(X_i)]) / n + o_p(1/\sqrt{n}), \tag{A.31}$$

with the functions specified as follows:

- For $j = 1$, $f_1(X_i) = \frac{\mathbb{1}_{(X_i > q_X(p))}}{f_X(q_X(p))}$ – which follows from the Bahadur representation of the sample quantile, see e.g. Wendler (2011).
- For $j = 2$, $f_2(X_i) = \frac{(X_i - q_X(p)) \mathbb{1}_{(X_i > q_X(p))}}{1 - p}$ – which follows from the Bahadur representation for \widehat{ES}_n , see (A.1).
- For $i = 3$, $f_3(X_i) = \frac{1}{k} \sum_{l=1}^k \frac{\mathbb{1}_{(X_i > q_X(p_l))}}{f_X(q_X(p_l))}$ – recalling the definition of the corresponding estimator, (4), and using the case $i = 1$.
- For $i = 4$, $f_4(X_i) = \frac{\mathbb{1}_{(X_i > q_X(\kappa^{-1}(p)))}}{f_X(q_X(\kappa^{-1}(p)))}$ – recalling the definition of the corresponding estimator, (8), and using the case $j = 1$.

Analogously, we know from (A.2) (based on Proposition 4.8 in Bräutigam, 2020) that

$$\hat{m}(X, n, r) = \sum_{i=1}^n (g(X_i) - \mathbb{E}[g(X_i)]) / n + o_p(1/\sqrt{n}), \tag{A.32}$$

with $g(X_i) = |X_i - \mu|^r - r \mathbb{E}[(X - \mu)^{r-1} \text{sgn}(X - \mu)^r] (X_i - \mu)$.

These representations (A.31) and (A.32) hold as, by assumption in Theorem 1, the conditions for the bivariate asymptotics between $\zeta_n^{(j)}(p)$ and $\hat{m}(X, n, r)$ are fulfilled, as we explain in the following for each risk measure estimator separately:

- For $j = 1$, the set of conditions (S_1^*) includes all conditions of Theorem 4 to conclude to the bivariate FCLT.
- For $j = 2$, the set of conditions (S_2^*) includes all conditions of Proposition 5 to conclude to the bivariate FCLT.
- For $j = 3$, the set of conditions (S_3^*) includes all conditions of Theorem 4 to apply a multivariate version of it at all the quantile levels p_1, \dots, p_k and, using the continuous mapping theorem, conclude to the bivariate FCLT.

- For $j = 4$, the set of conditions (S_4^*) includes all conditions of Theorem 4 to apply it at the quantile level k^{-1} , recalling (8).

Then, we consider Theorem 9 for each choice of $f_j, j = 1, \dots, 4$, as defined above, combined with g , so that the bivariate asymptotics (A.10) then holds.

Further, we can identify, by our construction,

$$\zeta_{n/2, t+n/2}^{(j)}(p) - \zeta^{(j)}(p) = 2\bar{Q}_n + o_p(1/\sqrt{n}), \tag{A.33}$$

$$\zeta_{n/2, t}^{(j)}(p) - \zeta^{(j)}(p) = 2\bar{Y}_n + o_p(1/\sqrt{n}), \tag{A.34}$$

$$\hat{m}(X, n/2, r, t) - m(X, r) = 2\bar{Z}_n + o_p(1/\sqrt{n}), \tag{A.35}$$

using the definitions (A.11) and (A.12).

To conclude to the trivariate asymptotics (A.14), we need the strong mixing and the moment condition (A.13). But, recalling the set of assumptions (S_1^*), . . . (S_4^*), we see that strong mixing holds by assumption as well as ($M_{r+\tau}$) (see (M_k)) such that (A.13) holds.

Step 2: Concluding the bivariate asymptotics

By Slutsky theorem, we know that adding a rest which converges in probability to 0, does not change the limiting distribution, thus, from Equations (A.34)–(A.35) and (A.14), it follows that, as $n \rightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \zeta_{n/2, t+n/2}^{(j)}(p) - \zeta^{(j)}(p) \\ \zeta_{n/2, t}^{(j)}(p) - \zeta^{(j)}(p) \\ \hat{m}(X, n/2, r, t) - m(X, r) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, 4\Sigma), \tag{A.36}$$

with the covariance matrix Σ being related to Γ as described in Theorem 9 and 4Σ is to be understood as elementwise multiplication. By the multivariate Delta method, we can deduce from (A.36) that, as $n \rightarrow \infty$,

$$\sqrt{n} \begin{pmatrix} \log|\zeta_{n/2, t+n/2}^{(j)}(p)| - \log|\zeta^{(j)}(p)| \\ \log|\zeta_{n/2, t}^{(j)}(p)| - \log|\zeta^{(j)}(p)| \\ \hat{m}(X, n/2, r, t) - m(X, r) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \tilde{\Sigma}), \tag{A.37}$$

where $\tilde{\Sigma}_{jk} = \begin{cases} 4\Sigma_{jk} / (\zeta^{(j)}(p))^2 & \text{for } j, k \in \{1, 2\}, \\ 4\Sigma_{jk} & \text{for } j = k = 3, \\ 4\Sigma_{jk} / \zeta^{(j)}(p) & \text{else.} \end{cases}$

Applying the continuous mapping theorem to (A.27) with the function $f(x, y, z) = (x - y, z)$, we obtain

$$\sqrt{n} \begin{pmatrix} \log|\zeta_{n/2, t+n/2}^{(j)}(p)| - \log|\zeta_{n/2, t}^{(j)}(p)| \\ \hat{m}(X, n/2, r, t) - m(X, r) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \hat{\Sigma}),$$

where $\hat{\Sigma}_{jk} = \begin{cases} \tilde{\Sigma}_{11} + \tilde{\Sigma}_{22} & \text{for } j = k = 1, \\ \tilde{\Sigma}_{33} & \text{for } j = k = 2, \\ \tilde{\Sigma}_{13} - \tilde{\Sigma}_{23} & \text{else.} \end{cases}$

By tracing back the definitions of Σ (see Theorem 9), we see that $\hat{\Sigma}$ equals $\tilde{\Gamma}$ as defined in Theorem 1, and thus concludes the proof.

Cite this article: Bräutigam M and Kratz M (2023). How do empirical estimators of popular risk measures impact procyclicality?, *Annals of Actuarial Science*, 17, 547–579. <https://doi.org/10.1017/S1748499523000039>