Pro-cyclicality of risk measurements. Empirical quantification and theoretical confirmation

Marcel Bräutigam

École Doctorale de Sciences Mathématiques de Paris Centre
Laboratoire de Probabilités, Statistique et Modélisation - LPSM, Sorbonne Université

THÈSE DE DOCTORAT

Discipline : Mathématiques
Spécialité : Statistique

présenté par

Marcel BRÄUTIGAM

Pro-cyclicality of Risk Measurements - Empirical Quantification and Theoretical Confirmation

---

dirigée par Prof. Marie KRATZ (ESSEC CREAR)

Soutenue le 27 février 2020 devant le jury composé de :

Prof. Patrice BERTAIL	Université Paris Nanterre	Rapporteur
Prof. Valérie CHAVEZ-DEMOULIN	Université de Lausanne	Rapporteur
Prof. Rama CONT	University of Oxford	Président du jury
Dr. Michel DACOROGNA	Prime Re Solutions	Examinateur
Prof. Liudas GIRAITIS	Queen Mary University of London	Examinateur
Prof. Marie KRATZ	ESSEC Business School	Directrice de thèse
Prof. Olivier WINTENBERGER	Sorbonne Université	Examinateur

Declaration of Funding

I, Marcel BRÄUTIGAM, acknowledge the funding throughout the 3 years of my research from a doctoral fellowship by Labex MME-DII (Modèles Mathématiques et Économiques de la Dynamique, de l'Incertitude et des Interactions) and ESSEC Business School.

Abstract

This thesis examines, empirically and theoretically, the pro-cyclicality of risk measurements made on historical data. Namely, the effect that risk measurements overestimate the future risk in times of crisis, while underestimating it in quiet times.

As starting point, we lay down a methodology to empirically evaluate the amount of pro-cyclicality when using a sample quantile (Value-at-Risk) process to measure risk. Applying this procedure to 11 stock indices, we identify two factors explaining the pro-cyclical behavior: The clustering and return-to-the-mean of volatility (as modeled by a GARCH$(1,1)$) and the very way of estimating risk on historical data (even when no volatility dynamics are present).

To confirm these claims theoretically, we proceed in two steps. First, we derive bivariate (functional) central limit theorems for quantile estimators with different measure of dispersion estimators. We establish them for sequences of iid random variables as well as for the class of augmented GARCH(p,q) processes.

Then, we use these asymptotics to theoretically prove the pro-cyclicality observed empirically. Extending the setting of the empirical study, we show that no matter the choice of risk measure (estimator), measure of dispersion estimator or underlying model considered, pro-cyclicality will always exist.

Keywords: pro-cyclicality; quantile estimator; Value-at-Risk; measure of dispersion; volatility; asymptotic distribution; (functional) central limit theorem; correlation; financial markets; risk measure process; GARCH

Résumé

Cette thèse analyse, d'un point de vue empirique et théorique, la procyclicité des mesures de risque sur les données historiques, i.e. l'effet de surestimation du risque futur en temps de crise, et sa sous-estimation en temps normal.

Nous développons une méthodologie pour évaluer empiriquement le degré de procyclicité, en introduisant un processus de quantiles ('Value-at-Risk') historiques pour mesurer le risque. En appliquant cette procédure à 11 indices boursiers, nous identifions deux facteurs expliquant la procyclicité: le 'clustering' et le retour à la moyenne de la volatilité (tel que modélisée par un GARCH(1,1)), mais aussi la façon intrinsèque d'estimer le risque sur des données historiques (même en l'absence de dynamique de la volatilité).

Pour confirmer théoriquement ces arguments, nous procédons en deux étapes. Premièrement, nous démontrons des théorèmes bivariés (fonctionnels) de limite centrale pour les estimateurs de quantiles avec différents estimateurs de dispersion. Comme modèles de base, nous considérons les suites de variables aléatoires iid, ainsi que la classe des processus GARCH(p,q) augmentés.

Enfin, ces résultats asymptotiques permettent de valider théoriquement la procyclicité observée empiriquement. Généralisant cette étude à d'autres mesures de risque et de dispersion, nous concluons que la procyclicité persistera quel que soit le choix de ces mesures.

mots clés: procycliclité; estimateur de quantile; Value-at-Risk; mesure de dispersion; volatilité; distribution asymptotique; théorème (fonctionnel) de limite centrale; corrélation; marchés financiers; processus de mesure des risques; GARCH

Acknowledgements

Not only the readers of this thesis (if any) but also the people who know me in general will confirm my tendency to long and intricate sentences and texts. To avoid a text longer than the thesis itself, I decided to be concise and factual (but please don't see this brevity as a form of disrespect or ignorance). I hope to be the contrary when having the opportunity to thanking you in person.

I would like to thank:

- Prof. Marie Kratz as the supervisor of my thesis,
- Dr. Michel Dacorogna for the joint work and being part of the jury,
- Prof. Patrice Bertail and Prof. Valérie Chavez-Demoulin for having accepted to referee the thesis and being part of the jury,
- Prof. Rama Cont, Prof. Liudas Giraitis and Prof. Olivier Wintenberger for being part of the jury,
- CREAR (ESSEC) and LPSM (Sorbonne University) and their administrative staff for their support and providing me workplaces,
- and, finally, also you.

Contents

List of Abbreviations

ACF	Autocorrelation function
ARCH	autoregressive conditional heteroscedasticity (model)
a.s.	almost surely
cdf	cumulative distribution function
CI	confidence interval
CLT	Central Limit Theorem
df	degrees of freedom
ES	Expected Shortfall
FCLT	Functional Central Limit Theorem
FHS	Filtered historic simulation
GARCH	generalized autoregressive conditional heteroskedasticity (model)
iid	independently, indentically distributed
IQR	Interquartile range
pdf	probability density function
MAD	Mean absolute deviation around the mean
MedianAD	Median absolute deviation around the median
NED	near-epoch dependent
RMS	root mean square
RMSE	root mean square error
rv	random variable
SI	stock index
std	standard deviation
SQP	Sample Quantile Process
VaR	Value-at-Risk
vs	versus
wlog	without loss of generality

List of Notations

Symbol	Definition	Name
$\mathbb{N}$		the set of natural numbers (including zero)
$\mathbb{R}$		the set of real numbers
$\xrightarrow{d}$		convergence in distribution
$\xrightarrow{P}$		convergence in probability
$\xrightarrow{a.s.}$		almost sure convergence
$\xrightarrow{D_d[0,1]}$		convergence in distribution in the d-dimensional Skorohod space $D_d[0,1]$
$X_n = o_P(a_n)$		convergence of a sequence of random variables X_n / a_n to 0 in probability (for a_n a series of constants)
a^T		the transpose of a vector a
$\min(a,b)$		the minimum of a and b
$\max(a,b)$		the maximum of a and b
$\mathbb{1}_A$		the indicator function of the set A
$\forall$		universal quantifier ('for all')
$\lvert x\rvert$	$\left(\sum_{i=1}^{d} x_i^2\right)^{1/2}$	euclidean norm for a vector $x \in \mathbb{R}^d$ with elements $x_1, ..., x_d$
$\lceil x \rceil$	$\min\{m \in \mathbb{Z} : m \geq x\}$	nearest integer bigger than x
$\lfloor x \rfloor$	$\max\{m \in \mathbb{Z} : m \leq x\}$	nearest integer smaller than x
$f(x) = O(g(x))$	$\lvert f(x)\rvert \leq Mg(x)\ \forall x \geq x_0$ with $M > 0, x \in \mathbb{R}$	Big-O
$f(x) = o(g(x))$	for every $\varepsilon > 0\ \lvert f(x)\rvert \leq \varepsilon g(x)\ \forall x \geq N, N \in \mathbb{R}$	small-o
$\operatorname{sgn}(x)$	$\mathbb{1}_{(x \geq 0)} - \mathbb{1}_{(x \leq 0)}$	the sign function
μ	$\mathbb{E}[X]$	mean
σ^2	$\operatorname{Var}(X)$	variance
θ	$\mathbb{E}\lvert X - \mu\rvert$	mean absolute deviation (MAD)
μ_r	$\mathbb{E}[(X-\mu)^r]$	r-th centred/central moment
$\lVert X\rVert_p$	$\mathbb{E}^{1/p}[\lvert X\rvert^p]$	L^p-norm of a random vector $X \in \mathbb{R}^d$
$m(X,r)$	$\mathbb{E}[\lvert X-\mu\rvert^r]$	r-th absolute centred/central moment
$F_X(\cdot)$	$\mathbb{P}(X \leq \cdot)$	cumulative distribution function
$f_X(\cdot)$		probability density function
$q_X(p)$	$F_X^{-1}(p)$	quantile of order p
ν	$q_X(1/2)$	median
ξ	$q_{\lvert X-\nu\rvert}(1/2)$	Median absolute deviation around the median (MedianAD)

Symbol	Definition	Name
$\bar{X}_n$	$\frac{1}{n}\sum_{i=1}^{n} X_i$	sample mean
$\hat{\sigma}_n^2$	$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$	sample variance (unknown μ)
$\tilde{\sigma}_n^2$	$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$	sample variance (known μ)
$\hat{\theta}_n$	$\frac{1}{n}\sum_{i=1}^{n}\lvert X_i - \bar{X}_n\rvert$	Sample MAD (unknown μ)
$\tilde{\theta}_n$	$\frac{1}{n}\sum_{i=1}^{n}\lvert X_i - \mu\rvert$	Sample MAD (known μ)
$\hat{m}(X,n,r)$	$\frac{1}{n}\sum_{i=1}^{n}\lvert X_i - \bar{X}_n\rvert^r$	r-th absolute central sample moment (unknown μ)
$\tilde{m}(X,n,r)$	$\frac{1}{n}\sum_{i=1}^{n}\lvert X_i - \mu\rvert^r$	r-th absolute central sample moment (known μ)
$F_n(x)$	$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(X_i \le x)}$	empirical cdf
$V_{k,n}(t)$	$\frac{1}{n-1}\sum_{i=t+1-n}^{t}\lvert X_i - \frac{1}{n}\sum_{i=t+1-n}^{t} X_i\rvert^k$	r-th absolute central sample moment (unknown μ)
$q_n(p)$	$X_{(\lceil np \rceil)}$	sample quantile of order p, where $X_{(1)} \le ... \le X_{(n)}$ is the ordered sample of $X_1, ..., X_n$
$q_{n,\hat{\mu},\hat{\sigma}}(p)$	$\hat{\mu}_n + \hat{\sigma}_n q_Y(p)$	parametric location-scale quantile estimator (unknown μ)
$q_{n,\hat{\sigma}}(p)$	$\mu + \hat{\sigma}_n q_Y(p)$	parametric location-scale quantile estimator (known μ)
$\hat{v}_n$	$\frac{1}{2}(X_{(\lfloor \frac{n+1}{2} \rfloor)} + X_{(\lfloor \frac{n+2}{2} \rfloor)})$	sample median
$\hat{\xi}_n$	$\frac{1}{2}(W_{(\lfloor \frac{n+1}{2} \rfloor)} + W_{(\lfloor \frac{n+2}{2} \rfloor)})$ where $W_j = \lvert X_j - \hat{v}_n\rvert$, $j = 1, ..., n$	sample MedianAD

List of Conditions

(M_k) $\mathbb{E}[X^{2k}] < \infty$

(C_0) F_X is continuous

$(C_l^{'})$ F_X is l-times differentiable

(C_l) the l-th derivative of F_X is continuous

(P) f_X is positive

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

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

(P_r) $\sum_{i=1}^{p} \|g_i(\varepsilon_0)\|_r < \infty, \quad \sum_{j=1}^{q} \|c_j(\varepsilon_0)\|_r < 1,$

(L_r) $\mathbb{E}[\exp(4r \sum_{i=1}^{p} |g_i(\varepsilon_0)|^2)] < \infty, \quad \sum_{j=1}^{q} |c_j(\varepsilon_0)| < 1.$

Chapter 1

Introduction

1.1 Evolution of the thesis

The starting points of this thesis are two questions tackled in an empirical study on pro-cyclicality, which subsequently led to the development of results on asymptotic bivariate distributions between quantile and measure of dispersion estimators.

Before presenting these questions, we want to contextualize their background first: Since the introduction of risk based capital regulations in banking and insurance, there has been a need for financial institutions to evaluate their risk on the basis of probabilistic models. A first milestone followed from the introduction of the RiskMetrics system by JP Morgan in 1994 (see [97]). JP Morgan identified the capital needed by companies to cover their risk, to be the quantile at a certain threshold of the return distribution of the portfolio. This risk measure was called Value-at-Risk (VaR) and was shortly after included in the banking regulation (see [15]). Subsequently, the axiomatic approach to risk measurement, initiated in the seminal paper by Artzner, Delbaen, Eber, and Heath [8], led, inter alia, to the development of alternative risk measures. Among them the most prominent example is the Expected Shortfall (ES) (see [8], [7], [1], [111]), another risk measure which is widely used nowadays. In view of VaR, ES and other proposed alternatives (as e.g. expectiles, see [100], [87]), the question of the appropriateness of the risk measure to use for evaluating the risk of financial institutions has been heavily debated, especially after the financial crisis of 2008/2009. For a review of the arguments on this subject, we refer e.g. to [37] and [53].

Independently from the choice of adequate risk measures, there is an accepted idea that in times of crisis, risk measurements overestimate the future risk, while they underestimate it in quiet times. This effect is called pro-cyclicality (of risk measurements) and is the phenomenon we are interested in studying.

Note that we have to be careful with the term "pro-cyclicality" as there are different meanings to it. In the literature it is usually viewed from a macro-economic perspective: Pro-cyclicality refers to the effect that the macro-economy is affected by the banks reaction to changing macro-economic conditions, which in return affects the banks performance again (having reinforcing effects on cyclical fluctuations of the macro-economy). For a more thorough literature rewiew from this point of view, see the introduction in [29] (as well as the references in [9], [109]).

Here, we are not looking for modelling the pro-cyclical behavior of risk measurements macro-economically, but rather to drawing conclusions from statistical analyses on the underlying financial time series itself. While the pro-cyclicality of risk measurements is a known fact, there have been little attempts to study empirically on real financial data the direct relation between risk measure estimates and the state of the financial market (among the few, see Bec and Gollier [17]), and even less to quantify the pro-cyclicality. Thus, the initial questions at the beginning of the thesis are:

Q1 How can we measure and subsequently quantify the pro-cyclicality of risk estimation (on empirical data)?

Q2 What are the factors explaining this effect (using the tool set of statistics, instead of building macro-economic models)?

To answer the questions **Q1** and **Q2**, we set up an empirical study on pro-cyclicality; see [28], [29] for two subsequent versions of this study. Therein we consider the VaR as risk measure and estimate it by a sample quantile, as such a 'historical estimation' is still widely used in practice (see e.g. [103], [58] for surveys on VaR estimation in banks). Then, we introduce a performance indicator to assess the predictive quality of risk estimation, namely (the logarithm of) a ratio of, in this case, sample quantiles (VaR estimates) computed on subsequent samples. To quantify the pro-cyclicality present in the data we look at the behavior of the ratio conditioned on a proxy for the state of the financial market: In our first study, [28], we observe a negative correlation between (the logarithm of) this ratio and the sample standard deviation (used as a proxy for the market state), not only on real data, but also in simulations from identically and independently distributed (iid) and GARCH(1,1) models.

These empirical findings incentivised us to consider this correlation from a theoretical point of view: First, we look at the joint distribution of the sample quantile with the sample standard deviation (std). But then we realize that this could be embedded in a more general framework. Apart from the (non-parametric) sample quantile, we consider an additional quantile estimator, namely the parametric location-scale quantile estimator. This enables us to relate our work to the empirical findings of Zumbach in [131] and [129] where the author considers the correlation of 'the realized volatilities with the centred volatility increment' (for different underlying data sets and processes) - which can be seen as the correlation between the sample standard deviation and the difference of two location-scale quantile estimators. Similarly, we take other measure of dispersion estimators than the sample standard deviation into account, e.g. the mean absolute deviation around the sample mean (MAD) or median absolute deviation around the sample median (MedianAD). As mentioned, we consider in the empirical study, simulations from iid as well as GARCH$(1,1)$ models. Hence, the first two theoretical questions arising from this study:

Q3 What is the joint distribution of quantile estimators with measure of dispersion estimators for iid models?

Q4 What is the joint distribution of quantile estimators with measure of dispersion estimators for a GARCH(1,1) model?

Independently from the fact that these two open questions are motivated from the empirical quantification of pro-cyclicality, we see them as a natural complement to the literature on joint distributions of quantile estimators with measure of *location* estimators:

For an iid sample, we found that there are two cases of joint *asymptotic* distributions of the sample quantile with location estimators treated in the literature. On the one hand, there is the case of the sample median. But as the sample median is a sample quantile itself, its joint asymptotics with the sample quantile are included in the well-known asymptotics of a vector of sample quantiles. On the other hand, the joint asymptotics of the sample quantile with the sample mean. They were treated first by Lin (see [91]) and later, with another approach, by Ferguson (see [62]). The latter result has then been used in [20] to introduce a new characterization of the normal distribution and a normality test.

In contrast, joint asymptotics between quantile estimators and measure of dispersion estimators have not yet been considered in full generality in the literature. Nevertheless a few examples exist. For instance, for symmetric location-scale distributions, the MedianAD equals half the interquartile range (IQR) (see e.g [72]), and their sample estimators are asymptotically equivalent, as shown in [71]. Thus, the joint asymptotics of the sample quantile and sample MedianAD for such distributions follow from the asymptotics of a vector of sample quantiles. There are two further implicit contributions of joint asymptotics of the sample quantile with measure of dispersion estimators: Under some symmetry-type

conditions on the underlying distribution, Falk proved in [61] the asymptotic independence of the sample median (i.e. a sample quantile at level 0.5) and the sample MedianAD. Moreover, in [44], the joint asymptotics between the interquartile range and the standard deviation are shown. An explicit contribution can be found in the work of Bos and Janus, [25]. Therein the authors provide the joint asymptotic distribution of the sample quantile with the r-th absolute sample moment in the case of an underlying Gaussian distribution with known mean. Even if they restrict their work to the Gaussian distribution and the mean to be known, this can be considered as an extension to higher moments of the results of [44]. These results of [71], [61], [44], [25], can be seen as special cases of our treatment of joint asymptotic distributions of quantile estimators and measure of dispersion estimators.

Apart from theoretical results in the iid case, we are interested in corresponding results for the GARCH$(1,1)$. We use in our empirical study a GARCH$(1,1)$ process to model the volatility and then assess its pro-cyclicality in simulations. However, since the introduction of the ARCH and GARCH processes in the seminal papers by Engle, [54], and Bollerslev, [23], respectively, various GARCH modifications and extensions have been proposed (see e.g. [24] for a non-exhaustive (G)ARCH glossary). Thus, it is appealing to provide a more general theoretical framework that includes a bigger family of GARCH-type processes. We decide to focus on the family of augmented GARCH(p,q) processes, introduced by Duan in [49], as they contain many well-known GARCH processes as special cases.

Conditions for the stationarity of such GARCH-type processes, as well as (functional) central limit theorems (CLT), have been obtained in various ways by exploiting different dependence concepts (see e.g. the introduction in [89] for additional references). Further, the limit theorems also extend to different sample estimators, as for example: sample moments (e.g. [77] for augmented GARCH(1,1); [21], [89] for augmented GARCH(p,q)), sample autocovariance and sample variance (e.g. [94] for the GARCH(1,1); [10] for augmented GARCH(1,1)), or the sample quantile (see e.g. [88] and references therein). Still, joint asymptotics of quantile estimators with measure of dispersion estimators have not been considered yet.

Establishing such bivariate asymptotics, providing answers to the questions **Q3** and **Q4**, is a theoretical contribution on its own. By presenting joint asymptotic results in the iid and augmented GARCH(p,q) case of quantile estimators (and, more generally, other risk measure estimators), not with measures of location, but measures of dispersion estimators, we hope to offer a useful complement not yet present in the statistical literature.

At the same time, it gives us a foundation to study, from a theoretical point of view, the pro-cyclicality of the VaR observed in our empirical study ([28], [29]). Knowing the joint asymptotics between the sample quantile and the r-th absolute central sample moment, could we deduce the asymptotic distribution of *the log-ratio* of sample quantiles with the r-th absolute sample moment? Stated more generally:

Q5 Can we mathematically prove the pro-cyclicality of risk measurements for underlying iid and augmented GARCH(p,q) models?

Q6 How does the degree of pro-cyclicality in such models depend on the choice of risk measure (estimator) and measure of dispersion estimator?

With these questions, we come back to the starting point of the thesis, but now from a theoretical point of view. Answering them proves the pro-cyclicality beyond the VaR (i.e. the sample quantile), extending the results to other risk measures like the ES and expectile.

Clearly, it is also of interest to us to see if we could use such theoretical results to add theoretical evidence to the empirical findings of the pro-cyclicality of the VaR in our empirical study. And finally, to see if with all the gained theoretical and empirical insight on pro-cyclicality, we could propose alternative (counter-cyclical) risk measures, that do not suffer from this phenomenon:

Q7 Can we use the results obtained when answering **Q5** to strengthen our claims on the reasons for the pro-cyclicality observed in real data (c.f. question **Q2**)?

Q8 Can we propose a (counter-cyclical) risk measure to overcome the pro-cyclicality of historical risk measurement?

We answer in this thesis the first seven questions, **Q1**-**Q7**, giving an empirical quantification and theoretical confirmation (for the group of stochastic models under consideration) of the pro-cyclicality of historical risk measurement. It also opens up the path for new approaches in counter-cyclical risk management and risk measures. It is ongoing work, which should answer **Q8**.

1.2 Structure and content of the thesis

After having motivated the evolution of the eight questions presented, we lay down how they are tackled in this thesis. The thesis has four main chapters corresponding, each, in big parts, to one of the four papers written during my PhD. Nevertheless, these chapters contain more material than the papers, namely, analyses, remarks and links to the appendix for supplementary material, which were not presented in the paper versions, but were part of the study to assess the main results. Further, to outline the main findings and facilitate the transition from one chapter to another without loosing the connecting ideas, we present at the end of each chapter a short, separate summary page with key points and key questions that create a link to subsequent results. We also provide a list indicating related content in the Appendix.

The structure of the thesis reflects the evolution described in the previous section. As explained, my investigations started with an empirical study on the pro-cyclicality of risk measurements, trying to answer questions **Q1** and **Q2**. Consequently, Chapter 2 presents this study whose goals are threefold: First, to introduce a way to measure pro-cyclicality. Then, to quantify it on real data and, thirdly, to find explanations for this effect. Therefore, considering the VaR as risk measure, we start by defining a new indicator, called look-forward ratio to assess the accuracy of risk estimation. It quantifies the difference between the historically predicted VaR and the estimated realized future risk (measured *ex post* by the VaR realized one year later). If the predicted VaR is close to the realized VaR one year later, then the ratio will be around 1. In the two other cases, it will be either smaller than one (if the prediction is too conservative) or bigger than one (if the predicted VaR is insufficient).

To quantify the pro-cyclicality, we analyse the look-forward ratio conditioned on the current market state. As a proxy for qualifying the market state, we choose the volatility, since the realized volatility is high in times of crisis and low when there is little information hitting the market. Consequently, we assess the pro-cyclicality by computing the correlation between the look-forward ratio and the realized volatility. The first version of our study was based on the sample standard deviation (std) as classical estimator of the realized volatility, see [28]. Here we prefer considering alternatively the sample MAD as measure of realized volatility. The reason to favour the latter is of theoretical nature: While a sample correlation can always be computed, it is only meaningful if its theoretical counterpart is well defined, i.e. the variance of the estimators is finite. In the case of the sample std, this implies the existence of a fourth moment of the underlying distribution, while only a second moment when using the sample MAD. As financial data is known to be heavy tailed (potentially without a finite 4th moment), choosing MAD over std seems more prudent.

Having layed down a methodology to evaluate the amount of pro-cyclicality in the way financial institutions measure risk, we apply it on 11 stock indices of major economies: We observe a negative correlation between the (logarithm of the) look-forward ratio and the sample MAD, confirming and, at the same time, quantifying pro-cyclicality. Put differently, we find that in times of high volatility, the ratio is significantly below 1, meaning that in times of crisis future risk is overestimated, while it can go even over 3 in times of low volatility, signaling a strong underestimation of future risk in quiet market conditions.

In a third step, we look for factors to characterize this effect. Therefore, we try to find models which, each, can explain a different component of the observed pro-cyclicality: First, we consider an

iid model to check if there is an intrinsic part to the pro-cyclicality that is present, even when the model does not exhibit any form of dynamics and dependence. On the other hand, we choose a GARCH$(1,1)$ model (fitted on the data). With this choice, we purposely do not aim at fitting as accurately as possible the data, but choosing a simple process to isolate and model the volatility clustering present in the data. Simulating from the iid model, we observe that it produces negative correlation between the log-ratio and the volatility, but with less amplitude than in the real data. Meaning that only part of the pro-cyclicality in the real data can be explained by this. Turning to simulations from the GARCH$(1,1)$ model, we see that it presents a very similar clustering behavior as in the data and also gives a strong negative correlation between the log-ratio and the volatility. This negative correlation is stronger than that obtained on the real data.

From the analysis of those two simple models, we conclude that the pro-cyclicality can be explained by two factors: (i) the way risk is measured, as a function of quantiles estimated on past observations, and (ii) the clustering and return-to-the-mean of volatility. This chapter formed the basis for the joint works in [28] and [29].

The empirical study was the starting point for the development of the theoretical results in Chapters 3 and 4. Having considered in the previous chapter the correlation between a ratio of VaR estimates and the sample MAD, in these two chapters we look more generally at the joint asymptotics of quantile estimators with measure of dispersion estimators (the MAD and std being examples of it) for the two models considered in the empirical study.

In Chapter 3 we treat the iid case. We consider the sample quantile and the parametric location-scale quantile estimator as two different ways of estimating the quantile. The interest in using these two consistent estimators is their different speed of convergence, which consequently should affect their joint asymptotics with measure of dispersion estimators. Alike, we look at two classes of measure of dispersion estimators, namely, the r-th absolute central sample moment for any integer r (which includes the sample MAD and the sample variance) and the sample MedianAD. We then derive the bivariate asymptotic normality between functions of such quantile estimators and functions of these measure of dispersion estimators. While we treat in this chapter, first the case of the sample quantile, then the one of the parametric location-scale quantile estimator, the approach is in all cases the same: We use the classical bivariate CLT for iid samples. For it to be applicable, the corresponding estimators need to have an (asymptotic) representation in terms of a sum of iid rv's. Such a representation is trivial for the sample mean and sample variance, thus also for the location-scale quantile. For the sample quantile, it is called Bahadur representation (and we use it in the version of [68]). An analogous representation exists for the sample MedianAD ([93]) and sample MAD ([116]). However, for the general r-th absolute central sample moment, we need to establish it, generalising the results of [116] (case $r = 1$).

Further, recalling the choice of sample MAD over sample std in the previous chapter to have weaker moment conditions, we compare and discuss the (moment and smoothness) conditions on the underlying distribution depending on the choice of quantile and measure of dispersion estimator. This is especially relevant for the choice of corresponding estimators in practice. Besides, we compare the impact of the choice of the quantile estimator and of the measure of dispersion estimator, on the strength of the correlations in their asymptotic distribution when considering, exemplary, the Gaussian and Student-t distributions. Finally, simulations show a good finite sample performance of the asymptotic results. This chapter formed the basis for the joint works in [32] and [31].

Then, in Chapter 4, we treat the joint asymptotics in the case of a family of GARCH models. While we simulated from a GARCH$(1,1)$ process in the empirical study, from a mathematical perspective, we want to consider a more general framework. At the same time, to establish such joint asymptotics, we need some representation and certain dependence properties to be available or deducable for the estimators considered. The class of augmented GARCH(p,q) processes seems a good compromise. It was introduced in [49] and includes different GARCH models used in practice.

Analogously to the iid case, to establish any kind of limit theory, we need a sum representation (in terms of functionals of the underlying process) to hold: For the Bahadur representation of the sample quantile, we prove that the representation in [121] is applicable to the class of augmented GARCH(p,q) processes. Then, as in the iid case, we need to establish the representation for the r-th absolute central sample moment. We prove a representation which holds more generally for ergodic, stationary, short-memory processes, once again generalising the results of [116] for the MAD. Finally, to establish the bivariate functional central limit theorem (FCLT) between these estimators, we rely on a multivariate FCLT from [11].

The central condition in this multivariate FCLT is the existence of a Δ-dependent (also called 'm-dependent') approximation for the representations of our estimators. We show that L_2-near epoch dependence (NED) (a dependence concept dating back to Ibragimov in [80]) is a sufficient condition to guarantee such Δ-dependent approximations. The reason for relying on the concept of L_2-NED, is that we can use results by Lee, [89], and Wendler, [121], to show that the representations of our estimators indeed are L_2-NED. As the results in [89] hold exactly for augmented GARCH(p,q) processes, this is, retrospectively, one of the reasons to consider this class of GARCH processes.

While the univariate asymptotics of the r-th absolute central sample moment is a novelty on its own, we keep the focus on the bivariate asymptotics - as generalization of the corresponding results in the iid case of Chapter 3. We illustrate the general results by showing, for selected examples of augmented GARCH(p,q) processes, under which conditions on the moments and parameters of the process the joint asymptotics holds. This chapter formed the basis for the joint work in [30].

In Chapter 5, we bring together the results of the empirical study in Chapter 2 and the theoretical results from Chapters 3 and 4, answering questions **Q5** to **Q7**. In Chapter 2, we observed a negative correlation between the logarithm of the look-forward ratio and the sample MAD on real data and in simulations. Here, we give a mathematical proof confirming this by establishing the joint asymptotic distribution of the logarithm of the look-forward ratio with the measure of dispersion estimator - when assuming either an iid or augmented GARCH(p,q) model. However, this effect is not only valid for the VaR or MAD: We consider estimators of three different well-known risk measures (VaR, ES, expectile), and the r-th absolute centred sample moment (for any integer $r > 0$), as measure of dispersion estimator, and the two models mentioned. We show that the strength of pro-cyclicality depends on these three factors: The choice of risk measure (estimator), the measure of dispersion estimator and the model considered. But, no matter the choices, the pro-cyclicality will always be present.

In the iid case, because of the construction of the look-forward ratio, these results seem intuitively clear: The realized future risk is estimated on a sample that is disjoint from the one of the historical risk estimator (and the measure of dispersion estimator). Thus, the former estimator should be uncorrelated to the latter two. This implies that the asymptotics with the logarithm of the look-forward ratio can be recovered from the asymptotics between the risk measure estimator and measure of dispersion estimator itself. Nevertheless, anticipating the more involved treatment for dependent processes, we prove this claim formally, even for iid samples. The asymptotic disjointness of two estimators is modeled by considering the first half of the sample to be equal to 0 for the one, and the second half to be 0 for the other estimator. Consequenty, we need to establish the asymptotics with a CLT for non-identically but independently distributed samples. Using this framework, we show that a necessary and sufficient condition to prove the pro-cyclicality is the validity of the asymptotics between the risk measure estimator and the measure of dispersion estimator. This means that for the VaR and our choice of expectile estimator (which is a VaR estimator), we can use the results in Chapter 3. On the contrary, such a CLT between an ES estimator and the r-th absolute central sample moment needs to be established. Using the representation of the ES in [38], we prove this by correspondingly adapting the reasoning of the proof in Chapter 3 for VaR to ES.

For the family of augmented GARCH(p,q) processes, the results need some more work. In contrast to the iid case, the bivariate FCLT's are not anymore a sufficient (but still necessary) condition for proving

the pro-cyclicality. Indeed, we need two additional assumptions: First, another bound on the dependence behaviour of the underlying process by assuming the process to be strongly mixing with geometric rate. This is due to the fact that, for dependent processes, estimators over disjoint samples are a priori not uncorrelated anymore. To prove an asymptotic uncorrelatedness, we use results on covariance bounds for strongly mixing processes given in [113]. Note that this latter condition on the process is not really restrictive as it has been shown that many GARCH processes fulfill the strong mixing (see e.g. [27], [36]). Second, we need slightly stronger moment conditions on the innovations process (which are imposed due to the use of a CLT for non-stationary, strongly mixing processes).

We show that the results on pro-cyclicality are structurally the same as in the iid case. Also, as in the iid case, we need to prove a FCLT between the ES and the r-th central absolute sample moment. The proof uses the representation for the ES in [38] and an analogous approach as for the VaR in Chapter 4 (needing to prove a Delta-dependent approximation for the ES representation).

We consider two applications in this chapter. As we can derive closed form expressions for the pro-cyclicality in the iid case, we assess how the degree of pro-cyclicality varies for different estimator and distribution choices. Considering, exemplary, the Gaussian and different Student-t distributions, we compare the pro-cyclicality of the VaR, expectile and ES estimators when considering the sample variance or sample MAD as measure of dispersion estimators.

Secondly, we provide additional evidence, using the theoretical results from this chapter, for the empirical observations in Chapter 2. Namely, that, apart from the iid effect, the characteristics of the GARCH$(1,1)$ model (i.e. the return-to-the-mean and clustering of volatility) create the pro-cyclicality in the data. For this, we consider the empirical residuals of the GARCH$(1,1)$ process fitted to the data of Chapter 2. Such residuals should form in theory an iid sample. Indeed, we show that its empirically measured pro-cyclicality values fall most of the times in the 95% confidence intervals for the theoretical iid pro-cyclicality values. In this way we give an answer to **Q7**. This chapter formed the basis for the joint work in [33].

We conclude and discuss further research perspectives in Chapter 6.

The thesis is supplemented by an extensive Appendix. Foremost, it includes the detailed results of the empirical study and simulations done to assess the obtained results. Further, especially (but not only) the study in Chapter 2 made it necessary to empirically assess various additional claims (e.g. how do the results change when considering weekly instead of daily data, the ES instead of the VaR etc.). Moreover, we present and derive the closed form expressions of the pro-cylicality for Gaussian and Student iid models, as well as a comparative study on GARCH optimization using the programming language R.

The full material is available as online appendix, see [34], and we selected a part of it for the appendix in this document. While it is usually mentioned throughout the text, we also provide at the end of each chapter an overview about which parts of the supplementary material in the appendix are linked to the current chapter.

Chapter 2

Quantifying pro-cyclicality: An empirical study

This chapter formed the basis for the joint work in [29] and its previous version [28].

2.1 Introduction

The pro-cyclicality of risk measurement is a known fact, it means that in times of crisis, the measurements overestimate the future risk, while they underestimate it in quiet times. Still, there have been little attempts to study empirically on real financial data the pro-cyclicality of risk measure estimates themselves (among the few, see Bec and Gollier [17]), and even less to quantify it. Indeed, this pro-cyclicality is usually assumed to be a consequence of the volatility clustering and its return to the mean. Yet, we may question if this mean-reverting is fast enough to produce pro-cyclicality on a one year horizon, the typical time horizon used in risk management when measuring the risk (see the insurance and banking regulation, i.e. [59], [16]). Thus, our interest and the need to analyse further this effect.

Reiterating the questions from the introduction, this study aims at answering the two questions **Q1** and **Q2**:

Q1 How can we measure and subsequently quantify the pro-cyclicality of risk estimation (on empirical data)?

Q2 What are the factors explaining this effect (using the tool set of statistics, instead of building macro-economic models)?

The novelty of this work is to propose alternative methods, on the one hand, to quantify the predictive power of risk measurements, and, on the other hand, to explore the issue of pro-cyclicality by quantifying it, then identifying some of its main factors.

Such a study requires a dynamic reformulation of risk measurements departing from a pure static approach. While such an approach holds for any risk measure in question, e.g. be it VaR or ES, we focus here on the example of VaR (nevertheless, a short analysis in the case of the ES can be found in the online-appendix, [34], Appendix A.4). This means departing from estimating quantiles of an *unconditional* distribution. Clearly, different dynamic extensions already exist in the literature, see e.g. [56], [47]. Here, we focus on the simplest dynamic version of the 'regulatory' risk measure VaR, namely the sample quantile process (SQP) considering the measurement itself as a random process. This notion has been proposed in the context of mathematical finance for options pricing, introduced by Miura [96] in a general context while Akahori [2] looked at it for Brownian Motion, and Embrechts and Samorodnitsky [52] for the class of infinitely divisible stochastic processes. Using the SQP, we estimate it on financial time series and explore its dynamics and ability to predict correctly the future risk.

We conduct our study on 11 stock indices (SI) of major economies to analyse the common behaviors and consider them as realizations of the generating process of financial returns, which we are exploring.

We assume that the sample measurement is an estimation of the future risk of the time series, as commonly done in solvency regulations. We define a new indicator, a look-forward ratio that quantifies the difference between the historically predicted risk and the estimated realized future risk (measured *ex post*), to assess the accuracy of risk estimation.

Moreover, to take into account the market state, we choose the volatility as a proxy for qualifying it, since the realized volatility is high in times of crisis, and low when there is little information hitting the market. That is why we analyse the look-forward SQP ratio conditioned to the realized volatility. As estimator for the realized volatility, and to ensure a good rate of convergence for it without requiring strong conditions on the underlying process, we use the mean absolute deviation instead of the root mean square deviation (note that the first version of this study was using the std and can be found in the online-appendix, [34], Appendix A.3). We find a relatively strong (among the indices considered, between -46% and -57% for a VaR of 99%) negative linear correlation between the logarithm of the look-forward ratio and the volatility, as well as a negative rank correlation of the same order (an average over all 11 indices of -49% for the Spearman ρ_S). We observe that in times of high volatility, the ratio is significantly below 1, while it can go even over 3, in times of low volatility, signaling a strong underestimation of future risk. With these empirical observations, we give an answer to question **Q1**. Note that [107] shows that historically estimated VaR underestimates the future risk in a high volatility scenario - when looking at short horizons (next day). This is the opposite conclusion to our findings when looking at long horizons (1 year).

To answer question **Q2**, i.e. looking for factors to characterize this effect, we proceed in two steps. First, we simulate identically and independently distributed (iid) random variables (rv's) and show that it produces a negative correlation between the log-ratio and the volatility as in real data but with less amplitude. Second, we model the stock indices log-returns by GARCH(1,1), a process that models the volatility clustering present in the data. Note that we use on purpose a simple GARCH to isolate the volatility clustering effect (and not to fit as accurately as possible the data); we take first normal innovations, neglecting the tails, then Student-t innovations, to take them into account. Using Monte Carlo simulations, we show that GARCH presents a very similar clustering behavior as in the data, and gives also a strong negative correlation between the log-ratio and the volatility. This negative correlation is stronger than that obtained on the data. From the analysis of those two simple models, we conclude that the pro-cyclicality can be explained by two factors: (i) the way risk is measured as a function of quantiles estimated on past observations, and (ii) the clustering and return-to-the-mean of volatility.

This chapter is organized in three main sections. In Section 2.2, we quantify the predictive power of risk measures. We start by formally treating the risk measure as a stochastic process introducing the SQP. Then, we explore its empirical behavior on 11 stock indices. Subsequently, we define a new statistics, the look-forward SQP ratio, to explore the ability of the SQP to predict the risk for the future year. Building upon this, in Section 2.3 we quantify the pro-cyclicality on these 11 stock indices. Conditioning the SQP on the realized annual volatility allows us to distinguish between the two main market states: normal versus crisis time. We show quantitatively that the SQP underestimates the risk in times of low volatility, while overestimating it in times of high volatility, confirming the pro-cyclicality of risk measurements. Then, Section 2.4 looks for explanations of this behaviour. We subsequently analyse the results on pro-cyclicality when, first, simulating data from an iid model and, then, from a GARCH(1,1) model - and compare them to the results obtained from real data. Through this comparison, we identify two factors producing the (pro-cyclicality) effect with the help of these two models. An intrinsic part due to the risk measurement, as seen with iid rv's, the other, due to clustering and return-to-the-mean of the volatility as produced by a GARCH(1,1) process. We also discuss the influence of the sample size and the data frequency. We summarize the results and discuss their consequences in the conclusion, Section 2.5. Therein we also offer a short list of the takeaways of this chapter, the key questions which follow for us from this study and a list of the content in the appendix which is related to this chapter.

2.2 Quantifying the predictive power of risk measures

2.2.1 Sample Quantile Process

Since the risk is changing with time, we choose to model it dynamically with a risk measure that is defined as a stochastic process. We use the simplest dynamic extension of the VaR, namely the Sample Quantile Process, first introduced in the context of options pricing (see [96], [2], [52]). Here we do not assume any underlying model, and use directly this dynamic risk measure in its broad definition to check its performance in evaluating the risk on financial data.

Definitions

Let us quickly recall the definition of the VaR. For a loss random variable L having a continuous distribution function F_L, the VaR at level α of L is simply the quantile of order α of L:

$$VaR_\alpha(L) = \inf\left\{x : \mathbb{P}[L \leq x] \geq \alpha\right\} = q_L(\alpha). \tag{2.1}$$

The VaR is still predominantly estimated on historical data (see e.g. [103] or [58] for quantitative surveys on this matter), using the empirical quantile $q_n(\alpha)$ associated with a n-loss sample $(L_1, \ldots, L_n)$ with $\alpha \in (0,1)$ defined by:

$$\widehat{\mathrm{VaR}}_n(\alpha) = q_n(\alpha) = \inf\left\{x : \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(L_i \leq x)} \geq \alpha\right\}. \tag{2.2}$$

Note that our choice of definition (2.2) is such that the VaR equals an order statistics of the data. It would also be possible to obtain the VaR by linear interpolation between two order statistics. This might provide a slightly different value: the smaller the data set and/or the higher the threshold, the more the two quantile definitions usually differ. By construction and given our choice of definition of a quantile, the empirical quantile will then be larger or equal to that obtained by linear interpolation.

For our study we need a dynamic version of the VaR, considering the loss as a stochastic process $L = (L_t)_t$: (2.3) corresponds to the continuous version of the VaR defined in (2.2) on the interval $[t-T, t)$:

$$Q_{\alpha,T}(t) = \inf\left\{x : \frac{1}{T}\int_{t-T}^{t} \mathbb{1}_{(L_s \leq x)} ds \geq \alpha\right\}. \tag{2.3}$$

Note that $Q_{\alpha,T}(t)$ is a special case of the Sample Quantile Process $(Q_{\mu,\alpha,T,t}(L))_{t\geq 0}$ of L at threshold α which is defined with respect to a random measure μ on $\mathbb{R}^+$, for $0 \leq T < t$, by (see e.g. [2], [52]):

$$Q_{\mu,\alpha,T,t}(L) = \inf\left\{x : \frac{1}{\int_{t-T}^{t} d\mu(s)} \int_{t-T}^{t} \mathbb{1}_{(L_s \leq x)} d\mu(s) \geq \alpha\right\}. \tag{2.4}$$

Indeed, choosing μ as the Lebesgue measure gives back (2.3). This is the case on which we focus in this chapter. Nevertheless, note that the introduction of μ in the general definition (2.4) offers ways to explore the SQP as a predictor of the risk. For instance, in [29], we considered $\mu(s) = \mu_p(s) = |L_s|^p$, varying $p \in \mathbb{R}$ to study the impact of extreme movements. Here we take $p = 0$ to focus on risk measures used in risk management and regulation.

We introduce the empirical estimator of the SQP, $(\widehat{Q}_{\alpha,T,t})_{t\geq 0}$, defined, for any $t \geq T$, by

$$\widehat{Q}_{\alpha,T}(t) = \inf\left\{x : \frac{1}{T}\sum_{t_i \in [t-T,t)} \mathbb{1}_{(L_{t_i} \leq x)} \geq \alpha\right\}. \tag{2.5}$$

Note that this choice of estimator corresponds to the 'rolling window VaR' (with rolling window size T).

Empirical exploration of the SQP

Given this formalism, we are now going to build realized time series of SQP's computed over various stock indices, which we consider as various sample paths of the process. Since our goal is to look at the appropriateness of the VaR calculations done by financial institutions, it seems natural to use stock indices and see how our various SQP's perform on this data.

Data - We consider data from 11 different stock indices which we have retrieved mainly through Bloomberg (and, for the time periods which were not available, through other global finance sources). 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 Table 2.1. The reason for such a large set of indices is the need to see if the process has similar behavior over most of the developed economies. Moreover, we want a relatively long dataset, which is offered by all these stock indices.

TABLE 2.1: For each index we provide the country, country code, the Bloomberg index (the name under which one can find the index on Bloomberg) and a short description of the index.

Country	Country Code	Bloomberg Index	Description
Australia	AUS	AS30	The All Ordinaries Index (AOI) is made up of the 500 largest companies listed on the Australian Securities Exchange (ASX)
Canada	CAN	SPTSX	The Canada S&P/TSX Composite Index comprises around 70% of market capitalization for Toronto Stock Exchange listed companies
France	FRA	SBF250	The CAC All-Tradable Index is a French stock market index representing all sectors of the French economy (traded on Euronext Paris)
Germany	DEU	CDAX	CDAX is a composite index of all stocks traded on the Frankfurt Stock Exchange in the Prime Standard and General Standard market segment
Great Britain	GBR	ASX	The FTSE All-Share Index (ASX) comprises more than 600 (out of 2000+) companies traded on the London Stock Exchange
Italy	ITA	ITSMBCIG	This index is called Comit Globale R it includes all the shares listed on the MTA and is calculated using reference prices
Japan	JPN	TPX	The Tokyo Price Index (TPX) is a capitalization-weighted price index of all First Section stocks (1600+ companies) of the Tokyo Stock Exchange
Netherlands	NLD	AEX	The AEX index is a stock market index composed of a maximum of 25 of the most actively traded shares listed on Euronext Amsterdam
Singapore	SGP	STI	FTSE Straits Times Index (STI) tracks the performance of the top 30 companies listed on the Singapore Exchange
Sweden	SWE	OMXAFGX	Sweden OMX Affärsvärldens General Index is a stock market index designed to measure the performance of the Stockholm Stock Exchange
United States of America	USA	SPX	S&P 500 is an American stock market index based on 500 large companies listed on the NYSE or NASDAQ (around 80% market capitalization)

As commonly done, we refer to the S&P 500 index as our main example and focus on this specific index whenever showing figures. Selected results for the other indices are available in the Appendix A.1, all other results in the online-appendix, [34]. For the quantitative evaluation we will usually show both,

the (average) results for each index (considering it as one realization of a stochastic process) as well as the average over all 11 indices with the corresponding standard deviation over the 11 results of each index.

For any time series, let us denote the closing price at time t by $S(t)$ and by Δt the interval between two consecutive prices. We focus on daily intervals. We then define the log-return $X_{\Delta t}(t)$ by:

$$X_{\Delta t}(t) = \ln\left(\frac{S(t)}{S(t-\Delta t)}\right) \quad . \tag{2.6}$$

If not specified otherwise, for ease of notation, we will refer to $X_i = X_{\Delta t}(t_i)$ as the daily log-return (with $\Delta t = 1$ day) at time t_i. As the risk measures presented are based on the losses, it is, in some situations, of interest to work with the log-return losses instead of the log-returns, simply defined as: $L_i = -X_i$, the daily log-return loss at time t_i.

Visualization of the SQP- At first, we focus on a rather qualitative evaluation, illustrated on the S&P 500. In Figure 2.1, we plot the rolling-window VaR_α, at monthly frequency, which are moving around the unconditional VaR (i.e. computed on the overall sample) at the same level α, which is a good approximation of the regulatory benchmark. Note that we see the rolling-window VaR as a sample path not only of $(\hat{Q}_{\alpha,T}(t))$, but also of the SQP $(Q_{\alpha,T}(t))$ itself. The sample size, T, is taken to be 1 year and we consider thresholds $\alpha = 95\%$ and 99%, respectively. We see that the fluctuations of $(Q_{\alpha,T}(t))$ around the unconditional VaR can be quite large. Particularly, on the upper side, the movements are more abrupt.

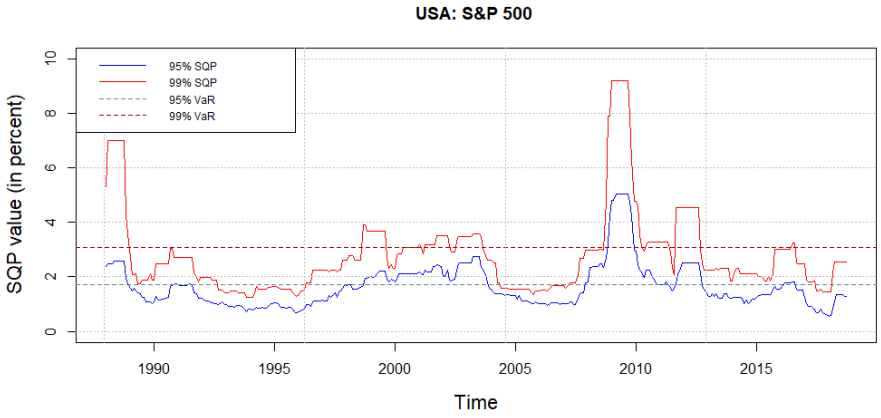


FIGURE 2.1: Sample Quantile Processes (SQP) rolling every month with $T = 1$ year, and thresholds $\alpha = 95\%$ and 99%, respectively. Dashed lines correspond to $\widehat{VaR}_n(\alpha)$ computed over the whole sample.

2.2.2 Predictive power for the risk using the SQP

We now focus on the main goal namely testing the SQP's ability to assess the estimated future risk correctly - what we call predictive power. Let us briefly point out why this differs from the classical backtesting, which designates a statistical procedure that compares realizations with forecasts. First, recall that, for the VaR, some of the backtesting methods used by European banks and the Basel committee are based on the so-called violation process counting how many times the estimated/forecast $\widehat{VaR}_n(\alpha)$ has been violated by realized losses during the following year; those VaR 'violations' should form a sequence of iid Bernoulli variables with probability $(1-\alpha)$ (see e.g. [40] for simple binomial test, [85] for multinomial one). Here instead, we want to assess the risk, thus the quality of estimation. While the

violation ratio in [43] gives a measure of the quality of the estimation (based on the violation process), here we do not question at all if the violation process is in line with the acceptable number of exceptions.

Taking the risk management point of view, we consider the estimated VaR directly as a capital amount (without referring to the confidence interval associated with the estimate). This is why, as pointed out earlier, we propose to interpret this estimate (a statistical number) as an observation of a SQP sample path. We will then judge the relevance of the obtained VaR by testing directly if the capital computed last year (via a VaR estimate) corresponds to that needed this year (the realized future value of the VaR).

Note that it does not mean denying the uncertainty: To cope with it, as well as to show financial strength, a buffer on top of the risk adjusted capital is usually introduced by financial institutions (its evaluation depends on the institution own assessment and not by regulation).

This alternative way of looking at risk (additionally to the violation process), leads subsequently to a new way of assessing the quality of the risk estimation. Moreover, this method for risk quantification can be applied to any other risk measure besides the VaR (as for instance the Expected Shortfall).

Look-forward SQP ratio

To define a look-forward ratio, we consider the SQP estimator $(\widehat{Q}_{\alpha,T}(t))_t$, computed according to (2.5) during a period (of length T years) that ends at time t; for ease of notation, we omit the hat above in the following.

We measure the quality of the risk estimation by comparing it to the VaR, estimated one year later, i.e. $Q_{\alpha,T_s}(t+T_s)$ with $T_s = 1$ year (the index s stands for 'solvency'). This time period corresponds to that required by most regulators: Explicitly for the next twelve month as in insurance, see [59], and implicitly, as an observation period, for banks, see e.g. [16]. It means to check if the estimate $Q_{\alpha,T}(t)$ is a good estimate at time t of the unknown future value $Q_{\alpha,T_s}(t+T_s)$ of the realized risk during that period of time T_s.

This leads us to introduce a new rv, the look-forward ratio, denoted by $R_{\alpha,T}(t)$, which is a ratio of SQP's and quantifies the difference between the historically predicted risk $Q_{\alpha,T}(t)$ and the estimated realized future risk $Q_{\alpha,T_s}(t+T_s)$ computed using the period of time $[t, t+T_s)$:

$$R_{\alpha,T}(t) = \frac{Q_{\alpha,T_s}(t+T_s)}{Q_{\alpha,T}(t)}. \tag{2.7}$$

To be clear, let us look at a simple example. Choosing $\alpha = 95\%$, $t =$ January 2014, and $T = T_s$, the denominator $Q_{95\%,1}(t)$ of the ratio $R_{95\%,1}(t)$ corresponds then to the empirical estimate of the SQP computed on a 1 year sample from January 2013 ($= t - T$) to December 2013 (*i.e.* the month before t, as we are considering the interval $[t-T, t)$), which acts as an estimate for the future risk in 2014 (if T would be 3 years, $Q_{95\%,3}(t)$ would be computed from January 2011 up to December 2013). The numerator corresponds to the future risk $Q_{\alpha,T_s}(t+T_s)$ and is defined as the rolling VaR estimated on the one year sample from January 2014 (t) until December 2014 (last month before $t+T_s$).

Consequently, if we look at the values of the ratio $R_{\alpha,T}(t)$, a value near 1 means that the prediction $Q_{\alpha,T}(t)$ correctly assesses the future risk estimated by $Q_{\alpha,T_s}(t+T_s)$. Similarly, if the ratio is above 1, we can say that we have an under-estimation of the future risk, i.e. the observed future risk. And to the contrary, if it is below 1, we have an over-estimation of the future risk, which would lead to an over-evaluation of the needed capital.

Assessing the average SQP ratio

Taking a monthly rolling window for the ratios $R_{\alpha,T_s}(t)$ (i.e. updating t every month), we compute the average SQP ratio over the whole sample. Here we focus on looking at the overall averages (over the

11 different indices) of the monthly SQP ratios. But results for the 11 different indices on their own are reported in Table 2.2, too.

TABLE 2.2: SQP ratios (defined in (2.7)) average over the whole historical sample, for T equal to 1 year, for each index and for four different thresholds (95%, 97.5%, 99 % and 99.5%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG** ($\pm\,\sigma$)
95.0%	1.02	1.06	1.06	1.09	1.07	1.08	1.07	1.05	1.07	1.06	1.05	**1.06** $\pm$ 0.02
97.5%	1.01	1.08	1.05	1.08	1.07	1.10	1.07	1.05	1.06	1.06	1.06	**1.06** $\pm$ 0.02
99.0%	1.02	1.08	1.04	1.08	1.08	1.11	1.09	1.11	1.07	1.06	1.07	**1.07** $\pm$ 0.03
99.5%	1.06	1.09	1.04	1.07	1.08	1.17	1.07	1.11	1.08	1.05	1.08	**1.08** $\pm$ 0.04

We see, in Table 2.2, that all the averages over the 11 indices of the look-forward ratios are clearly different from 1, telling us that the future risk is, on average, not well predicted. When comparing these averages for different thresholds, we see that they increase with the threshold but do not change a lot. Furthermore, we clearly see for all four choices of α that the average ratio is above 1, even when considering two standard deviations around the average value. We find back the well known result that the historical estimation of VaR underestimates the risk (on average).

To complete this analysis, we look at the Root Mean Square (RMS) distance and compute the associated Root Mean Square Error (RMSE), i.e. the average RMS distance of the ratio from a perfect prediction (i.e. ratio of 1):

$$RMSE := \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(R_{\alpha,T_s}(t_j)-1\right)^2}, \tag{2.8}$$

where $t_1, ..., t_N$ are the N time points the ratio is evaluated at. Recall that, as we are working with a monthly rolling window, N equals 12 times the amount of years on which the ratio is tested on. This RMSE measures the "error" of the forecasted risk for a certain sample path.

TABLE 2.3: Average RMSE (defined in (2.8)) for the SQP ratios (eq. (2.7), with $T = T_s$) - using historical data for each index. In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG** ($\pm\,\sigma$)
95.0%	0.36	0.45	0.44	0.50	0.48	0.54	0.51	0.46	0.47	0.43	0.44	**0.46** $\pm$ 0.05
97.5%	0.34	0.52	0.43	0.51	0.47	0.57	0.53	0.46	0.46	0.44	0.45	**0.47** $\pm$ 0.06
99%	0.36	0.55	0.39	0.50	0.51	0.60	0.57	0.70	0.51	0.47	0.51	**0.52** $\pm$ 0.09
99.5%	0.48	0.58	0.40	0.49	0.51	0.73	0.55	0.71	0.53	0.45	0.54	**0.54** $\pm$ 0.10

When considering the RMSE of the SQP ratio, see Table 2.3, we observe that, on average over the 11 different indices, the error increases with the threshold. Again, the average RMSE is clearly above 0 reaching 50% of the right value 1 of the SQP ratio. This means that there are high fluctuations of the ratio, for the underestimation of slightly less than 10% (recall the last column of Table 2.2). This analysis shows that the ratio alone cannot give a fine enough picture of the quality of risk estimation. Indeed, we

want to understand why and in which circumstances we over- or underestimate the risk measured by the SQP throughout the sample. It is what we tackle next.

2.3 Pro-cyclicality results on real data

As mentioned in the introduction, pro-cyclicality of risk measurements implies that, when the market is quiet, the future risk is under-estimated, while, during crisis, the future risk is over-estimated. We deal with this problem by first defining an indicator of market states, then evaluating the performance of the risk measure when conditioning it on this indicator.

2.3.1 Realized volatility as an indicator of market states

To better understand the dynamic behavior of the ratios, we condition them on past volatility. This is inspired by the seminal empirical study on Foreign Exchange Rates done in [41], where the authors conditioned the returns on past volatility. On the one hand, they showed that after a period of high volatility, the returns tend to be negatively correlated. It means that the volatility should diminish in the future, with a distribution more concentrated around its mean; thus the quantiles should be smaller at given thresholds than current ones. It implies that the future risk would be overestimated when evaluated on this high volatility period. On the other hand, they showed that in times of low volatility, consecutive returns tend to be positively auto-correlated; thus the volatility should increase on the long term, with the consequence in the future of quantiles shifted to the right, meaning that our estimation would underestimate the future risk.

To be able to condition on volatility, we define an empirically measured annualized volatility, $v_{k,T}(t)$, taken over a rolling sample of size T year(s), corresponding to n business days, up to but not including a time t. By abuse of notation, we refer with $t-j$ to a point in time j days before time t (for any integer-valued j), which gives us:

$$v_{k,T}(t) = \sqrt{n} \times v_{k,n}(t-1)\,, \quad \text{where} \quad v_{k,n}(t-1) \text{ denotes a realization of} \tag{2.9}$$

$$V_{k,n}(t-1) := \left\{ \frac{1}{n-1} \sum_{i=t-n}^{t-1} \left| X_i - \frac{1}{n} \sum_{i=t-n}^{t-1} X_i \right|^k \right\}^{1/k}, \tag{2.10}$$

$$\text{and, accordingly, } V_{k,T}(t) := \sqrt{n} \times V_{k,n}(t-1), \tag{2.11}$$

the X_i's denoting the daily log-returns defined in (2.6), and k playing the role of emphasizing more or less large events. We consider $k = 1, 2$. If $k = 2$, $v_{2,T}(t)$ is nothing else than the usual empirical standard deviation (denoted std), called in the literature 'realized volatility', whereas if $k = 1$, it is the empirical Mean Absolute Deviation around the sample mean (denoted MAD). We may generalize the notion of realized volatility to any k but use it here for $k = 1, 2$. Note that the notation is chosen such that the annual realized volatility $v_{k,T}(t)$ at time t (January 1, 2014 for instance) is computed on the previous T year(s) until $t-1$ included (i.e. until December 31, 2013, in our previous example). We evaluate $v_{k,T}(t)$ on a rolling basis every month. Hereby, we obtain a monthly time series of annual realized volatility $(v_{k,T}(t))_t$, which can be used as a benchmark of the current market state.

In Figure 2.2, we present $v_{k,T}(t)$ for the S&P 500 (log-)returns between January 1987 and September 2018 for both a rolling sample of $T = 1$ year and of 3 years, as well as for $k = 1$ and $k = 2$. We also indicate the average annual realized volatility of the index calculated over the whole sample for both $k = 1$ and $k = 2$ (represented by the dashed lines). Their values are 11.9% for $k = 1$ and 18.1% for $k = 2$. We can see that the volatilities above these benchmarks mostly stand for periods of high instability or crisis (and not only in the USA). The high volatility of the period 1987-1989 is for instance explained

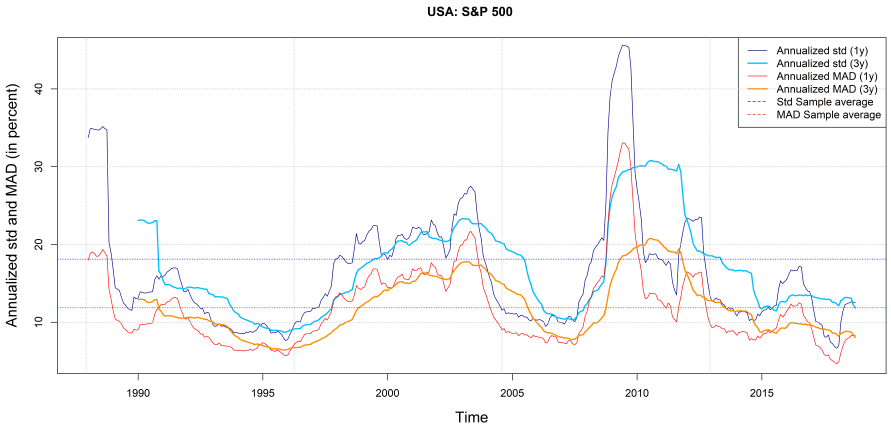


FIGURE 2.2: Annual realized volatility $v_{k,T}(t)$, defined in (2.9), in percent, for $k = 1$ (MAD) and $k = 2$ (std), of the S&P 500 returns over a rolling sample of size $T = 1, 3$ year(s). The dashed lines represent the sample average (over the whole sample) of the two different measures of the volatility (with the same color code).

by the New York Stock Exchange crash in October 1987. In 1997, Asia was hit by a crisis, as well as Russia in 1998 and Argentina in 1999-2000. In 2001, the United States experienced the bursting of the internet bubble. Following the Lehman Brother's bankruptcy, the period of 2008-2009 was a period of very high volatility. And finally, the sovereign debt crisis in Europe also impacted the S&P 500 Index in 2011-2012; this is an illustration of the increased dependence between markets during times of crisis. Observing those co-movements in various indices during crises is important for portfolios of financial assets, when looking at pro-cyclicality of the global market. Moreover, we see that the dynamics of both definitions ($k = 1$ and $k = 2$) in (2.9) are quite similar, while the size of the movements are more pronounced in the case of the estimated std.

We conclude from the analysis above that the annual realized volatility $v_{k,T}(t)$ for $k = 1, 2$ is a good indicator that reacts well and quickly to various market states. It is thus a reasonable proxy to qualify the times of high risks and to use it to discriminate between quiet periods and periods of crisis. We use it to condition our statistics (the look-forward SQP ratio (2.7)) and see if we can detect different behaviors of the price process (which underlies the risk measure) during these periods in comparison to quiet times, as it was the case for returns in [41]. We now turn to this statistics to look at the ability of the SQP risk measures to correctly predict the future risks, whatever the market state, and to capture the dynamics. Note that to ease the reading in the rest of the chapter, we will use abusively in the text the term 'volatility' for the expression 'annualized realized volatility', $v_{k,T}$ (as well as for the rv $V_{k,T}$).

2.3.2 SQP predictive power as a function of volatility

In the case of the S&P 500, we represent in the same graph the time series of both the volatility $v_{k,T}(t)$ and the SQP ratios, separately for each of the thresholds 95% and 99%, to compare their behavior over time. We illustrate in Figure 2.3 these dynamics when considering e.g. the case of a 1-year sample and $k = 1$ for the volatility (MAD). One can easily observe on these graphs a strong opposite behavior for SQP ratios (in dark) and for volatility (in light). Also, the dynamic behavior appears clearly.

The realized SQP ratios can differ from 1 either on the upper side or on the lower side. The negative dependence between the realized SQP ratios and the volatility is obvious on both graphs. This particular feature is of importance in our study because it means that when the volatility in year t is high in the market, the realized SQP ratio in year t is quite low. Recalling the definition of the ratio, this situation

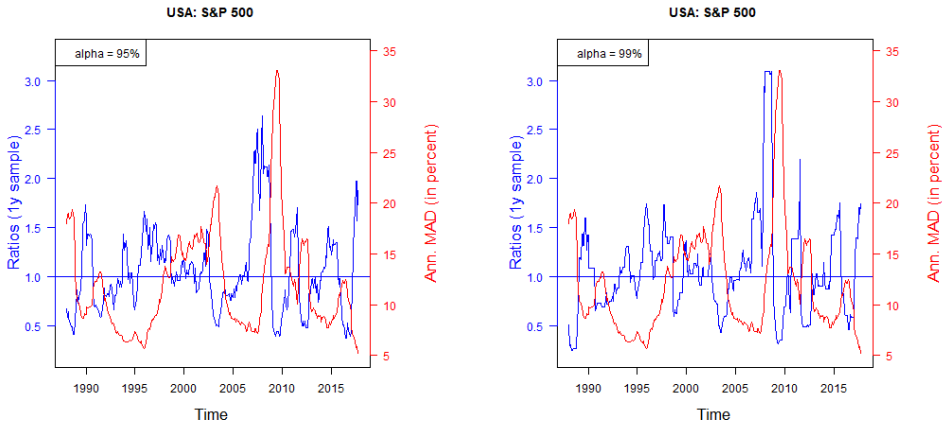


FIGURE 2.3: Time series of SQP Ratios (2.7) (left y-axis in dark) and of annualized realized volatility (2.9) for $k = 1$ (MAD; right y-axis in light) computed over 1 year ($T = 1y$) for $\alpha = 95\%$ (left plot) and $\alpha = 99\%$ (right plot)

means that the risks have been over-evaluated by the SQP calculated in year t (in comparison to the realized risk in year $t+1$). Conversely, when the volatility is low in year t, the realized SQP ratio is often much higher than 1 in this year, which means that the risks for the year $t+1$ have been under-evaluated by the calculation of the SQP in year t.

Another way of visualizing the pro-cyclicality is to plot the various realized SQP ratios as a function of volatility. We illustrate the obtained behavior in Figure 2.4, choosing the same parameters as in Figure 2.3. Such figures highlight better the existing negative dependence between these two quantities. The plots of the first row of Figure 2.4 point out that the (negative) dependence is non linear, hence we consider the log-ratios instead (see the plots on the second row), which provides a higher dependence than without the log: We observe that the more volatility there is in year t, the lower are the ratios $R_{\alpha,T}(t)$. When the log-ratios are negative, it means that losses in year $t+1$ have been overestimated with the measures calculated in year t. In the original scale, we can better quantify the magnitude of the effect: When looking at the case $\alpha = 99\%$, with respect to the overestimation, we can observe ratios below 0.5 for high volatilities (above twice the average of 12%, with MAD), *i.e.* the risk computed at the height of the crisis is more than twice the size of the risk measured a year later. On the other hand, we see that the underestimation can be very big since the realized ratios for the SQP can be even larger than 3; in other words, the risk next year is more than 3 times the risk measured during the current year for this sample. We also observe that the overestimation is very systematic for high volatility, whereas it is not for the underestimation at low volatility. Indeed, in this latter case, we see also values below 1, while we do not see any value above 1 when the volatility is high (above 20%, close to twice the average over the whole sample). Although we present here only the S&P 500, very similar behaviors are exhibited by all the other indices, see e.g. the Appendix A.1.

Given the highly non-linear behavior of the dependence (as illustrated in Figure 2.4), we choose the *logarithm* of the SQP ratio (instead of the SQP ratio) to look at the linear Pearson correlation $\rho\left(\log\left(R_{\alpha,T}(t)\right), V_{k,T}(t)\right)$ considering the volatility choosing $k = 1$ (MAD). We will keep this choice of volatility (MAD) in the rest of the paper. Indeed, an important motivation for this choice, in particular for practical use, is that it implies weaker conditions on the moments of the underlying distribution than when using the classical standard deviation ($k = 2$). For the correlation between the two quantities (the log-ratio and the volatility respectively), the existence of the variance of those estimators is needed. I.e. the existence of a second moment of the underlying distribution in the case of the MAD instead of a fourth moment for the standard deviation. So, from now on, we use the name 'volatility' for 'annualized

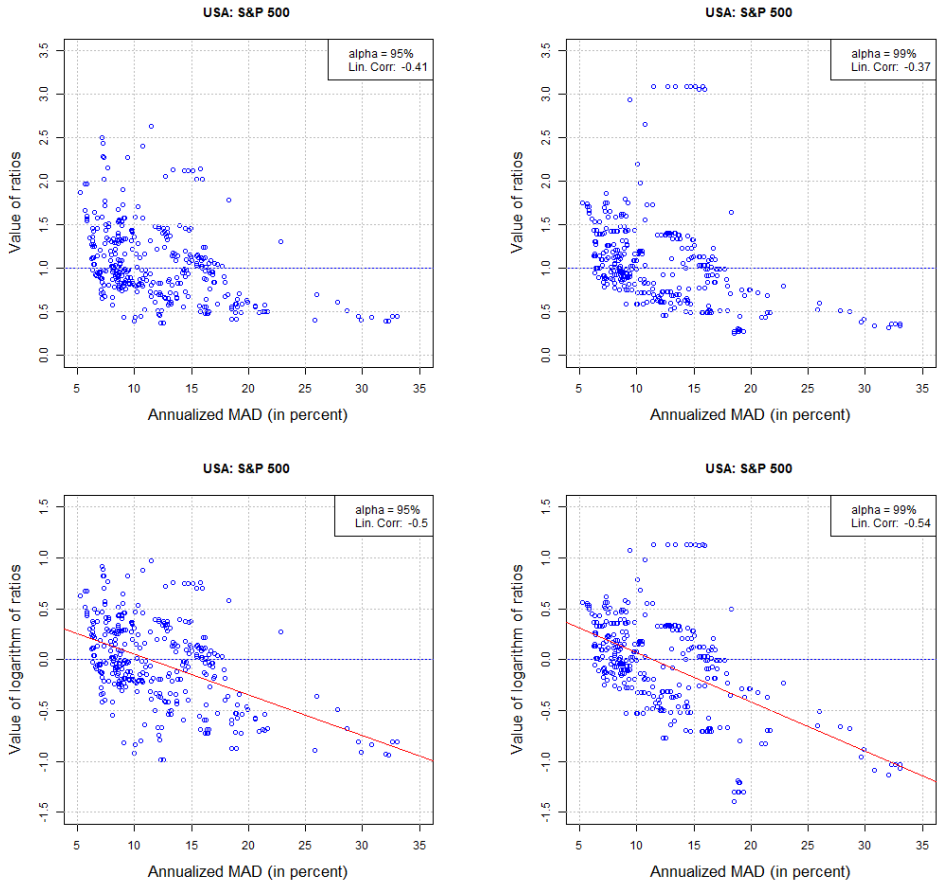


FIGURE 2.4: SQP ratios as a function of volatility (MAD): On the left $\alpha = 95\%$ and on the right $\alpha = 99\%$; first row SQP ratios, second row logarithm of SQP ratios

realized MAD' ($v_{1,T}$ and the rv $V_{1,T}$).

In Table 2.4, we present the value of $\rho(\log(R_{\alpha,T}(t)), V_{1,T}(t))$ for all indices and as the average over all indices. We see that the correlation is always significantly negative. The average values for the four different thresholds α are very similar and around 50% (in absolute value).

Another way to present this pro-cyclicality analysis, more in line with the risk management point of view, may be through the volatility binning. Indeed, since we have seen that the SQP is underestimated for low volatility and overestimated for high one, we can show by how much the SQP is over- or underestimated for a hypothetical (represented by the stock index) portfolio of a bank, depending on the volatility bins. To do so, we compute the average SQP ratios in each bin, choosing 5 and 10 uniform volatility bins. Results are illustrated for 5 bins and all indices in Table 2.5.

We observe that the lowest bin underestimates on average the capital for next year by 35%, whereas the highest one overestimates it by 22%. The same holds true for 10 bins (these results can be found in the online-appendix, [34]), but with increasing gaps between the extreme bins (overestimation of 44% and underestimation of 33%). Those numbers are significant from the risk management prospective; it means increasing by more than a third (35 or 44) the capital amount in periods of low volatility, while decreasing it by more than a fifth (22 or 33) during crisis periods.

To explore further the non-linear dependence, beyond the logarithm transformation, we also consider the notion of rank correlation, looking for instance at the Spearman ρ_S.

TABLE 2.4: Pearson correlation $\rho(\log(R_{\alpha,T}(t)), V_{1,T}(t))$ between the logarithm of the SQP ratios and the volatility, for each index and for $T = T_s = 1$ year, over the whole historical sample, and for four thresholds (95%, 97.5%, 99% and 99.5%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG** ($\pm$ **sd**)
95%	-0.51	-0.46	-0.50	-0.49	-0.52	-0.58	-0.52	-0.50	-0.51	-0.50	-0.50	**-0.51** $\pm$ 0.03
95%	-0.55	-0.47	-0.49	-0.44	-0.52	-0.58	-0.53	-0.50	-0.50	-0.53	-0.52	**-0.51** $\pm$ 0.04
99%	-0.57	-0.46	-0.51	-0.47	-0.52	-0.52	-0.55	-0.54	-0.48	-0.50	-0.54	**-0.51** $\pm$ 0.04
99.5%	-0.52	-0.46	-0.47	-0.45	-0.51	-0.48	-0.54	-0.50	-0.45	-0.49	-0.53	**-0.49** $\pm$ 0.03

TABLE 2.5: The average SQP ratios (defined in (2.7)) within 5 uniform bins ($bin_1, \cdots, bin_5$ ordered from lowest to highest) of volatility over the whole historical sample, for each index. In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG** ($\pm \sigma$)
bin_1	1.21	1.23	1.35	1.35	1.36	1.64	1.36	1.30	1.38	1.33	1.31	**1.35** $\pm$ 0.11
bin_2	1.09	1.03	1.06	1.13	1.17	1.11	1.15	1.33	1.24	1.09	1.09	**1.14** $\pm$ 0.08
bin_3	1.07	1.20	1.12	1.10	1.22	0.95	1.11	0.98	1.11	1.04	1.01	**1.08** $\pm$ 0.08
bin_4	0.96	0.99	0.99	1.11	0.94	0.93	1.04	0.82	0.86	1.03	1.11	**0.98** $\pm$ 0.09
bin_5	0.77	0.87	0.78	0.74	0.68	0.79	0.73	0.83	0.79	0.81	0.74	**0.78** $\pm$ 0.05

TABLE 2.6: Average over all indices of the Spearman ρ_S between the volatility and the SQP ratios computed on a 1-year sample for four thresholds $\alpha = 95\%, 97.5\%, 99\%$ and 99.5%.

$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$	$\alpha = 99.5\%$
-0.48 $\pm$ 0.05	-0.49 $\pm$ 0.03	-0.49 $\pm$ 0.03	-0.48 $\pm$ 0.03

In Table 2.6, we display only the average Spearman ρ_S over all indices as we have seen in the other tables that indices do not differ significantly from one another in their behaviors. In this table, the results confirm those obtained when considering the Pearson correlation between the logarithm of the SQP ratios and the volatility, only the magnitude is slightly lower than for the Pearson correlation.

Now let us turn to the question of why we observe such (negative) dependence. It is the target of the next section, to look for factors that explain empirically the pro-cyclical behavior observed so far.

2.4 Pro-cyclicality results on different models

We want to find out empirically the reasons for the under-/over-estimation of risk. To do so, we need to use models able to isolate effects.

The first and simplest model we consider is iid rv's. The reason of such choice is to check if the very way we measure risk, independently of volatility models, creates some negative dependence between the log-ratio and the realized volatility.

Then we explore if the volatility clustering present in the data is fast enough to produce pro-cyclicality. To this aim, we choose the simple GARCH(1,1) model, to isolate the clustering effect by itself. Indeed, there are many GARCH type models in the literature (e.g. EGARCH, HARCH, HAR, etc...) that might reproduce the empirical regularities of the data better than this choice, but they would mix effects (e.g. EGARCH mixes clustering and skewness, HARCH and HAR mix various clustering at different frequencies); it would render the interpretation much more difficult: from which effect would the pro-cyclicality come? We then calibrate the GARCH(1,1) to the data, considering first normal innovations, then Student ones to see if there would be any additional effect to the clustering when better calibrating the tails.

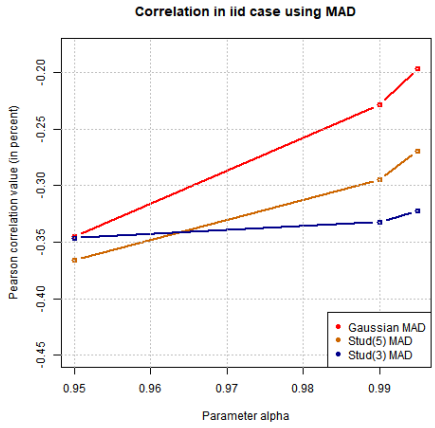
2.4.1 IID model

Let us consider a rv X, with mean μ, finite variance σ^2, distribution function F_X. We simulate 100'000 samples each of size N, $(X_1, \cdots, X_N)$, with parent rv X, from which we estimate, on rolling subsamples of size n, the statistics considered so far: the logarithm of the look-forward ratio (2.7), and the volatility (2.10) for $k = 1$. To compute those statistics, we produce time series of empirical estimators $(Q_{\alpha,T}(t))_{t\geq 0}$ (see (2.5)), with $T = 1$ year, corresponding to $n = 252$ business days, denoted, for simplicity, by $(Q_{\alpha,n}(t))_{t\geq 0}$ and $(V_{1,n}(t))_t$ (defined in (2.10)), computed on rolling samples updating t every month. We evaluate the empirical Pearson and Spearman correlations between the time series of the logarithm of a ratio of sample quantiles, $\log\left(Q_{\alpha,n}(t+1y)/Q_{\alpha,n}(t)\right)$, and the time series of the sample volatility $V_{1,n}(t)$. Note that the sample quantiles $Q_{\alpha,n}(t)$ and $Q_{\alpha,n}(t+1y)$ are computed on subsequent but disjoint samples, i.e. the points in time are t and $t+1y$ (t plus 1 year), and the sample size n is chosen as $n = 252$ (which corresponds to 1 year of data). Hence, those samples are disjoint by construction.

We consider two standard cases for F_X, the Gaussian and the Student (with 3 or 5 degrees of freedom) distributions, to study the effect of both light and heavy tail distributions.

TABLE 2.7: Averages over 100'000 repetitions of Pearson and Spearman correlation between $\log\left(Q_{\alpha,n}(t+1y)/Q_{\alpha,n}(t)\right)$ and $V_{1,n}(t)$ for the IID case using $\alpha = 95\%, 99\%, 99.5\%$, $n = 252$ and updating t every month. Two asterisks indicate a negative value for the correlation at the 99% confidence level, one at 95%, none at the 90%, and † at the 85%. On the right plot, depiction of the empirical averages of the Pearson correlation when using $V_{1,n}(t)$.

	Normal sample $X \sim \mathcal{N}(\mu,\sigma)$	Student-t sample $X \sim t(5,\mu,\sigma)$	$X \sim t(3,\mu,\sigma)$
Pearson			
$\alpha = 95\%$	$-0.34^{\star\star}$	$-0.37^{\star\star}$	$-0.35^{\star\star}$
$\alpha = 99\%$	-0.23	$-0.29^{\star}$	$-0.33^{\star\star}$
$\alpha = 99.5\%$	$-0.20^{\dagger}$	$-0.27^{\star}$	$-0.32^{\star}$
Spearman			
$\alpha = 95\%$	$-0.33^{\star\star}$	$-0.35^{\star\star}$	$-0.34^{\star\star}$
$\alpha = 99\%$	-0.22	$-0.28^{\star}$	$-0.32^{\star\star}$
$\alpha = 99.5\%$	$-0.19^{\dagger}$	$-0.26^{\star}$	$-0.31^{\star\star}$



The results are displayed in Table 2.7, when taking $N \simeq 8000$ in the simulations (to match the real data sample sizes), thresholds $\alpha = 95\%$, 99% and 99.5%, and the definitions of the volatility for the

MAD ($k = 1$). We observe that, at high confidence levels, both Pearson and Spearman correlations are negative and of the same order. Finally, we also observe that the correlation decreases in absolute value when the threshold α increases, as illustrated in the plot of Table 2.7.

Comparing the results in Table 2.7 with those obtained with historical data in Table 2.4, the correlation appears less pronounced in the former (for $\alpha = 99\%$, the Pearson correlation is -0.51 for the average over all the stock indices, compared to -0.29 for Student-t ($\nu = 5$) iid rv's). So, the pro-cyclicality is only partially explained by the way risk is measured. Some part of the negative correlation stays yet not explained. In the next section, we turn to the GARCH model to see if volatility clustering might explain the remaining effect.

2.4.2 GARCH$(1,1)$ model

Fitting the GARCH - In our empirical analysis, we observe volatility clustering that goes, of course, hand in hand with a return to the mean of volatility. Thus, part of the pro-cyclical behavior with volatility might be explained by this clustering, since the measuring is done on a longer time period than the typical clustering size. A natural model, namely the GARCH, comes to mind as it has been developed precisely for modelling volatility clustering. Recall that our goal is not to fit as accurately as possible the data (choosing the latest generation of GARCH type models) but to isolate causes for the pro-cyclical behavior.

Given this aim, we turn to the simplest version of GARCH, the GARCH(1,1), introduced in [23], to model the log- returns $(X_t)_t$ as: $X_{t+1} = \varepsilon_t \sigma_t$, where the variance σ_t^2 of X_t satisfies

$$\sigma_t^2 = \omega + \alpha X_t^2 + \beta \sigma_{t-1}^2, \quad \text{with} \quad \omega > 0,\ \alpha \geq 0,\ \beta \geq 0, \tag{2.12}$$

where, for our purposes, the error terms $(\varepsilon_t)_t$ constitute either a Gaussian or a Student-t with ν degrees of freedom white noise (with mean 0 and variance 1).

To estimate our model, we fit the parameters ω, α, β to each full sample of the 11 indices, using a robust optimization method proposed by Zumbach [130], which is more exact in parameter estimation and more precise in reproducing the average volatility compared to some of the other standard methods used. As a side note, we remark that the robustness is especially important as other optimization methods may not always fulfill the stationarity condition ($\alpha + \beta < 1$) for the GARCH process. A deeper analysis on this issue, comparing the performance of different statistical packages used for GARCH fitting with the software R, can be found in Appendix E.

Even though we consider a simple model, some care is needed to find a good way to optimize the GARCH(1,1) parameters with the right choice of innovations, in order to reproduce both the clustering effect and the accurate tail behavior present in the data. To do so, we argue on both the maximum likelihood and the τ_{cor} criteria, where τ_{cor} is a parameter specifying the time decay δt of the autocorrelation of the squared log returns, expressed in terms of (α, β): $\tau_{\text{cor}} := \delta t / |\log(\alpha + \beta)|$ (see Eq. (12) in Zumbach [130]). Let us explain how we arrive at the optimal model: We start with (2.12) optimized with, respectively, Gaussian and Student innovations on the full samples. We observe that the model with Gaussian innovations provides a good fit for the τ_{cor} and the mean volatility, whereas the tail is underestimated. Switching from Gaussian to Student innovations increases the likelihood, which is good, but also the τ_{cor}, worsening the fit of the autocorrelation of the squared log returns. Moreover, the tail is still underestimated particularly for the negative returns (losses). Therefore, we fit the GARCH with Student innovations *on the losses only*, obtaining reasonable tails, with degrees of freedom ν around 5-6, but obviously too short τ_{cor} and smaller normalized likelihood. Hence, to optimize both the τ_{cor} and the tail, we choose the model combining the GARCH (1,1) parameters obtained with Gaussian innovations on full samples, with Student innovations fitted on the losses only. We check that this model has a better likelihood than the one with Gaussian innovations, and a similar one as for the model optimized on full

samples with Student innovations. Moreover, this model combines a good fit for the τ_{cor} and the mean volatility, and a good tail estimation.

TABLE 2.8: Fitting results of the GARCH models, defined in (2.12), for the 11 indices

	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA
Gaussian innovations											
ω $[10^{-6}]$	3.45	1.11	2.54	2.18	2.42	5.29	2.21	2.78	3.29	1.81	1.70
α $[10^{-1}]$	1.58	1.03	0.93	0.87	0.80	1.42	1.07	1.23	1.05	1.06	0.99
β $[10^{-1}]$	8.06	8.85	8.90	8.99	9.06	8.27	8.81	8.60	8.75	8.77	8.88
$\alpha+\beta$	*0.96*	*0.99*	*0.98*	*0.99*	*0.99*	*0.97*	*0.99*	*0.98*	*0.98*	*0.98*	*0.99*
τ_{cor} (days)	28	86	60	71	70	32	80	60	51	58	76
Fitting Results											
Normalized Likelihood	3.37	3.45	3.12	3.15	3.05	3.04	3.13	3.16	3.09	3.34	3.27
Volatility [%]	15.7	15.5	19.5	19.7	20.6	20.9	20.9	20.4	20.6	16.3	17.9
Historical [%]	15.7	15.5	19.6	19.7	20.7	21.0	21.1	20.6	20.7	16.4	18.1
Student innovations (fitted on losses) with Gaussian GARCH parameters											
ν	5.15	4.75	5.97	6.00	5.98	5.89	5.77	5.09	6.44	6.25	4.75
Fitting Results											
Normalized Likelihood	3.40	3.47	3.14	3.17	3.07	3.07	3.15	3.20	3.12	3.36	3.30
Volatility [%]	15.4	14.9	19.5	19.5	20.5	20.8	20.6	19.7	20.6	16.2	17.5

In Table 2.8, we report the GARCH parameters, as well as the results for τ_{cor} and the normalized likelihood (for fair comparison between the samples of different sizes) - for both, the GARCH model with Gaussian innovations as well as the optimized model keeping the Gaussian GARCH parameters (ω, α, β), but considering now Student innovations fitted on the losses only. We see that the optimization gives, in all cases, parameters that fulfill the GARCH stationarity condition. The estimated parameters do not vary much from one index to the other, except for AUS and JPN. Those two indices exhibit the shortest τ_{cor}. The typical clustering obtained with this fit is between 5 weeks (shortest period for AUS) and 4 months, with an average value of 3 months (a business month contains 21 days), which is short enough to produce pro-cyclicality on a yearly horizon. As can be seen in Figure 2.5, the τ_{cor} for the Gaussian innovations reproduces better the clustering of the data than for the Student ones, which present a too long extent.

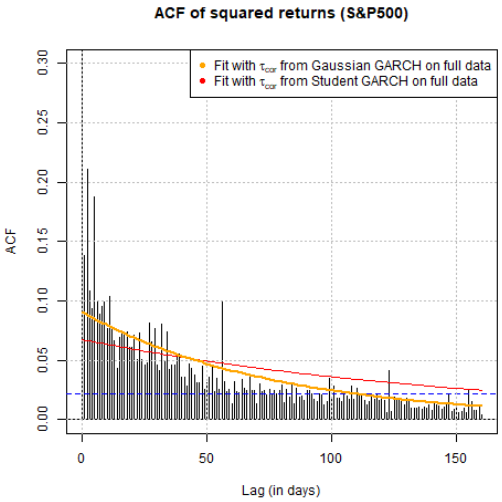


FIGURE 2.5: Autocorrelation function (ACF) of the squared returns of the S&P500 with the fitted τ_{cor} for Gaussian (yellow) and, respectively, Student (red) innovations. The (blue) dashed line represents the 95% confidence level for no autocorrelation.

Although we do not expect a large variation for the likelihood, as it is not so sensitive to tails, we observe in Table 2.8 a slight increase of the normalized likelihood for all indices, when going from Gaussian to Student innovations; it indicates a consistent improvement in the description of the data.

As an additional check for the goodness of fit, we report in Table 2.8, for each index, the average volatility obtained by simulation of the fitted GARCH (with 1000 replications for each index, considering the same sample size as the original data sample), and compare it with the average annualized historical volatility over the real data. We see that the volatilities of the GARCH reproduce very well (slightly lower) the historical ones on average, when estimated over 1000 replications of the GARCH process, whatever the innovation, although somewhat better for the Gaussian innovations.

Dependence between SQP ratios and annualized realized volatility - Now we consider both the GARCH model with Gaussian innovations and the optimized one with Student innovations (with Gaussian GARCH parameters). Doing so allows us to study separately the effect of clustering from the tail influence. We apply the same statistical analyses as those done on the historical data on each of the 1000 replications for both models, focussing on the dependence between the volatility and the logarithm of the SQP-ratios: All in all, we observe qualitatively similar results with the GARCH simulations as with the historical data and the values are closer to those of the indices than in the case of iid rv's.

As in Section 2.3.2, we first look at the Pearson correlation between the annualized daily volatility (recall, we focus on the MAD) and the logarithm of the SQP ratios computed on a 1-year sample. Those are depicted in Table 2.9: We observe negative correlations of the GARCH for the averages, as for historical data. Yet, they are higher (in absolute value) than for the data (see Table 2.4). Furthermore, the negative correlation is systematically (on average for all indices) decreasing (in absolute value) with the threshold, as in the case of historical data, but more pronounced. Note that the choice of innovations has a weak impact on the correlation values, almost vanishing with higher thresholds. It shows that the tail fatness is not a discriminating factor for pro-cyclicality.

TABLE 2.9: Average (over all indices and its 1000 repetitions each) Pearson correlation between volatility and logarithm of SQP ratios computed on a 1-year sample, using a GARCH(1,1) model, and for four thresholds (95, 97.5, 99, and 99.5%). Standard deviation is computed over the 11 averages of the values obtained for each index.

AVG$\pm\sigma$	$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$	$\alpha = 99.5\%$
Gaussian	-0.63 $\pm$ 0.01	-0.61 $\pm$ 0.01	-0.58 $\pm$ 0.01	-0.57 $\pm$ 0.01
Student-t	-0.63 $\pm$ 0.01	-0.62 $\pm$ 0.01	-0.59 $\pm$ 0.01	-0.57 $\pm$ 0.01

As in Figure 2.4, we illustrate the obtained results with the S&P 500 index when considering one (simulated) sample path of the fitted GARCH(1,1) model (with Gaussian GARCH parameters and Student innovations). The simulated sample path depicted was chosen such that it is representative of the average behavior observed in the simulation. We look at the various realized SQP ratios (as well as its logarithm), as a function of the volatility; see Figure 2.6. For a better comparison between the GARCH realization and the historical data, we use the same y-scale for both. We observe a similar behavior for GARCH (Fig.2.6) and historical data (Fig.2.4):

- The correlation is negative in both cases and the slopes of the linear regression lines look similar.
- The more volatility there is in year t, the lower than 1 are the ratios $R_{\alpha,T}(t)$, which means that losses in year $t+1$ have been overestimated with the measures calculated in year t.
- This overestimation can result in ratios below 0.5: The risk computed at the height of the crisis, in the realization below, is more than double the size of the risk measured a year later.

- For the underestimation, we see that the realized ratios can be larger than 3; in other words, the risk next year is more than three times the risk measured during the current year for this sample.
- As with the historical data, this overestimation is systematic for high volatility, whereas it is less the case for low volatility.

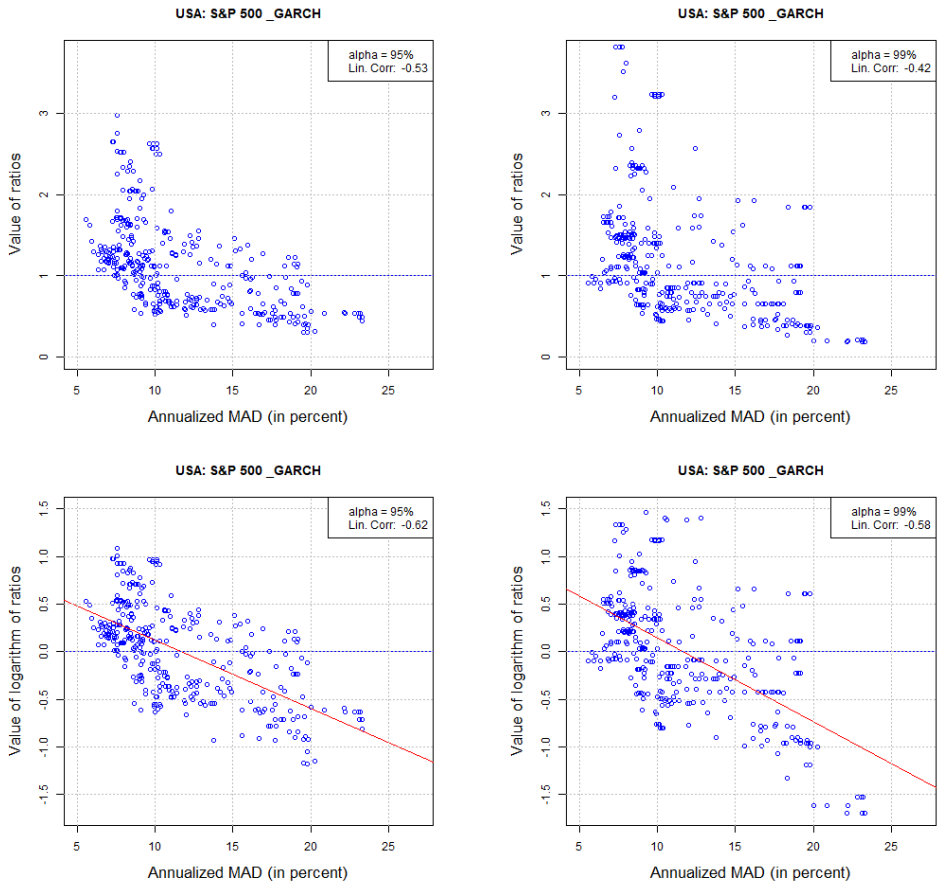


FIGURE 2.6: SQP ratios as a function of volatility (MAD) for one simulated sample path of the GARCH(1,1) (that fits the S&P500): On the left $\alpha = 95\%$ and on the right $\alpha = 99\%$; first row SQP ratios, second row logarithm of SQP ratios

We notice in Figure 2.6, that the linear correlation is not an optimal approximation of the dependence structure (as already observed for historical data, although the correlation is slightly stronger for GARCH). Thus, we consider the logarithm of the SQP ratio to take into account the non linear dependence and, additionally, compute a rank correlation (Spearman)- see Table 2.10.

TABLE 2.10: Average (over all indices and its 1000 repetitions each) Spearman's ρ_S rank correlation between the volatility and the SQP ratios computed on a 1-year sample, using a GARCH(1,1) model. Standard deviation is computed over the 11 averages of the values obtained for each index.

AVG$\pm\sigma$	$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$	$\alpha = 99.5\%$
Gaussian	-0.61 $\pm$ 0.01	-0.59 $\pm$ 0.00	-0.56 $\pm$ 0.00	-0.55 $\pm$ 0.00
Student-t	-0.61 $\pm$ 0.00	-0.60 $\pm$ 0.00	-0.57 $\pm$ 0.01	-0.56 $\pm$ 0.01

For the GARCH(1,1) model, the rank correlation is of the same order (slightly lower) than the Pearson correlation. This is similar to the historical data (there the differences were a bit more pronounced, especially for low α). On average (i.e. when looking at the average values obtained for the 11 indices considered) we see, in this case, that the behavior remains similar with respect to α for the Pearson and Spearman correlation. Here also, tail fatness does not matter.

Overall, a simple GARCH model captures pretty well the main features we have seen in the data, with some differences discussed above. We conclude that the clustering of volatility and its return to the mean appear as an important factor of the pro-cyclicality, contrary to the tail fatness. It reinforces the effect of the risk measurement itself, discussed in the iid case.

2.4.3 Other influences

Let us eventually discuss three other possible influences on pro-cyclicality: sample sizes, data frequency and type of risk measures.

First, we look at the impact of sample sizes. Recall that we estimate the correlation between the log-ratio and the realized volatility, $\rho(\log(R_{\alpha,T}(t)), V_{k,T}(t))$, which we can generalize for different sample sizes T_1, T_2, T_3 to

$$\rho\left(\log\left(\frac{Q_{\alpha,T_1}(t+T_1)}{Q_{\alpha,T_3}(t)}\right), V_{k,T_2}(t)\right).$$

Nevertheless, the setting of our study gives us natural restrictions on the choice of these three sample sizes, which we discuss in the following. In the study, we chose $T_1 = T_s$ ($= 1$ year), the time frame prescribed by regulation in order to assess the risk of the next year. Hence, increasing or decreasing this sample size T_1 does not make sense when wanting to assess the pro-cyclicality in regulatory risk measurement. The volatility estimator $V_{k,T_2}(t)$ acts as a proxy for the market state. We have seen, in Figure 2.2, that choosing $T_2 = 1,3$ (years) work well for this. However, choosing longer sample sizes than 3 years to evaluate the volatility does not answer any longer our problem as we would not capture anymore the dynamics of the market state: Indeed, the longer the sample size T_2, the faster V_{k,T_2} will converge to the overall sample average killing the dynamics. As a third component, we can vary the sample size T_3 in the estimate $Q_{\alpha,T_3}(t)$ for the future risk. Again, increasing T_3 too much looses the dynamics, missing the main target of correctly estimating $Q_{\alpha,T_s}(t+T_s)$. Also, note that choosing different sample sizes $T_2 \neq T_3$ would deprive the interpretability of our analysis: The volatility estimator at time t should be computed over the same sample as the one for the estimate for the future risk; otherwise, it makes no sense to set these two quantities into relation with each another. Therefore, we keep $T_1 = T_s$ and choose a reasonable and identical sample size for the two other quantities under study, $T_2 = T_3$, no longer than 3 years.

The different claims can be confirmed empirically, see the online-appendix, [34], for details. E.g. when running the method for 5, 7 or 10 years we see that the longer the sample size is, the faster the dynamics and the clustering disappear. Similar patterns for 1 and 3 years can be observed in the analysis for the historical data. However, in the GARCH case, we can clearly observe the effects of the clustering: As expected, the longer the sample, the less visible is the clustering, which means a lower dependency of the ratio with volatility than in the case of $T = 1$ year. Note also that, the higher the threshold is, the longer the sample needs to be, to see this effect of decreasing correlation.

The second influence we measure is the impact of the data frequency on our statistics. We run the same analysis on weekly data and obtain results for the negative correlation of the same order, see in the online-appendix, [34], Appendix A.2. For the iid case, only the number of points matters, which implies that the longer the frequency is, the larger the time window must be to keep the same number of points for estimation.

The third influence we analyse is the definition of risk measure. When turning from VaR to another popular risk measure such as Expected Shortfall (ES), similar statements can be made. We have seen

above that the volatility clustering reinforces the negative correlation between the SQP ratio and the volatility: There is no reason why this would reverse in the case of ES given the fact that ES can be seen as a weighted average of VaR's (see Emmer, Kratz, and Tasche [53]). Results on the ES may be found in Appendix A.4.

2.5 Conclusion

In this empirical study, we explore the appropriateness of measuring risk with regulatory risk measures on historical data, to forecast the risk faced by financial institutions, thereby addressing the pro-cyclicality of risk measurements. We adopt the point of view of risk management and regulation, by looking at the pro-cyclicality of risk measurements/estimations themselves. We do not look for modelling such behavior macro-economically, but rather to draw conclusions from statistical analyses on the underlying financial time series itself. It is thus a quite different point of view compared to the pro-cyclicality analyses in economics. We think that this novel approach complements the various economic studies performed on the topic.

First, we check and quantify empirically, considering 11 stock indices, that risk measures based on VaR are in fact pro-cyclical: In times of low volatility, the risk is underestimated, while in times of crises the risk is overestimated. Although pro-cyclicality of risk models has widely been assumed by market actors, to our knowledge, it was never clearly quantified, nor shown empirically with enough evidence. The identification and quantification of pro-cyclicality is made possible by modelling the risk measure as a stochastic process and conditioning it on volatility to be able to discriminate the different market states. For this, we use the Sample Quantile Process (SQP), a 'dynamic generalization of VaR'. To measure its predictive power, we introduce a new statistics, a look-forward ratio of risk measure estimators. While we only used it for the VaR/SQP, it may be applied to any other risk measure like Expected Shortfall, without restrictions. It is a novel way of measuring the accuracy of risk estimation in line with what is done in risk management practice and complementing the violation process method.

Second, we look for factors explaining this pro-cyclical behavior. To do so, we introduce simple models to isolate effects at the source of pro-cyclicality. We find firstly that measuring risk on past historical data with a quantile, or a quantiles function, is intrinsically pro-cyclical, since even iid random variables present a negative dependence between the look-forward SQP ratio and the volatility. Moreover, the clustering of volatility modelled by a GARCH(1,1) process with Gaussian or Student innovations reinforces this negative dependence, as the volatility will tend to return to its mean in a much shorter period than a one year horizon (τ_{cor} is typically of only a few weeks). We thus relate secondly and partly this pro-cyclical effect to the volatility clustering present in financial markets.

While answering the two questions layed out in the introduction (**Q1** and **Q2**), this study paves also the way for further explorations, both empirical and theoretical, which naturally led to all further results of the thesis presented in the following chapters. Having presented a very specific setup in the empirical study, the natural question arises of how results change if one modifes one of the parameters: How do the results vary when considering another risk measure (ES), another volatility measure (std), another frequency of data (weekly) or longer time periods T (3,5 or 7 years). While we have discussed this briefly in Section 2.4.3, these are all questions dealt with, in the Appendix B. Further, having assessed empirically the correlation between the look-forward ratio and the volatility for an iid and a GARCH(1,1) model, we might ask what we can say from a theoretical perspective: Can we theoretically prove the pro-cyclicality observed empirically? Answering this question means, in a first step, to consider the joint distribution of the sample quantile with the MAD (or std) (recall **Q3** and **Q4**). For the iid model, this is tackled in Chapter 3, generalising the framework by considering also an alternative quantile estimator (apart from the sample quantile) and various estimators of measure of dispersion (other than MAD and

std). Then, in Chapter 4, corresponding results for a big class of GARCH processes, called augmented GARCH(p,q) processes, are presented.

We then connect in Chapter 5 the empirical findings of the study and the theoretical results on bivariate asymptotic distributions. We theoretically evaluate the correlation of look-forward ratio and volatility measure (when having an underlying iid or augmented GARCH(p,q) model) and thus prove the existence of pro-cyclicality, irrespective of the risk or volatility measures considered (but, of course, to a varying degree depending on these choices) - in this way coming back to **Q5** and **Q6**.

Takeaways

- Proposition of a novel way to quantify the predictive power of risk measurements: The look-forward ratio, (2.7), is the ratio between historically predicted risk and (ex post) estimated realized future risk.
- Empirical Quantification of pro-cyclicality by conditioning this look-forward ratio on a measure of volatility (representing the market state): A negative (Pearson and rank) correlation is observed betwee (the logarithm) of the look-forward ratio and the volatility estimator; see Figures 2.3, 2.4 as well as Tables 2.4, 2.6.
- Explanations for the pro-cyclicality by comparing pro-cyclicality in real data with simulation results from two models: There is an intrinsic part due to the risk measurement, as seen with iid rv's, the other part is due to clustering and return-to-the-mean of the volatility as produced by a GARCH(1,1) process.

Key questions (to be followed up in the thesis)

- How does the degree of pro-cyclicality depend on our setup choices (risk measure, volatility measure, sample size, data frequency)? Empirical analyses with some alternative choices can be found in the Appendix A (see below for a more precise list).
- Can we mathematically prove this pro-cyclicality of risk measurements ? This is treated in Chapter 5.

Related contributions in the appendix

- Selected additional tables and figures of this study can be found in Appendix A.1, further material in Appendix A.1 of the online-appendix.
- The empirical analysis when using weekly instead of daily data can be found in Appendix A.2 of the online-appendix.
- The empirical analysis when using the std as volatility measure can be found in Appendix A.3 of the online-appendix.
- Part of the empirical analysis when using the ES as risk measure can be found in Appendix A.4, the full study in Appendix A.4 of the online-appendix.
- Simulated results with another model (LM-ARCH) can be found in Appendix A.5 of the online-appendix.
- An overview how the empirically measured correlation changes with the use of longer sample sizes can be found in Appendix A.6 of the online-appendix.
- A discussion about the different packages used for fitting a GARCH process with the statistical software R - in comparison with the method of [130] - can be found partially in Appendix E. The full study is in Appendix E of the online-appendix.

Chapter 3

Estimators in the IID case: Asymptotic theory

This chapter formed the basis for the joint work [31] and its previous version [32].

3.1 Introduction

In the previous chapter, we empirically observed the negative correlation between the look-forward ratio (a ratio of sample quantiles) with the sample MAD: On log-returns from 11 stock indices, recall Table 2.4, as well as when using an iid or GARCH(1,1) model; see Tables 2.7 and 2.9.

Consequently, we want to answer the underlying question if we can confirm these empirical observations theoretically. Thus, we focus in a first step on the joint distributions of quantile estimators with measure of dispersion estimators (question **Q3**). Indeed, we establish such joint asymptotics in bigger generality than only looking at the sample quantile and sample MAD. While we consider in this chapter iid samples, this will be extended to dependent ones in the next chapter.

As quantile estimators we consider, apart from the (non-parametric) sample quantile, also the parametric location-scale quantile estimator. The interest in considering two quantile estimators lies in the fact that, although being both consistent estimators, they have different speeds of convergence. As a result, this impacts the joint asymptotic dependence with the measure of dispersion estimators. By measures of dispersion, we mean well-known quantities as the variance or standard deviation, but also less frequently used ones as, for instance, the mean absolute deviation around the mean (MAD) which we already used in Chapter 2 or the median absolute deviation around the median (MedianAD). The latter two have the advantage of relaxing the asymptotic constraints coming with the use of the sample variance (such as the existence of the fourth moment of the underlying distribution). For a more general and historical overview of measures of dispersion, we refer e.g. to [45].

For a recap of the literature on joint asymptotics of sample quantiles and measures of *location*, as well as the existing contributions to the asymptotics of quantile estimators and measure of dispersion estimators, we refer back to Section 1.1. We see these bivariate asymptotic results on their own as a complement to the existing statistical literature.

But, clearly, for us these asymptotics will always be naturally linked to the question of risk measure estimation and pro-cyclicality: The sample quantile can be seen as a Value-at-Risk estimator and the functional framework allows us to extend the results to some Expected Shortfall estimators - we will come back to this and the pro-cyclicality issue from a theoretical point of view in Chapter 5.

The structure of this chapter is as follows. We present in Sections 3.2 to 3.4 the main theoretical results about the asymptotic joint distribution and dependence between functions of quantile estimators and functions of measure of dispersion estimators. First, the asymptotics of the sample quantile with the r-th absolute central sample moment or sample MedianAD, respectively in Sections 3.2 and 3.3. Then, the analogous results when now using the parametric location scale quantile estimator instead of the

sample quantile, see Sections 3.4.1 and 3.4.2. In Section 3.5 we discuss the conditions on the underlying distribution under which the theorems hold and address different points regarding the results presented.

In Section 3.6, we focus on two different applications. First, we examine the different behavior of the asymptotics depending on the choice of the quantile estimator and the measure of dispersion. Therefore, we compare the correlations of the asymptotic distribution with the sample quantile versus the parametric location-scale quantile estimator (for each corresponding measure of dispersion). We discuss this in the cases of two frequently used location-scale distributions, the Gaussian and the Student-t, and when considering the variance, MAD and MedianAD as measures of dispersion. Second, we evaluate the finite sample approximation of the theoretical asymptotics: In a simulation study, we compare the sample correlation between quantile estimators and measure of dispersion estimators (each on a finite sample) to the theoretical correlation of the asymptotic distribution- considering distributions with light and heavy tails, respectively. We conclude in Section 3.7, presenting also an overview of related material which can be found in the Appendix B.

Notation

Let $(X_1, \cdots, X_n)$ be a sample of size n, with parent random variable (rv) X. We denote its ordered sample by $X_{(1)} \leq ... \leq X_{(n)}$. Whenever it exists, we introduce the standardised version (with mean 0 and variance 1) of X, namely $Y := \frac{X-\mu}{\sigma}$, and correspondingly the cdf, pdf and quantile of order p as F_Y, f_Y and $q_Y(p)$.

We consider two different types of dispersion estimators. First, the r-th absolute central sample moment, for any integer $r > 0$, $\hat{m}(X,n,r) = \frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_n|^r$. It includes as special cases quantities already introduced in Chapter 2 (the sample MAD $\hat{m}(X,n,1) = \hat{\theta}_n$ and the sample variance $\frac{n-1}{n}\hat{m}(X,n,2) = \hat{\sigma}_n^2$). Note that we chose $\hat{\theta}_n$ with a factor of $\frac{1}{n}$ instead of $\frac{1}{n-1}$ to be in line with the literature (see e.g. [70],[104],[116]), and since it does not matter asymptotically. Second, the sample median absolute deviation around the sample median (MedianAD) $\hat{\xi}_n = \frac{1}{2}(W_{(\lfloor\frac{n+1}{2}\rfloor)} + W_{(\lfloor\frac{n+2}{2}\rfloor)})$ where $W_j = |X_j - \hat{\nu}_n|, j = 1, ..., n$, and $\hat{\nu}_n = \frac{1}{2}(X_{(\lfloor\frac{n+1}{2}\rfloor)} + X_{(\lfloor\frac{n+2}{2}\rfloor)})$ (sample median of the original sample).

Then, we consider two quantile estimators. Those estimators are defined as follows, for any order $p \in (0,1)$,

$$q_n(p) = X_{(\lceil np\rceil)}, \tag{3.1}$$

$$q_{n,\hat{\mu},\hat{\sigma}}(p) = \hat{\mu}_n + \hat{\sigma}_n q_Y(p) \text{ (and } q_{n,\hat{\sigma}}(p) = \mu + \hat{\sigma}_n q_Y(p), \text{ respectively)}, \tag{3.2}$$

where $\hat{\mu}_n$ and $\hat{\sigma}_n$ (by abuse of notation) are any estimators of the mean μ and standard deviation σ . We choose them to be the sample mean $\bar{X}_n$ and the square root of the sample variance $\sqrt{\hat{\sigma}_n^2}$, respectively.

Recall that the location-scale family of distributions $\mathcal{F}$ is the proper subset of all distributions such that

$$\text{if } F \in \mathcal{F}, \text{ then for any } 0 < a < \infty, b \in \mathbb{R},\ G(x) := F(ax+b) \in \mathcal{F}. \tag{3.3}$$

To have a unified notation, we introduce the quantile estimator $\hat{q}_n$ that may represent either the sample quantile q_n, or the location-scale quantile estimator $q_{n,\hat{\mu},\hat{\sigma}}$ (or $q_{n,\hat{\sigma}}$ for μ known) whenever F_X belongs to the location-scale family of distributions.

Note that we refer to 'historical estimation' when estimating the quantile with the sample quantile, as it is evaluated on the historical data sample. We introduce on purpose two different ways of estimating the quantile as it has some impact on the covariance and correlation of its asymptotic distribution with the corresponding measure of dispersion (e.g. sample variance, sample MAD or sample MedianAD), thus in practice too. Although both quantile estimators converge to the same quantity, the theoretical quantile, they do not have the same rate of convergence. Using the location-scale quantile estimator, we obtain a better rate of convergence than with the historical estimation, as expected. Hence the interest of investigating this second way of estimation, is to see how this impacts the dependence structure.

In addition, to be consistent in the notation with related results in the literature, we generalise a notation used in [20] and [62]: Assuming that η is a continuous real-valued function and that the rv $\eta(X)$ has finite moments up to order l, we set, for $1 \leq k \leq l$ and $p \in (0,1)$,

$$\tau_k(\eta(X), p) = \mathrm{Cov}\left(\eta^k(X), \mathbb{1}_{(X>q_X(p))}\right) \tag{3.4}$$

$$= p(1-p)\left(\mathbb{E}[\eta^k(X)|X > q_X(p)] - \mathbb{E}[\eta^k(X)|X < q_X(p)]\right), \tag{3.5}$$

where the expression (3.5) points out that this quantity involves the truncated moments of both tails. When η is the identity function, we abbreviate $\tau_k(X,p)$ as $\tau_k(p)$.

Conditions used Before stating the main results, let us present the different conditions the underlying random variable X may need to fulfil. Depending on the choice of quantile estimator and measure of dispersion estimator, we impose three different types of conditions: The existence of a finite $2k$-th moment for any $k > 0$, the continuity of F_X (or its l-th derivative), the l-fold differentiability of F_X (at a given point or neighbourhood) for any integer $l > 0$, and the positivity of the density (at a given point or neighbourhood). Those conditions are named as:

(M_k) $\quad \mathbb{E}[X^{2k}] < \infty$,

(C_0) $\quad F_X$ is continuous,

(C_l') $\quad F_X$ is l-times differentiable,

(C_l) $\quad$ the l-th derivative of F_X is continuous,

(P) $\quad f_X$ is positive.

Note that a standard condition often stated in the literature is F_X to be absolutely continuous and strictly monotonically increasing. Clearly, this latter requirement is more general than our conditions (C_0) and (P) almost everywhere.

Finally, to have results as general as possible (in view of statistical applications), we consider functions h_1, h_2 of the estimators that we assume to be continuous real-valued functions with existing derivatives denoted by h_1' and h_2' respectively. Note that in fact, to apply the Delta method, it suffices for the derivatives to exist only at the point where they are evaluated at. We will omit recalling it in the conditions of the results. Now we are ready to present the different bivariate asymptotics in the upcoming sections.

3.2 Sample quantile and r-th absolute centred sample moment

Let us start with the joint bivariate asymptotics of the sample quantile and the r-th absolute central sample moment.

Theorem 3.1 *Consider an iid sample with parent rv X having existing (unknown) mean μ and variance σ^2. Assume conditions (C_1'), (P) at $q_X(p)$ each, (M_r) for the correponding integer r, as well as (C_0) at μ for $r = 1$. Then the joint behaviour of the functions h_1 of the sample quantile $q_n(p)$ (defined in (3.1)), for $p \in (0,1)$, and h_2 of the r-th sample absolute central moment $\hat{m}(X,n,r)$ (see List of Notations), is asymptotically normal:*

$$\sqrt{n}\begin{pmatrix} h_1(q_n(p)) - h_1(q_X(p)) \\ h_2(\hat{m}(X,n,r)) - h_2(m(X,r)) \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \Sigma^{(r)}),$$

where the covariance matrix $\Sigma^{(r)} = (\Sigma_{ij}^{(r)}, 1 \le i,j \le 2)$ *of the asymptotic distribution satisfies*

$$\Sigma_{11}^{(r)} = \frac{p(1-p)}{f_X^2(q_X(p))}\left(h_1'(q_X(p))\right)^2; \quad \Sigma_{22}^{(r)} = \left(h_2'(m(X,r))\right)^2 \operatorname{Var}\left(|X-\mu|^r - r(X-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\right); \tag{3.6}$$

$$\Sigma_{12}^{(r)} = \Sigma_{21}^{(r)} = h_1'(q_X(p))\,h_2'(m(X,r)) \times \frac{\tau_r(|X-\mu|,p) - r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\tau_1(p)}{f_X(q_X(p))}, \tag{3.7}$$

$m(X,r)$ *being defined in the List of Notations and* τ_r *in* (3.4).

The correlation of the asymptotic distribution between the functional h_1 *of the sample quantile and the functional* h_2 *of the r-th absolute sample moment is - up to its sign* $a_\pm = \operatorname{sgn}(h_1'(q_X(p)) \times h_2'(m(X,r)))$ *- the same whatever the choice of* h_1, h_2*:*

$$\frac{\Sigma_{12}^{(r)}}{\sqrt{\Sigma_{11}^{(r)}\Sigma_{22}^{(r)}}} = a_\pm \times \frac{\tau_r(|X-\mu|,p) - r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\tau_1(p)}{\sqrt{p(1-p)\operatorname{Var}\left(|X-\mu|^r - r(X-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\right)}}. \tag{3.8}$$

3.2.1 Auxiliary results

To prove Theorem 3.1 we need some auxiliary results first. Namely, we need to find the asymptotics of $\hat{m}(X,n,r) = \frac{1}{n}\sum_{i=1}^n |X_i - \bar{X}_n|^r$ for any integer $r \ge 1$. As such a result might be of interest in its own right, we give it separately in Proposition 3.3. Its proof is based on the following Lemma, which is an adaption from Lemma 2.1 in [116].

Lemma 3.2 *Consider an iid sample with parent rv* X *with finite second moment. Then, for any integer* $v \ge 1$*, given that, additionally, the* v*-th moment of* X *exists, letting* $n \to \infty$*, it holds that*

$$\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^v\left(|X_i-\bar{X}_n| - |X_i-\mu|\right) = (\bar{X}_n-\mu) \times \mathbb{E}\left[(X-\mu)^v \operatorname{sgn}(\mu - X)\right] + o_P(1/\sqrt{n}). \tag{3.9}$$

Proof The proof follows the lines of the proof of Lemma 2.1 in [116]. Therein the case $v = 0$ was treated and almost sure convergence concluded. Here we consider $v \ge 1$ and only demand convergence to 0 in probability. For consistency, we also use the notation of [116]. Namely, we need the rv's $A_n := \min(\bar{X}_n, \mu), B_n := \max(\bar{X}_n, \mu)$ as well as a partition of $\{1,...,n\} = \mathcal{K}_n \cup \mathcal{L}_n$ with

$$\begin{aligned}\mathcal{K}_n &:= \{i = 1,...,n : A_n < X_i < B_n\},\\ \mathcal{L}_n &:= \{i = 1,...,n\}\backslash\mathcal{K}_n.\end{aligned}$$

As, it holds that $(x-\mu)^v\left(|x-\bar{X}_n| - |x-\mu|\right) = (x-\mu)^v(\bar{X}_n-\mu)\operatorname{sgn}(\mu-x)$, for any $x \in \mathbb{R}\backslash(A_n, B_n)$, we rewrite the left hand side of (3.9) (ignoring the factor $1/n$ for the moment) as

$$\begin{aligned}\sum_{i=1}^n (X_i-\mu)^v\left(|X_i-\bar{X}_n| - |X_i-\mu|\right) &= (\bar{X}_n-\mu)\sum_{i\in\mathcal{L}_n}(X_i-\mu)^v\operatorname{sgn}(\mu-X_i)\\ &\quad + \sum_{i\in\mathcal{K}_n}(X_i-\mu)^v\left(|X_i-\bar{X}_n| - |X_i-\mu|\right)\\ &= (\bar{X}_n-\mu)\sum_{i=1}^n (X_i-\mu)^v\operatorname{sgn}(\mu-X_i) + \tilde{R}_n,\end{aligned} \tag{3.10}$$

$$\text{where } \tilde{R}_n := \sum_{i\in\mathcal{K}_n} (X_i-\mu)^v \left(|X_i-\bar{X}_n| - |X_i-\mu|\right) - (\bar{X}_n-\mu)\sum_{i\in\mathcal{K}_n}(X_i-\mu)^v \operatorname{sgn}(\mu-X_i)$$
$$= \sum_{i\in\mathcal{K}_n}(X_i-\mu)^v\Big(|X_i-\bar{X}_n| - |X_i-\mu| - (\bar{X}_n-\mu)\operatorname{sgn}(\mu-X_i)\Big).$$

Note that, by construction, for $i\in\mathcal{K}_n$, it holds that $|X_i-\mu|\le|\bar{X}_n-\mu|$. Further, for any $x\in\mathbb{R}$, we have that $|\operatorname{sgn}(x)|\le 1$ and $||x-\bar{X}_n|-|x-\mu||\le|\bar{X}_n-\mu|$. Thus, we can bound $\tilde{R}_n$ as

$$|\tilde{R}_n| \le \operatorname{card}(\mathcal{K}_n)\times|\bar{X}_n-\mu|^v\left(|\bar{X}_n-\mu|+|\bar{X}_n-\mu|\right) = 2\operatorname{card}(\mathcal{K}_n)|\bar{X}_n-\mu|^{v+1},$$

where $\operatorname{card}(\mathcal{K}_n)$ denotes the cardinality of the set $\mathcal{K}_n$. It is shown in [116] that $\frac{\operatorname{card}(\mathcal{K}_n)}{n} \xrightarrow[n\to\infty]{a.s.} 0$. Since we know that $\sqrt{n}|\bar{X}_n-\mu|^{v+1} \xrightarrow[n\to\infty]{P} 0$ for any integer $v\ge 1$, we have $\sqrt{n}\frac{1}{n}|\tilde{R}_n| \xrightarrow{P} 0$, i.e. $\frac{\tilde{R}_n}{n} = o_P(1/\sqrt{n})$.

Going back to the first term in (3.10), and multiplying it by $\frac{1}{n}$, we obtain by the strong law of large numbers

$$\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^v\operatorname{sgn}(\mu-X_i) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[(X-\mu)^v\operatorname{sgn}(\mu-X)].$$

Thus, the result in (3.9) follows. □

Now we are ready to state the asymptotic relation between the r-th absolute central sample moment with known and unknown mean, respectively.

Proposition 3.3 *Consider an iid sample with parent rv X. Then, for any integer $r\ge 1$, given that the r-th moment of X exists, and (C_0) at μ for $r=1$, it holds that, as $n\to\infty$*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n|X_i-\bar{X}_n|^r\right)$$
$$= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n|X_i-\mu|^r\right) - r\sqrt{n}(\bar{X}_n-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r] + o_P(1). \quad (3.11)$$

Proof We distinguish three different cases for r: Even integers r, $r=1$ and odd integers $r>1$.

Even integers r - In such a case $|X_i-\mu|^r = (X_i-\mu)^r$ such that we can simply consider the known asymptotics for central moments. E.g. in [90], Example 5.2.7, they conclude that for any even integer $r>1$, as $n\to\infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X}_n)^r - \mathbb{E}[(X-\mu)^r]\right)$$
$$= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^r - \mathbb{E}[(X-\mu)^r]\right) - r\sqrt{n}(\bar{X}_n-\mu)\left(\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^{r-1}\right) + o_P(1). \quad (3.12)$$

Case $r = 1$ - This case is known too. E.g. we can deduce from Lemma 2.1 in [116], that if $\mathbb{E}[X] < \infty$ and F_X is continuous at μ, it holds that, as $n \to \infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_n| - \mathbb{E}[|X-\mu|]\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i-\mu| - \mathbb{E}[|X-\mu|]\right) + \sqrt{n}(2F_X(\mu)-1)(\bar{X}_n - \mu) + o_P(1). \quad (3.13)$$

Note that the continuity of F_X at μ, i.e. condition (C_0) at μ, is needed to have the asymptotic normality of $\hat{m}(X,n,r)$, as explained in [116]. This is specific to the case $r = 1$.

Odd integer $r > 1$ - This is the only case requiring some work. Set $r = 2u+1$, for any integer $u \geq 1$.

By binomial expansion we get $\sum_{i=1}^{n}|X_i - \bar{X}_n|^{2u+1} =$

$$\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X}_n)^{2u}|X_i - \bar{X}_n| = \sum_{k=0}^{2u}(-1)^k\binom{2u}{k}(\bar{X}_n-\mu)^k\left(\sum_{i=1}^{n}(X_i-\mu)^{2u-k}|X_i-\bar{X}_n|\right),$$

which we write out for $k = 0, 1$ (and keeping $k \geq 2$ as sum) as:

$$= \sum_{i=1}^{n}|X_i-\bar{X}_n|(X_i-\mu)^{2u} - 2u(\bar{X}_n-\mu)\sum_{i=1}^{n}|X_i-\bar{X}_n|(X_i-\mu)^{2u-1} + \sum_{k=2}^{2u}(-1)^k\binom{2u}{k}(\bar{X}_n-\mu)^k\left(\sum_{i=1}^{n}(X_i-\mu)^{2u-k}|X_i-\bar{X}_n|\right). \quad (3.14)$$

Recall that, for the asymptotics, we need to multiply (3.14) by $\frac{1}{\sqrt{n}}$. Following the analogous argumentation as in [90] for even integers, we conclude that all terms in (3.14) apart from the first two ($k = 0, 1$) vanish as $\sqrt{n}(\bar{X}_n - \mu)^v \xrightarrow{P} 0$ for $v \geq 2$. Hence, we are left with the analysis of the first two terms of (3.14).

For the first term of (3.14), we multiply it by $1/n$ and apply Lemma 3.2 with $v = 2u$ such that we have, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^{2u}|X_i-\bar{X}_n| = \frac{1}{n}\sum_{i=1}^{n}\left(|X_i-\mu|^{2u+1} + (X_i-\mu)^{2u}\left(|X_i-\bar{X}_n| - |X_i-\mu|\right)\right) = \frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^{2u+1} + (\bar{X}_n-\mu)\,\mathbb{E}\left[(X-\mu)^{2u}\operatorname{sgn}(\mu - X)\right] + o_P(1/\sqrt{n}). \quad (3.15)$$

Analgously, for the second term of (3.14), by applying Lemma 3.2 with $v = 2u-1$ (as well as the weak law of large numbers), we get

$$\begin{aligned}
&\frac{1}{n}(-2u)(\bar{X}_n-\mu)\sum_{i=1}^{n}(X_i-\mu)^{2u-1}|X_i-\bar{X}_n| \\
&\quad=\frac{1}{n}(-2u)(\bar{X}_n-\mu)\sum_{i=1}^{n}\Big((X_i-\mu)^{2u}\operatorname{sgn}(X_i-\mu)+(X_i-\mu)^{2u-1}\left(|X_i-\bar{X}_n|-|X_i-\mu|\right)\Big) \\
&\quad=-2u(\bar{X}_n-\mu)\left(\mathbb{E}[(X-\mu)^{2u}\operatorname{sgn}(X-\mu)]+o_P(1)\right) \\
&\qquad-2u(\bar{X}_n-\mu)\left((\bar{X}_n-\mu)\,\mathbb{E}\left[(X-\mu)^{2u-1}\operatorname{sgn}(\mu-X)\right]+o_P(1/\sqrt{n})\right). \qquad (3.16)
\end{aligned}$$

Putting (3.15) and (3.16) together, we get from (3.14), for $n\to\infty$, that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}|X_i-\bar{X}_n|^{2u+1}=&\frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^{2u+1}-(2u+1)(\bar{X}_n-\mu)\Big(\mathbb{E}[(X-\mu)^{2u}\operatorname{sgn}(X-\mu)]+o_P(1)\Big) \\
&-2u(\bar{X}_n-\mu)\left((\bar{X}_n-\mu)\,\mathbb{E}\left[(X-\mu)^{2u-1}\operatorname{sgn}(\mu-X)\right]+o_P(1/\sqrt{n})\right) \\
&+o_P(1/\sqrt{n}).
\end{aligned}$$

As $\sqrt{n}(\bar{X}_n-\mu)^2\xrightarrow{P}0$, we can conclude, for $n\to\infty$,

$$\begin{aligned}
&\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}|X_i-\bar{X}_n|^{2u+1} \\
&\qquad=\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^{2u+1}-(2u+1)\sqrt{n}(\bar{X}_n-\mu)\,\mathbb{E}[(X-\mu)^{2u}\operatorname{sgn}(X-\mu)]+o_P(1). \qquad (3.17)
\end{aligned}$$

A statement for any integer r. To conclude, we can summarize the different cases in (3.12), (3.13) and (3.17) as follows. For any integer $r\geq 1$, it holds, as $n\to\infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i-\bar{X}_n|^{r}\right)=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^{r}\right)-r\sqrt{n}(\bar{X}_n-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^{r}]+o_P(1).$$

□

3.2.2 Proof of Theorem 3.1

The main proof, using the Bahadur representation of the sample quantile, is presented first. An alternative proof which uses the Taylor expansion instead of the Bahadur representation can be found subsequently.

Finally, note that the various results in this chapter are proved without introducing the functions h_1, h_2. We refer to Part 3 in this proof of Theorem 3.1 as an illustration of the application of the Delta method. This latter step being common to all proofs of joint asymptotic distributions, it is not repeated.

Using Bahadur's method

The proof of Theorem 3.1 consists of three parts. In the first part, we assume the mean μ to be known. We show the asymptotic joint normality of the sample quantile and the sample measure of dispersion with known mean, $\tilde{m}(X,n,r)$ (see List of Notations), for any integer $r>0$. To apply the bivariate Central Limit Theorem (CLT) in this case we need to use the Bahadur representation of the sample

quantile. While this first part is a straightforward extension of the Gaussian case proposed in [25], the second part when considering an unknown mean, involves more work. The challenge is to establish a Bahadur-like representation for the r-th absolute central sample moment $\hat{m}(X,n,r)$. This having been previously proven as auxiliary result in Proposition 3.3, we use it along with the Bahadur representation of the sample quantile to apply the bivariate CLT once again, now for $\hat{m}(X,n,r)$. Finally, Part 3 shows the application of the Delta method, which concludes the proof.

Part 1 - Known Mean -

- *Bahadur representation*

  We use the Bahadur representation for sample quantiles from an iid sample given in [68], as the needed conditions (C_1') and (P) at $q_X(p)$ each are fulfilled by assumption,

$$q_n(p) = q_X(p) + \frac{1 - F_n(q_X(p)) - (1-p)}{f_X(q_X(p))} + R_{n,p}, \text{ where } R_{n,p} = o_P(n^{-1/2}). \tag{3.18}$$

  Note that the original representation goes back to [12] where he showed that, as $n \to \infty$, $R_{n,p} = O(n^{-3/4}\log(n))$ almost surely. But the conditions in the version of [68] are milder and the convergence in probability to 0 of $R_{n,p}$ sufficient for our case: With this Bahadur representation, we are able to use the bivariate CLT for the sample quantile $q_n(p)$ and the sample measure of dispersion with known mean μ, $\tilde{m}(X,n,r)$.

- *Central Limit Theorem*

  Under condition (M_r), $r > 0$, we obtain

$$n^{-1/2}\sum_{i=1}^{n}\left(\begin{pmatrix}\mathbb{1}_{(X_i>q_X(p))}\\ |X_i-\mu|^r\end{pmatrix} - \begin{pmatrix}1-p\\ m(X,r)\end{pmatrix}\right) = n^{1/2}\left(\begin{pmatrix}1-F_n(q_X(p))\\ \frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^r\end{pmatrix} - \begin{pmatrix}1-p\\ m(X,r)\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}(0,\hat{\Sigma}^{(r)}), \tag{3.19}$$

  where $\hat{\Sigma}^{(r)} = \begin{pmatrix}\mathrm{Var}(\mathbb{1}_{(X>q_X(p))}) & \mathrm{Cov}(\mathbb{1}_{(X>q_X(p))}, |X-\mu|^r)\\ \mathrm{Cov}(\mathbb{1}_{(X>q_X(p))}, |X-\mu|^r) & \mathrm{Var}(|X-\mu|^r)\end{pmatrix}$.

  Then, we need to pre-multiply (i.e. from the left side) equation (3.19) by $\begin{bmatrix}1/(f_X(q_X(p))) & 0\\ 0 & 1\end{bmatrix}$ to use the Bahadur representation (3.18) of the sample quantile. One gets (as in [25], just with a different notation),

$$n^{1/2}\begin{pmatrix}\frac{1-F_n(q_X(p))-(1-p)}{f_X(q_X(p))}\\ \frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^r - m(X,r)\end{pmatrix} = n^{1/2}\begin{pmatrix}q_n(p) - q_X(p) - R_{n,p}\\ \tilde{m}(X,n,r) - m(X,r)\end{pmatrix} \xrightarrow{d} \mathcal{N}(0,\tilde{\Sigma}^{(r)}) \tag{3.20}$$

  where now

$$\tilde{\Sigma}^{(r)} = \begin{pmatrix}\frac{\mathrm{Var}(\mathbb{1}_{(X>q_X(p))})}{f_X^2(q_X(p))} & \frac{\mathrm{Cov}(\mathbb{1}_{(X>q_X(p))}, |X-\mu|^r)}{f_X(q_X(p))}\\ \frac{\mathrm{Cov}(\mathbb{1}_{(X>q_X(p))}, |X-\mu|^r)}{f_X(q_X(p))} & \mathrm{Var}(|X-\mu|^r)\end{pmatrix}. \tag{3.21}$$

  As $R_{n,p} = o_P(n^{-1/2})$, we can ignore it in an asymptotic analysis, as it follows from Slutsky's theorem that the distribution does not change. Now let us compute the covariance matrix $\tilde{\Sigma}^{(r)}$ of the asymptotic distribution. As we assume (C_1') and (P) at $q_X(p)$, we have $F_X(q_X(p)) = p$, hence

$$\mathbb{E}[\mathbb{1}_{(X>q_X(p))}] = 1-p \text{ and } \mathrm{Var}(\mathbb{1}_{(X>q_X(p))}) = p(1-p).$$

  Therefore, using $\tau_k(\eta(X),p)$, defined in (3.4), we can write, by definition,

$$\mathrm{Cov}(\mathbb{1}_{(X>q_X(p))}, |X-\mu|^r) = \tau_r(|X-\mu|,p), \tag{3.22}$$

which concludes to the asymptotic joint distribution of the sample quantile and the sample measure of dispersion with known μ for any integer $r > 0$.

Part 2 - Unknown Mean -

To show the bivariate CLT in the case of an unknown mean we need to replace in Part 1 $\tilde{m}(X,n,r)$ by $\hat{m}(X,n,r)$. From Proposition 3.3 we know that, for $n \to \infty$,

$$\hat{m}(X,n,r) = \tilde{m}(X,n,r) - r(\bar{X}_n - \mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r] + o_P(1/\sqrt{n})$$

Also, from Part 1 the covariance of the asymptotic distribution between the sample quantile and the r-th absolute sample moment with known mean μ equals

$$\tilde{\Sigma}_{12}^{(r)} = \frac{\tau_r(|X-\mu|,p)}{f_X(q_X(p))}.$$

Further, in [62], it is shown that the covariance (scaled by $\sqrt{n}$) of the asymptotic distribution between the sample quantile and the sample mean equals to $\frac{\tau_1(p)}{f_X(q_X(p))}$. Hence, recalling that a rest of $o_P(1)$, by Slutsky's theorem, will not change the asymptotic distribution, we can write for the covariance of the asymptotic distribution between the sample quantile and the r-th absolute central sample moment with unknown mean:

$$\Sigma_{12}^{(r)} = \Sigma_{21}^{(r)} = \tilde{\Sigma}_{12}^{(r)} - r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\frac{\tau_1(p)}{f_X(q_X(p))}.$$

In comparison to Part 1, $\Sigma_{11}^{(r)}$ remains unchanged: $\Sigma_{11}^{(r)} = \tilde{\Sigma}_{11}^{(r)} = \frac{p(1-p)}{f_X^2(q_X(p))}$. Further, the variance of $\hat{m}(X,n,r)$ follows directly from the representation in Proposition 3.3:

$$\Sigma_{22}^{(r)} = \operatorname{Var}(|X-\mu|^r - r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r](X-\mu)).$$

Thus, we have completely specified the covariance matrix $\Sigma^{(r)}$.

Part 3 - Delta Method -

Getting the expressions involving the functions h_1, h_2 is an application of the bivariate Delta method. For a given function $h(x,y) = \begin{pmatrix} h_1(x) \\ h_2(y) \end{pmatrix}$, such that the Jacobian of $h(x,y)$ exists at the point $x = q_X(p), y = m(X,r)$, namely

$$J\left(h\left(q_X(p), m(X,r)\right)\right) := \begin{bmatrix} \frac{\partial h_1(x)}{\partial x} & \frac{\partial h_1(x)}{\partial y} \\ \frac{\partial h_2(y)}{\partial x} & \frac{\partial h_2(y)}{\partial y} \end{bmatrix}_{x=q_X(p),y=m(X,r)} = \begin{bmatrix} h_1'(q_X(p)) & 0 \\ 0 & h_2'(m(X,r)) \end{bmatrix},$$

we can apply the Delta method. So, using the asymptotics proved above in Part 2 (covariance matrix $\Sigma^{(r)}$), it follows that

$$\sqrt{n}\,h\left(q_n(p), \hat{m}(X,n,r)\right) - h\left(q_X(p), m(X,r)\right) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, J\left(h(q_X(p), m(X,r))\,\Sigma^{(r)} J\left(h(q_X(p), m(X,r)\right)^T\right).$$

It means to replace $\Sigma^{(r)}$ by $J\left(h(q_X(p), m(X,r))\,\Sigma^{(r)} J\left(h(q_X(p), m(X,r)\right)^T$. That is why the factors $h_1'(q_X(p)), h_2'(m(X,r))$ appear in the covariance terms of (3.6) and (3.7).

Using Taylor's method

As it may have interest on its own, we offer an additional proof for Theorem 3.1, which is based on a Taylor expansion and extends the ideas of [62] who used this technique to prove the asymptotic normality between the sample quantile and the sample mean. The proof, as in the main proof of Theorem 3.1, consists of three parts. In the first and principal part, we show the Taylor expansion and asymptotic normality in the case of estimating the r-th absolute central sample moment with known mean μ for any integer valued r (in analogy to the first part in the proof of Theorem 3.1 with Bahadur's method). The second part consists of extending the previous result to the case where we estimate the r-th absolute central sample moment in the case of an unknown mean μ, and the third in the application of the Delta method. As they are identical to Part 2 and 3 in the proof of Theorem 3.1 when using Bahadur's method, we do not repeat them here and focus on Part 1:

Proof We start establishing the asymptotic normality in the case of estimating the r-th absolute central sample moment by $\tilde{m}(X,n,r) = \frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^r$. This is done in three steps. The first step is to provide a representation such that our quantities of interest, the sample quantile and the measure of dispersion estimator, are functions of the uniform order statistics. Then, we use the Taylor expansion to prove the asymptotic normality of each of the estimators. This is the step which requires more extensive differentiability and continuity conditions than the proof of Theorem 3.1. Finally, in a third step, we compute the covariance (and then the correlation) of the asymptotic distribution between the r-th absolute central sample moment and the sample quantile.

Step 1: Functions of the uniform order statistics

Recall that for a standard exponentially distributed iid sample $(Z_1,...,Z_{n+1})$, defining $U_j := \frac{\sum_{i=1}^{j} Z_i}{\sum_{k=1}^{n+1} Z_k}$, for $j = 1,...,n$, we have that $(U_1,...,U_n)$ has the same distribution as the order statistics from a sample of size n from a standard uniform distribution (see e.g. [108]). This allows us to express the sample quantile $q_n(p)$ and the r-th absolute central sample moment $\tilde{m}(X,n,r)$ as follows:

$$q_n(p) = X_{(\lceil np \rceil)} = q_X(U_{\lceil np \rceil}), \tag{3.23}$$

$$\tilde{m}(X,n,r) = \frac{1}{n}\sum_{i=1}^{n}|q_X(U_i)-\mu|^r. \tag{3.24}$$

Step 2: Taylor expansions

Using this, we can proceed with the Taylor expansion. Only some work is needed for $\tilde{m}(X,n,r)$ as we can use the result of [62] for the sample quantile: By expanding the sample quantile $q_n(p) = q_X(U_{\lceil np \rceil})$ around p, Ferguson gets $q_n(p) = q_X(p) + q'_X(p)(U_{\lceil np \rceil} - p) + o_P(1)$. And following equations (11), (13) and (15) in [62], one concludes to the asymptotic normality of the sample quantile: For $n \to \infty$, it holds:

$$\sqrt{n}(q_n - q_X(p)) = \sqrt{n}q'_X(p)\left(\frac{\sum_{j=1}^{\lceil np \rceil} Z_j}{\sum_{k=1}^{n+1} Z_k} - p\right) + o_P(1) \xrightarrow[n\to\infty]{d} q'_X(p)B(p),$$

where $B(t) := W(t) - tW(1)$ is the Brownian bridge, W denoting the standard Wiener process. Then, expanding each $q_X(U_i)$ in (3.24) around $i/(n+1)$, $i = 1,...,n$, we obtain, as $n \to \infty$:

$$\tilde{m}(X,n,r) = \frac{1}{n}\sum_{i=1}^{n}\left(\left|q_X\left(\frac{i}{n+1}\right)-\mu\right|^r + r\left|q_X\left(\frac{i}{n+1}\right)-\mu\right|^{r-1} \times q'_X\left(\frac{i}{n+1}\right)\operatorname{sgn}\left(q_X\left(\frac{i}{n+1}\right)-\mu\right)\left(U_i - \frac{i}{n+1}\right)\right) + o_P(1/\sqrt{n}),$$

as again, higher order terms vanish asymptotically (in probability). Further, in analogy to μ_n in [62], we define $\mu_n(X,r) := \frac{1}{n}\sum_{i=1}^{n}\left|q_X\left(\frac{i}{n+1}\right)-\mu\right|^r$. We can interpret it as the right Riemann sum:

$$\mu_n(X,r) = \frac{n+1}{n}\times\frac{1}{n+1}\sum_{i=1}^{n}\left|q_X\left(\frac{i}{n+1}\right)-\mu\right|^r \underset{n\to\infty}{\to} \int_0^1 |q_X(t)-\mu|^r dt.$$

Using the transformation $t = F_X(x)$, we obtain:

$$\int_0^1 |q_X(t)-\mu|^r dt = \int_{-\infty}^{+\infty} |q_X(F_X(x))-\mu|^r dF_X(x) = \int_{-\infty}^{+\infty} |x-\mu|^r dF_X(x) = m(X,r),$$

from which we conclude that $\lim_{n\to\infty}\mu_n(X,r) = m(X,r)$.

Also, by the order of the error term of the right Riemann sum approximation, $O(n^{-1})$, we know that $\lim_{n\to\infty}\sqrt{n}\left(\mu_n(X,r)-m(X,r)\right)=0$. Hence, $m(X,r)$ can be replaced by $\mu_n(X,r)$, even in asymptotics when multiplied by $\sqrt{n}$, and we can write, as $n\to\infty$,

$$\begin{aligned}&\sqrt{n}\left(\tilde{m}(X,n,r)-\mu_n(X,r)\right)\\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} r\left|q_X\left(\frac{i}{n+1}\right)-\mu\right|^{r-1} q'_X\left(\frac{i}{n+1}\right)\operatorname{sgn}\left(q_X\left(\frac{i}{n+1}\right)-\mu\right)\left(U_i-\frac{i}{n+1}\right)\right)+o_P(1).\end{aligned}$$

By using the asymptotics calculated in [62] (see eq. (12),(14) and (16) therein), we can then conclude to the following convergence in distribution

$$\sqrt{n}(\tilde{m}(X,n,r)-\mu_n(X,r)) \underset{n\to\infty}{\overset{d}{\to}} \int_0^1 r\,|q_X(t)-\mu|^{r-1}q'_X(t)\,\operatorname{sgn}\left(q_X(t)-\mu\right)B(t)dt.$$

Hence the asymptotic normality of the measure of dispersion (in the case of known mean μ).

Finally, we can conclude to the normal joint distribution by using the Cramér-Wold device (the increments of the Brownian motion being independent and normally distributed).

Step 3: Covariance and Correlation of the asymptotic distribution

We can compute the covariance of the asymptotic distribution, using the first two moments of the Brownian bridge, as:

$$\begin{aligned}\tilde{\Sigma}_{12}^{(r)} &= \operatorname{Cov}\left(q'_X(p)B(p), \int_0^1 r|q_X(t)-\mu|^{r-1}q'_X(t)\operatorname{sgn}(q_X(t)-\mu)B(t)dt\right)\\ &= q'_X(p)\int_0^1 r|q_X(t)-\mu|^{r-1}q'_X(t)\operatorname{sgn}(q_X(t)-\mu)\,\mathbb{E}[B(p)B(t)]dt\\ &= q'_X(p)\int_0^1 r|q_X(t)-\mu|^{r-1}q'_X(t)\operatorname{sgn}(q_X(t)-\mu)q'_X(t)(\min(p,t)-pt)dt.\end{aligned}$$

Hence, we are left with computing the integral:

$$\begin{aligned}
&\int_0^1 r|q_X(t)-\mu|^{r-1}q'_X(t)\,\mathrm{sgn}(q_X(t)-\mu)q'_X(t)(\min(p,t)-pt)dt\\
&=\int_0^p r|q_X(t)-\mu|^{r-1}q'_X(t)\,\mathrm{sgn}(q_X(t)-\mu)q'_X(t)t(1-p)dt+\int_p^1 r|q_X(t)-\mu|^{r-1}q'_X(t)\,\mathrm{sgn}(q_X(t)-\mu)q'_X(t)p(1-t)dt\\
&=(1-p)\left((|q_X(t)-\mu|^r t)|_0^p-\int_0^p|q_X(t)-\mu|^r dt\right)+p\left((|q_X(t)-\mu|^r(1-t))|_p^1+\int_p^1|q_X(t)-\mu|^r dt\right)\\
&=p\int_p^1|q_X(t)-\mu|^r dt-(1-p)\int_0^p|q_X(t)-\mu|^r dt=p\int_{q_X(p)}^{\infty}|x-\mu|^r dF_X(x)-(1-p)\int_{-\infty}^{q_X(p)}|x-\mu|^r dF_X(x).
\end{aligned}$$

For the second equality, we use partial integration for help functions u, v: Note that (by triple chain rule) for $u' = r|q_X(t)-\mu|^r q'_X(t)\,\mathrm{sgn}(q_X(t)-\mu)q'_X(t)$, we have $u = |q_X(t)-\mu|^r$ and v being t or $1-t$ respectively. Further, we used the change of variables $t = F_X(x)$ in the last equation. Thus, we have overall, recalling the definition of τ_r in (3.5),

$$\tilde{\Sigma}_{12}^{(r)}=\tilde{\Sigma}_{21}^{(r)}=q'_X(p)\tau_r(|X-\mu|,p)=\frac{1}{f_X(q_X(p))}\tau_r(|X-\mu|,p),$$

from which we can deduce the correlation of the asymptotic distribution, namely

$$\frac{\tilde{\Sigma}_{12}^{(r)}}{\sqrt{\tilde{\Sigma}_{11}^{(r)}\tilde{\Sigma}_{22}^{(r)}}}=\frac{\tau_r(|X-\mu|,p)}{\sqrt{p(1-p)}\sqrt{\mathrm{Var}(|X-\mu|^r)}}.$$

□

3.2.3 Extensions

There are straightforward extensions to the joint bivariate asymptotics presented. Similarly, we could also apply the same reasoning to present such extensions for any other of the forthcoming results (Theorem 3.6, Propositions 3.7 and 3.8), but restrict ourselves, exemplarily, to this case.

Clearly, as e.g. in [25], we can look at the joint distribution of a vector of sample quantiles, instead of only at one sample quantile, with the r-th absolute central sample moment. Also, we could consider a more general function $h(x,y)$ as up to now our function was of the form $h(x,y)=\begin{bmatrix}h_1(x)\\h_2(y)\end{bmatrix}$. These two ideas are combined in Theorem 3.4.

For this, denote by $\mathbf{q}_X(\mathbf{p})$ the m-vector of quantiles evaluated at $p_i, i=1,...,m$, where $0<p_1<...<p_m<1$, and by $\mathbf{q}_n(\mathbf{p})$ the corresponding m-vector of sample quantiles $q_n(p_i), i=1,...,m$. Recall that we denote by z^T the transpose of a vector z.

Theorem 3.4 *Consider an iid sample with parent rv X having existing (unknown) mean μ and variance σ^2. Further, consider a function $h:\mathbb{R}^{m+1}\mapsto\mathbb{R}^{m+1}$, i.e. $h(x_1,...,x_m,y)=\begin{pmatrix}h_1(x_1,...,x_m,y)\\...\\h_{m+1}(x_1,...,x_m,y)\end{pmatrix}$ with continuous real-valued components $h_i(x_1,...,x_m,y), i=1,...,m+1$, and existing partial derivatives denoted by $\partial_i h_j, i,j\in\{1,...,m+1\}$. Assume conditions (C_1'), (P) at $q_X(p_i), i=1,...,m$ each, and (M_r) for the corresponding integer $r>0$, respectively, as well as (P) at μ for $r=1$. Then, the joint behaviour of the functional h of the sample quantile vector $\boldsymbol{q}_n(\boldsymbol{p})$ and of the r-th absolute central sample moment $\hat{m}(X,n,r)$ is asymptotically normal:*

$$\sqrt{n}\,h\begin{pmatrix}\boldsymbol{q}_n(\boldsymbol{p})\\\hat{m}(X,n,r)\end{pmatrix}-h\begin{pmatrix}\boldsymbol{q}_X(\boldsymbol{p})\\m(X,r)\end{pmatrix}\xrightarrow[n\to\infty]{d}\mathcal{N}\left(0,J\left(h\begin{pmatrix}\boldsymbol{q}_X(\boldsymbol{p})\\m(X,n,r)\end{pmatrix}\right)\Sigma^{(m,r)}J\left(h\begin{pmatrix}\boldsymbol{q}_X(\boldsymbol{p})\\m(X,n,r)\end{pmatrix}\right)^T\right),$$

where $J\left(h\begin{pmatrix}\boldsymbol{q}_X(\boldsymbol{p})\\ m(X,n,r)\end{pmatrix}\right)$ *is the Jacobian matrix of* $h(x_1,...,x_m,y)$ *evaluated at* $\begin{pmatrix}\boldsymbol{q}_X(\boldsymbol{p})\\ m(X,n,r)\end{pmatrix}$ *and the covariance matrix* $\Sigma^{(m,r)}$ *of dimension* $(m+1)\times(m+1)$ *of the asymptotic distribution can be written as*

$$\Sigma^{(m,r)} = \begin{bmatrix}\Sigma^{(m)} & \boldsymbol{s}(X,r)\\ \boldsymbol{s}(X,r)^T & \operatorname{Var}\left((|X-\mu|^r - r(X-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\right)\end{bmatrix}$$

with $\Sigma^{(m)}_{ij} = \frac{p_i(1-p_j)}{f_X(q_X(p_i))f_X(q_X(p_j))}$ *for* $i,j\in\{1,...,m\}$ *s.t.* $i\le j$ *and the i-th element of the m-vector* $\boldsymbol{s}(X,r)$ *being* $\frac{\tau_r(|X-\mu|,p_i)-r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\tau_1(p_i)}{f_X(q_X(p_i))}, i=1,...,m.$

As a corollary of the theorem, we can state how the result explicitly looks like if we go back to the one-dimensional sample quantile case, as in Theorem 3.1 but with a general function $h(x,y)$.

Corollary 3.5 *Consider an iid sample with parent rv X having existing (unknown) mean* μ*, variance* σ^2 *and a function* $h(x,y)=\begin{pmatrix}h_1(x,y)\\ h_2(x,y)\end{pmatrix}$ *with continuous real-valued components* $h_1(x,y),h_2(x,y)$ *and existing partial derivatives denoted by* $\partial_i h_j, i,j\in\{1,2\}$*. Assume conditions* $(C_1^{'}),(P)$ *at* $q_X(p)$ *each, and* (M_r) *for the corresponding integer* $r>0$*, respectively, as well as* (P) *at* μ *for* $r=1$*. Then, the joint behaviour of the functional h of the sample quantile* $q_n(p)$ *(for* $p\in(0,1)$*) and of the r-th absolute central sample moment* $\hat{m}(X,n,r)$*is asymptotically normal:*

$$\sqrt{n}\,h\begin{pmatrix}q_n(p)\\ \hat{m}(X,n,r)\end{pmatrix} - h\begin{pmatrix}\hat{q}_X(p)\\ m(X,r)\end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,\Sigma^{(h,r)}),$$

where the covariance matrix $\Sigma^{(h,r)} = (\Sigma^{(h,r)}_{ij}, 1\le i,j\le 2)$ *of the asymptotic distribution satisfies,*

$$\begin{aligned}
\Sigma^{(h,r)}_{11} &= \operatorname{Var}(q_n(p))\ (\partial_1 h_1)^2 + 2\partial_1 h_1\partial_2 h_1 \operatorname{Cov}(q_n(p),\hat{m}(X,n,r)) + \operatorname{Var}(\hat{m}(X,n,r))\ (\partial_2 h_1)^2;\\
\Sigma^{(h,r)}_{22} &= \operatorname{Var}(\hat{m}(X,n,r))\ (\partial_2 h_2)^2 + 2\partial_1 h_2\partial_2 h_2 \operatorname{Cov}(q_n(p),\hat{m}(X,n,r)) + \operatorname{Var}(q_n(p))\ (\partial_1 h_2)^2;\\
\Sigma^{(h,r)}_{12} &= \Sigma^{(h,r)}_{21}\\
&= \operatorname{Cov}(q_n(p),\hat{m}(X,n,r))\,(\partial_1 h_1\partial_2 h_2 + \partial_2 h_1\partial_1 h_2) + \operatorname{Var}(q_n(p))\,\partial_1 h_1\partial_1 h_2 + \operatorname{Var}(\hat{m}(X,n,r))\,\partial_2 h_1\partial_2 h_2,
\end{aligned}$$

denoting, by abuse of notation, $\partial_i h_j, i,j\in\{1,2\}$ *the corresponding partial derivative evaluated at* $(q_X(p),m(X,r))$*, and the variances of the asymptotic distribution of* $\sqrt{n}q_n(p)$ *as* $\operatorname{Var}(q_n(p))$*, of* $\sqrt{n}\hat{m}(X,n,r)$ *as* $\operatorname{Var}(\hat{m}(X,n,r))$ *and the covariance of their joint asymptotic distribution as* $\operatorname{Cov}(q_n(p),\hat{m}(X,n,r))$*. Those quantities equal to*

$$\begin{aligned}
\operatorname{Var}(q_n(p)) &= \frac{p(1-p)}{f_X^2(q_X(p))},\\
\operatorname{Var}(\hat{m}(X,n,r)) &= \operatorname{Var}\left(|X-\mu|^r - r(X-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\right),\ \textit{and}\\
\operatorname{Cov}(q_n(p),\hat{m}(X,n,r)) &= \frac{\tau_r(|X-\mu|,p) - r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\tau_1(p)}{f_X(q_X(p))}.
\end{aligned}$$

The correlation of the asymptotic distribution between the functional h of the measure of dispersion and the sample quantile can be deduced from the above expressions. In the specific case of having

$\partial_2 h_1 = \partial_1 h_2 = 0$, *it is identical -up to its sign - whatever the choice of* h *(under that restriction), namely*

$$\frac{\Sigma_{12}^{(h,r)}}{\sqrt{\Sigma_{11}^{(h,r)}\Sigma_{22}^{(h,r)}}} = \operatorname{sgn}(\partial_1 h_1 \partial_2 h_2) \times \frac{\tau_r(|X-\mu|,p) - r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\tau_1(p)}{\sqrt{p(1-p)\operatorname{Var}\left(|X-\mu|^r - r(X-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\right)}}.$$

3.3 Sample quantile and MedianAD

Turning to the case with the sample MedianAD, the different dependence structure appears clearly in the expressions of the covariance and correlation, see (3.26), (3.27) below, when compared to Theorem 3.1 (e.g. involving maxima - something we do not have in Theorem 3.1).

Theorem 3.6 *Consider an iid sample with parent rv* X *with (unknown) median* ν*, MedianAD* ξ *and, if existing, mean* μ *and variance* σ^2*. Assume conditions* (C_0) *in neighbourhoods of* $\nu \pm \xi$*,* (C_1') *at* $q_X(p), \nu$ *and* $\nu \pm \xi$*, and* (P) *at* $\nu, q_X(p)$ *and at least at one of* $\nu \pm \xi$ *each. Then the joint behaviour of the functions* h_1 *of the sample quantile* $q_n(p)$ *(for* $p \in (0,1)$*) and* h_2 *of the sample MedianAD* $\hat{\xi}_n$ *(defined in the List of Notations) is asymptotically normal:*

$$\sqrt{n}\begin{pmatrix} h_1(q_n(p)) - h_1(q_X(p)) \\ h_2(\hat{\xi}_n) - h_2(\xi) \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,\Gamma),$$

where the covariance matrix $\Gamma = (\Gamma_{ij}, 1 \le i,j \le 2)$ *of the asymptotic distribution satisfies*

$$\Gamma_{11} = \frac{p(1-p)}{f_X^2(q_X(p))}\left(h_1'(q_X(p))\right)^2; \qquad \Gamma_{22} = \frac{1+\gamma/f_X^2(\nu)}{4(f_X(\nu+\xi)+f_X(\nu-\xi))^2}\left(h_2'(\xi)\right)^2; \tag{3.25}$$

$$\Gamma_{12} = \Gamma_{21} = h_1'(q_X(p))\,h_2'(\xi)\times \tag{3.26}$$
$$\frac{-\max\left(0,\, F_X(\nu+\xi) - \max\left(F_X(\nu-\xi), p\right)\right) + \frac{1-p}{2} + \frac{f_X(\nu+\xi)-f_X(\nu-\xi)}{f_X(\nu)}\max\left(-\frac{p}{2}, \frac{p-1}{2}\right)}{f_X(q_X(p))\,\left(f_X(\nu+\xi)+f_X(\nu-\xi)\right)}$$

with $\gamma := (f_X(\nu+\xi) - f_X(\nu-\xi))^2 - 4f_X(\nu)\,(f_X(\nu+\xi) - f_X(\nu-\xi))\,(1 - F_X(\nu-\xi) - F_X(\nu+\xi))$.

The correlation of the asymptotic distribution between the two functions is - up to its sign $a_\pm :=$ $\operatorname{sgn}(h_1'(q_X(p))\,h_2'(\xi))$ *- the same whatever the choice of* h_1, h_2*:* $\dfrac{\Gamma_{12}}{\sqrt{\Gamma_{11}\Gamma_{22}}} = a_\pm \times$

$$\frac{-\max\left(0, F_X(\nu+\xi) - \max\left(F_X(\nu-\xi), p\right)\right) + \frac{1-p}{2} + \frac{f_X(\nu+\xi)-f_X(\nu-\xi)}{f_X(\nu)}\max\left(-\frac{p}{2}, \frac{p-1}{2}\right)}{\sqrt{\frac{p(1-p)}{4}}\sqrt{1+\frac{\gamma}{f_X^2(\nu)}}}. \tag{3.27}$$

Two remarks can be made with respect to the result presented. First, the asymptotic dependence with the sample MedianAD does not even require a finite mean. Second, for symmetric distributions, it holds that $f_X(\nu+\xi) = f_X(\nu-\xi)$, thus $\gamma = 0$, so the expressions of the covariance matrix Γ of the asymptotic distribution and (3.27) simplify a lot: In such a case the correlation of the asymptotic distribution is independent of the specific underlying distribution (we have written this out explicitly in [32]). This behavior is to be expected from the results in [71]: They state that in such cases the sample MedianAD is asymptotically equivalent to half the sample IQR, i.e. we are computing a correlation between sample quantiles.

The proof of Theorem 3.6 relies on using the Bahadur representations for both quantities involved.

Proof As in the proof of Theorem 3.1, conditions (C_1') and (P) hold at $q_X(p)$ by assumption, such that the Bahadur representation of the sample quantile holds, recall (3.18). We also use a Bahadur

representation (version of [93]) for the sample MedianAD $\hat{\xi}_n$. Namely, as the conditions are fulfilled by assumption, we have, for $n \to \infty$,

$$\hat{\xi}_n - \xi = \frac{1/2 - (F_n(\nu+\xi) - F_n(\nu-\xi))}{f_X(\nu+\xi) + f_X(\nu-\xi)} - \frac{f_X(\nu+\xi) - f_X(\nu-\xi)}{f_X(\nu+\xi) + f_X(\nu-\xi)} \frac{1/2 - F_n(\nu)}{f_X(\nu)} + \Delta_n, \quad (3.28)$$

where $\Delta_n = o_P(n^{-1/2})$. Clearly, (3.28) can be rewritten in terms of an iid sum as

$$\hat{\xi}_n - \xi = \frac{\frac{1}{n}\sum_{i=1}^n \left(\alpha \mathbb{I}_{(X_i \le \nu)} - f_X(\nu)\mathbb{I}_{(\nu-\xi<X_i\le\nu+\xi)}\right) - \frac{1}{2}(\alpha - f_X(\nu))}{\beta f_X(\nu)} + \Delta_n, \quad (3.29)$$

where, for notational simplification, $\alpha := f_X(\nu+\xi) - f_X(\nu-\xi)$ and $\beta := f_X(\nu+\xi) + f_X(\nu-\xi)$, respectively.

Using the respective Bahadur representations (3.18) and (3.29), and the fact that, by definition of ν and ξ, $\mathbb{P}(X \le \nu) = F_X(\nu) = 1/2$ and $\mathbb{P}(\nu - \xi < X \le \nu + \xi) = F_{|X-\nu|}(\xi) = 1/2$, we apply the bivariate CLT and obtain:

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^n \left(\begin{pmatrix} \mathbb{I}_{(X_i > q_X(p))} \\ \alpha \mathbb{I}_{(X_i \le \nu)} - f_X(\nu)\mathbb{I}_{(\nu-\xi<X_i\le\nu+\xi)} \end{pmatrix} - \begin{pmatrix} 1-p \\ 1/2(\alpha - f_X(\nu)) \end{pmatrix} \right) \\ & = n^{1/2} \left(\begin{pmatrix} 1 - F_n(q_X(p)) \\ \frac{1}{n}\sum(\alpha \mathbb{I}_{(X_i \le \nu)} - f_X(\nu)\mathbb{I}_{(\nu-\xi<X_i\le\nu+\xi)}) \end{pmatrix} - \begin{pmatrix} 1-p \\ 1/2(\alpha - f_X(\nu)) \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, \tilde{\Gamma}), \end{aligned} \quad (3.30)$$

where $\tilde{\Gamma} = \begin{pmatrix} p(1-p) & cov_{ind:q_n,\hat{\xi}_n} \\ cov_{ind:q_n,\hat{\xi}_n} & \mathrm{Var}(\alpha \mathbb{I}_{(X_i \le \nu)} - f_X(\nu)\mathbb{I}_{(\nu-\xi<X_i\le\nu+\xi)}) \end{pmatrix}$, with, as we are going to prove,

$$cov_{ind:q_n,\hat{\xi}_n} := \alpha \max(0, p - 1/2) - f_X(\nu)\left(\max\left(0, F_X(\nu+\xi) - \max(F_X(\nu-\xi), p)\right) - (1-p)/2\right).$$

Then, we need to pre-multiply (i.e. from the left side) equation (3.30) by $\begin{bmatrix} 1/(f_X(q_X(p))) & 0 \\ 0 & 1/(\beta f_X(\nu)) \end{bmatrix}$ to use the Bahadur representation of the sample quantile and of the sample MedianAD (recall (3.18), (3.29)). We obtain:

$$n^{1/2} \begin{pmatrix} \frac{1-F_n(q_X(p))-(1-p)}{f_X(q_X(p))} \\ \frac{\frac{1}{n}\sum(\alpha \mathbb{I}_{(X_i\le\nu)} - f_X(\nu)\mathbb{I}_{(\nu-\xi<X_i\le\nu+\xi)}) - 1/2(\alpha - f_X(\nu))}{\beta f_X(\nu)} \end{pmatrix} = n^{1/2} \begin{pmatrix} q_n(p) - q_X(p) - R_{n,p} \\ \hat{\xi}_n - \xi - \Delta_n \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \Gamma),$$

where, ignoring $R_{n,p}$ and Δ_n since they are $o_P(n^{-1/2})$ (same argumentation as for $R_{n,p}$ in the proof of Theorem 3.1, Part 1), the covariance matrix is as follows

$$\Gamma = \begin{pmatrix} \frac{p(1-p)}{f_X^2(q_X(p))} & \frac{cov_{ind:q_n,\hat{\xi}_n}}{\beta f_X(\nu) f_X(q_X(p))} \\ \frac{cov_{ind:q_n,\hat{\xi}_n}}{\beta f_X(\nu) f_X(q_X(p))} & \frac{\mathrm{Var}(\alpha \mathbb{I}_{(X_i\le\nu)} - f_X(\nu)\mathbb{I}_{(\nu-\xi<X_i\le\nu+\xi)})}{\beta^2 f_X^2(\nu)} \end{pmatrix}.$$

We are left with computing the covariance $cov_{ind:q_n,\hat{\xi}_n}$ and, for the variance of the asymptotic distribution of the MedianAD, the variance $\mathrm{Var}(\alpha \mathbb{I}_{(X_i\le\nu)} - f_X(\nu)\mathbb{I}_{(\nu-\xi<X_i\le\nu+\xi)})$. Recalling that $\mathbb{P}(X \le \nu) = 1/2$

and alike $\mathbb{P}(\nu-\xi < X \le \nu+\xi) = F_{|X-\nu|}(\xi) = 1/2$, we get for the latter:

$$\begin{aligned}
&\mathrm{Var}(\alpha \mathbb{1}_{(X_i\le\nu)} - f_X(\nu)\mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)}) \\
&\quad= \alpha^2 \mathrm{Var}(\mathbb{1}_{(X_i\le\nu)}) + f_X^2(\nu)\,\mathrm{Var}(\mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)}) + 2\alpha f_X(\nu)\,\mathrm{Cov}(\mathbb{1}_{(X_i\le\nu)}, -\mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)}) \\
&\quad= \frac{1}{4}\left(\alpha^2 + f_X^2(\nu) - 8\alpha f_X(\nu)\left(\mathbb{E}[\mathbb{1}_{(\nu-\xi<X_i\le\nu)}] - 1/4\right)\right) \\
&\quad= \frac{1}{4}\left(\alpha^2 + f_X^2(\nu) - 4\alpha f_X(\nu)\left(1/2 - 2F_X(\nu-\xi)\right)\right) \\
&\quad= \frac{1}{4}(f_X^2(\nu)+\gamma), \quad \text{where} \quad \gamma := \alpha^2 - 4\alpha f_X(\nu)\left(1 - F_X(\nu-\xi) - F_X(\nu+\xi)\right).
\end{aligned}$$

Let us turn to the computation of $cov_{ind:q_n,\hat{\xi}_n}$:

$$\begin{aligned}
&\mathrm{Cov}(\mathbb{1}_{(X_i>q_X(p))}, \alpha\mathbb{1}_{(X_i\le\nu)} - f_X(\nu)\mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)}) = \\
&\alpha\,\mathbb{E}[\mathbb{1}_{(X_i>q_X(p))}\mathbb{1}_{(X_i\le\nu)}] - f_X(\nu)\,\mathbb{E}[\mathbb{1}_{(X_i>q_X(p))}\mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)}] - (1-p)(\alpha - f_X(\nu))/2. \qquad (3.31)
\end{aligned}$$

Let us consider one after the other the two expectations in (3.31). Note that we can write (using the definition of ν)

$$\mathbb{1}_{(X_i>q_X(p))}\mathbb{1}_{(X_i\le\nu)} = \begin{cases} 0 & \text{if } \nu \le q_X(p) \quad (\Leftrightarrow p \ge 1/2) \\ \mathbb{1}_{(q_X(p)<X_i\le\nu)} & \text{if } \nu > q_X(p) \quad (\Leftrightarrow p < 1/2) \end{cases},$$

from which we deduce $\mathbb{E}[\mathbb{1}_{(X_i>q_X(p))}\mathbb{1}_{(X_i\le\nu)}] = \max(1/2-p, 0)$. Analogously,

$$\mathbb{1}_{(X_i>q_X(p))}\mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)} = \begin{cases} 0 & \text{if } q_X(p) > \nu+\xi \quad (\Leftrightarrow p > F_X(\nu+\xi)), \\ \mathbb{1}_{(q_X(p)<X_i\le\nu+\xi)} & \text{if } \nu-\xi \le q_X(p) \le \nu+\xi \quad (\Leftrightarrow F_X(\nu-\xi) \le p \le F_X(\nu+\xi)), \\ \mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)} & \text{if } q_X(p) < \nu-\xi \quad (\Leftrightarrow p < F_X(\nu-\xi)). \end{cases}$$

Thus we have $\mathbb{E}[\mathbb{1}_{(X_i>q_X(p))}\mathbb{1}_{(\nu-\xi<X_i\le\nu+\xi)}] = \max\left(0, F_X(\nu+\xi) - \max\left(F_X(\nu-\xi), p\right)\right)$.
Combining these two expressions in (3.31) provides:

$$\begin{aligned}
cov_{ind:q_n,\hat{\xi}_n} &= \alpha\max\left(1/2-p, 0\right) - f_X(\nu)\max\left(0, F_X(\nu+\xi) - \max\left(F_X(\nu-\xi), p\right)\right) - (1-p)(\alpha - f_X(\nu))/2 \\
&= \alpha\max\left(-p/2, (p-1)/2\right) - f_X(\nu)\left(\max\left(0, F_X(\nu+\xi) - \max\left(F_X(\nu-\xi), p\right)\right) - (1-p)/2\right).
\end{aligned}$$

This concludes the computations. Nevertheless, to be explicit, let us write out the overall covariance and correlation of the asymptotic distribution:

$$\begin{aligned}
\Gamma_{12} = \Gamma_{21} &= \frac{-\max\left(0, F_X(\nu+\xi) - \max\left(F_X(\nu-\xi), p\right)\right) + \frac{1-p}{2} + \frac{f_X(\nu+\xi)-f_X(\nu-\xi)}{f_X(\nu)}\max\left(-\frac{p}{2}, \frac{p-1}{2}\right)}{(f_X(\nu+\xi)+f_X(\nu-\xi))f_X(q_X(p))}, \\
\frac{\Gamma_{12}}{\sqrt{\Gamma_{11}\Gamma_{22}}} &= \frac{-\max\left(0, F_X(\nu+\xi) - \max\left(F_X(\nu-\xi), p\right)\right) + \frac{1-p}{2} + \frac{f_X(\nu+\xi)-f_X(\nu-\xi)}{f_X(\nu)}\max\left(-\frac{p}{2}, \frac{p-1}{2}\right)}{\sqrt{\frac{p(1-p)}{4}}\sqrt{1+\frac{\gamma}{f_X^2(\nu)}}},
\end{aligned}$$

which are exactly the covariance in (3.26) and correlation as in (3.27), respectively, for the case $h_1(x) = h_2(x) = x$ (the case with general functions h_1, h_2 follows directly by the application of the Delta method).

As expected, the above computed asymptotic variance of the sample MedianAD, i.e.

$$\Gamma_{22} = \frac{1 + \gamma / f_X^2(\nu)}{4\left(f_X(\nu+\xi) + f_X(\nu-\xi)\right)^2}, \tag{3.32}$$

exactly equals the variance of the asymptotic distribution of the sample MedianAD as in equation (11) of [117] (while in [93] they seem to have some typos in their definition of the quantity γ such that one does not get the same result). □

3.4 Location-scale quantile and measures of dispersion

As a comparison to using historical estimation via sample quantiles, we alternatively estimate the quantile via the known analytical formula for the quantile of the model, considering a given location-scale distribution with unknown but finite mean μ and variance σ^2 as defined in (3.3) (meaning that we are considering only a sub-class of all distribution functions). Consequently, we can write the quantile of order p in such cases as

$$q_X(p) = \mu + \sigma q_Y(p), \tag{3.33}$$

where Y is the corresponding rv with standardised distribution having mean 0 and variance 1.

Hence, if we estimate μ by the sample mean $\bar{X}_n$ and σ by the square-root of the sample variance, $\sqrt{\hat{\sigma}_n^2}$, the estimator of (3.33) defined in (3.2), can be written as

$$q_{n,\hat{\mu},\hat{\sigma}}(p) = \bar{X}_n + \hat{\sigma}_n q_Y(p).$$

First, we present a result for the dependence of (here) the location-scale quantile estimator with the r-th absolute central sample moment, the analogon to Section 3.2. Then, we present the asymptotics of the location-scale quantile when using the sample MedianAD, analogously to Section 3.3.

Note that in the case μ known, the estimator reduces to $q_{n,\hat{\sigma}}(p) = \mu + \hat{\sigma}_n q_Y(p)$ (as given in (3.2)), and studying the dependence with the dispersion measure estimators becomes simpler; we refer to our work in [32] for this subcase.

3.4.1 Location-scale quantile and r-th absolute centred sample moment

Let us start with presenting the analogon of Theorem 3.1 for functions of the location-scale quantile estimator.

Proposition 3.7 *Consider an iid sample with parent rv X having existing (unknown) mean μ and variance σ^2. For any integer $r > 0$ assume conditions (M_r), and, if $r = 1$, (C_0) at μ and (M_2). Then, the joint behaviour of the functions h_1 of the quantile estimator $q_{n,\hat{\mu},\hat{\sigma}}(p)$ from a location-scale model (for $p \in (0,1)$) and h_2 of the r-th absolute central sample moment $\hat{m}(X,n,r)$ (defined in the List of Notations) is asymptotically normal:*

$$\sqrt{n}\begin{pmatrix} h_1(q_{n,\hat{\mu},\hat{\sigma}}(p)) - h_1(q_X(p)) \\ h_2(\hat{m}(X,n,r)) - h_2(m(X,r)) \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \Lambda^{(r)}), \tag{3.34}$$

where the covariance matrix $\Lambda^{(r)} = (\Lambda_{ij}^{(r)}, 1 \le i,j \le 2)$ *of the asymptotic distribution satisfies*

$$\Lambda_{11}^{(r)} = \sigma^2 \left(h_1'(q_X(p))\right)^2 \Big(1 + q_Y(p)\Big(q_Y(p)(\mathbb{E}[Y^4]-1)/4 + \mathbb{E}[Y^3]\Big)\Big); \tag{3.35}$$

$$\Lambda_{22}^{(r)} = \sigma^{2r} \left(h_2'(m(X,r))\right)^2 \operatorname{Var}\Big(|Y|^r - rY\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]\Big); \tag{3.36}$$

$$\Lambda_{12}^{(r)} = \Lambda_{21}^{(r)} = \sigma^{r+1} h_1'(q_X(p))\, h_2'(m(X,r)) \times \tag{3.37}$$

$$\left(\mathbb{E}[Y^{r+1}\operatorname{sgn}(Y)^r] + \frac{q_Y(p)}{2}(\mathbb{E}[|Y|^{r+2}] - \mathbb{E}[|Y|^r]) - r\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]\left(1 + \frac{q_Y(p)}{2}\,\mathbb{E}[Y^3]\right)\right).$$

The correlation of the asymptotic distribution between the functional h_1 *of the location-scale quantile estimator and the functional* h_2 *of the measure of dispersion is - up to its sign* $a_\pm = \operatorname{sgn}(h_1'(q_X(p)) \times h_2'(m(X,r)))$ *- the same whatever the choice of* h_1, h_2*:* $\frac{\Lambda_{12}^{(r)}}{\sqrt{\Lambda_{11}^{(r)}\Lambda_{22}^{(r)}}} = a_\pm \times$

$$\frac{\mathbb{E}[Y^{r+1}\operatorname{sgn}(Y)^r] + \frac{q_Y(p)}{2}(\mathbb{E}[|Y|^{r+2}] - \mathbb{E}[|Y|^r]) - r\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]\left(1 + \frac{q_Y(p)}{2}\,\mathbb{E}[Y^3]\right)}{\sqrt{\left(1 + q_Y(p)\left(q_Y(p)\frac{\mathbb{E}[Y^4]-1}{4} + \mathbb{E}[Y^3]\right)\right)\operatorname{Var}\left(|Y|^r - rY\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]\right)}}. \tag{3.38}$$

Note that using the location-scale quantile model implies assuming the existence of a finite fourth moment with any measure of dispersion estimator, even with the sample MAD - this is in contrast to the historical estimation with the sample quantile.

Proof As in the proofs with the sample quantile q_n we want to apply the CLT to the estimators involved. For this we need iid sum like representations. Recall, from Proposition 3.3 we have that

$$\hat{m}(X,n,r) = \tilde{m}(X,n,r) - r\sqrt{n}(\bar{X}_n - \mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r] + o_P(1/\sqrt{n})$$

As $q_{n,\hat{\mu},\hat{\sigma}}(p) = \bar{X}_n + q_Y(p)\hat{\sigma}_n = \bar{X}_n + q_Y(p)\tilde{\sigma}_n + o_P(1/\sqrt{n})$ includes the square-root of an iid-sum, we will apply a multivariate CLT to the vector $(\tilde{\sigma}_n^2, \tilde{m}(X,n,r), \bar{X}_n)^T$ and then deduce the bivariate asymptotics of $q_{n,\hat{\mu},\hat{\sigma}}(p)$ and $\hat{m}(X,n,r)$ by the continuous mapping theorem.

By the trivariate CLT it holds that

$$n^{1/2}\left(\begin{pmatrix}\bar{X}_n\\ \tilde{\sigma}_n^2\\ \tilde{m}(X,n,r))\end{pmatrix} - \begin{pmatrix}\mu\\ \sigma^2\\ m(X,r)\end{pmatrix}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \widetilde{\Lambda}^{(r)}) \tag{3.39}$$

where the covariance matrix $\widetilde{\Lambda}^{(r)} = (\widetilde{\Lambda}_{ij}^{(r)}), i,j = 1,2,3$, has to be determined. The components can be computed directly. First, the variances: $\widetilde{\Lambda}_{11}^{(r)} = \operatorname{Var}(X) = \sigma^2, \widetilde{\Lambda}_{22}^{(r)} = \operatorname{Var}(X^2) = \mu_4 - \sigma^4, \widetilde{\Lambda}_{33}^{(r)} = \operatorname{Var}(|X|^r) = \sigma^{2r}\operatorname{Var}(|Y|^r)$. Then, the covariances

$$\begin{aligned}
\widetilde{\Lambda}_{21}^{(r)} &= \operatorname{Cov}((X-\mu)^2, X) = \operatorname{Cov}((X-\mu)^2, X-\mu) = \mu_3 = \sigma^3\,\mathbb{E}[Y^3],\\
\widetilde{\Lambda}_{31}^{(r)} &= \operatorname{Cov}(|X-\mu|^r, X) = \sigma^{r+1}\,\mathbb{E}[Y|Y|^r] = \sigma^{r+1}\,\mathbb{E}[Y^{r+1}\operatorname{sgn}(Y)^r],\\
\widetilde{\Lambda}_{32}^{(r)} &= \operatorname{Cov}(|X-\mu|^r, (X-\mu)^2) = \sigma^{r+2}\left(\mathbb{E}[|Y|^{r+2}] - \mathbb{E}[|Y|^r]\right).
\end{aligned}$$

Applying the multivariate Delta method to (3.39), we directly deduce the following CLT:

$$n^{1/2}\left(\begin{pmatrix}\bar{X}_n\\ \sqrt{\tilde{\sigma}_n^2}\\ \tilde{m}(X,n,r))\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma\\ m(X,r)\end{pmatrix}\right)\xrightarrow[n\to\infty]{d}\mathcal{N}(0,\widehat{\Lambda}^{(r)}), \tag{3.40}$$

where $\widehat{\Lambda}_{ij}^{(r)}=\begin{cases}\widetilde{\Lambda}_{ij}^{(r)} & \text{for } i,j\in\{1,3\}\\ \widetilde{\Lambda}_{ij}^{(r)}/(4\sigma^2) & \text{for } i=j=2\\ \widetilde{\Lambda}_{ij}^{(r)}/(2\sigma) & \text{else}\end{cases}$. Then, we apply the continuous mapping theorem to (3.40) using $f(x,y,z)\mapsto(x+ay,z+bx)$ with $a=q_Y(p), b=-r\sigma^{r-1}\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]$ to conclude

$$n^{1/2}\left(\begin{pmatrix}\bar{X}_n+a\sqrt{\tilde{\sigma}_n^2}\\ \tilde{m}(X,n,r))+b\bar{X}_n\end{pmatrix}-\begin{pmatrix}q_X(p)\\ m(X,r)\end{pmatrix}\right)\xrightarrow[n\to\infty]{d}\mathcal{N}(0,\Lambda^{(r)}), \tag{3.41}$$

with

$$\begin{aligned}
\Lambda_{11}^{(r)}&=\operatorname{Var}(X+\frac{q_Y(p)}{2\sigma}X^2)=\sigma^2\left(1+q_Y(p)\left(q_Y(p)(\mathbb{E}[Y^4]-1)/4+\mathbb{E}[Y^3]\right)\right),\\
\Lambda_{22}^{(r)}&=\sigma^{2r}\operatorname{Var}\left(|Y|^r-rY\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]\right),\\
\Lambda_{12}^{(r)}&=\operatorname{Cov}(X+\frac{q_Y(p)}{2\sigma}X^2,\sigma^r\left(|Y|^r-rY\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]\right))\\
&=\sigma^{r+1}\left(\mathbb{E}[Y^{r+1}\operatorname{sgn}(Y)^r]+\frac{q_Y(p)}{2}(\mathbb{E}[|Y|^{r+2}]-\mathbb{E}[|Y|^r])-r\,\mathbb{E}[Y^{r-1}\operatorname{sgn}(Y)^r]\left(1+\frac{q_Y(p)}{2}\,\mathbb{E}[Y^3]\right)\right)
\end{aligned}$$

By Slutsky's theorem, adding a rest which converges to zero in probability, does not change the limiting distribution. Thus, the limit of (3.41) holds also when considering $\hat{m}(X,n,r)$ and $q_{n,\hat{\mu},\hat{\sigma}}(p)$. □

3.4.2 Location-scale quantile and MedianAD

We now consider the joint asymptotics of functions of the location-scale quantile estimator with functions of the sample MedianAD.

Proposition 3.8 *Consider an iid sample with parent rv X from a location-scale distribution having (unknown) median ν, MedianAD ξ, mean μ and variance σ^2. Under (M_2), (C_0) in neighbourhoods of $\nu\pm\xi$, (C_1') at $\nu,\nu\pm\xi$, (P) at ν, and at least at one of $\nu\pm\xi$ each, the joint behaviour of the functions h_1 of the quantile estimator $q_{n,\hat{\mu},\hat{\sigma}}(p)$ from a location-scale model (for $p\in(0,1)$) and h_2 of the sample MedianAD $\hat{\xi}_n$ (defined in the List of Notations) is asymptotically normal:*

$$\sqrt{n}\begin{pmatrix}h_1(q_{n,\hat{\mu},\hat{\sigma}}(p))-h_1(q_X(p))\\ h_2(\hat{\xi}_n)-h_2(\xi)\end{pmatrix}\xrightarrow[n\to\infty]{d}\mathcal{N}(0,\Pi), \tag{3.42}$$

where the covariance matrix $\Pi = (\Pi_{ij}, 1 \le i,j \le 2)$ *of the asymptotic distribution satisfies*

$$\Pi_{11} = \sigma^2 \left(h_1'(q_X(p))\right)^2 \left(1 + q_Y(p)\left(q_Y(p)(\mathbb{E}[Y^4]-1)/4 + \mathbb{E}[Y^3]\right)\right); \tag{3.43}$$

$$\Pi_{22} = \frac{1+\gamma/f_Y^2(\frac{\nu-\mu}{\sigma})}{4(f_Y(\frac{\nu+\xi-\mu}{\sigma}) + f_Y(\frac{\nu-\xi-\mu}{\sigma}))^2} \left(h_2'(\xi)\right)^2; \tag{3.44}$$

$$\Pi_{12} = \Pi_{21} = \frac{h_1'(q_X(p))\, h_2'(\xi)\, \sigma^2}{2\left(f_Y(\frac{\nu+\xi-\mu}{\sigma}) + f_Y(\frac{\nu-\xi-\mu}{\sigma})\right)} \times \left(- \mathbb{E}\left[(Y^2 q_Y(p) + 2Y) \mathbb{1}_{\left(\frac{\nu-\xi-\mu}{\sigma} < Y \le \frac{\nu+\xi-\mu}{\sigma}\right)}\right] \right. \tag{3.45}$$

$$\left. + \frac{f_Y(\frac{\nu+\xi-\mu}{\sigma}) - f_Y(\frac{\nu-\xi-\mu}{\sigma})}{f_Y(\frac{\nu-\mu}{\sigma})} \mathbb{E}\left[(Y^2 q_Y(p) + 2Y) \mathbb{1}_{\left(Y \le \frac{\nu-\mu}{\sigma}\right)}\right] + \frac{q_Y(p)}{2}\left(1 - \frac{f_Y(\frac{\nu+\xi-\mu}{\sigma}) - f_Y(\frac{\nu-\xi-\mu}{\sigma})}{f_Y(\frac{\nu-\mu}{\sigma})}\right)\right),$$

γ *being defined in Theorem 3.6. The correlation of the asymptotic distribution remains independent - up to its sign* $a_\pm = \operatorname{sgn}(h_1'(q_X(p))h_2'(\xi))$ *- of the specific choice of* h_1, h_2*:*

$$\frac{\Pi_{12}}{\sqrt{\Pi_{11}\Pi_{22}}} = a_\pm \times \tag{3.46}$$

$$\frac{- \mathbb{E}\left[(Y^2 q_Y(p) + 2Y) \mathbb{1}_{\left(\frac{\nu-\xi-\mu}{\sigma} < Y \le \frac{\nu+\xi-\mu}{\sigma}\right)}\right] + \frac{f_Y(\frac{\nu+\xi-\mu}{\sigma}) - f_Y(\frac{\nu-\xi-\mu}{\sigma})}{f_Y(\frac{\nu-\mu}{\sigma})} \mathbb{E}\left[(Y^2 q_Y(p) + 2Y) \mathbb{1}_{\left(Y \le \frac{\nu-\mu}{\sigma}\right)}\right] + \frac{q_Y(p)}{2}\left(1 - \frac{f_Y(\frac{\nu+\xi-\mu}{\sigma}) - f_Y(\frac{\nu-\xi-\mu}{\sigma})}{f_Y(\frac{\nu-\mu}{\sigma})}\right)}{\sqrt{1+\gamma/f_Y^2(\frac{\nu-\mu}{\sigma})} \times \sqrt{1 + q_Y^2(p)\frac{\mathbb{E}[Y^4]-1}{4} + q_Y(p)\,\mathbb{E}[Y^3]}}.$$

While in the case of the asymptotics of the sample MedianAD with the sample quantile, we did not even require a finite mean of the underlying distribution, here, with the location-scale quantile estimator, we need a finite fourth moment.

Proof We proceed analogously to the proof of Proposition 3.7. First we establish a trivariate CLT, then by the multivariate Delta-method and continuous mapping theorem the wanted asymptotics.

To use the trivariate CLT in this case, recall the Bahadur representation (3.29) for the sample MedianAD, and the asymptotic equivalence, i.e., as $n \to \infty$, $\hat{\sigma}_n^2 = \tilde{\sigma}_n^2 + o_P(1/\sqrt{n})$.
Thus, as we have iid sums (and finite fourth moment by assumption), we can apply the CLT and obtain:

$$n^{1/2}\left(\begin{pmatrix} \bar{X}_n \\ \tilde{\sigma}_n^2 \\ \frac{\frac{1}{n}\sum_{i=1}^n \left(\alpha \mathbb{1}_{(X_i \le \nu)} - f_X(\nu)\mathbb{1}_{(\nu-\xi < X_i \le \nu+\xi)}\right) - \frac{1}{2}(\alpha - f_X(\nu))}{\beta f_X(\nu)} \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \\ \xi \end{pmatrix}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \widetilde{\Pi}),$$

with $\alpha = f_X(\nu+\xi) - f_X(\nu-\xi)$, $\beta = f_X(\nu+\xi) + f_X(\nu-\xi)$ and $\widetilde{\Pi} = (\widetilde{\Pi}_{ij}), i,j = 1,2,3$ to be computed.
Since $\widetilde{\Pi}_{ij}$ for $i,j \in \{1,2\}$ have been already computed, see (3.39), and we know also $\widetilde{\Pi}_{33}$, see Γ_{22} in

(3.25), we are left with computing the covariances $\widetilde{\Pi}_{31}, \widetilde{\Pi}_{32}$.

Step 1: Covariance with the sample mean.
Recall that $\mathbb{P}(X \leq \nu) = 1/2$ and $\mathbb{P}(|X - \nu| \leq \xi) = 1/2$. Then,

$$\begin{aligned}
\widetilde{\Pi}_{31} &= \mathrm{Cov}\left(X_i, \frac{\alpha \mathbb{1}_{(X_i \leq \nu)} - f_X(\nu)\mathbb{1}_{(\nu-\xi<X_i\leq\nu+\xi)}}{\beta f_X(\nu)}\right) \\
&= \frac{1}{\beta f_X(\nu)}\left(\alpha\,\mathbb{E}[X_i \mathbb{1}_{(X_i\leq\nu)}] - f_X(\nu)\,\mathbb{E}[X_i \mathbb{1}_{(\nu-\xi<X_i\leq\nu+\xi)}] - \frac{\mu}{2}(\alpha - f_X(\nu))\right) \\
&= \frac{\sigma}{\beta f_X(\nu)}\left(\alpha\,\mathbb{E}[Y_i \mathbb{1}_{(Y_i\leq\frac{\nu-\mu}{\sigma})}] - f_X(\nu)\,\mathbb{E}[Y_i \mathbb{1}_{(\frac{\nu-\xi-\mu}{\sigma}<Y_i\leq\frac{\nu+\xi-\mu}{\sigma})}]\right) \\
&= \frac{\sigma}{\beta}\left(\frac{\alpha}{f_X(\nu)}\,\mathbb{E}[Y \mathbb{1}_{(Y\leq\frac{\nu-\mu}{\sigma})}] - \mathbb{E}[Y \mathbb{1}_{(\frac{\nu-\xi-\mu}{\sigma}<Y\leq\frac{\nu+\xi-\mu}{\sigma})}]\right), \qquad (3.47)
\end{aligned}$$

using $X_i = \mu + \sigma Y_i$ for the second equality.

Step 2: Covariance with the sample variance.

$$\begin{aligned}
\widetilde{\Pi}_{32} &= \mathrm{Cov}\left((X_i-\mu)^2, \frac{\alpha \mathbb{1}_{(X_i \leq \nu)} - f_X(\nu)\mathbb{1}_{(\nu-\xi<X_i\leq\nu+\xi)}}{\beta f_X(\nu)}\right) \\
&= \frac{1}{\beta f_X(\nu)}\left(\alpha\,\mathbb{E}[(X_i-\mu)^2 \mathbb{1}_{(X_i\leq\nu)}] - f_X(\nu)\,\mathbb{E}[(X_i-\mu)^2 \mathbb{1}_{(\nu-\xi<X_i\leq\nu+\xi)}] - \frac{\sigma^2}{2}(\alpha - f_X(\nu))\right) \\
&= \frac{\sigma^2}{\beta f_X(\nu)}\left(\alpha\,\mathbb{E}[Y_i^2 \mathbb{1}_{(Y_i\leq\frac{\nu-\mu}{\sigma})}] - f_X(\nu)\,\mathbb{E}[Y_i^2 \mathbb{1}_{(\frac{\nu-\xi-\mu}{\sigma}<Y_i\leq\frac{\nu+\xi-\mu}{\sigma})}] - \frac{1}{2}(\alpha - f_X(\nu))\right) \\
&= \frac{\sigma^2}{\beta}\left(\frac{\alpha}{f_X(\nu)}\,\mathbb{E}[Y^2 \mathbb{1}_{(Y\leq\frac{\nu-\mu}{\sigma})}] - \mathbb{E}[Y^2 \mathbb{1}_{(\frac{\nu-\xi-\mu}{\sigma}<Y\leq\frac{\nu+\xi-\mu}{\sigma})}] - \frac{1}{2}\left(\frac{\alpha}{f_X(\nu)} - 1\right)\right). \qquad (3.48)
\end{aligned}$$

By the Delta-method, we then have

$$n^{1/2}\left(\begin{pmatrix} \bar{X}_n \\ \sqrt{\tilde{\sigma}_n^2} \\ \frac{\frac{1}{n}\sum_{i=1}^n\left(\alpha \mathbb{1}_{(X_i\leq\nu)} - f_X(\nu)\mathbb{1}_{(\nu-\xi<X_i\leq\nu+\xi)}\right) - \frac{1}{2}(\alpha - f_X(\nu))}{\beta f_X(\nu)} \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma \\ \xi \end{pmatrix}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \widehat{\Pi}), \qquad (3.49)$$

where $\widehat{\Pi}_{ij}^{(r)} = \begin{cases} \widetilde{\Pi}_{ij}^{(r)} & \text{for } i,j \in \{1,3\} \\ \widetilde{\Pi}_{ij}^{(r)}/(4\sigma^2) & \text{for } i=j=2 \\ \widetilde{\Pi}_{ij}^{(r)}/(2\sigma) & \text{else} \end{cases}$. Finally, we apply the continuous mapping theorem with the function $f(x,y,z) \mapsto (x + q_Y(p)y, z)$ to (3.49). Recalling again that a rest of $o_P(1)$ does not change the limiting distribution by Slutsky's theorem, we conclude

$$\begin{aligned}
&n^{1/2}\left(\begin{pmatrix} q_{n,\hat{\mu},\hat{\sigma}}(p) \\ \hat{\xi}_n \end{pmatrix} - \begin{pmatrix} q_X(p) \\ \xi \end{pmatrix}\right) \\
&\quad = n^{1/2}\left(\begin{pmatrix} \bar{X}_n + q_Y(p)\tilde{\sigma} + o_P(1) \\ \frac{\frac{1}{n}\sum_{i=1}^n\left(\alpha \mathbb{1}_{(X_i\leq\nu)} - f_X(\nu)\mathbb{1}_{(\nu-\xi<X_i\leq\nu+\xi)}\right) - \frac{1}{2}(\alpha - f_X(\nu))}{\beta f_X(\nu)} + \Delta_n \end{pmatrix} - \begin{pmatrix} q_X(p) \\ \xi \end{pmatrix}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \Pi),
\end{aligned}$$

where we only need to compute Π_{12}.

$$\Pi_{12} = \widehat{\Pi}_{13} + q_Y(p)\widehat{\Pi}_{23} = \frac{\sigma}{2\beta}\left(\frac{\alpha}{f_X(\nu)}\mathbb{E}[2Y\mathbb{1}_{(Y\leq\frac{\nu-\mu}{\sigma})}] - \mathbb{E}[2Y\mathbb{1}_{(\frac{\nu-\xi-\mu}{\sigma}<Y\leq\frac{\nu+\xi-\mu}{\sigma})}]\right.$$
$$\left. + q_Y(p)\left(\frac{\alpha}{f_X(\nu)}\mathbb{E}[Y^2\mathbb{1}_{(Y\leq\frac{\nu-\mu}{\sigma})}] - \mathbb{E}[Y^2\mathbb{1}_{(\frac{\nu-\xi-\mu}{\sigma}<Y\leq\frac{\nu+\xi-\mu}{\sigma})}] - \frac{1}{2}\left(\frac{\alpha}{f_X(\nu)} - 1\right)\right)\right)$$

from which (3.45) follows, by plugging in the explicit expressions for $\beta, \alpha, f_X(\nu)$ in terms of f_Y (e.g. $f_X(\nu) = \frac{1}{\sigma}f_Y(\frac{\nu-\mu}{\sigma})$). □

3.5 Discussion

Let us address different points regarding the main results presented in Sections 3.2 to 3.4. First, we collect some remarks and implications, which directly follow from the asymptotic theorems. Then, we review and compare the conditions imposed by the different results on the underlying distribution F_X.

Remarks - First, regarding the generality of the results presented and their proofs:

- Theorem 3.1 and Proposition 3.7 treat the case of the r-th absolute sample moment as measure of dispersion. We saw in Chapter 2 that the original interest was about the sample MAD ($r = 1$) and the sample std (whose results follow from the case of the sample variance, $r = 2$). We show the results for the r-th absolute sample moment for any integer $r \geq 1$ for its mathematical generality. A presentation of the results (and proof) only treating specifically the cases $r = 1, 2$ can be found in [32].

- While the results for the asymptotic Pearson correlation between these quantile and measure of dispersion estimators are shown, the joint normality of these estimators allows us to directly deduce the asymptotic rank correlations between these estimators (in the case of an existing Pearson correlation) using the results of [86]. More details on this can be found in the online-appendix, [34], Appendix B.3.

- Also, note that we established the different asymptotics using a bivariate CLT. Because of the iid-sum structure of the representations of the estimators, we could have also concluded the joint asymptotics, after establishing the univariate asymptotics, by applying the Cramér-Wold device.

- Finally, note that an alternative to the Bahadur representation to establishing iid-sum representations for the estimators might be by remarking that the estimators considered are Hadamard differentiable (see e.g. [120] for details on this concept). Still, this might ask for additional conditions on the distribution function, for instance in the case of the MedianAD, as remarked in [93].

Further, it is worth to point out the following properties of the asymptotic distributions presented when X follows a location-scale distribution:

- For the class of location-scale distributions, all the correlations of the asymptotic distributions neither depend on the location parameter μ, nor on the scale parameter σ.

- The asymptotics when considering a location-scale quantile estimator with known mean μ, i.e. $q_{n,\hat{\sigma}_n}$, can be deduced from Propositions 3.7 and 3.8. The explicit results (for the sample MAD, sample variance and sample MedianAD as measures of dispersion) can be found in [32]. Therein one can see, for instance, that the correlation of the asymptotic distribution of $q_{n,\hat{\sigma}}$ with any of these mentioned measure of dispersion estimators is independent (up to its sign) of the order p of the quantile.

- Further, if we assume the location-scale distribution to be symmetric, all the correlations of the asymptotic distribution have their minimum (of value 0) at $p = 0.5$ (in the case of considering the sample quantile at level 0.5, i.e. the sample median, with the sample MedianAD this was proved by [61]) and are point symmetric around $p = 0.5$.

- Additionally, in such a symmetric case the correlation of the asymptotic distribution of the sample quantile with the MedianAD is completely independent of the choice of the underlying symmetrical location-scale distribution, see [32]. As remarked earlier, this follows from the asymptotic equivalence of sample MedianAD and half the sample IQR in such a case (shown in [71]) as this means we are considering the correlation of the asymptotic distribution between sample quantiles of different orders.

Last but not least, we comment on the functions h_1, h_2 used in the results.

- Regarding the choice of functions h_1 and h_2, some care has to be taken when applying the Delta method, as the conditions will not always be satisfied: E.g. when using the logarithm, $h_1 = \log$, the quantity $h_1^{'}(q_X(p)) = 1/q_X(p)$ is not defined at $p = F_X(0)$. In such a case, if the left-sided and right-sided limits (for the covariance and correlation of the asymptotic distribution respectively) coincide, we simply set the value at the point itself, by continuity of the limit, to be the left-sided limit.

- If the functions h_1, h_2 are such that $h_1'(q_X(p)) \times h_2'(m(X,r)) = 0$ (or $h_1'(q_X(p)) \times h_2'(\xi) = 0$ in the case of the MedianAD), then we have asymptotic linear independence in any of the results presented: The covariances and correlations of the asymptotic distributions will equal zero (as $\operatorname{sgn}(0) = 0$, by definition).

Conditions on the underlying distribution - While we specified in each theorem and proposition which moment and smoothness conditions the underlying parent random variable has to fulfil, we offer in Table 3.1 an overview when considering, exemplary, the three measures of dispersion which are most well known/offer the least constraints: The sample MAD, the sample variance and the sample MedianAD. For each estimator, we present separately the conditions needed for having a Bahadur/ iid-sum representation (second column) as well as the conditions for the univariate asymptotics (third column). These latter conditions are also sufficient for the joint asymptotics of any measure of dispersion and quantile estimator, as found in the theorems.

We see from Table 3.1 that, in most cases, the conditions for the Bahadur representation and the asymptotic normality are the same. The two exceptions are the location-scale quantile estimator and the sample variance. Also, as already mentioned, the main differences in the choice of estimators lie in the moment conditions: Using the location scale quantile requires the existence of a fourth moment (in contrast to no moment condition for the sample quantile). Also, for the measure of dispersion estimators, a fourth moment is needed when using the sample variance, whereas only a finite second moment with the sample MAD, and no moment conditions at all are imposed by the sample MedianAD. This is a very important remark as it needs to be taken into account when choosing estimators in practice. Lastly, we see that for all different estimators the continuity and differentiability conditions on F_X are not very restrictive.

Apart from comparing the theoretical conditions of these asymptotics, another important point in applications is to assess the finite sample performance; this will be done in Section 3.6.2, via a simulation study.

TABLE 3.1: Conditions needed for the representation as Bahadur/iid-sum representation of the sample estimators considered (second column), and for their use in computation of univariate or bivariate asymptotic normality (third column)

Quantile Estimator	Bahadur/ iid-sum Representation	Asymptotic Normality of Estimator / Joint Asymptotics (with a Measure of Dispersion Estimator)
$q_n(p)$	(QE1): (C_1') and (P) at $q_X(p)$ each	(QE1)
$q_{n,\hat{\mu},\hat{\sigma}}(p)$ or $q_{n,\hat{\sigma}}(p)$	–	(QE2) : $\begin{cases}(M_2)\\ (X-\mu)^2 \text{ not constant}\end{cases}$

Measure of Dispersion Estimator	Bahadur / iid-sum Representation	Asymptotic Normality of Estimator / Joint Asymptotics (with a Quantile Estimator)
$\hat{\sigma}_n^2$	–	(MD1) : $\begin{cases}(M_2)\\ (X-\mu)^2 \text{ not constant}\end{cases}$
$\hat{\theta}_n$	(MD2) : $\begin{cases}(M_1)\\ (C_0) \text{ at } \mu\end{cases}$	(MD2)
$\hat{\xi}_n$	(MD3) : $\begin{cases}(C_0) \text{ in a neighbourhood of } \nu \pm \xi\\ (D_1) \text{ at } \nu, \nu \pm \xi,\\ (P) \text{ at } \nu, \text{ and at least one of } \nu - \xi, \nu + \xi\end{cases}$	(MD3)

3.6 Examples/ finite sample performance

We consider two different applications in this section. Both aim at further understanding the dependence behaviour of the asymptotic distribution between the different estimators and the implications for their use in practice. First, we bring together the asymptotic results of the historical estimation with the sample quantile and the usage of the location-scale quantile estimator: We simply compare the strength of correlation of the asymptotic distribution for the two different quantile and selected measure of dispersion estimators. To do so, we focus on the two examples of typical location-scale distributions, one with light tails (Gaussian), the other with heavy tails (Student-t). Second, we evaluate empirically in a simulation study for Gaussian and Student distributions how well the finite sample results approximate the theoretical asymptotics of Sections 3.2 to 3.4.

We will consider the three most common examples for the measure of dispersion estimators: The sample MAD, the sample variance and the sample MedianAD. These were the three for which we already presented and compared the associated conditions on the underlying distribution, recall Table 3.1.

For notational convenience, we use a framework that incorporates all different cases. Recall from the introduction in this chapter that we denote the quantile estimator simply by $\hat{q}_n$ (representing either the sample quantile q_n or the location-scale quantile model $q_{n,\hat{\mu},\hat{\sigma}}$, $q_{n,\hat{\sigma}}$ respectively). Additionally, we write $\hat{D}_{i,n}$, $i = 1,2,3$, for the measures of dispersion of sample size n which we consider in this section, i.e.

referring by $\hat{D}_{1,n} = \hat{\theta}_n$ to the sample MAD, $\hat{D}_{2,n} = \hat{\sigma}_n^2$ to the sample variance, and $\hat{D}_{3,n} = \hat{\xi}_n$ to the sample MedianAD.

For the ease of readibility, by abuse of notation, the term 'sample' in the context of estimators may be omitted in this section as we will be exclusively referring to sample quantities throughout: We will use variance, MAD and MedianAD synonymously for sample variance, sample MAD and sample MedianAD, respectively.

3.6.1 The impact of the choice of the quantile estimator

While it is known that both estimators are consistent, the asymptotic distribution of the parametric location-scale quantile estimator has smaller variance than the sample quantile. Here we look at how this influences the correlations of their asymptotic distribution with the measure of dispersion estimators. Clearly, these correlations can be deduced from the theorems presented in Sections 3.2 to 3.4. For the sake of conciseness, we present the expressions and their derivation in the Appendix B.1. An extended analysis also including the covariances and ratios of covariances or correlations of the asymptotic distributions, can be found in [32].

We consider the correlations of the asymptotic distribution when using, respectively, the sample quantile and the location-scale quantile estimator assuming, for simplicity, μ to be known .

As the correlations (up to their sign) do not depend on the functions h_1, h_2 of the corresponding quantities, we focus here on the case where h_1, h_2 are the identity functions. This will make the results more traceable. Further, we comment only on $p \geq 0.5$, as the case of $p < 0.5$ can be deduced by the corresponding symmetry around $p = 0.5$ (as both distributions considered, Gaussian and Student, are symmetric).

In Figure 3.1, we plot the correlations of the asymptotic distributions for the different quantile and measure of dispersion estimators ($\hat{q}_n$ and $\hat{D}_{i,n}, i = 1,2,3$). The left plot corresponds to a Gaussian distribution, the two on the right to Student distributions with decreasing degrees of freedom ($\nu = 10$ then 5).

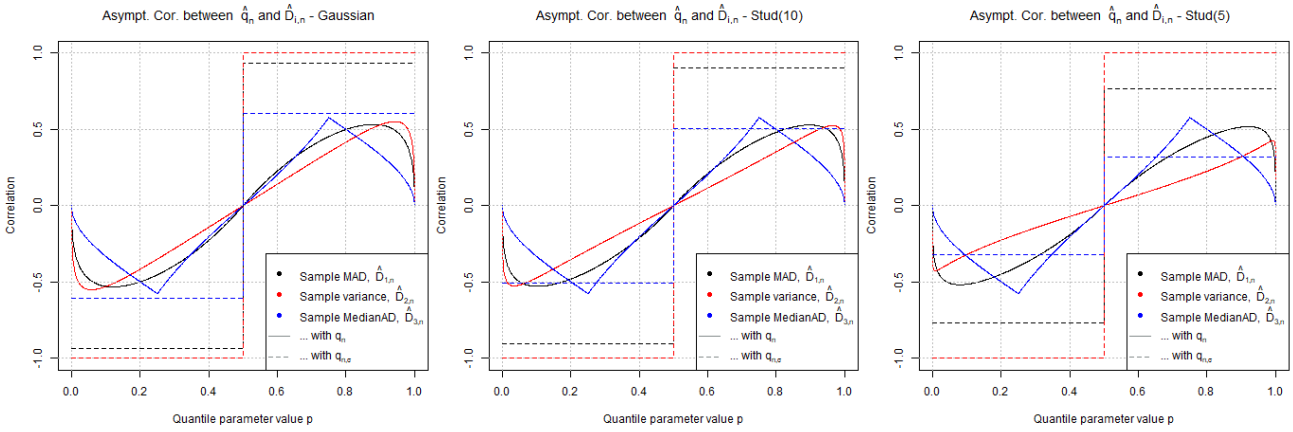


FIGURE 3.1: Comparison of correlations of the asymptotic distributions between the different quantile and measure of dispersion estimators; From left to right: Gaussian, Student(10), Student(5).

Let us start with two general observations on the plots (also mentioned in the remarks in Section 3.5):

- The different correlations are point symmetric around $p = 0.5$ where they attain the value of 0 (as for any symmetric location-scale distribution).

- The correlation with the location-scale quantile estimator is constant up to the sign of $(p - 0.5)$. The reason is that, in $q_{n,\hat{\sigma}}(p)$, only σ needs to be estimated - which does not depend on the chosen order of the quantile p.

Concerning the correlations with the sample quantile q_n (solid lines), we see that:

- All three correlations have a similar range.
- In the tails (i.e. for small and big values of p), the correlations are somewhat similar for variance (red) and MAD (black), but clearly the lowest for the MedianAD (blue). This may be explained by the fact that for a very robust measure of dispersion (as e.g. the MedianAD), an extreme value in a sample does not influence the measure so much as for the variance (which incorporates every deviation from its mean to the square).
- The distinctive shape of the correlation with the MedianAD (with its two peaks at $p = 0.25$ and $p = 0.75$) is related to the sample MedianAD being asymptotically equivalent to half of the sample interquartile range, $\frac{q_n(0.75)-q_n(0.25)}{2}$, for symmetric distributions (see [71]).
- All three correlations with q_n, tend to 0 for $p \to 1$ (but they are only defined for $p \in (0, 1)$). These empirical observations are confirmed theoretically in the literature: The sample maximum, $q_n(1)$, is independent from the sample variance (see [92]). Corresponding results for the MedianAD and MAD can be expected. Because of asymptotic equivalence of sample MedianAD and half the sample IQR for symmetric distributions, the asymptotic independence can be explained by the corresponding results on order statistics, see e.g. [60].

With respect to the location-scale quantile estimator $q_{n,\hat{\sigma}}$ (dotted lines), we notice that:

- The correlation is the strongest for the variance (red) and of similar magnitude as with the MAD (black). The correlation with the MedianAD (blue) is at least one third weaker.
- For the Gaussian distribution, the correlations (dotted lines) are stronger than with the sample quantile (solid lines) whatever the measure of dispersion: The correlation with the MedianAD bounds the correlations of the different measures of dispersion with the sample quantile.

Finally, the following changes can be observed with heavier tailed distributions:

- With increasing heavyness of the tails, the correlation of the location-scale quantile estimator $q_{n,\hat{\sigma}}$ decreases significantly for $\nu = 5$ in the case of the MAD, and for both $\nu = 5$ and 10 for the MedianAD. As for such a quantile estimator we are effectively talking about the correlation of $\hat{\sigma}_n$ with the measures of dispersion, it seems logic that the heavier the distribution, the less correlation between a robust and non-robust measure of dispersion.
- The correlation between $q_{n,\hat{\sigma}}$ and the MedianAD does not bound the correlations with q_n anymore.
- With increasing heavyness of the tail, the correlation of the sample quantile q_n with the sample variance decreases for tail values of p (while staying similar for the MAD).

3.6.2 The effect of sample size in estimation

We want to assess the finite sample performance, in view of the asymptotic results we obtained for the joint distribution of the quantile and dispersion measure estimators. When working with data, we estimate the quantile and measure of dispersion estimators on finite samples of size n. To subsequently evaluate the corresponding covariance and correlation empirically we need, say l, independent realizations of these quantile and measure of dispersion estimators.

To assess the finite sample performance, we conduct a simulation study in the following way. 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 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 quantile or the dispersion measure, with $n = 126, 252, 504, 1008$ (being multiples or fractions of one year of data, i.e. 252 working days/ data points), and different lengths l corresponding to the independent realizations of the quantile or dispersion measure estimator to estimate the sample correlation of interest. Taking the example of an extreme quantile with $p = 0.95$, we compute l independent realizations of the quantile estimator $(\hat{q}_{n,k}(p))_{k=1,\dots,l}$ on disjoint samples and, accordingly, l realizations of the measure of dispersion estimator $(\hat{D}_{i,n,k})_{k=1,\dots,l}$. We then estimate $\text{Cor}\left(\hat{q}_n(p), \hat{D}_{i,n}\right)$ by its sample correlation, using these l pairs of independent realizations 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 '$(n \to \infty)$' 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 Table 3.2 could be derived from the theorems presented in Sections 3.2 to 3.4 (and can be found, as mentioned, in the Appendix B.1). Further, we have seen that the correlation results for location-scale distributions are independent of its parameters, hence the choice of μ, σ^2 does not matter.

As theoretical counterpart to the empirical confidence intervals, we also provide confidence intervals for the theoretical correlation of the asymptotic distribution. They are to be understood in the following way: The correlation is estimated on a sample of size l. Thus, for such a given sample size, we can build confidence intervals for the sample Pearson correlation coefficient around the theoretical value (using the classical variance-stabilizing Fisher transform of the correlation coefficient for a bivariate normal distribution to compute the confidence intervals -see the original paper [63] or e.g. a standard encyclopedia entry [112]). Note that the comparison of these theoretical and the empirical confidence interval values has to be considered with care: Recall that the bivariate normality of the quantile estimator and measure of dispersion estimator holds asymptotically. Hence, it is not clear if for the sample sizes n considered, we can assume bivariate normality (this could be tested). Still, we provide those theoretical confidence intervals as approximate guidance for the empirical confidence intervals.

In Table 3.2, we focus on the approximation of the correlation of the asymptotic distribution by its sample correlation as a function of the sample size n, the different dispersion estimators and the three different distributions considered. Thus, we only consider the sample quantile (not the location-scale quantile estimator) and 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, c.f. online-appendix, [34], Appendix B.2). We present here the case for $p = 0.95$. Clearly, for a higher quantile, as e.g. $p = 0.99$, but with the same sample size n for the estimation of the quantile, the sample correlation will be less precise. More extensive results, considering varying sample sizes l, additionally the location-scale quantile estimator and the case $p = 0.99$, are available in the online-appendix, [34], Appendix B.2. Therein, also a simulation study for the rank correlations can be found (Appendix B.3).

Recall that when working with the sample standard deviation, the existence of the fourth moment is a necessary condition. Thus, as it does not exist for a Student distribution with 3 degrees of freedom, we simply write 'NA' as theoretical value instead.

Let us look at the results in Table 3.2. First we consider the Gaussian case. For the three dispersion measures, we see that a sample size of $n = 126$ suffices to estimate on average the correlation of the asymptotic distribution well enough. Also, the theoretical confidence intervals coming from a sample correlation of size $l = 50$ are well captured by the empirical confidence intervals. Moving to heavier tailed distributions, the picture changes a bit. The sample correlation with the sample variance does not

estimate on average accurately the theoretical value. For increasing n, it approaches the theoretical value. This can be explained by the fact that the theoretical correlation values come from the underlying asymptotic bivariate normal distribution. Hence, for a small n the corresponding sample quantities are not yet bivariate normally distributed and one would need a larger sample for this. This different behaviour is not observed for the MAD or MedianAD with more accurate results, comparable to the Gaussian case. While the average with the MAD is slightly (one percent point) below the theoretical value for most values of n, it equals the theoretical value exactly in the case of the MedianAD. In both cases, the sample confidence intervals correspond quite well to the theoretical ones, potentially indicating that the sample quantities converge faster to a bivariate normal distribution.

In summary, the superior finite sample performance (for the heavier-tailed distributions considered) makes the use of the MAD or MedianAD more favourable than the variance. Especially keeping in mind that for heavy distributions where the fourth moment does not exist, the correlation with the variance is not defined theoretically. As can be seen in Table 3.2, the correlation value with the MedianAD is the same for all three distributions considered (and in general for any symmetric location-scale distribution as remarked in Section 3.5). Thus, to better discriminate the results according to the distribution, we recommend the usage of the MAD over the MedianAD.

TABLE 3.2: Average values from a 1'000-fold repetition. Comparing the sample Pearson correlations of the sample measure of dispersion with the sample quantile, 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). 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 '$(n \to \infty)$' and a 95% confidence interval for the sample correlation ($l = 50$ fixed), are provided as benchmark in the last column. We consider the threshold $p = 0.95$.

$p = 0.95$	$n = 126$	$n = 252$	$n = 504$	$n = 1008$	theoretical value $(n \to \infty)$
Gaussian distr.					
$\widehat{\mathrm{Cor}}(\hat{\sigma}_n^2, q_n(p))$	55 (33;71)	55 (34;73)	55 (34;73)	55 (34;71)	55 (32;72)
$\widehat{\mathrm{Cor}}(\hat{\theta}_n, q_n(p))$	48 (26;66)	48 (26;69)	48 (25;69)	48 (26;66)	48 (23;67)
$\widehat{\mathrm{Cor}}(\hat{\xi}_n, q_n(p))$	23 (-4;48)	23 (-3;48)	23 (-4;49)	23 (-4;45)	23 (-5;48)
Student(5) distr.					
$\widehat{\mathrm{Cor}}(\hat{\sigma}_n^2, q_n(p))$	51 (19;75)	49 (19;71)	47 (19;68)	46 (20;67)	43 (17;63)
$\widehat{\mathrm{Cor}}(\hat{\theta}_n, q_n(p))$	50 (27;71)	50 (27;69)	50 (27;70)	51 (27;69)	51 (27;69)
$\widehat{\mathrm{Cor}}(\hat{\xi}_n, q_n(p))$	23 (-6;50)	23 (-6;47)	23 (-6;47)	23 (-6;48)	23 (-5;48)
Student(3) distr.					
$\widehat{\mathrm{Cor}}(\hat{\sigma}_n^2, q_n(p))$	25 (-8;55)	22 (-9;52)	19 (-9;47)	17 (-14;44)	NA
$\widehat{\mathrm{Cor}}(\hat{\theta}_n, q_n(p))$	48 (21;68)	47 (23;67)	47 (20;68)	47 (23;67)	48 (23;67)
$\widehat{\mathrm{Cor}}(\hat{\xi}_n, q_n(p))$	23 (-4;49)	22 (-7;48)	22 (-7;47)	23 (-7;49)	23 (-5;48)

3.7 Conclusion

In this chapter we answered the question **Q3**, about the joint asymptotics of quantile estimators with measure of dispersion estimators for the iid case. We showed the joint asymptotic bivariate normality of functions of two quantile estimators (the sample quantile and the parametric location-scale quantile estimator) and two different types of measure of dispersion estimators (the r-th absolute central sample moment, for any integer $r > 0$, and sample MedianAD) each.

We considered two different quantile estimators, as their different speed in convergence also affects their asymptotic joint distribution with the measure of dispersion estimators, respectively. We analysed this in further detail by looking at the correlations of the asymptotic distributions, when using the sample quantile versus the location-scale quantile estimator (for each of the different measures of dispersion respectively), considering the two main location-scale distributions, the Gaussian and Student ones.

Apart from the speed of convergence, the use of the different quantile estimators also implies different moment conditions. So does also using the sample variance (finite fourth moment needed), versus the sample MAD (finite second moment needed) or the sample MedianAD (not even a finite mean is necessary), which certainly play a role in practice.

Besides, we also verified through simulation on Gaussian and Student distributions a good finite sample performance of the theoretical asymptotic results presented.

After having established asymptotic results for underlying iid distributions, this will be extended for dependent processes in the next chapter. Further, these results will be a basis to consider the procyclicality from a theoretical perspective in Chapter 5.

Takeaways

- We establish a Bahadur-like (iid sum) representation for the r-th absolute central sample moment; see Proposition 3.3.
- We prove the asymptotic normality between quantile estimators (sample quantile, location-scale quantile) and measure of dispersion estimators (r-th absolute central sample moment and the sample MedianAD; note that the former includes, as special cases, the sample MAD and sample variance); see Theorems 3.1 and 3.6, Propositions 3.7 and 3.8.
- The choice of measure of dispersion estimator implies different moment conditions (4th moment for the sample variance, second moment for the sample MAD, no moment condition for sample MedianAD). This should be an important decision guidance in applications.
- We show a good finite sample performance of the asymptotics in a simulation study; see Section 3.6.2.

Key questions (to be followed up in the thesis)

- Can these results be extended for other classes of underlying processes? We consider in Chapter 4 these asymptotics for the family of augmented GARCH(p,q) processes.
- Can these results be extended for other risk measures? We consider asymptotics for other risk measures in Chapter 5.
- What do these results imply for the pro-cyclicality observed in Chapter 2? This question is tackled in Chapter 5.

Related contributions in the appendix

- The analytical expressions and their derivation (for an underlying Gaussian or Student-t distribution) for the correlation of the asymptotic distribution between the sample or location-scale quantile estimator and the sample MAD, variance and MedianAD can be found in the Appendix B.1.
- A more detailed simulation study of the finite sample behavior of the Pearson Correlation of quantile estimators with measure of dispersion estimators can be found in Appendix B.2 of the online-appendix.
- An overview of the known theoretical relations between Pearson and rank correlation as well as a simulation study for the latter can be found in Appendix B.3 of the online-appendix.

Chapter 4

Estimators in the case of augmented GARCH processes: Asymptotic theory

This chapter formed the basis for the joint work [30], albeit we have corrected some mistakes.

4.1 Introduction

This chapter can be seen as extension of the results in Chapter 3. While we looked at the bivariate asymptotic distribution of quantile estimators and measure of dispersion estimators in the iid case there, here we consider an underlying family of GARCH processes. Recall that we showed in Chapter 2, amongst others, in simulations the negative correlation between the (logarithm of the) look-forward ratio and the sample MAD for a GARCH(1,1) process. We chose in the empirical study this simple process because we wanted to isolate certain effects relevant to this model, namely the clustering and return-to-the-mean of volatility.

Nonetheless, since the introduction of the ARCH and GARCH processes in the seminal papers by Engle, [54], and Bollerslev, [23], respectively, various GARCH modifications and extensions have been proposed and used (see e.g. [24] for a non-exhaustive (G)ARCH glossary).

Thus, it is of interest to cover a broad range of GARCH processes when considering the bivariate asymptotics between quantile and measure of dispersion estimators. We focus on so called augmented GARCH(p,q) processes, which were introduced by Duan in [49]. One reason for this choice is that it is a family of processes which includes frequently used GARCH variations as e.g. AGARCH, EGARCH or TGARCH. Another reason is that, for our bivariate asymptotics, we will need certain results on univariate (representations leading to) asymptotics. Recall that we have given in the introduction in Chapter 1 a short literature review of results on asymptotics that extend to different estimators apart from the underlying process itself. Among them, we find the results of Lee in [89] for powers of augmented GARCH(p,q) processes, which will be of use to us in this chapter: As in the iid case, Bahadur-like representations for the quantile and measure of dispersion estimators are necessary in a first step. While such a representation exists for the sample quantile, see [121], it needs to be developed for the r-th absolute central sample moment (analogously to Chapter 3). To be able to conclude a univariate FCLT for this estimator from such a representation, we will need to employ the results of [89].

In contrast to the iid case, there is no Bahadur-like representation for the sample MedianAD in the literature. Thus, we focus on the r-th absolute central sample moment as measure of dispersion (similarly, a location-scale quantile estimator does not make sense for this class of distributions). Thus, in this chapter, we develop a bivariate functional central limit theorem for the sample quantile with the r-th absolute centred sample moment, tackling question **Q4**.

The structure in this chapter is as follows. We present in Section 4.2 the main theorem (and its proof) producing the bivariate FCLT for the sample quantile and the r-th absolute centred sample moment for

augmented GARCH(p, q) processes. The part of the proof for establishing a representation for the r-th absolute central sample moment follows, in its structure, the iid case. Thus, we simply refer to the corresponding proofs in Chapter 3 and only comment on the differences. As a short illustration we present in Section 4.3 specific examples of well-known GARCH models which belong to the family of augmented GARCH(p,q) processes. We also show how the general conditions in the main result translate for these specific cases. We conclude in Section 4.4.

4.2 The bivariate FCLT

Let us introduce the augmented GARCH(p,q) process $X = (X_t)_{t\in\mathbb{Z}}$, due to Duan in [49], namely, for integers $p, q \geq 1$, X_t satisfies

$$X_t = \sigma_t \, \varepsilon_t \tag{4.1}$$

$$\text{with} \quad \Lambda(\sigma_t^2) = \sum_{i=1}^{p} g_i(\varepsilon_{t-i}) + \sum_{j=1}^{q} c_j(\varepsilon_{t-j})\Lambda(\sigma_{t-j}^2), \tag{4.2}$$

where (ε_t) is a series of iid rv's with mean 0 and variance 1, $\sigma_t^2 = \mathrm{Var}(X_t)$ and $\Lambda, g_i, c_j, i = 1, ..., p, j = 1, ..., q$, are real-valued measurable functions. As in [89], we restrict the choice of Λ to the so-called group of either polynomial GARCH(p,q) or exponential GARCH(p,q) processes:

(Lee) $\Lambda(x) = x^\delta$, for some $\delta > 0$, or $\Lambda(x) = \log(x)$.

Before discussing needed conditions for our results on this class of processes, let us recall two familiar notions: We call a process $(X_n, n \geq 1)$ stationary if, for any integers $k \geq 1, h \geq 0$, the joint distribution of $(X_{1+h}, ..., X_{k+h})$ does not depend on the choice of h. Note that this condition is sometimes also called strict stationarity. To ergodicity we refer to in the classical sense (see e.g. [35] for a definition), such that a stationary and ergodic sequence fulfills the Birkhoff ergodic theorem (the 'generalization' of the law of large numbers for ergodic, stationary processes).

Clearly, for a stationary solution to (4.1) and (4.2) to exist, the functions Λ, g_i, c_j as well as the innovation process $(\varepsilon_t)_{t\in\mathbb{Z}}$ have to fulfill some regularity conditions (see e.g. [89], Lemma 1). Alike, for the bivariate FCLT to hold, certain conditions need to be fulfilled; we list them in the following.
First, conditions concerning the dependence structure of the process X. We use the concept of L_p-near-epoch dependence (L_p-NED) which goes back to [80]. Therein it was introduced for mixing processes, here we use a general definition as in [46] but restricted to stationary processes. Let $(Z_n)_{n\in\mathbb{Z}}$, be a sequence of rv's and $\mathcal{F}_s^t = \sigma(Z_s, ..., Z_t)$, for $s \leq t$, the corresponding σ-algebra. The usual L_p-norm is denoted by $\|\cdot\|_p := \mathbb{E}^{1/p}[|\cdot|^p]$. Let us recall the L_p-NED definition.

Definition 4.1 (L_p-NED, [46]) *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 $k \geq 0$, it holds that*

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

for non-negative constants $\nu(k)$ such that $\nu(k) \to 0$ as $k \to \infty$.
If $\nu(k) = O(k^{-\tau-\varepsilon})$ for some $\varepsilon > 0$, we say that X_n is L_p-NED of size $(-\tau)$.
If $\nu(k) = O(e^{-\delta k})$ for some $\delta > 0$, we say that X_n is geometrically L_p-NED.

The second set of conditions concerns the distribution of the augmented GARCH(p,q) process. We impose the same type of conditions as in the iid case, recall Chapter 3: First, the existence of a finite $2k$-th moment for any integer $k > 0$ for the process itself, (M_k). Then, given that the process X is stationary, the continuity, (C_0), or l-fold differentiability of its distribution function F_X, (C_l'), or continuity of the l-th derivative of the distribution function, (C_l) (each at a given point or neighbourhood) for any integer $l > 0$, and the positivity of its density f_X, (P), (at a given point or neighbourhood).

The third type of conditions is set on the functions $g_i, c_j, i = 1,...,p, j = 1,...,q$ of the augmented GARCH(p,q) process of the (Lee) family: Positivity of the functions used and boundedness in L_r-norm for either the polynomial GARCH, (P_r), or exponential GARCH, (L_r), respectively, for a given integer $r > 0$,

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

$$(P_r) \qquad \sum_{i=1}^{p} \|g_i(\varepsilon_0)\|_r < \infty, \sum_{j=1}^{q} \|c_j(\varepsilon_0)\|_r < 1,$$

$$(L_r) \qquad \mathbb{E}[\exp(4r \sum_{i=1}^{p} |g_i(\varepsilon_0)|^2)] < \infty, \sum_{j=1}^{q} |c_j(\varepsilon_0)| < 1.$$

Note that condition (L_r) requires the c_j to be bounded functions.

Remark 4.2 *By construction from (4.1) and (4.2) σ_t and ε_t are independent (and σ_t a functional of $(\varepsilon_{t-j})_{j=1}^{\infty}$). Thus, the conditions on the moments, distribution and density could be formulated in terms of ε_t only. At the same time this might impose some conditions on the functions $g_i, c_j, i = 1,...,p, j = 1,...,q$ (which might not be covered by (A), (P_r) or (L_r)). Thus, we keep the conditions on X_t even if they might not be minimal.*

Now, let us state the main result. To ease its presentation we introduce a trivariate normal random vector (functionals of X), denoted by $(U, V, W)^T$, with mean zero and the following covariance matrix:

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

Theorem 4.3 (bivariate FCLT) *Consider an augmented GARCH(p,q) process X as defined in (4.1) and (4.2). Introduce the random vector $T_{n,r}(X) = \begin{pmatrix} q_n(p) - q_X(p) \\ \hat{m}(X,n,r) - m(X,r) \end{pmatrix}$, for an integer $r > 0$. Assume the process satisfies condition (Lee), (C_0) at the mean μ for $r = 1$, and both conditions (C_2'), (P) at $q_X(p)$. Further, also conditions (M_r), (A), and either $(P_{max(1,r/\delta)})$ for X belonging to the group of polynomial GARCH, or (L_r) for the group of exponential GARCH. Then, we have the following FCLT: For $t \in [0,1]$, as $n \to \infty$,*

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

where $(\boldsymbol{W}_{\Gamma^{(r)}}(t))_{t\in[0,1]}$ *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* $\mathrm{Cov}(\boldsymbol{W}_{\Gamma^{(r)}}(t), \boldsymbol{W}_{\Gamma^{(r)}}(s)) = \min(s,t)\Gamma^{(r)}$, *where*

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

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

Remark 4.4 *Note that the bivariate FCLT between the sample quantile and the r-th absolute centred sample moment requires exactly the same conditions in comparison to the respective univariate convergence (which might be apparent after having gone through Remark 4.10 and the proof of Theorem 4.3).*

Requiring $(C_2^{'})$, (P) *at* $q_X(p)$ *exactly correspond to the conditions for the CLT of the sample quantile of a stationary* L_1*-NED with polynomial rate process - see [121]. Further,* $(P_{\max(1,r/\delta)})$ *or* (L_r) *respectively, together with* (M_r), (A) *and* (C_0) *at* μ *for* $r = 1$, *are the conditions for the univariate CLT of the r-th centred sample moment for augmented GARCH(p,q) processes - which is a new result we establish in Proposition 4.8.*

Choosing $t = 1$ in Theorem 4.3 provides the usual CLT that we state for completeness:

Corollary 4.5 *Consider an augmented GARCH(p,q) process as defined in (4.1) and (4.2). Under the same conditions as in Theorem 4.3, the joint behaviour of the sample quantile* $q_n(p)$ *(for* $p \in (0,1)$*) and the r-th absolute centred sample moment* $\hat{m}(X,n,r)$, *is asymptotically bivariate normal:*

$$\sqrt{n}\begin{pmatrix} q_n(p) - q_X(p) \\ \hat{m}(X,n,r) - m(X,r)\end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,\Gamma^{(r)}), \tag{4.3}$$

where the covariance matrix $\Gamma^{(r)} = (\Gamma_{ij}^{(r)}, 1 \le i,j \le 2)$ *of the asymptotic distribution is as in Theorem 4.3.*

As special case we can also recover the CLT between the sample quantile and the r-th absolute centred sample moment in the iid case, proved earlier in Theorem 3.1 (even if the conditions in an augmented GARCH(p,q) setting will be slightly more restrictive):

Corollary 4.6 *Consider an augmented GARCH(p,q) process as defined in (4.1) and (4.2), choosing* g_i, c_j, Λ *such that* $\sigma_t^2 = \sigma^2 > 0$ *is a positive constant for all t. Under the same conditions as in Theorem 4.3, the joint behaviour of the sample quantile* $q_n(p)$ *(for* $p \in (0,1)$*) and the r-th absolute centred sample moment* $\hat{m}(X,n,r)$, *is asymptotically bivariate normal:*

$$\sqrt{n}\begin{pmatrix} q_n(p) - q_X(p) \\ \hat{m}(X,n,r) - m(X,r)\end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,\Gamma^{(r)}), \tag{4.4}$$

where the covariance matrix $\Gamma^{(r)} = (\Gamma_{ij}^{(r)}, 1 \le i,j \le 2)$ *of the asymptotic distribution simplifies to*

$$\begin{aligned}
\Gamma_{11}^{(r)} &= \frac{p(1-p)}{f_X^2(q_X(p))};\\
\Gamma_{22}^{(r)} &= r^2\,\mathbb{E}[X_0^{r-1}\,\mathrm{sgn}(X_0)^r]^2\sigma^2 + \mathrm{Var}(|X_0|^r) - 2r\,\mathbb{E}[X_0^{r-1}\,\mathrm{sgn}(X_0)^r]\,\mathrm{Cov}(X_0,|X_0|);\\
\Gamma_{12}^{(r)} &= \Gamma_{21}^{(r)} = \frac{1}{f_X(q_X(p))}\left(r\,\mathbb{E}[X_0^{r-1}\,\mathrm{sgn}(X_0)^r]\,\mathrm{Cov}(\mathbb{1}_{(X_0\le q_X(p))}, X_0) - \mathrm{Cov}(\mathbb{1}_{(X_0\le q_X(p))}, |X_0|^r)\right).
\end{aligned}$$

4.2.1 Proof of Theorem 4.3

Before stating the proof of the main theorem, let us start with two auxiliary results which parallel the according auxiliary results in the iid case. We will provide a Bahadur-like representation of $\hat{m}(X,n,r) = \frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_n|^r$ for any integer $r \geq 1$. As such a result is of interest in its own right, we give it separately in Proposition 4.8. To prove it, we need the following Lemma, which extends Lemma 2.1 in [116] (case $v = 1$) to any moment $v \in \mathbb{N}$, and Lemma 3.2 presented in the iid case. Additionally, we present the multivariate FCLT from [11] as a Lemma, such that we can more easily refer to it.

Lemma 4.7 *Consider a stationary and ergodic time-series $(X_n, n \geq 1)$ which has 'short-memory', i.e. $\sum_{i=0}^{\infty}|\mathrm{Cov}(X_0,X_i)| < \infty$. Then, for $v = 1$ or 2, given that the 2nd moment of X_0 exists, or, for any integer $v > 2$, given that the v-th moment of X_0 exists, letting $n \to \infty$, it holds that*

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^v\left(|X_i-\bar{X}_n| - |X_i-\mu|\right) = (\bar{X}_n-\mu)\times\mathbb{E}[(X_0-\mu)^v\,\mathrm{sgn}(\mu-X_0)] + o_P(1/\sqrt{n}). \tag{4.5}$$

Proof The proof follows its equivalent in the iid case, the proof of Lemma 3.2. The argumentation needs to be adapted only at the end in two points, using the stationarity, ergodicity and short-memory of the process. Here, it follows by these three properties that $\sqrt{n}|\bar{X}_n-\mu|^{v+1} \xrightarrow[n\to\infty]{P} 0$ holds for any integer $v \geq 1$. Further, as a last step, we use the ergodicity of the process, instead of the strong law of large numbers, to conclude that

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^v\,\mathrm{sgn}(\mu-X_i) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}[(X-\mu)^v\,\mathrm{sgn}(\mu-X)].$$

□

Now we are ready to state the asymptotic relation between the r-th absolute centred sample moment with known and unknown mean, respectively. This enables us to compute the asymptotics of $\hat{m}(X,n,r)$ (given that the necessary moments exist). As for Lemma 4.7, it is an extension to the stationary, ergodic and short-memory case of Proposition 3.3 in the iid case.

Proposition 4.8 *Consider a stationary and ergodic time-series $(X_n, n \geq 1)$ which has 'short-memory', i.e. $\sum_{i=0}^{\infty}|\mathrm{Cov}(X_0,X_i)| < \infty$. Then, for any integer $r \geq 1$, given that the r-th moment of X_0 exists and (C_0) at μ for $r = 1$, it holds, as $n \to \infty$, that*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i-\bar{X}_n|^r\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^r\right) - r\sqrt{n}(\bar{X}_n-\mu)\,\mathbb{E}[(X_0-\mu)^{r-1}\,\mathrm{sgn}(X_0-\mu)^r] + o_P(1). \tag{4.6}$$

Proof Analogously to the proof of Lemma 4.7, the proof can be extended from the proof of Proposition 3.3 in the iid case. We comment on the differences compared to the iid case for the three different cases of r:

Even integers r - Recall that for the result in Proposition 3.3 we refered to the example 5.2.7 in [90]. Therein they only consider the iid case but in this case, (3.12) still holds as $\sqrt{n}(\bar{X}_n-\mu)^v \xrightarrow{P} 0$, for $v \geq 2$, holds for an ergodic, stationary, short-memory process too.

Case $r = 1$ - The result cited in the iid case, (3.13), holds for ergodic, stationary time-series too, see Lemma 2.1 in [116].

Odd integer $r > 1$ - We point out the three differences to the corresponding proof in the iid case. First, as remarked above for even integers r, $\sqrt{n}(\bar{X}_n-\mu)^v \xrightarrow{P} 0$, for $v \geq 2$, follows from the stationarity, ergodicity and short-memory of the process. Second, we use the ergodicity instead of the law of large numbers. Third, we use Lemma 4.7 instead of its counterpart in the iid case, Lemma 3.2. □

Finally, we state a multivariate FCLT which we will use to prove the theorem, and which is from [11] (adapted to our needs):

Lemma 4.9 (Theorem A.1 in [11]) *Consider a d-dimensional random process $(u_j, j \in \mathbb{Z})$, which is centered and has finite variance, i.e.*

$$\mathbb{E}[u_j] = 0, \quad \|u_j\|_2^2 < \infty \; \forall j \in \mathbb{Z}, \tag{4.7}$$

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

$$u_j = f(\varepsilon_j, \varepsilon_{j-1}, ...), \tag{4.8}$$

where $f : \mathbb{R}^{1\times\infty} \to \mathbb{R}^d$ is a measurable function and $(\varepsilon_j, j \in \mathbb{Z})$ 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 $\left(u_j^{(\Delta)}, j \in \mathbb{Z}\right)$ such that, for any $\Delta \geq 1$, we have

$$u_j^{(\Delta)} = f^{(\Delta)}(\varepsilon_{j-\Delta}, ..., \varepsilon_j, ..., \varepsilon_{j+\Delta}) \tag{4.9}$$

$$\text{and} \sum_{\Delta\geq 1} \|u_0 - u_0^{(\Delta)}\|_2 < \infty, \tag{4.10}$$

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

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

$$\sqrt{n}tU_{[nt]} \overset{D_d[0,1]}{\to} W_\Gamma(t),$$

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 $\mathrm{Cov}(W_\Gamma(s), W_\Gamma(t)) = \min(s,t)\Gamma$.

Remark 4.10 *This multivariate FCLT extends the univariate counterpart from, e.g., Billingsley in [22]. For that, they proof in [11] why the univariate FCLT's are sufficient to establish the multivariate version, using Cramér-Wold and the univariate tightness of the corresponding processes.*

The cited version in Lemma 4.9 differs in two small details from the original Theorem A.1 in [11]. First, it is less general as they assume $f : \mathbb{R}^{d'\times\infty} \to \mathbb{R}^d$ (and consequently $f^\Delta : \mathbb{R}^{d'\times\infty} \to \mathbb{R}^d$) as well as $(\varepsilon_j, j \in \mathbb{Z})$ to be an iid sequence of random vectors with values in $\mathbb{R}^{d'}$. In our case $d' = 1$ is sufficient.

Further, note that we adapted (4.9) from originally being $u_j^{(\Delta)} = f^{(\Delta)}(\varepsilon_j, ..., \varepsilon_{j-\Delta})$. Indeed, it is straightforward to show that the proof of [11] still holds with this modification.

Now we finally proceed with the proof of the main result.

Proof of Theorem 4.3. The proof consists of four steps. We first show that the process (X_t) fulfils the conditions required for having a Bahadur representation of the sample quantile, second, that it also fulfils the conditions for a similar representation for the r-th absolute centred sample moment, third, that the conditions for an FCLT (Lemma 4.9) are fulfilled, which we then use in the fourth step to conclude the multivariate FCLT.

Step 1: Bahadur representation of the sample quantile - conditions.

The Bahadur representation of the sample quantile for a GARCH(p,q) process is well known and can be obtained as a special case for Bahadur representations for processes with a certain dependence structure,

see e.g. [88] and references therein. Here we want to establish the Bahadur representation for sample quantiles from augmented GARCH(p,q) processes of the (Lee) family. We will use the Bahadur representation for general NED processes (see Theorem 1 in [121]). It holds under some conditions that we need to verify. For the ease of comparison, we adapt some of the notation of Theorem 1 in [121]. We have the following:

- Choosing the bivariate function $g(x,t) := \mathbb{I}_{(x\leq t)}$, the non-negativity, boundedness, measurability, and non-decreasingness in the second variable, are straightforward. The function g also satisfies the variation condition uniformly in some neighbourhood of $q_X(p)$ if it is Lipschitz-continuous (see Example 1.5 in [121]). But the latter follows from condition $(C_2^{'})$.

- The differentiability of $\mathbb{E}[g(X,t)] = F_X(t)$ and positivity of its derivative at $t = q_X(p)$ are given by condition (P) at $q_X(p)$.

- The condition

$$|F_X(x) - F_X(q_X(p)) - f_X(q_X(p))(x - q_X(p))| = o\left(|x - q_X(p)|^{3/2}\right) \quad \text{as } x \to q_X(p)$$

  is fulfilled as, by our assumption $(C_2^{'})$, F_X is twice differentiable in $q_X(p)$ (see Remark 2, [121]).

- The stationarity of the process follows from assumption $(P_{\max(1,r/\delta)})$ or (L_r), respectively, and Lemma 1 of [89].

- Lastly, let us verify that the process (X_t) is L_1-NED with polynomial rate. Denoting, for $s \leq t$, the sigma-algebra $\mathcal{F}_s^t = \sigma(\varepsilon_s, ..., \varepsilon_t)$, we can write for any integer $\Delta \geq 1$

$$\|X_t - \mathbb{E}[X_t|\mathcal{F}_{t-\Delta}^{t+\Delta}]\|_2^2 = \|\sigma_t - \mathbb{E}[\sigma_t|\mathcal{F}_{t-\Delta}^{t+\Delta}]\|_2^2 \ \sqrt{\mathbb{E}[\varepsilon_t^2]}.$$

  But $\mathbb{E}[\varepsilon_t^2] < \infty$ since (M_r) holds. Notice that the property of being geometrically L_2-NED, $\|\sigma_t - \mathbb{E}[\sigma_t|\mathcal{F}_{t-\Delta}^{t+\Delta}]\|_2 = O(e^{-\kappa\Delta})$ for some $\kappa > 0$, implies L_1-NED with polynomial rate, as

$$\|\sigma_t - \mathbb{E}[\sigma_t|\mathcal{F}_{t-\Delta}^{t+\Delta}]\|_1 \leq \|\sigma_t - \mathbb{E}[\sigma_t|\mathcal{F}_{t-\Delta}^{t+\Delta}]\|_2 = O(e^{-\kappa\Delta}) = O(\Delta^{-(\beta+3)}), \tag{4.11}$$

  for some $\beta > 3$. So it suffices showing that σ_t is geometrically L_2-NED. For the polynomial GARCH, it follows from Corrollary 1 in [89], which can be applied as (A) and (P_1) hold. For the exponential GARCH case, it follows from Corrollary 3 in [89] as (A) and (L_r) hold.

Thus, we can use Theorem 1 of [121] and write, as $n \to \infty$,

$$q_n(p) - q_X(p) + \frac{F(q_X(p)) - F_n(q_X(p))}{f_X(q_X(p))} = o_P(1/\sqrt{n}). \tag{4.12}$$

Note that we do not use the exact remainder bound as in [121] as for our purposes $o_P(1/\sqrt{n})$ is enough.

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

The representation being given in Proposition 4.8 under some conditions, we only need to check that we fulfil them.

- The stationarity of the process is satisfied under $(P_{max(1,r/\delta)})$ or (L_r) as observed in Step 1.

- 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 $(M_r), (A), (P_{max(1,r/\delta)})$ in the polynomial case, and $(M_r), (A), (L_r)$ in the exponential case respectively, imply the CLT, using Corollary 2 and 3 in [89], respectively.

Thus, the representation for $\hat{m}(X,n,r)$ as in (4.6) holds.

Step 3: Conditions for applying the FCLT

In our case we want to apply Lemma 4.9 in a three-dimensional version. This simplifies the computation, and by applying the continuous mapping theorem we will finally get back a two-dimensional representation. This will be made explicit in Step 4.

Therefore, let us define, anticipating its use in Step 4 for the FCLT of $U_n(X) := \frac{1}{n}\sum_{j=1}^{n} u_j$,

$$u_j = \begin{pmatrix} X_j \\ |X_j|^r - m(X,r) \\ \frac{p - \mathbb{1}_{(X_j \le q_X(p))}}{f_X(q_X(p))} \end{pmatrix}.$$

We need to verify that u_j fulfils (4.7): $\mathbb{E}[u_j] = 0$ holds by construction. $\mathbb{E}[|X_j|^{2r}] < \infty$ is guaranteed since $|X_t|^r$ satisfies a CLT (see Step 2), thus also $\mathbb{E}[u_j^2] < \infty$. Finally $X_j = f(\varepsilon_j, \varepsilon_{j-1}, ...)$ follows from Lemma 1 in [89], as we assume (A). Hence, this latter relation also holds for functionals of X_j, so for u_j, i.e. (4.8) is fulfilled.

Then, we define a Δ-dependent approximation $u_0^{(\Delta)}$ satisfying (4.9) and (4.10). Denote, for the ease of notation, $X_{0\Delta} := \mathbb{E}[X_0|\mathcal{F}_{-\Delta}^{+\Delta}]$, and set $u_0^{(\Delta)} = \begin{pmatrix} X_{0\Delta} \\ \mathbb{E}[|X_0|^r|\mathcal{F}_{-\Delta}^{+\Delta}] - m(X,r) \\ \frac{p - \mathbb{1}_{(X_{0\Delta} \le q_X(p))}}{f_X(q_X(p))} \end{pmatrix}$ with $\mathcal{F}_s^t = \sigma(\varepsilon_s, ..., \varepsilon_t)$ for $s \le t$. Thus, (4.9) is fulfilled by construction. Let us verify (4.10). We can write

$$\begin{aligned}
&\sum_{\Delta \ge 1} \|u_0 - u_0^{(\Delta)}\|_2 \\
&= \sum_{\Delta \ge 1} \mathbb{E}\left[(X_0 - X_{0\Delta})^2 + \left(|X_0|^r - \mathbb{E}[|X_0|^r|\mathcal{F}_{-\Delta}^{+\Delta}] \right)^2 + \frac{1}{f_X^2(q_X(p))} \left(-\mathbb{1}_{(X_0 \le q_X(p))} + \mathbb{1}_{(X_{0\Delta} \le q_X(p))} \right)^2 \right]^{1/2} \\
&\le \sum_{\Delta \ge 1} \left(\|X_0 - X_{0\Delta}\|_2 + \| |X_0|^r - \mathbb{E}[|X_0|^r|\mathcal{F}_{-\Delta}^{+\Delta}] \|_2 + \frac{1}{f_X(q_X(p))} \left\| -\mathbb{1}_{(X_0 \le q_X(p))} + \mathbb{1}_{(X_{0\Delta} \le q_X(p))} \right\|_2 \right).
\end{aligned} \tag{4.13}$$

Obviously, a sufficient condition for (4.13) is the finiteness of its summands. 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 $\sum_{\Delta \ge 1} \|X_0 - X_{0\Delta}\|_2 < \infty$.

The condition of geometric L_2-NED of X_0 and $|X_0^r|$ is satisfied, on the one hand in the polynomial case under $(M_r), (A)$ and $(P_{\max(1,r/\delta)})$ via Corollary 2 in [89], on the other hand in the exponential case under $(M_r), (A)$ and (L_r) via Corollary 3 in [89]. Thus, as X_0 is geometric L_2-NED this follows also for its bounded functional $\mathbb{1}_{(X_0 \le q_X(p))}$ using Lemma 3.5 in [121] as we showed already in Step 1 that this functional satisfies the variation condition (note that the result in the case of an indicator function goes back to [105]).

Step 4: Multivariate FCLT

Having checked the conditions for the FCLT of Lemma 4.9 in Step 3, we can apply a trivariate FCLT for u_j:

Using the Bahadur representation (4.12) of the sample quantile (ignoring the rest term for the moment), we can state:

$$\sqrt{n}\frac{1}{n}\sum_{j=1}^{[nt]} u_j = \sqrt{n}\, t \begin{pmatrix} \bar{X}_{[nt]} \\ \frac{1}{[nt]}\sum_{j=1}^{[nt]}|X_j|^r - m(X,r) \\ \frac{p-F_{[nt]}(q_X(p))}{f_X(q_X(p))} \end{pmatrix} \overset{D_3[0,1]}{\to} \mathbf{W}_{\tilde{\Gamma}^{(r)}}(t) \quad \text{as } n \to \infty, \tag{4.14}$$

where $\mathbf{W}_{\tilde{\Gamma}^{(r)}}(t), t \in [0,1]$ is the 3-dimensional Brownian motion with covariance matrix $\tilde{\Gamma}^{(r)} \in \mathbb{R}^{3\times 3}$, i.e. the components $\tilde{\Gamma}^{(r)}_{ij}, 1 \le i,j \le 3$, satisfy the same dependence structure as for the random vector $(U,V,W)^T$ described in (D), with all series being absolutely convergent. By the multivariate Slutsky theorem, we can add $\begin{pmatrix} 0 \\ 0 \\ R_{[nt],p} \end{pmatrix}$ to the asymptotics in (4.14) without changing the resulting distribution (as $\sqrt{n}R_{[nt],p} \overset{p}{\to} 0$). Hence, as $n \to \infty$,

$$\sqrt{n}\, t \begin{pmatrix} \bar{X}_{[nt]} - \mu \\ \frac{1}{[nt]}\sum_{j=1}^{[nt]}|X_j|^r - m(X,r) \\ \frac{p-F_{[nt]}(q_X(p))}{f_X(q_X(p))} \end{pmatrix} + \sqrt{n}\, t \begin{pmatrix} 0 \\ 0 \\ R_{[nt],p} \end{pmatrix} = \sqrt{n}\, t \begin{pmatrix} \bar{X}_{[nt]} - \mu \\ \frac{1}{[nt]}\sum_{j=1}^{[nt]}|X_j|^r - m(X,r) \\ q_{[nt]}(p) - q_X(p) \end{pmatrix} \overset{D_3[0,1]}{\to} \mathbf{W}_{\tilde{\Gamma}^{(r)}}(t). \tag{4.15}$$

Recalling the representation of $\hat{m}(X,n,r)$ as in (4.6), we apply to (4.15) the multivariate continuous mapping theorem using the function $f(x,y,z) \mapsto (ax+y,z)$ with $a = -r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]$. Further, by Slutsky's theorem once again, we can add to $ax+y$ a rest of $o_P(1/\sqrt{n})$ without changing the limiting distribution, to obtain, as $n \to \infty$,

$$\begin{aligned} &\sqrt{n}\, t \begin{pmatrix} a(\bar{X}_{[nt]} - \mu) + \frac{1}{[nt]}\sum_{j=1}^{[nt]}|X_j|^r - m(X,r) + o_P(1/\sqrt{n}) \\ q_{[nt]}(p) - q_X(p) \end{pmatrix} \\ &= \sqrt{n}\, t \begin{pmatrix} \hat{m}(X,[nt],r) - m(X,r) \\ q_{[nt]}(p) - q_X(p) \end{pmatrix} \overset{D_2[0,1]}{\to} \mathbf{W}_{\Gamma^{(r)}}(t), \end{aligned} \tag{4.16}$$

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

4.3 Examples

As we have already mentioned in the introduction of this chapter, the family of augmented GARCH(p,q) processes includes different GARCH processes frequently used in application. In contrast to the iid case, for such dependent processes we cannot provide analytical closed form expressions for the correlation of the asymptotic distribution. Thus, we rather focus in this section on reviewing some well-known examples of augmented GARCH(p,q) processes and discuss which conditions these models need to fulfill in order for the bivariate asymptotics of Theorem 4.3 to be valid. Still, the presented selection of augmented GARCH (p,q) processes is not exhaustive.

Recalling Theorem 4.3, note that the moment condition on the innovations, (M_r), the continuity and differentiability conditions, (C_2'), (P), each at $q_X(p)$, and (C_0) at μ for $r = 1$ as well as condition (A), remain the same for the whole class of augmented GARCH processes. But, depending on the specifications of the process, (4.1) and (4.2), the conditions, $(P_{\max(1,r/\delta)})$ for polynomial GARCH or (L_r) for exponential GARCH respectively, translate differently in the various examples. For each of these models, we provide the corresponding volatility equation, (4.2), and the specifications of the functions g_i

and c_j. We consider 10 models which belong to the group of polynomial GARCH ($\Lambda(x) = x^\delta$) and two examples of exponential GARCH ($\Lambda(x) = \log(x)$). Let us start by presenting the volatility equation, (4.2), for each of them in Table 4.1.

Note that in Table 4.1 the specification of g_i is the same for the whole APGARCH family (only the c_j change), whereas for the two exponential GARCH models, it is the reverse. The general restrictions on the parameters are as follows: $\omega > 0, \alpha_i \geq 0, -1 \leq \gamma_i \leq 1, \beta_j \geq 0$ for $i = 1, ..., p$, $j = 1, ..., q$. Further, the parameters in the GJR-GARCH (TGARCH) are denoted with an asterix (with a plus or minus) as they are not the same as in the other models.

TABLE 4.1: Presentation of the volatility equation (4.2) and the corresponding specifications of functions g_i, c_j for selected augmented GARCH models.

	standard formula for $\Lambda(\sigma_t^2)$	corresponding specifications of g_i, c_j in (4.2)
Polynomial GARCH		
APGARCH family	$\sigma_t^{2\delta} = \omega + \sum_{i=1}^{p} \alpha_i \left(\lvert y_{t-i}\rvert - \gamma_i y_{t-i}\right)^{2\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{2\delta}$	$g_i = \omega/p$ and $c_j = \alpha_j \left(\lvert\varepsilon_{t-j}\rvert - \gamma_j \varepsilon_{t-j}\right)^{2\delta} + \beta_j$
AGARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \left(\lvert y_{t-i}\rvert - \gamma_i y_{t-i}\right)^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$	$c_j = \alpha_j \left(\lvert\varepsilon_{t-j}\rvert - \gamma_j \varepsilon_{t-j}\right)^2 + \beta_j$
GJR-GARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{p} \left(\alpha_i^* + \gamma_i^* \mathbb{1}_{(y_{t-i}<0)}\right) y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$	$c_j = \beta_j + \alpha_j^* \varepsilon_{t-j}^2 + \gamma_j^* \max(0, -\varepsilon_{t-j})^2$
GARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$	$c_j = \alpha_j \varepsilon_{t-j}^2 + \beta_j$
ARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2$	$c_j = \alpha_j \varepsilon_{t-j}^2$
TGARCH	$\sigma_t = \omega + \sum_{i=1}^{p} \left(\alpha_i^+ max(y_{t-i}, 0) - \alpha_i^- min(y_{t-i}, 0)\right) + \sum_{j=1}^{q} \beta_j \sigma_{t-j}$	$c_j = \alpha_j \lvert\varepsilon_{t-j}\rvert - \alpha_j \gamma_j \varepsilon_{t-j} + \beta_j$
TSGARCH	$\sigma_t = \omega + \sum_{i=1}^{p} \alpha_i \lvert y_{t-i}\rvert + \sum_{j=1}^{q} \beta_j \sigma_{t-j}$	$c_j = \alpha_j \lvert\varepsilon_{t-j}\rvert + \beta_j$
PGARCH	$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i \lvert y_{t-i}\rvert^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta.$	$c_j = \alpha_j \lvert\varepsilon_{t-j}\rvert^\delta + \beta_j$
VGARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i (\varepsilon_{t-i} + \gamma_i)^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$	$g_i = \omega/p + \alpha_i (\varepsilon_{t-i} + \gamma_i)^2$ and $c_j = \beta_j$
NGARCH	$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i (y_{t-i} + \gamma_i \sigma_{t-i})^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$	$g_i = \omega/p$ and $c_j = \alpha_j (\varepsilon_{t-j} + \gamma_j)^2 + \beta_j$
Exponential GARCH		$c_j = \beta_j$ and
MGARCH	$\log(\sigma_t^2) = \omega + \sum_{i=1}^{p} \alpha_i \log(\varepsilon_{t-i}^2) + \sum_{j=1}^{q} \beta_j \log(\sigma_{t-j}^2)$	$g_i = \omega/p + \alpha_i \log(\varepsilon_{t-i}^2)$
EGARCH	$\log(\sigma_t^2) = \omega + \sum_{i=1}^{p} \alpha_i \left(\lvert\varepsilon_{t-i}\rvert - \mathbb{E}\lvert\varepsilon_{t-i}\rvert\right) + \gamma_i \varepsilon_{t-i} + \sum_{j=1}^{q} \beta_j \log(\sigma_{t-j}^2)$	$g_i = \omega/p + \alpha_i (\lvert\varepsilon_{t-i}\rvert - \mathbb{E}\lvert\varepsilon_{t-i}\rvert) + \gamma_i \varepsilon_{t-i}$

Having introduced these different models and their volatility equation, let us present the authors of as well as giving an explanation of the abbreviations for the different models.

- APGARCH: Asymmetric power GARCH, introduced by Ding et al. in [48]. One of the most general polynomial GARCH models.
- AGARCH: Asymmetric GARCH, defined also by Ding et al. in [48], choosing $\delta = 1$ in APGARCH.
- GJR-GARCH: This process is named after its three authors Glosten, Jaganathan and Runkle and was defined by them in [69]. For the parameters α_i^*, γ_i^* it holds that $\alpha_i^* = \alpha_i(1 - \gamma_i)^2$ and $\gamma_i^* = 4\alpha_i\gamma_i$.
- GARCH: Choosing all $\gamma_i = 0$ in the AGARCH model (or $\gamma_i^* = 0$ in the GJR-GARCH), gives back the well-known GARCH(p,q) process by Bollerslev in [23].

- ARCH: Introduced by Engle in [54]. We recover it by setting all $\gamma_i = \beta_j = 0, \forall i, j$.
- TGARCH: Choosing $\delta = 1/2$ in the APGARCH model leads us the so called threshold GARCH (TGARCH) by Zakoian in [128]. For the parameters α_i^+, α_i^- it holds that $\alpha_i^+ = \alpha_i(1-\gamma_i), \alpha_i^- = \alpha_i(1+\gamma_i)$.
- TSGARCH: Choosing $\gamma_i = 0$ in the TGARCH model we get, as a subcase, the TSGARCH model, named after its authors, i.e. Taylor, [118], and Schwert, [115].
- PGARCH: Another subfamily of the APGARCH processes is the Power-GARCH (PGARCH), also called sometimes NGARCH (i.e. non-linear GARCH) due to Higgins and Bera in [73].
- VGARCH: The volatility GARCH (VGARCH) model by Engle and Ng in [57] is also a polynomial GARCH model but is not part of the APGARCH family.
- NGARCH: This non-linear asymmetric model is due to Engle and Ng in [57], and sometimes also called NAGARCH.
- MGARCH: This model is called multiplicative or logarithmic GARCH and goes back to independent suggestions, in slightly different formulations, of Geweke in [66], Pantula in [102] and Milhøj in [95].
- EGARCH: This model is called exponential GARCH, introduced by Nelson in [98].

As the nesting of the different models presented is not obvious, we give a schematic overview in Figure 4.1.

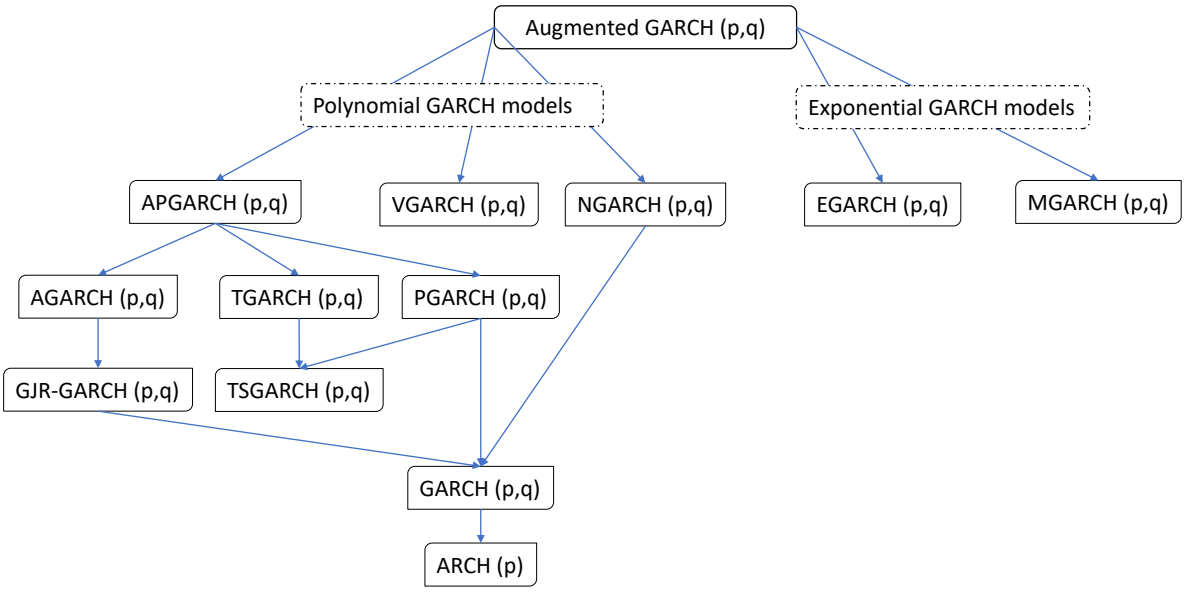


FIGURE 4.1: Schematic overview of the nesting of some augmented GARCH(p,q) models.

In Tables 4.2 we present how the conditions $(P_{\max(1,r/\delta)})$ or (L_r) translate for each model. Table 4.2 treats the specific case of an augmented GARCH(p,q) process with $p = q = 1$, and is presented here (whereas a table for the general case for arbitrary $p \geq 1, q \geq 0$ can be found in the online-appendix, [34]). In the first column we consider the conditions for the general r-th absolute centred sample moment,

$r \in \mathbb{N}$. Of strongest interest to us are the specific cases of the sample MAD ($r = 1$) and the sample variance ($r = 2$) as measure of dispersion estimators respectively, presented in the second and third column.

For the selected polynomial GARCH models the requirement $\sum_{i=1}^{p} \|g_i(\varepsilon_0)\|_{\max(1,r/\delta)} < \infty$ in condition $(P_{\max(1,r/\delta)})$ will always be fulfilled. Thus, we only need to analyse the condition $\sum_{j=1}^{q} \|c_j(\varepsilon_0)\|_{\max(1,r/\delta)} < 1$.

Note that in Table 4.2 the restrictions on the parameter space, given by $(P_{\max(1,r/\delta)})$ or (L_r) respectively, are the same as the conditions for univariate FCLT's of the process X_t^r itself (see [21], [77]). For $r = 1$, they coincide with the conditions for e.g. β-mixing with exponential decay (see [36]).

TABLE 4.2: Conditions $(P_{\max(1,r/\delta)})$ or (L_r) respectively translated for different augmented GARCH(1,1) models. Left column for the general r-th absolute centred sample moment, middle for the MAD ($r = 1$) and right for the variance ($r = 2$).

augmented GARCH (1,1)	$r \in \mathbb{N}$	$r = 1$	$r = 2$
APGARCH	$\mathbb{E}[\lvert\alpha_1 (\lvert\varepsilon_0\rvert - \gamma_1\varepsilon_{t-1})^{2\delta} + \beta_1\rvert^r] < 1$	$\alpha_1 \mathbb{E}\left[(\lvert\varepsilon_0\rvert - \gamma_1\varepsilon_{t-1})^{2\delta}\right] + \beta_1 < 1$	$\mathbb{E}[\lvert\alpha_1 (\lvert\varepsilon_0\rvert - \gamma_1\varepsilon_{t-1})^{2\delta} + \beta_1\rvert^2] < 1$
AGARCH	$\mathbb{E}[\lvert\alpha_1 (\lvert\varepsilon_0\rvert - \gamma_1\varepsilon_{t-1})^2 + \beta_1\rvert^r] < 1$	$\alpha_1 \mathbb{E}\left[(\lvert\varepsilon_0\rvert - \gamma_1\varepsilon_{t-1})^2\right] + \beta_1 < 1$	$\mathbb{E}[\lvert\alpha_1 (\lvert\varepsilon_0\rvert - \gamma_1\varepsilon_{t-1})^2 + \beta_1\rvert^2] < 1$
GJR-GARCH	$\mathbb{E}[\lvert\alpha_1^*\varepsilon_0^2 + \beta_1 + \gamma_1^* \max(0,-\varepsilon_0^2)\rvert^r] < 1$	$\alpha_1^* + \beta_1 + \gamma_1^* \mathbb{E}[\max(0,-\varepsilon_0)^2] < 1$	$\mathbb{E}[\lvert\alpha_1^*\varepsilon_0^2 + \beta_1 + \gamma_1^* \max(0,-\varepsilon_0^2)\rvert^2] < 1$
GARCH	$\mathbb{E}[(\alpha_1\varepsilon_0^2 + \beta_1)^r] < 1$	$\alpha_1 + \beta_1 < 1$	$\alpha_1^2 \mathbb{E}[\varepsilon_0^4] + \alpha_1\beta_1 + \beta_1^2 < 1$
ARCH	$\alpha_1^r \mathbb{E}[\varepsilon_0^{2r}] < 1$	$\alpha_1 < 1$	$\alpha_1^2 \mathbb{E}[\varepsilon_0^4] < 1$
TGARCH	$\mathbb{E}[\lvert\alpha_1\lvert\varepsilon_{t-1}\rvert - \alpha_1\gamma_1\varepsilon_{t-1} + \beta_1\rvert^r] < 1$	$\alpha_1 \mathbb{E}\lvert\varepsilon_{t-1}\rvert + \beta_1 < 1$	$\mathbb{E}[\lvert\alpha_1\lvert\varepsilon_{t-1}\rvert - \alpha_1\gamma_1\varepsilon_{t-1} + \beta_1\rvert^2] < 1$
TSGARCH	$\mathbb{E}[\lvert\alpha_1\lvert\varepsilon_{t-1}\rvert + \beta_1\rvert^r] < 1$	$\alpha_1 \mathbb{E}\lvert\varepsilon_{t-1}\rvert + \beta_1 < 1$	$\mathbb{E}[\lvert\alpha_1\lvert\varepsilon_{t-1}\rvert + \beta_1\rvert^2] < 1$
PGARCH	$\mathbb{E}[\lvert\alpha_1\lvert\varepsilon_0\rvert + \beta_1\rvert^{2r}] < 1$	$\alpha_1 + 2\alpha_1\beta_1 \mathbb{E}\lvert\varepsilon_0\rvert + \beta_1^2 < 1$	$\mathbb{E}[\lvert\alpha_1\lvert\varepsilon_0\rvert + \beta_1\rvert^4] < 1$
VGARCH		for any $r \in \mathbb{N}$: $\beta_1 < 1$	
NGARCH	$\mathbb{E}[\lvert\alpha_1(\varepsilon_0 + \gamma_1)^2 + \beta_1\rvert^r] < 1$	$\alpha_1(1 + \gamma_1^2) + \beta_1 < 1$	$\mathbb{E}[\lvert\alpha_1(\varepsilon_0 + \gamma_1)^2 + \beta_1\rvert^2] < 1$
MGARCH	for any $r \in \mathbb{N}$: $\mathbb{E}[\exp(4r\lvert\omega/p + \alpha_1 \log(\varepsilon_0^2)\rvert^2)] < \infty$ and $\lvert\beta_1\rvert < 1$		
EGARCH	for any $r \in \mathbb{N}$: $\mathbb{E}[\exp(4r\lvert\omega/p + \alpha_1(\lvert\varepsilon_0\rvert - \mathbb{E}\lvert\varepsilon_0\rvert) + \gamma_1\varepsilon_0\rvert^2)] < \infty$ and $\lvert\beta_1\rvert < 1$		

4.4 Conclusion

In this chapter, we showed the bivariate FCLT of the sample quantile with the r-th absolute central sample moment, for augmented GARCH(p,q) processes, answering question **Q4**. With this, we extended the results of Chapter 3 in the iid case. As quantile estimator, we considered only the sample quantile (as a location-scale quantile estimator does not make sense for this class of distributions).

To prove the bivariate asymptotics, we needed to have Bahadur representations for our estimators. In the case of the sample quantile, we showed why we can apply the representation of [121] for augmented GARCH(p,q) processes. For the r-th absolute central sample moment, we established the representation (analogously to the iid case in Chapter 3). We did so for ergodic, stationary, short-memory processes.

With the help of the multivariate FCLT in [11], we then deduced that the conditions of these univariate asymptotics suffice for the joint bivariate asymptotics.

As for this class of processes, the correlation of their asymptotic distributions have no analytical closed form solution, we restricted ourselves, as examples, to describe how the conditions imposed by the asymptotics translate for specific processes of the family of augmented GARCH(p,q) processes.

Having established joint asymptotics for quantile and measure of dispersion estimators for both type of models considered in the empirical study in Chapter 2, enables us to finally tackle in Chapter 5 the pro-cyclicality from a theoretical point of view.

Takeaways

- We establish a Bahadur-like (sum-like) representation for the r-th absolute central sample moment for ergodic, stationary, short-memory processes; see Proposition 4.8.
- We prove a bivariate FCLT between the sample quantile and the r-th absolute central sample moment (which includes the sample MAD and sample variance) for augmented GARCH(p,q) processes; see Theorem 4.3.

Key questions (to be followed up in the thesis)

- What do these results imply for the pro-cyclicality observed in Chapter 2? This is a question we tackle in Chapter 5.
- Can these results be extended for other risk measures? We consider asymptotics for other risk measures in Chapter 5.

Related contributions in the appendix

- A generalization of Table 4.2 for arbitrary $p \geq 1, q \geq 0$ can be found in the online-appendix, [34], Appendix C.

Chapter 5

Pro-cyclicality: Connecting empirics and theory

This chapter formed the basis for the joint work [33].

5.1 Introduction

This chapter links the empirical findings of Chapter 2 with the theoretical results in Chapters 3 and 4. Recall, once again, how the pro-cyclicality of the VaR was assessed in Chapter 2: We considered the correlation between the logarithm of a ratio of VaR estimates and the sample MAD, and showed that it was negative. In a next step, we then provided theoretical results on the asymptotic distribution of quantile estimators with measure of dispersion estimators for iid models (Chapter 3) and for the family of augmented GARCH(p,q) processes (Chapter 4).

Consequently, here we are interested in establishing theoretical results on the pro-cyclicality (or, in other words, the behaviour of the *measure* of pro-cyclicality we used in Chapter 2). Obviously, from a theoretical point of view, there is always the interest in very general and broad results. Here we consider, as already in Chapters 3, 4, other measure of dispersion estimators besides the std or MAD, namely their generalization, the r-th absolute central sample moment. Also, on the side of risk measures, we extend beyond the VaR and take into account ES and expectile estimators. To be clear, this is not done only for the sake of bigger theoretical generality. But because, as already mentioned in the empirical study in Chapter 2, the question naturally arises if such pro-cyclical effects depend on our choice of risk measure (or even risk measure estimator) and also on the measure of dispersion estimator.

Thus, the questions we tackle in this chapter are the following:

Q5 Can we mathematically prove the pro-cyclicality of risk measurements for underlying iid and augmented GARCH(p,q) models?

Q6 How does the degree of pro-cyclicality in such models depend on the choice of risk measure (estimator) and dispersion measure estimator?

Q7 Can we use the results obtained when answering **Q5** to strengthen our claims on the reasons for the pro-cyclicality observed in real data (c.f. question **Q2**)?

Anticipating the results, we show that the strength of pro-cyclicality depends on the choice of risk measure (estimator), the measure of dispersion estimator and the model considered. But, no matter the choices, the pro-cyclicality will always be present.

This chapter has four sections apart from this short introduction. We finish this introduction with the necessary notation and mathematical framework (formalizing the notion of pro-cyclicality in an asymptotic setting), as well as recalling the notions of the three risk measures under consideration, VaR, ES and expectile, and their corresponding estimators. In Section 5.2 we prove the pro-cyclicality of the different

risk and dispersion measures for an underlying iid model. To be able to do so, we first need to establish the joint asymptotics of estimators for risk measures other than VaR, with measure of dispersion estimators. Since the expectile estimator is essentially a VaR estimator, the only case we need to prove are the asymptotics for the ES estimator.

Note that assessing the pro-cyclicality in the iid case is intuitively clear: Recalling Chapter 2, the risk measure estimator at time $t+1$ year is computed on a sample disjoint from the sample for the risk measure estimator at time t (as well as the measure of dispersion estimator at time t). Hence, in an iid sample, those estimators should be uncorrelated. Thus, considering the correlation of the log-ratio of VaR estimators with the measure of dispersion estimator, comes back to the correlation of the logarithm of the VaR at time t with the measure of dispersion estimator at time t (multiplied by a constant). But these asymptotics are known already from Chapter 3. Nevertheless, we will consider this situation more formally, as the reasoning will be the basis for the case of augmented GARCH(p,q) processes.

The pro-cyclicality of the latter is treated in Section 5.3: We first prove a bivariate FCLT for the ES estimator with the measure of dispersion estimator. Then we assess the pro-cyclicality when considering such an augmented GARCH(p,q) process. As we do not have an underlying independent sample, two estimators computed on disjoint samples can still be dependent. But we show that, as in the iid case, asymptotically the risk measure estimator at time $t+1$ year will be uncorrelated to the risk measure estimator (and the measure of dispersion estimator) at time t.

The theoretical results are applied in Section 5.4. We compare the pro-cyclicality of the different risk measures with the sample variance and sample MAD. Since only in the iid case (and not for augmented GARCH(p,q) processes) closed-form analytical expressions are available, we focus on the former case, considering the Gaussian and the Student-t distribution as examples. As a second application, we discuss the relevance of these asymptotic results for our empirical results on real data obtained in Chapter 2. Looking at the residuals of the fitted GARCH(1, 1) process (c.f. Chapter 2) on each of the 11 indices considered, we compare their pro-cyclicality to the ones of iid realizations. We show that they are similar (in the sense that the empirical pro-cyclicality of the residuals often falls within the confidence interval of the iid pro-cyclicality). Thus, we provide additional arguments why we can relate the pro-cyclicality observed empirically partly to an intrinsic part as in the iid models and partly to the GARCH effects as claimed in Chapter 2. We conclude in Section 5.5.

Family of Processes Considered

As mentioned, the samples considered will be either realizations from an underlying iid distribution or from augmented GARCH(p,q) processes. For the definition of such an augmented GARCH(p,q) process $X = (X_t)_{t\in\mathbb{Z}}$, recall equations (4.1) and (4.2).

As in Chapters 3 and 4, we need certain conditions on the underlying process for the bivariate asymptotics between quantile estimators and measure of dispersion estimators to hold (in the iid as well as for the augmented GARCH(p,q) case), namely (M_k), (C_0), (C_l'), (C_l), (P) as well as (Lee),(A), (P_r), (L_r).

We used in Chapter 4 the concept of L_p-near-epoch dependence (L_p-NED) (see Definition 4.1). As additional requirement, we will need the concept of strong mixing (to prove a FCLT for the ES and also to establish the pro-cyclicality for augmented GARCH(p,q) models). Denote, for a sequence of random variables $(Z_n)_{n\in\mathbb{Z}}$, the corresponding σ-algebra as $\mathcal{F}_s^t = \sigma(Z_s, ..., Z_t)$ for $s \leq t$. Then, the definition of strong mixing is as follows:

Definition 5.1 (Strong mixing) *Define as measure of dependence, for any integer $n \geq 1$,*

$$\alpha(n) := \sup_{j\in\mathbb{Z}} \sup_{C\in\mathcal{F}_{-\infty}^{j}, D\in\mathcal{F}_{j+n}^{\infty}} |P(C\cap D) - P(C)P(D)|. \tag{5.1}$$

The sequence of rv's $(Z_n)_{n\in\mathbb{Z}}$ is called strongly mixing if $\alpha(n) \to 0$ as $n \to \infty$. 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.

Note that for a stationary time series we can omit in (5.1) the $\sup_{j\in\mathbb{Z}}$ and simply set wlog $j = 0$.

Risk Measures

Further, let us recall the definitions of the risk measures we consider. Value-at-Risk (VaR), one of the most used risk measures, is simply a quantile at a certain level of the underlying distribution (see Definition 2.1). Despite the availability of other approaches, the VaR is still usually estimated on historical data in practice, using the empirical quantile $\widehat{\mathrm{VaR}}_n(p)$ 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 measure, [7], contrary to Expected Shortfall (ES), introduced in slightly different formulations in [8], [7], [1], [111]. ES is defined as follows (e.g. [1]) for a loss random variable L and a level $p \in (0,1)$:

$$\mathrm{ES}_p(L) = \frac{1}{1-p}\int_p^1 q_L(u)du = \mathbb{E}[L|L \geq q_L(p)]. \tag{5.2}$$

While the first equality in (5.2) 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 (5.2) by averaging over k sample quantiles, i.e.

$$\widetilde{\mathrm{ES}}_{n,k}(p) := \frac{1}{k}\sum_{i=1}^{k} q_n(p_i), \tag{5.3}$$

for a specific choice of $p = p_1 < p_2 < \dots < p_k < 1$. This was e.g. proposed in [53] 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 [38] as

$$\widehat{\mathrm{ES}}_n(p) := \frac{1}{n - [np] + 1}\sum_{i=1}^{n} L_i\, \mathbb{1}_{(L_i \geq q_n(p))}. \tag{5.4}$$

It can be seen as a special case of $\widetilde{\mathrm{ES}}_{n,k}(p)$ choosing $k = n - [np] + 1$ and the p_i accordingly.

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 [100] and then as a risk measure in [87]. This risk measure satisfies many favourable properties (in particular for backtesting), making it appealing from a theoretical point of view (see e.g. [19], [18] and references therein) but not (yet?) in practice (see e.g. [53]). It is defined, for a square-integrable loss random variable L and level $p \in (0,1)$, by the following minimiser

$$e_p(L) = \operatorname*{argmin}_{x\in\mathbb{R}} \left(p\,\mathbb{E}[\max(L-x,0)^2] + (1-p)\,\mathbb{E}[\max(x-L,0)^2] \right). \tag{5.5}$$

While a natural estimator for the expectile is the empirical argmax of (5.5), there exists another way to define an estimator of e_p. Recall the relation between an expectile and quantile, see [127]: 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

$$\kappa(p) = \frac{pq_L(p) - \int_{-\infty}^{q_L(p)} x dF_L(x)}{\mathbb{E}[L] - 2\int_{-\infty}^{q_L(p)} x dF_L(x) - (1-2p)q_L(p)}. \tag{5.6}$$

Thus, such a sample quantile estimator for the expectile at level p, exploiting this relation, is denoted as

$$e_n(p) := q_n(\kappa^{-1}(p)). \tag{5.7}$$

As unified notation, representing these risk measures, and their estimators, we introduce, for $i = 1, ..., 4$:

$$\zeta_i(p) = \begin{cases} \mathrm{VaR}_p(L) & \text{for } i = 1, \\ \mathrm{ES}_p(L) & \text{for } i = 2, \\ \sum_{i=1}^{k} \mathrm{Var}_{p_i}(L)/k & \text{for } i = 3, \\ e_p(L) & \text{for } i = 4. \end{cases} \quad \text{with estimators } \zeta_{n,i}(p) = \begin{cases} \widehat{\mathrm{VaR}}_n(p) & \text{for } i = 1, \\ \widehat{\mathrm{ES}}_n(p) & \text{for } i = 2, \\ \widehat{\mathrm{ES}}_{n,k}(p) & \text{for } i = 3, \\ e_n(p) & \text{for } i = 4. \end{cases} \tag{5.8}$$

Setup of Statistical Framework

Lastly, we comment on the statistical framework needed to assess the pro-cyclicality. Following the empirical study developed in Chapter 2, the measure of interest is the linear correlation of the logarithm of a ratio of sample quantiles with the sample MAD ($\hat{\theta}_n$), namely

$$\mathrm{Cor}\left(\log \left| \frac{\widehat{\mathrm{VaR}}_{n,t+1y}(p)}{\widehat{\mathrm{VaR}}_{n,t}(p)} \right|, \hat{\theta}_{n,t} \right). \tag{5.9}$$

Here we extend this setup to a more general choice of dispersion measure and risk measure estimators. As measure of dispersion estimators, we consider the r-th absolute central sample moment, and as risk measures the ones presented in (5.8). For this, we need to introduce a time-series notation of our estimated quantities: Thus, by $\widehat{\mathrm{VaR}}_{n,t}, \widehat{\mathrm{ES}}_{n,t}, \widehat{\mathrm{ES}}_{n,k,t}, e_{n,t}, \zeta_{n,i,t}, \hat{m}(X, n, r, t)$ we denote, corresponding estimators estimated at time t over the last n observations before time t.

Above all, we are interested in the correlation of the asymptotic distribution corresponding to (5.9). But, analogously to the theoretical results in Chapters 3 and 4, we will present asymptotic distributional results from which we can deduce this correlation. Note that by the choice of the sample size n of $n = 252$ (in the empirical study of Chapter 2) in (5.9), the quantile estimator $\widehat{\mathrm{VaR}}_{n,t+1y}(p)$ is computed on disjoint samples with respect to the other two estimators, i.e. $\widehat{\mathrm{VaR}}_{n,t}(p)$ and $\hat{\theta}_{n,t}$.

Thus, some care has to be taken to translate the setting of (5.9) into an asymptotic one (where we let $n \to \infty$). For the asymptotic framework at a fixed time t, consider a sample of overall size n. Then, the trick to have the disjointness of estimators, as in (5.9), is to consider $\widehat{\mathrm{VaR}}_{n/2,t+n/2}(p), \widehat{\mathrm{VaR}}_{n/2,t}(p)$ and $\hat{\theta}_{n/2,t}$, where we assume wlog that $n/2$ is an integer. It means that the VaR and MAD estimators are estimated on a sample of size $n/2$ each.

More generally, we are interested in the joint asymptotic distribution of the log-ratio, i.e. $\log \left| \frac{\zeta_{n/2,i,t+n/2}(p)}{\zeta_{n/2,i,t}(p)} \right|$, with the r-th absolute central sample moment $\hat{m}(X, n/2, r, t)$.

Then, the generalized analogue to (5.9), i.e. the correlation of the *asymptotic distribution* of (these) two quantities, is denoted, to ease and by abuse of notation, as

$$\lim_{n\to\infty} \mathrm{Cor}\left(\log \left| \frac{\zeta_{n/2,i,t+n/2}(p)}{\zeta_{n/2,i,t}(p)} \right|, \hat{m}(X, n/2, r, t) \right), \tag{5.10}$$

for $i = 1, ..., 4$, and any integer $r > 0$. Consequently, in view of the empirical results in Chapter 2, our measure of the pro-cyclicality of risk measure estimators amounts to the degree of negative correlation of (5.10).

A more formal treatment of this will be given in the proofs of Theorem 5.4 (iid case) and Theorem 5.9 (augmented GARCH(p,q) processes).

5.2 Pro-cyclicality in IID models

The aim of this section is to theoretically assess the pro-cyclicality (of risk measure estimators) in iid models, i.e (5.10). For this, we establish the joint asympottics between the log-ratio of risk measure estimators and the r-th absolute centred sample moment estimators. To do so, we first provide, in Section 5.2.1, the joint asymptotics between the risk measure estimators $\zeta_{n,i}(p)$ and $\hat{m}(X,n,r)$. Then, in Section 5.2.2, we use these results to conclude to the sought-after asymptotics and quantify (5.10).

5.2.1 CLT's between risk and dispersion measure estimators

We want to establish bivariate CLT's between $\zeta_{n,i}$ and $\hat{m}(X,n,r)$. Note that most cases are already covered by our results in Chapter 3: The asymptotics for the $\widehat{\text{VaR}}_n(p)$ with $\hat{\mu}(X,n,r)$, are given by Theorem 3.1. As $e_n(p)$, by definition, is a sample quantile at level $\kappa^{-1}(p)$, we can use the same theorem assuming κ is given. Also, $\widehat{\widetilde{\text{ES}}}_{n,k}(p)$ is, for any finite choice of k, an average of k sample quantiles at different levels $p_i, i = 1,...,k$. Thus, its bivariate asymptotics follows from the extension of Theorem 3.1 to a vector of sample quantiles, Theorem 3.4 and the continuous mapping theorem.

Thus, only the case of $\widehat{\text{ES}}_n(p)$ needs to be considered. The approach is the same as in Theorem 3.1, only that $\widehat{\text{VaR}}_n(p)$ is replaced by $\widehat{\text{ES}}_n(p)$, and with it, the conditions required on the underlying distribution slightly change.

Proposition 5.2 *Consider an iid sample with parent rv X having mean μ, variance σ^2. For any integer $r > 0$, assume that (M_r) holds, F_X is absolutely continuous, (C_3) holds in a neighbourhood of $q_X(p)$, and, if $r = 1$, (C_0) at μ and $(M_{1+\delta})$ for some $\delta > 0$ hold. Then the joint asymptotic distribution of the historically estimated expected shortfall $\widehat{\text{ES}}_n(p)$, for $p \in (0,1)$, and the r-th absolute central sample moment $\hat{m}(X,n,r)$, for any integer r, is bivariate normal with the following correlation of the asymptotic distribution (again, by abuse of notation):* $\lim_{n\to\infty} \text{Cor}(\widehat{\text{ES}}_n(p), \hat{m}(X,n,r)) =$

$$\frac{\text{Cov}(\frac{1}{1-p}(X-q_X(p))\mathbb{1}_{(X\geq q_X(p))}, |X-\mu|^r - r\,\mathbb{E}[(X-\mu)^{r-1}\,\text{sgn}(X-\mu)^r](X-\mu))}{\sqrt{\text{Var}\left(|X-\mu|^r - r(X-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\,\text{sgn}(X-\mu)^r]\right)}\sqrt{\text{Var}(\frac{1}{p}(X-q_X(p))\mathbb{1}_{(X\geq q_X(p))})}}. \tag{5.11}$$

Remark 5.3 *Note that the conditions on the underlying distribution are stronger than in the case of the VaR. This comes from the use of the Bahadur representation of this ES estimator. We need absolute continuity of F_X and continuity of the second derivative of f_X in a neighbourhood of $q_X(p)$. In Theorem 3.1, we only needed differentiability of F_X and positivity of f_X at the point $q_X(p)$. Also, in the case of $r = 1$, we have an additional moment condition, which comes from the ES estimator, namely the existence of at least the $2 + 2\delta$th moment. A thorough examination of the proof in [38] (which is set out for strongly mixing time series) should make it possible to reduce the moment condition to (M_1).*

Proof The proof follows the same ideas as the CLT between the sample quantile and the r-th absolute centred sample moment (Theorem 3.1). Only that, instead of using a Bahadur representation for the sample quantile, we use the Bahadur representation for $\widehat{\text{ES}}_n(p)$ from [38]. Since by assumption, F_X is absolutely continuous, (C_3) holds in a neighbourhood of $q_X(p)$, as well as $(M_{1+\delta})$ (or even stronger moment conditions), we can use the ES representation from [38]:

$$\widehat{ES}_n(p) - ES_p(X) = \frac{1}{(1-p)n}\sum_{i=1}^{n}(X_i - q_X(p))\mathbb{1}_{(X_i\geq q_X(p))} - (ES_p(X) - q_X(p)) + o_P(n^{-3/4+\kappa}), \tag{5.12}$$

for an arbitrary $\kappa > 0$.

Accordingly, we know the representation for $\hat{m}(X,n,r)$ from Proposition 3.3. As both, (C_0) at μ for $r=1$ and (M_r) hold, we have, as $n\to\infty$,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i-\bar{X}_n|^r\right)=\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}|X_i-\mu|^r\right)-r\sqrt{n}(\bar{X}_n-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]+o_P(1). \tag{5.13}$$

Using these two representations, we apply the bivariate CLT. By Slutsky's theorem, we know that we can ignore the remainder terms, which converge in probability to 0, as they do not change the limiting distribution. The covariance of the asymptotic distribution then simply equals the covariance of the i-th term of (5.12) and (5.13), respectively,

$$\operatorname{Cov}\left(\frac{1}{1-p}\left(X-q_X(p)\right)\mathbb{1}_{(X\geq q_X(p))},|X-\mu|^r-r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r](X-\mu)\right),$$

which can be simplified in some cases (e.g. location-scale distributions).

As a last step, we need to identify the variances in the asymptotic distribution. The variance for $\hat{\mu}(X,n,r)$ follows from Proposition 3.3, the one for $\widehat{\mathrm{ES}}_n(p)$ from [38]. They are, respectively,

$$\operatorname{Var}\left(|X-\mu|^r-r(X-\mu)\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]\right),$$
$$\operatorname{Var}\left(\frac{1}{1-p}(X-q_X(p))\mathbb{1}_{(X\geq q_X(p))}\right).$$

Hence, the result (5.11) holds. □

5.2.2 Results on pro-cyclicality

After having established the bivariate CLT's between the risk measure estimators and the r-th absolute central sample moment, we are ready to assess the pro-cyclicality as measured in (5.10).

Before stating the proposition, let us come back to the informal explanation given in Section 5.1: Recall that for any risk measure estimator at time $t+n/2$, $\zeta_{n/2,i,t+n/2}(p)$, the sample used is, by construction, disjoint from the sample used at time t. Thus the estimator $\zeta_{n/2,t+n/2,i}(p)$ will be uncorrelated with the r-th absolute centred sample moment $\hat{m}(X,n/2,r,t)$, at time t, as well as with the risk measure estimator $\zeta_{n/2,i,t}(p)$ at time t.

Translating this for the correlation of the asymptotic distribution (again abusing the notation), i.e. (5.10), it should hold, for $i=1,...,4$,

$$\begin{aligned}
&\lim_{n\to\infty}\operatorname{Cor}\left(\log\left|\frac{\zeta_{n/2,i,t+n/2}(p)}{\zeta_{n/2,i,t}(p)}\right|,\hat{m}(X,n/2,r,t)\right)\\
&\quad=\lim_{n\to\infty}\frac{\operatorname{Cov}(-\log|\zeta_{n/2,i,t}(p)|,\hat{m}(X,n/2,r,t))}{\sqrt{2\operatorname{Var}(\log|\zeta_{n/2,i,t}(p)|)}\sqrt{\operatorname{Var}(\hat{m}(X,n/2,r,t))}}\\
&\quad=\frac{-1}{\sqrt{2}}\lim_{n\to\infty}\operatorname{Cor}(\log|\zeta_{n,i,t}(p)|,\hat{m}(X,n,r,t))=\frac{-1}{\sqrt{2}}\left|\lim_{n\to\infty}\operatorname{Cor}(\zeta_{n,i,t}(p),\hat{m}(X,n,r,t))\right|,
\end{aligned}\tag{5.14}$$

where the first equality follows by the uncorrelatedness, the second by the scale invariance of the correlation and the third is a consequence of the Delta-method with the logarithm. But, anticipating the more involved formal treatment needed for augmented GARCH(p,q) processes, we also present the result in the iid case in a precise way.

Theorem 5.4 *Consider a risk measure estimator $\zeta_{n,i}$, $i \in \{1,...,4\}$, and the r-th absolute central sample moment $\hat{m}(X,n,r)$, for a chosen integer $r > 0$. Asumme that the conditions for a bivariate FCLT between these estimators are fulfilled (Theorem 3.1 or Proposition 5.2 respectively).*

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 is bivariate normal too, i.e.

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

and it holds that $\tilde{\Gamma}_{jk} = \begin{cases} \Gamma_{jk}/\zeta_i^2(p) & \text{for } j=k=1, \\ \Gamma_{jk}/2 & \text{for } j=k=2, \text{ In particular, the correlation of this asymptotic} \\ -\Gamma_{jk}/\zeta_i(p) & \text{otherwise.} \end{cases}$
bivariate distribution equals

$$\frac{\tilde{\Gamma}_{12}}{\sqrt{\tilde{\Gamma}_{11}}\sqrt{\tilde{\Gamma}_{22}}} = \frac{-1}{\sqrt{2}}\operatorname{sgn}(\zeta_i(p))\frac{\Gamma_{12}}{\sqrt{\Gamma_{11}}\sqrt{\Gamma_{22}}} = \frac{-1}{\sqrt{2}}\frac{|\Gamma_{12}|}{\sqrt{\Gamma_{11}}\sqrt{\Gamma_{22}}},$$

where Γ is the covariance matrix of the asymptotic bivariate distribution between $\zeta_{n,i}$ and $\hat{m}(X,n,r)$.

To prove this theorem we first present and prove a lemma. This lemma is set in a more general way than the proposition. Then, we will prove the theorem by arguing why the setting of the lemma applies in this case.

Lemma 5.5 *Let $(X_1,...,X_n)$ be an iid sample of copies from a rv X. Assume that, for given functions f and g, we have $\mathrm{Var}(f(X)) < \infty$ and $\mathrm{Var}(g(X)) < \infty$, such that the bivariate CLT holds, i.e.*

$$\sqrt{n}\begin{pmatrix} \sum_{j=1}^n (f(X_j) - \mathbb{E}[f(X_j)])/n \\ \sum_{j=1}^n (g(X_j) - \mathbb{E}[g(X_j)])/n \end{pmatrix} \xrightarrow{d} \mathcal{N}(0,\Gamma), \tag{5.15}$$

for a covariance matrix $\Gamma = (\Gamma_{ij}, 1 \le i,j \le 2)$. Define

$$Q_j = \begin{cases} 0 & \text{for } j \le \lfloor n/2 \rfloor \\ f(X_j) & \text{for } j > \lfloor n/2 \rfloor \end{cases}, Y_j = \begin{cases} f(X_j) & \text{for } j \le \lfloor n/2 \rfloor \\ 0 & \text{for } j > \lfloor n/2 \rfloor \end{cases}, Z_j = \begin{cases} g(X_j) & \text{for } j \le \lfloor n/2 \rfloor \\ 0 & \text{for } j > \lfloor n/2 \rfloor \end{cases}. \tag{5.16}$$

Denote their sample averages (normalized to mean 0) as

$$\bar{Q}_n = \sum_{j=1}^n (Q_j - \mathbb{E}[Q_j])/n, \bar{Y}_n = \sum_{j=1}^n (Y_j - \mathbb{E}[Y_j])/n, \bar{Z}_n = \sum_{j=1}^n (Z_j - \mathbb{E}[Z_j])/n. \tag{5.17}$$

Then, it holds that

$$\sqrt{n}\begin{pmatrix} \bar{Q}_n \\ \bar{Y}_n \\ \bar{Z}_n \end{pmatrix} \xrightarrow{d} \mathcal{N}(0,\Sigma), \tag{5.18}$$

where the covariance matrix Σ satisfies $\Sigma_{ij} = \begin{cases} \Gamma_{11}/2 & \text{for } i=j \in \{1,2\}, \\ \Gamma_{22}/2 & \text{for } i=j=3, \\ \Gamma_{12}/2 & \text{for } i,j \in \{2,3\} \text{ with } i \ne j \\ 0 & \text{otherwise.} \end{cases}$

Proof The proof consists of two steps. As we do not work directly on the X_j's, the first step is to establish univariate CLT's for each of the components of the vector (5.18) using a CLT (Lindeberg-Feller theorem) for independent but not identically distributed rv's. Then, in a second step, we argue why we can deduce the trivariate asymptotics directly via Cramér-Wold.

Step 1: Univariate CLT's

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.

Denote $\mathbb{E}[Q_j] = \mu_j, \mathrm{Var}(Q_j) = \sigma_j^2$ (by assumption, they are finite) and $s_n^2 := \sum_{j=1}^n \sigma_j^2 = n\,\mathrm{Var}(f(X))/2$.

For $\frac{\sum_{j=1}^n (Q_j-\mu_j)}{s_n} \xrightarrow{d} \mathcal{N}(0,1)$ to hold, we need to verify the so called Lindeberg's condition: For all $\varepsilon > 0$, we need to show that

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{j=1}^n \mathbb{E}[(Q_j-\mu_j)^2 \times \mathbb{1}_{(|Q_j-\mu_j|>\varepsilon s_n)}] = 0.$$

In our case, this translates to

$$\begin{aligned}
&\frac{1}{s_n^2} \sum_{j=1}^n \mathbb{E}[(Q_j-\mu_j)^2 \times \mathbb{1}_{(|Q_j-\mu_j|>\varepsilon s_n)}] \\
&\qquad = \frac{2}{n\,\mathrm{Var}(f(X))} \sum_{j=1}^{n/2} \mathbb{E}[(f(X_j)-\mathbb{E}[f(X_j)])^2 \times \mathbb{1}_{\left(|f(X_j)-\mathbb{E}[f(X_j)]|>\varepsilon\, n\,\mathrm{Var}(f(X))/2\right)}] \\
&\qquad = \frac{1}{\mathrm{Var}(f(X))} \mathbb{E}[(f(X)-\mathbb{E}[f(X)])^2 \times \mathbb{1}_{(|f(X)-\mathbb{E}[f(X)]|>\varepsilon\, n\,\mathrm{Var}(f(X))/2)}].
\end{aligned}$$

As $\mathrm{Var}(f(X))$ is finite, we know that $\mathbb{1}_{(|f(X)-\mathbb{E}[f(X)]|/\,\mathrm{Var}(f(X))>\varepsilon n/2)} \underset{n\to\infty}{\to} 0$ almost surely. Further, $(f(X)-\mathbb{E}[f(X)])^2 \times \mathbb{1}_{(|f(X)-\mathbb{E}[f(X)]|>\varepsilon\, n\,\mathrm{Var}(f(X))/2)}$ is dominated by $(f(X)-\mathbb{E}[f(X)])^2$, which by assumption is integrable (as $\mathrm{Var}(f(X)) < \infty$). Thus, by dominated convergence, it follows that

$$\lim_{n\to\infty} \mathbb{E}[(f(X)-\mathbb{E}[f(X)])^2 \times \mathbb{1}_{(|f(X)-\mathbb{E}[f(X)]|>\varepsilon\, n\,\mathrm{Var}(f(X))/2)}] = 0.$$

Thus, $\overline{Q}_n$, defined in (5.17), satisfies $\sqrt{n}\overline{Q}_n \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(f(X))/2)$, i.e. $\Sigma_{11} = \mathrm{Var}(f(X))/2$.

Similarly, we can conclude that $\sqrt{n}\bar{Y}_n \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(f(X))/2)$, i.e. $\Sigma_{22} = \mathrm{Var}(f(X))/2$ and $\sqrt{n}\bar{Z}_n \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(g(X))/2)$, i.e. $\Sigma_{33} = \mathrm{Var}(g(X))/2$.

Step 2: Trivariate CLT

To conclude the trivariate normality, it suffices, using the Cramér-Wold Device, to show that all linear combinations of $\overline{Q}_n, \bar{Y}_n, \bar{Z}_n$ are normally distributed.

For any $a, b, c \in \mathbb{R}$, we establish the CLT for $U_j := a\left(Q_j - \mathbb{E}[Q_j]\right) + b\left(Y_j - \mathbb{E}[Y_j]\right) + c(Z_j - \mathbb{E}[Z_j])$, i.e.

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

with $\sigma^2 = \lim_{n\to\infty} s_n^2/n$ to be determined - analogously to Step 1. Note that $\mathbb{E}[U_j] = 0$ and

$$\begin{aligned}
s_n^2 &= \sum_{j=1}^{n} \operatorname{Var}(U_j) \\
&= \sum_{j=1}^{n} \Big(a^2 \operatorname{Var}(Q_j) + b^2 \operatorname{Var}(Y_j) + c^2 \operatorname{Var}(Z_j) + 2ab \operatorname{Cov}(Q_j, Y_j) + 2ac \operatorname{Cov}(Q_j, Z_j) + 2bc \operatorname{Cov}(Y_j, Z_j) \Big) \\
&= a^2 \frac{n}{2} \operatorname{Var}(f(X)) + b^2 \frac{n}{2} \operatorname{Var}(f(X)) + c^2 \frac{n}{2} \operatorname{Var}(g(X)) + 2bc \frac{n}{2} \operatorname{Cov}(f(X), g(X)) := n\, d(a,b,c),
\end{aligned}$$

where

$$d(a,b,c) := \frac{1}{2} \left(a^2 \operatorname{Var}(f(X)) + b^2 \operatorname{Var}(f(X)) + c^2 \operatorname{Var}(g(X)) \right) + bc \operatorname{Cov}(f(X), g(X)), \tag{5.19}$$

which is finite by assumption. Lindberg's condition is in this case

$$\begin{aligned}
&\frac{1}{s_n^2} \sum_{j=1}^{n} \mathbb{E}[U_j^2 \times \mathbb{1}_{(|U_j| > \varepsilon s_n)}] = \frac{1}{s_n^2} \sum_{j=1}^{n/2} \mathbb{E}[U_j^2 \times \mathbb{1}_{(|U_j| > \varepsilon s_n)}] + \frac{1}{s_n^2} \sum_{j=n/2+1}^{n} \mathbb{E}[U_j^2 \times \mathbb{1}_{(|U_j| > \varepsilon s_n)}] \\
&= \frac{1}{n\, d(a,b,c)} \frac{n}{2} \mathbb{E}[(b\,(f(X) - \mathbb{E}[f(X)]) + c\,(g(X) - \mathbb{E}[g(X)]))^2 \times \mathbb{1}_{|U_j| > \varepsilon n\, d(a,b,c))}] \\
&+ \frac{1}{n\, d(a,b,c)} \frac{n}{2} \mathbb{E}[a^2\,(f(X) - \mathbb{E}[f(X)])^2 \times \mathbb{1}_{(|a(f(X) - \mathbb{E}[f(X)])| > \varepsilon n\, d(a,b,c))}].
\end{aligned}$$

Again, by dominated convergence we can conclude that this quantity converges to zero and thus establish the CLT, i.e.

$$\sqrt{n}\bar{U}_n \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

with $\sigma^2 = d(a,b,c)$. From the knowledge of the univariate asymptotics of Q_j, Y_j and Z_j, respectively, we can deduce from (5.19) that it must hold $\Sigma_{12} = \Sigma_{13} = 0$ and $\Sigma_{23} = \operatorname{Cov}(f(X), g(X))/2$ 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.

As $\Gamma_{11} = \operatorname{Var}(f(X)), \Gamma_{12} = \operatorname{Cov}(f(X), g(X)), \Gamma_{22} = \operatorname{Var}(g(X))$, the claims on the relation of Σ and Γ follow directly. □

Now we can turn to the proof of Theorem 5.4.

Proof The proof consists of two parts. In the first part, we show why we can apply Lemma 5.5 to the setting of Theorem 5.4 to establish trivariate asymptotics.

The second part uses Slutsky's theorem, the Delta method and the continuous mapping theorem to deduce from these trivariate asymptotics the claimed bivariate asymptotics.

Step 1: Applicability of Lemma 5.5

Recall that we already know that, for $i = 1, ..., 4$,

$$\zeta_{n,i}(p) = \frac{1}{n} \sum_{j=1}^{n} (f_i(X_j) - \mathbb{E}[f_i(X_j)]) + o_P(1/\sqrt{n}), \tag{5.20}$$

with the functions being specified as follows:

- For $i = 1$, $f_1(X_j) = \frac{\mathbb{1}_{(X_j > q_X(p))}}{f_X(q_X(p))}$ - which follows from the Bahadur representation of the sample quantile, see (3.18).
- For $i = 2$, $f_2(X_j) = \frac{(X_j - q_X(p))\mathbb{1}_{(X_j > q_X(p))}}{1-p}$ - which follows from the Bahadur representation for $\widehat{ES}_n$, see (5.12).
- For $i = 3$, $f_3(X_j) = \frac{1}{k}\sum_{l=1}^{k} \frac{\mathbb{1}_{(X_j > q_X(p_l))}}{f_X(q_X(p_l))}$ - recalling the definition of the corresponding estimator, (5.3), and using the case $i = 1$.
- For $i = 4$, $f_4(X_j) = \frac{\mathbb{1}_{(X_j > q_X(\kappa^{-1}(p)))}}{f_X(q_X(\kappa^{-1}(p)))}$ - recalling the definition of the corresponding estimator, (5.7), and using the case $i = 1$.

Analogously, we know from (3.11) that

$$\hat{m}(X, n, r) = \frac{1}{n}\sum_{j=1}^{n}(g(X_j) - \mathbb{E}[g(X_j)]) + o_P(1/\sqrt{n}), \tag{5.21}$$

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

Thus, we consider Lemma 5.5 for each choice of f_i, $i = 1, ..., 4$, as defined above, combined with g. We can identify, by our construction

$$\zeta_{n/2,i,t+n/2}(p) - \zeta_i(p) = \bar{Q}_n + o_P(1/\sqrt{n}), \tag{5.22}$$
$$\zeta_{n/2,i,t}(p) - \zeta_i(p) = \bar{Y}_n + o_P(1/\sqrt{n}), \tag{5.23}$$
$$\hat{m}(X, n/2,\, r,\, t) - m(X, r) = \bar{Z}_n + o_P(1/\sqrt{n}), \tag{5.24}$$

using the definitions in (5.16) and (5.17).

By the assumption in Theorem 5.4, the bivariate CLT between $\zeta_{n,i}$ and $m(X, n, r)$ holds. This implies that $\mathrm{Var}(f_i(X)) < \infty$, and $\mathrm{Var}(g(X)) < \infty$ hold (for each $i = 1, ..., 4$). Thus, the conditions of Lemma 5.5 are fulfilled such that (5.18) 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 (5.22)-(5.24) and (5.18), it follows that, as $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \zeta_{n/2,i,t+n/2}(p) - \zeta_i(p) \\ \zeta_{n/2,i,t}(p) - \zeta_i(p) \\ \hat{m}(X, n/2,\, r,\, t) - m(X, r) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{5.25}$$

with the covariance matrix Σ being related to Γ as described in Lemma 5.5. By the multivariate Delta method, we can deduce from (5.25) that, as $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \log|\zeta_{n/2,i,t+n/2}(p)| - \log|\zeta_i(p)| \\ \log|\zeta_{n/2,i,t}(p)| - \log|\zeta_i(p)| \\ \hat{m}(X, n/2,\, r,\, t) - m(X, r) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \tilde{\Sigma}), \tag{5.26}$$

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

Applying the continuous mapping theorem to (5.26) with the function $f(x,y,z) = (x-y,z)$, we obtain

$$\sqrt{n}\begin{pmatrix} \log|\zeta_{n/2,i,t+n/2}(p)| - \log|\zeta_{n/2,i,t}(p)| \\ \hat{m}(X,n/2,\,r,\,t) - m(X,r) \end{pmatrix} \underset{n\to\infty}{\overset{d}{\to}} \mathcal{N}(0,\hat{\Sigma}),$$

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 Lemma 5.5), we see that $\hat{\Sigma}$ equals $\tilde{\Gamma}$ as defined in Theorem 5.4, and thus conclude the proof. □

5.3 Pro-cyclicality in augmented GARCH(p,q) models

As second model, we turn now to assessing the pro-cyclicality for the family of augmented GARCH(p,q) processes, as presented in Chapter 4 (defined in (4.1) and (4.2)). The structure remains the same as in Section 5.2 for iid models.

As a prerequisit to consider the pro-cyclicality, we first establish FCLT's between our risk measure estimators and the r-th absolute central sampe moment. Subsequently, we focus in Section 5.3.2 on the quantification of the pro-cyclicality, (5.10), by establishing a CLT between the log-ratio of the risk measure estimators and the r-th absolute central sample moment for this class of processes.

5.3.1 FCLT's between risk and dispersion measure estimators

We want to establish FCLT's between $\zeta_{n,i}(p)$, $i = 1,...,4$, and $\hat{m}(X,n,r)$. As in the iid case, the bivariate FCLT for the estimator $\widehat{\mathrm{VaR}}_n(p)$ is already stated in Theorem 4.3. This applies also to $e_n(p) = \widehat{\mathrm{VaR}}_n(\kappa^{-1}(p))$ for κ given. Again, Thereom 4.3 can be directly extended to a FCLT for a k-vector of estimators $\widehat{\mathrm{VaR}}_n(p_i), i = 1,...,k$. Applying then the continuous mapping theorem yields the case of $\widehat{\mathrm{ES}}_{n,k}(p)$.

To establish, analogously to Proposition 5.2 in the iid case, the asymptotics with $\widehat{\mathrm{ES}}_n(p)$ we will need here a further dependence condition on the underlying process, namely, strong mixing with a geometric rate (recall Definition 5.1).

To establish the bivariate FCLT for $\widehat{\mathrm{ES}}_n(p)$, we proceed similarly to the case of $\widehat{\mathrm{VaR}}_n(p)$ in Chapter 4 and 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:

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

Using this 4-dimensional vector, we can now describe the joint asymptotic distribution of $\widehat{\mathrm{ES}}_n(p)$ and $\hat{m}(X, n, r)$.

Proposition 5.6 *Consider an augmented GARCH(p,q) process X as defined in (4.1) and (4.2). Introduce the random vector $T_{n,r}(X) = \begin{pmatrix} \widehat{ES}_n(p) - \mathrm{ES}(p) \\ \hat{m}(X, n, r) - m(X, r) \end{pmatrix}$, for an integer $r > 0$. Assume the process satisfies condition (Lee), (C_0) at the mean μ for $r = 1$ and $(M_{r+\delta})$ for some $\delta > 0$. Further, also conditions (M_r), (A), and either $(P_{max(1,r/\delta)})$ for X belonging to the group of polynomial GARCH, or (L_r) for the group of exponential GARCH. Finally, that F_X is absolutely continuous, (C_3) holds in a neighbourhood of $q_X(p)$, and all the 2nd partial derivatives of the joint distribution of (X_1, X_{k+1}), for $k \geq 1$, are bounded in a neighbourhood of $q_X(p)$.*

***If the process is strongly mixing with geometric rate**, we have the following FCLT: For $t \in [0, 1]$, as $n \to \infty$,*

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

where $(\boldsymbol{W}_{\Gamma^{(r)}}(t))_{t\in[0,1]}$ 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 $\mathrm{Cov}(\boldsymbol{W}_{\Gamma^{(r)}}(t), \boldsymbol{W}_{\Gamma^{(r)}}(s)) = \min(s, t)\Gamma^{(r)}$, where

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

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

Remark 5.7 *How restrictive is the condition of strong mixing with geometric rate for the augmented GARCH(p,q) processes? While we cannot give a general result covering all cases, there exist different results in the literature linking GARCH processes and strong mixing: Boussama proves in [27], Theorem 3.4.2, the strong mixing with geometric rate of a GARCH(p,q) process. Carrasco and Chen in [36] 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 strong mixing with geometric rate for the power GARCH(p,q) (PGARCH).*

Remark 5.8 *Comparing the conditions in Proposition 5.6 with those for $\widehat{\mathrm{VaR}}_n(p)$ in Theorem 4.3, 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') and (P) at $q_X(p)$). Also, for $r = 1$, we need $(M_{1+\delta})$ instead of (M_1). These extra conditions are as in the iid case, see Remark 5.3. But in Proposition 5.6, we also need 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 using the Bahadur representation of the ES in [38].*

Proof The proof follows the lines of the corresponding FCLT between the sample quantile and the r-th absolute centred sample moment (Theorem 4.3), also keeping the same structure of the proof in four steps.

Step 1: Bahadur representation of the ES - conditions.

As in the proof of Proposition 5.2, we want to use the Bahadur representation of the ES. It holds under the necessary conditions (i) and (ii) as given in [38], which are fulfilled by assumption:

(i) The process X is strongly mixing with geometric rate.

(ii) The stationarity of the process follows from assumption $(P_{\max(1,r/\delta)})$ or (L_r), respectively, with Lemma 1 of [89]. The conditions on continuity and moments imposed by [38] 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 (X_1, X_{k+1}) for $k \geq 1$.

Thus, we can apply the Bahadur representation of the ES

$$\widehat{ES}_n(p) - ES(p) = \frac{1}{(1-p)n} \sum_{i=1}^{n} (X_i - q_X(p)) \mathbb{1}_{(X_i \geq q_X(p))} - (ES(p) - q_X(p)) + o_P(n^{-3/4+\kappa}), \tag{5.27}$$

for an arbitrary $\kappa > 0$.

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

This step is exactly the same as in the proof of Theorem 4.3.

Step 3: Conditions for applying the FCLT

This step follows closely Step 3 in the proof of Theorem 4.3, adapted to the ES instead of the VaR. Here we are using a four-dimensional version of the FCLT (Lemma 4.9, choosing $d = 4$) - in contrast to a three-dimensional in Theorem 4.3.

Anticipating the use of Lemma 4.9 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 the conditions (4.7) to (4.9) of Lemma 4.9 hold: u_j fulfills (4.7) 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 [89] that $X_j = f(\varepsilon_j, \varepsilon_{j-1}, ...)$. This latter relation also holds for functionals of X_j, i.e. u_j, thus (4.8) holds.

Then, we define a Δ-dependent approximation $u_0^{(\Delta)}$ satisfying (4.9) and (4.10). Denote, for the ease of notation, $X_{0\Delta} := \mathbb{E}[X_0|\mathcal{F}_{-\Delta}^{+\Delta}]$, and set

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

with $\mathcal{F}_s^t = \sigma(\varepsilon_s, ..., \varepsilon_t)$ for $s \leq t$. Thus, (4.9) is fulfilled by construction. Let us verify (4.10). We can write

$$\begin{aligned} \sum_{\Delta \geq 1} \|u_0 - u_0^{(\Delta)}\|_2 \leq \sum_{\Delta \geq 1} \Big(& \|X_0 - X_{0\Delta}\|_2 + \||X_0|^r - \mathbb{E}[|X_0|^r|\mathcal{F}_{-\Delta}^{+\Delta}]\|_2 \\ & + q_X^2(p)) \left\| \mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))} \right\|_2 + \left\| X_0 \mathbb{1}_{(X_0 \geq q_X(p))} - X_{0\Delta} \mathbb{1}_{(X_{0\Delta} \geq q_X(p))} \right\|_2 \Big). \end{aligned} \tag{5.28}$$

Since we have already shown the finiteness for the first three parts of the sum in (5.28) (in Step 3 of the proof of Theorem 4.3), we only need to consider the fourth sum. This follows directly by a small 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

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

Choosing $p = 1 + \delta$, for δ as in Proposition 5.6, $\|X_0\|_{2+2\delta}$ is finite by assumption. Further, note that we can write, for any q,

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

Then, recall that we know from Step 3 in the proof of Theorem 4.3 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 $\kappa > 0$. Thus, $\sum_{\Delta \geq 1} \| \mathbb{1}_{(X_0 \geq q_X(p))} - \mathbb{1}_{(X_{0\Delta} \geq q_X(p))} \|_2^{1/q}$

is finite. Hence, we can conclude

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

which means that (4.10) is fulfilled.

Step 4: Multivariate FCLT

Having checked the conditions for the FCLT of Lemma 4.9 in Step 3, we can apply a 4-dimensional FCLT for u_j

$$\sqrt{n}\frac{1}{n}\sum_{j=1}^{[nt]} u_j = \sqrt{n}\, t \begin{pmatrix} \bar{X}_{[nt]} \\ \frac{1}{[nt]}\sum_{j=1}^{[nt]}|X_j|^r - m(X,r) \\ \frac{q_X(p)}{[nt]}\sum_{j=1}^{[nt]}(\mathbb{1}_{X_j\geq q_X(p))} - (1-p)) \\ \frac{1}{[nt]}\sum_{j=1}^{[nt]}(X_j\mathbb{1}_{X_j\geq q_X(p))} - \mathbb{E}[X_j\mathbb{1}_{X_j\geq q_X(p))}]) \end{pmatrix} \overset{D_4[0,1]}{\to} \mathbf{W}_{\tilde{\Gamma}^{(r)}}(t) \quad \text{as } n\to\infty, \tag{5.29}$$

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 $\tilde{\Gamma}^{(r)}_{ij}, 1\leq i,j\leq 4$, satisfy the dependence structure $(\tilde{D})$, with all series being absolutely convergent.

Recalling the representation of $\hat{m}(X,n,r)$ in Proposition 4.8 and the Bahadur representation (5.27) of the sample ES (ignoring the remainder terms for the moment), we apply to (5.29) the multivariate continuous mapping theorem using the function $f(w,x,y,z)\mapsto(aw+x, b(z-y))$ with $a=-r\,\mathbb{E}[(X-\mu)^{r-1}\operatorname{sgn}(X-\mu)^r]$, $b=1/(1-p)$, and obtain

$$\sqrt{n}\, t \begin{pmatrix} a(\bar{X}_{[nt]}) + \frac{1}{[nt]}\sum_{j=1}^{[nt]}|X_j|^r - m(X,r) \\ \frac{1}{1-p}\left(\frac{1}{[nt]}\sum_{j=1}^{[nt]}\mathbb{1}_{(X_j\geq q_X(p))}(X_j - q_X(p)) - (1-p)(ES_X(p) - q_X(p))\right) \end{pmatrix} \overset{D_2[0,1]}{\to} \mathbf{W}_{\Gamma^{(r)}}(t). \tag{5.30}$$

As by Slutsky's theorem, a remainder term that converges in probability to 0, does not change the limiting distribution, we get from (5.30),

$$\sqrt{n}\, t \begin{pmatrix} \hat{m}(X,[nt],r) - m(X,r) \\ \widehat{ES}_{[nt]}(p) - \mathrm{ES}_X(p) \end{pmatrix} \overset{D_2[0,1]}{\to} \mathbf{W}_{\Gamma^{(r)}}(t),$$

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

5.3.2 Results on pro-cyclicality

Finally, we can consider the pro-cyclicality for augmented GARCH(*p*,*q*) processes. As these processes exhibit dependence, two estimators, even if computed over disjoint samples, might be correlated (in contrast to the iid case). But it turns out that in our specific case the condition of strong mixing with geometric rate will make the estimators on disjoint samples asymptotically uncorrelated. Thus, we recover, structurally, the pro-cyclicality behaviour as in the iid case (recall our informal reasoning, (5.14)). Let us now state, as a theorem, the analogous result to Theorem 5.4.

Theorem 5.9 *Consider an augmented GARCH(p,q) process X as defined in (4.1) and (4.2), a risk measure estimator $\zeta_{n,i}$, $i\in\{1,\dots,4\}$, and the r-th absolute central sample moment $\hat{m}(X,n,r)$, for a given integer $r>0$. Asumme that the conditions for a bivariate FCLT between these estimators are fulfilled (Theorem 4.3 or Proposition 5.6, respectively).*

If, moreover, X ***is strongly mixing with geometric rate and additionally*** *$(M_{r+\delta})$* ***holds for some*** *$\delta > 0$, the asymptotic distribution of the logarithm of the look-forward ratio of the risk measure estimators with the r-th absolute central sample moment is bivariate normal too, i.e.*

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

and it holds that $\tilde{\Gamma} = \begin{cases}\Gamma_{jk}/\zeta_i^2(p) & \text{for } j=k=1,\\ \Gamma_{jk}/2 & \text{for } j=k=2,\\ -\Gamma_{jk}/\zeta_i(p) & \text{otherwise.}\end{cases}$ *In particular, the correlation of this asymptotic bivariate distribution equals to*

$$\frac{\tilde{\Gamma}_{12}}{\sqrt{\tilde{\Gamma}_{11}}\sqrt{\tilde{\Gamma}_{22}}} = \frac{-1}{\sqrt{2}}\operatorname{sgn}(\zeta_i(p))\frac{\Gamma_{12}}{\sqrt{\Gamma_{11}}\sqrt{\Gamma_{22}}} = \frac{-1}{\sqrt{2}}\frac{|\Gamma_{12}|}{\sqrt{\Gamma_{11}}\sqrt{\Gamma_{22}}},$$

where Γ is the covariance matrix of the asymptotic distribution between $\zeta_{n,i}(p)$ and $\hat{m}(X,n,r)$.

Remark 5.10 *Let us comment on the two additional conditions, with respect to those of Theorem 4.3, introduced in the Theorem 5.9, namely the strong mixing with geometric rate and $(M_{r+\delta})$. We need this dependence condition to make sure that the estimators we consider are asymptotically uncorrelated when computed over disjoint samples. The moment condition $(M_{r+\delta})$ comes from the fact that we use a CLT for non-stationary, strong mixing processes ([106], [51]), which requires a stronger condition than the classical (M_r).*

Contrary to these observations, recall that in Proposition 5.6, when establishing a FCLT with $\widehat{\mathrm{ES}}_n(p)$, we already needed the strong mixing condition. Thus, for the estimator $\widehat{\mathrm{ES}}_n(p)$, Theorem 5.9 imposes only in the case $r > 1$ stronger conditions, namely a stronger moment condition, $(M_{r+\delta})$, needed to prove the pro-cyclicality.

As in the iid case, we will establish a slightly more general result in a lemma, on which the proof of the theorem will be based. We present the more general lemma only as the side result, as we are interested specifically in the pro-cyclicality for augmented GARCH(p,q) processes.

Lemma 5.11 *Consider a univariate, stationary stochastic process $(X_j, j \in \mathbb{Z})$. Assume the conditions for the FCLT of Lemma 4.9 to hold such 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.*

$$\sqrt{nt}\begin{pmatrix} \sum_{j=1}^{[nt]}(f(X_j) - \mathbb{E}[f(X_j)])/[nt] \\ \sum_{j=1}^{[nt]}(g(X_j) - \mathbb{E}[g(X_j))/[nt] \end{pmatrix} \xrightarrow{D_2[0,1]} \boldsymbol{W}_\Gamma(t), \quad as \quad n \to \infty, \tag{5.31}$$

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

$$Q_j = \begin{cases} 0 & \text{for } j \le \lfloor n/2\rfloor \\ f(X_j) & \text{for } j > \lfloor n/2\rfloor \end{cases}, Y_j = \begin{cases} f(X_j) & \text{for } j \le \lfloor n/2\rfloor \\ 0 & \text{for } j > \lfloor n/2\rfloor \end{cases}, Z_j = \begin{cases} g(X_j) & \text{for } j \le \lfloor n/2\rfloor \\ 0 & \text{for } j > \lfloor n/2\rfloor \end{cases}. \tag{5.32}$$

Denote their sample averages (normalized to mean 0) as

$$\bar{Q}_n = \sum_{j=1}^{n}(Q_j - \mathbb{E}[Q_j])/n, \quad \bar{Y}_n = \sum_{j=1}^{n}(Y_j - \mathbb{E}[Y_j])/n, \quad \bar{Z}_n = \sum_{j=1}^{n}(Z_j - \mathbb{E}[Z_j])/n. \tag{5.33}$$

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

$$\mathbb{E}[|Q_j - \mathbb{E}[Q_j]|^{2+2\delta}] < \infty, \quad \mathbb{E}[|Y_j - \mathbb{E}[Y_j]|^{2+2\delta}] < \infty, \quad \mathbb{E}[|Z_j - \mathbb{E}[Z_j]|^{2+2\delta}] < \infty \,, \forall j, \tag{5.34}$$

it holds that

$$\sqrt{n}\begin{pmatrix} \bar{Q}_n \\ \bar{Y}_n \\ \bar{Z}_n \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma), \tag{5.35}$$

where the covariance matrix Σ *satisfies* $\Sigma_{ij} = \begin{cases} \Gamma_{11}/2 & \text{for } i = j \in \{1,2\}, \\ \Gamma_{22}/2 & \text{for } i = j = 3, \\ \Gamma_{12}/2 & \text{for } i, j \in \{2,3\} \text{ with } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$

Proof The idea of the proof is the same as in the iid case (this was the goal of providing a proof in the iid case). It consists of two steps. First, we need to establish univariate CLT's for each of the components of the vector in (5.35), using a CLT for non-stationary strongly mixing sequences. Secondly, we argue why we can deduce the trivariate asymptotics directly via Cramér-Wold. 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 [106], [51], 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* $\sigma_n^2 = \mathrm{Var}(\sqrt{n}\bar{W}_n)$*. If the following three conditions hold,*

$$\mathbb{E}[|W_j - E[W_j]|^{2+2\delta}] \leq c, \quad \forall j \tag{5.36}$$

$$\sigma^2 := \lim_n \sigma_n \in (0, \infty) \tag{5.37}$$

$$\sum_{k=0}^{\infty} (k+1)^2 \alpha(k)^{\delta/(4+\delta)} \leq d, \text{ for a finite constant } d \text{ independent of } k, \tag{5.38}$$

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

Let us explain how the stated version differs from [106] and [51]. Note that in [106] a stronger condition than (5.37) is introduced, namely

$$\forall (d_n) \text{ s.t. } d_n \to \infty : \sup_t |\mathrm{Var}(\sqrt{d_n}\frac{1}{d_n}\sum_{j=t}^{t+d_n-1} W_j) - \sigma^2| \to 0, \text{ as } n \to \infty, \tag{5.39}$$

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 $\sigma^2 := \lim_{n\to\infty} \sigma_n$) for *any* sequence $d_n \leq n$ such that $d_n \to \infty$ as $n \to \infty$. To ensure this general convergence, (5.39) 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).

In contrast to this, the version of the CLT in [51] shows that (5.39) is actually superfluous, but at the price of accepting potentially degenerate limiting distributions. As a compromise between the two, we demand (5.37), 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 (5.36) to (5.38) so that we can apply the CLT. First, we note that (5.36) corresponds, in our case, to our assumption (5.34), hence is satisfied. Direct computations lead to (5.37):

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

which is non-degenerate by (5.31).

We are left with showing (5.38). 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. $\alpha_Q(k) \le \alpha(k)$. As we know that X_j is strongly mixing with geometric rate, we have (see (5.1)) that $\alpha_Q(k) \le C\lambda^k$ for some constants $C > 0$ and $\lambda \in (0,1)$, which implies:

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

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

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

Thus, the series is convergent, from which we deduce (5.38). We conclude to the CLT, as $n \to \infty$

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

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

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

where

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

Step 2: Trivariate CLT

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

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

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

i.e.

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

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 (5.36) to (5.38). 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, (5.36) is fulfilled by assumption, (5.34).

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, (5.33) holds by the same argumentation as in the univariate case.

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

$$\begin{aligned}
\sigma_n^2 &= \operatorname{Var}(\sqrt{n}\sum_{j=1}^{n} U_j/n) = \frac{1}{n}\operatorname{Var}(\sqrt{n}\sum_{j=1}^{n}(aQ_j + bY_j + cZ_j)/n) \\
&= a^2\operatorname{Var}(\sqrt{n}\sum_{j=1}^{n} Q_j/n) + b^2\operatorname{Var}(\sqrt{n}\sum_{j=1}^{n} Y_j/n) + c^2\operatorname{Var}(\sqrt{n}\sum_{j=1}^{n} Z_j/n) \\
&+ 2ab\operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n} Q_j/n, \sqrt{n}\sum_{i=1}^{n} Y_i/n) + 2ac\operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n} Q_j/n, \sqrt{n}\sum_{i=1}^{n} Z_i/n) \\
&+ 2bc\operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n} Y_j/n, \sqrt{n}\sum_{i=1}^{n} Z_i/n).
\end{aligned} \tag{5.40}$$

As this expression for σ_n^2 is quite long and some computations will be involved, we split the computation into different parts. First, note that the respective variances in (5.40) are known from the univariate asymptotics:

$$\lim_{n\to\infty}\operatorname{Var}(\sqrt{n}\sum_{j=1}^{n} Q_j/n) = \sigma_Q^2, \quad \lim_{n\to\infty}\operatorname{Var}(\sqrt{n}\sum_{j=1}^{n} Y_j/n) = \sigma_Y^2, \quad \lim_{n\to\infty}\operatorname{Var}(\sqrt{n}\sum_{j=1}^{n} Z_j/n) = \sigma_Z^2. \tag{5.41}$$

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

- Computation of the first covariance of (5.40)

$$\begin{aligned}
&\operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n} Q_j/n, \sqrt{n}\sum_{i=1}^{n} Y_i/n) \\
&\qquad = \frac{1}{n}\sum_{j=n/2+1}^{n}\sum_{i=1}^{n/2}\operatorname{Cov}(f(X_j), f(X_i)) \\
&\qquad = \frac{1}{n}\sum_{j=n/2+1}^{n}\sum_{i=1}^{n/2}\operatorname{Cov}(f(X_{j-i}), f(X_0)) \\
&\qquad = \frac{1}{n}\left(\sum_{k=1}^{n/2} k\operatorname{Cov}(f(X_k), f(X_0)) + \sum_{k=n/2+1}^{n-1}(n-k)\operatorname{Cov}(f(X_k), f(X_0))\right) \\
&\qquad = \frac{1}{n}\left(\sum_{k=1}^{n/2} k\operatorname{Cov}(f(X_k), f(X_0)) + \sum_{k=1}^{n/2-1}(\frac{n}{2}-k)\operatorname{Cov}(f(X_{k+n/2}), f(X_0))\right),
\end{aligned} \tag{5.42}$$

where we used the stationarity of the underlying process X.

To bound the two sums in (5.42), we use covariance bounds provided in [113], Theorem 7.3. We recall them here, for convenience, for a process $(X_j, j \in \mathbb{Z})$:

- *If $f(X_k)$ is $\mathcal{F}_{l+k}^{\infty}$ measurable and $f(X_0)$ is $\mathcal{F}_{-\infty}^{l}$ measurable (for a chosen integer l and $k > 0$),*
- *if $\mathbb{E}[|f(X_0)|^p] < \infty$ and $\mathbb{E}[|f(X_k)|^q] < \infty$ for some $p, q > 1$ s.t. $\frac{1}{p} + \frac{1}{q} < 1$,*
- *if $X_j, j \in \mathbb{Z}$, is strongly mixing, with mixing coefficient $\alpha(k)$*

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

Choosing $q = 2$ and $p = 2 + 2\delta$ (as, by (5.34), those moments will exist), and $l = 0$, we can write the inequality above as

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

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

Recall, as the process is strong mixing with geometric rate, 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 (5.42):

$$\begin{aligned}\sum_{k=1}^{n/2} k\,\mathrm{Cov}(f(X_k), f(X_0)) \leq \sum_{k=1}^{n/2} k|\mathrm{Cov}(f(X_k), f(X_0))| &\leq \sum_{k=1}^{n/2} kM\alpha(k)^{1-\frac{1}{p}-\frac{1}{q}} \\ &\leq MC^{1-\frac{1}{p}-\frac{1}{q}} \sum_{k=1}^{n/2} k\lambda^{k(1-\frac{1}{p}-\frac{1}{q})}.\end{aligned}$$

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

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

we deduce that

$$\lim_n \frac{1}{n} \sum_{k=1}^{n/2} k\,\mathrm{Cov}(f(X_k), f(X_0)) = 0. \tag{5.43}$$

Now we need to look at the second sum of (5.42). We proceed in the same way using the strong mixing rate as well as the covariance bounds:

$$\begin{aligned}\frac{1}{n} \sum_{k=1}^{n/2-1} (\frac{n}{2} - k)\,\mathrm{Cov}(f(X_{k+n/2}), f(X_0)) &\leq \frac{1}{n} \sum_{k=1}^{n/2-1} (\frac{n}{2} - k)|\mathrm{Cov}(f(X_{k+n/2}), f(X_0))| \\ &\leq \frac{1}{n} \sum_{k=1}^{n/2-1} (\frac{n}{2} - k)M\alpha(k+n/2)^{1-\frac{1}{p}-\frac{1}{q}} \\ &\leq \frac{1}{n} \sum_{k=1}^{n/2-1} (\frac{n}{2} - k)M(C\lambda^{k+n/2})^{1-\frac{1}{p}-\frac{1}{q}}. \end{aligned} \tag{5.44}$$

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 (5.44)

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

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

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n/2-1}(\frac{n}{2}-k)\operatorname{Cov}(f(X_{k+n}),f(X_0)) = 0. \tag{5.45}$$

Combining (5.42) with (5.43) and (5.45), we conclude for the first covariance sum of (5.40) that:

$$\lim_n \operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n}Q_j/n, \sqrt{n}\sum_{i=1}^{n}Y_i/n) = 0. \tag{5.46}$$

- Computation of the second covariance of (5.40)

The computation of the limit of the second covariance of (5.40) 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 (5.42) we deduce that

$$\begin{aligned}\operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n}Q_j/n,\ \sqrt{n}\sum_{i=1}^{n}Z_i/n) &= \frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\operatorname{Cov}(Q_j,Z_i) = \cdots \\ &= \frac{1}{n}\left(\sum_{k=1}^{n/2}k\operatorname{Cov}(f(X_k),g(X_0)) + \sum_{k=1}^{n/2-1}(\frac{n}{2}-k)\operatorname{Cov}(f(X_{k+n/2}),g(X_0))\right).\end{aligned} \tag{5.47}$$

The covariance bounds are again applicable. Choosing $p = 2$ and $q = 2+2\delta$, those moments exist by (5.34). Thus, we obtain analogous results to (5.43) and (5.45) and can conclude, as for the first covariance of (5.40), that

$$\lim_{n\to\infty}\operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n}Q_j/n, \sqrt{n}\sum_{i=1}^{n}Z_i/n) = 0. \tag{5.48}$$

- Computation of the third covariance of (5.40)

We are left with

$$\begin{aligned}\operatorname{Cov}(\sqrt{n}\sum_{j=1}^{n}Y_j/n, \sqrt{n}\sum_{i=1}^{n}Z_i/n) &= \frac{1}{n}\sum_{j=1}^{n/2}\sum_{i=1}^{n/2}\operatorname{Cov}(f(X_j),f(X_i)) \\ &= \frac{1}{n}\left(\frac{n}{2}\operatorname{Cov}(f(X_0),f(X_0)) + 2\sum_{i=1}^{n/2-1}(\frac{n}{2}-i)\operatorname{Cov}(f(X_i),f(X_0))\right).\end{aligned} \tag{5.49}$$

Thus, we have, by direct calculation and recalling (5.43), that

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\sum_{j=1}^{n} Y_j/n, \sqrt{n}\sum_{i=1}^{n} Z_i/n) = \mathrm{Var}(f(X_0))/2 + \sum_{i=1}^{\infty} \mathrm{Cov}(f(X_i), f(X_0)). \tag{5.50}$$

Therefore, we can finally compute σ_n^2. We get from (5.50), recalling the expressions for the variances in (5.41) and for the covariances in (5.46), (5.48) and (5.50), that

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

Hence, we have in the limit

$$\lim_{n\to\infty} \sigma_n^2 = a^2\sigma_Q^2 + b^2\sigma_Y^2 + c^2\sigma_Z^2 + 2bc\left(\mathrm{Var}(f(X_0))/2 + \sum_{i=1}^{\infty}\mathrm{Cov}(f(X_i), f(X_0))\right). \tag{5.51}$$

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 (5.51) that it must hold $\Sigma_{12} = \Sigma_{13} = 0$ and $\Sigma_{23} = \mathrm{Var}(f(X_0))/2 + \sum_{i=1}^{\infty}\mathrm{Cov}(f(X_i), f(X_0))$ 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. □

After having proved Lemma 5.11, which was the main work, we can proceed with the proof of Theorem 5.9.

Proof The proof is structurally the same as in the iid case (on purpose, that is why whe chose to proceed this way for the iid case), we only have to update the references to the corresponding ones for augmented GARCH(p,q) processes. Still, we present the proof briefly. It consists of two parts. In the first part, we show that we can apply Lemma 5.11 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.

Step 1: Applicability of Lemma 5.11

Recall that we already know that, for $i = 1, ..., 4$,

$$\zeta_{n,i}(p) = \sum_{j=1}^{n}(f_i(X_j) - \mathbb{E}[f_i(X_j)])/n + o_P(1/\sqrt{n}), \tag{5.52}$$

with the functions specified as follows:

- For $i = 1$, $f_1(X_j) = \frac{\mathbb{1}_{(X_j > q_X(p))}}{f_X(q_X(p))}$ - which follows from the Bahadur representation of the sample quantile, see (4.12).
- For $i = 2$, $f_2(X_j) = \frac{(X_j - q_X(p))\mathbb{1}_{(X_j > q_X(p))}}{1-p}$ - which follows from the Bahadur representation for $\widehat{ES}_n$, see (5.27).
- For $i = 3$, $f_3(X_j) = \frac{1}{k}\sum_{l=1}^{k} \frac{\mathbb{1}_{(X_j > q_X(p_l))}}{f_X(q_X(p_l))}$ - recalling the definition of the corresponding estimator, (5.3), and using the case $i = 1$.
- For $i = 4$, $f_4(X_j) = \frac{\mathbb{1}_{(X_j > q_X(\kappa^{-1}(p)))}}{f_X(q_X(\kappa^{-1}(p)))}$ - recalling the definition of the corresponding estimator, (5.7), and using the case $i = 1$.

Analogously, we know from (4.6) that

$$\hat{m}(X, n, r) = \sum_{j=1}^{n} (g(X_j) - \mathbb{E}[g(X_j)])/n + o_P(1/\sqrt{n}), \tag{5.53}$$

i.e. $g(X_j) = |X_j - \mu|^r - r\,\mathbb{E}[(X-\mu)^{r-1}\,\mathrm{sgn}(X-\mu)^r](X_j - \mu)$.

We know that the representations (5.52) and (5.53) hold as, by assumption in Theorem 5.9, the conditions for the bivariate asymptotics between $\zeta_{n,i}$ and $m(X, n, r)$ are fulfilled.

Then, we consider Lemma 5.11 for each choice of f_i, $i = 1, ..., 4$, as defined above combined with g. We can identify, by our construction,

$$\zeta_{n/2,\,i,\,t+n/2}(p) - \zeta_i(p) = \bar{Q}_n + o_P(1/\sqrt{n}), \tag{5.54}$$

$$\zeta_{n/2,\,i,\,t}(p) - \zeta_i(p) = \bar{Y}_n + o_P(1/\sqrt{n}), \tag{5.55}$$

$$\hat{m}(X, n/2,\, r,\, t) - m(X, r) = \bar{Z}_n + o_P(1/\sqrt{n}), \tag{5.56}$$

using the definitions (5.32) and (5.33). Again, by assumption in Theorem 5.9, the bivariate CLT, i.e. (5.31), between $\zeta_{n,i}$ and $m(X, n, r)$ holds. As the strong mixing and the moment condition, (5.34), hold by assumption too, by Lemma 5.11, the claimed trivariate asymptotics (5.35) hold.

Step 2: Concluding the bivariate asymptotics

This is exactly the same as the Step 2 in the proof of Theorem 5.4, only replacing the use of Lemma 5.5 by Lemma 5.11, and the covariance matrix from the iid case with the one from the GARCH case. □

5.4 Application

In this section we consider two different applications of the theoretical results established on the pro-cyclicality of risk measures in Sections 5.2 and 5.3.

First, in Section 5.4.1, we want to assess the pro-cyclicality, i.e. (5.10), explicitly. This means to compute and compare the pro-cyclicality of five risk measure estimators ($\mathrm{VaR}_n(p), e_n(p), \widetilde{\mathrm{ES}}_{n,k}(p)$ for $k = 4$, $k = 50$ and $k = \infty$) with the two most used central absolute sample moments, the sample MAD ($\hat{m}(X, n, 1)$) and the sample variance ($\hat{m}(X, n, 2)$). In contrast to models from augmented GARCH(p,q) processes, for the iid case the closed form expressions of (5.10) can be computed. Thus, we only consider the latter case and look at, as two exemplary distributions, the Gaussian distribution and a Student distribution with varying degrees of freedom. This way we can compare how the degree of pro-cyclicality varies for different choices of risk measures, dispersion measures and underlying distributions.

As a second application, we use the result on these theoretical pro-cyclicalities for the two models (Section 5.2 and 5.3), to see if we can add evidence to our empirical claims on the pro-cyclicality (measured on stock indices) of Chapter 2. Recall that we claimed that part of the pro-cyclicality in the real data should be due to the GARCH effects (as the pro-cyclicality of simulated GARCH$(1,1)$ values was similar to the one in the real data), while the other part should be due to the very way risk is estimated (as observed in the iid case). From Theorem 5.9, we know that pro-cyclicality in augmented GARCH(p,q) processes is not an artificial artefact. Still, we cannot easily use these results to compute the theoretical pro-cyclicality for such processes (and compare it with the degree of pro-cyclicality in real data).

Instead, we consider the residuals of the GARCH(1,1) process fitted to the data from Chapter 2. If the pro-cyclicality in the data is due to the GARCH effects, the pro-cyclical behaviour of these residuals should be as the one from iid samples. With this procedure we provide an additional, alternative argumentation why the pro-cyclicality effects in the data are related partly to an intrinsic part (as observed in iid models) and partly to the volatility behaviour represented by a GARCH(1,1) model.

5.4.1 Comparing pro-cyclicality in IID models

In the following, we consider the pro-cyclicality as in (5.10) (i.e. the correlation in the asymptotic distribution of the log-ratio of risk measure estimators with measure of dispersion estimators) for underlying iid models. We consider as risk measure estimator one VaR estimator ($\widehat{\mathrm{VaR}}_n(p)$), one expectile estimator ($e_n(p)$) and three ES estimators ($\widetilde{\mathrm{ES}}_{n,4}$, $\widetilde{\mathrm{ES}}_{n,50}$, $\widetilde{\mathrm{ES}}_{n,\infty}$). As measure of dispersion estimator, we focus on the sample MAD ($\hat{m}(X,n,1)$) and the sample variance ($\hat{m}(X,n,2)$). The closed form solutions follow from Theorem 5.4 and the corresponding bivariate CLT's, and their derivations can be found in Appendix D.1. Here we focus on plotting and comparing them.

We start by presenting the results for the Gaussian distribution, $\mathcal{N}(0,1)$, and then the Student-t distributions with ν degrees of freedom, choosing $\nu = 3,4,5,10$ or 50 but always normalized to have mean 0 and variance 1.

Gaussian Distribution In Figure 5.1 we plot the correlations in the asymptotic distribution of the different risk measure estimators with the sample variance (left column) and the sample MAD (right column), respectively. In the second row, we zoom into the tail as, from a risk management point of view, we are interested in the behaviour for high values of p.

Looking at the plots in the first row, we see that we have the same tendencies of the correlation of the asymptotic distribution (for VaR, ES and expectile respectively), irrespectively of the choice of the dispersion measure (left plot with the variance, right with the MAD). Let us take a closer look at the correlation of the asymptotic distribution with the sample variance. VaR and expectile have a similar behaviour, being symmetric around $p = 0.5$ (where the correlation equals zero), then increasing to a maximum (in absolute values) and for tail values again, decreasing in direction of 0 correlation. The ES, being an integral/sum over the VaR, is not symmetric around $p = 0.5$. The location of its zero depends on the estimation method. The correlation increases (in absolute value) from its zero on, until it reaches its maximum for an upper tail value of p, then decreases again when p tends to 1. Further, we see that $\widetilde{\mathrm{ES}}_{n,4}$ is quite different from $\widetilde{\mathrm{ES}}_{n,\infty}$, while $\widetilde{\mathrm{ES}}_{n,50}$ approximates the latter already well. For $p \geq 0.5$, the ES has clearly higher correlation of the asymptotic distribution than the VaR (except in the tail where they are quite similar). The correlation of the asymptotic distribution of the expectile is lower than with VaR and ES, except in the tail where it is highest. For the MAD in the right plot, the same observations hold, only that the maximum value of correlation decreases (slightly) and the location of these maxima is further away from the boundary values of p (especially for the ES estimators).

Looking at the second row of Figure 5.1, we see a zoom of the correlation plots for high values of the quantile level ($p > 0.8$). In the case with the sample variance, we see that for values of $p < 0.97$, in absolute values, the correlation with the expectile is lowest while the one with the ES (irrespective of

the choice of estimator) is the highest. For values further in the tail, the behaviour is inverted and the correlation with the ES and VaR are very similar. Further, all correlations seem to tend to 0 for $p \to 1$. On the right plot, in the case of the MAD as dispersion measure, we see the same behaviour, only that the threshold at which the behaviour is inverted is already at $p = 0.92$.

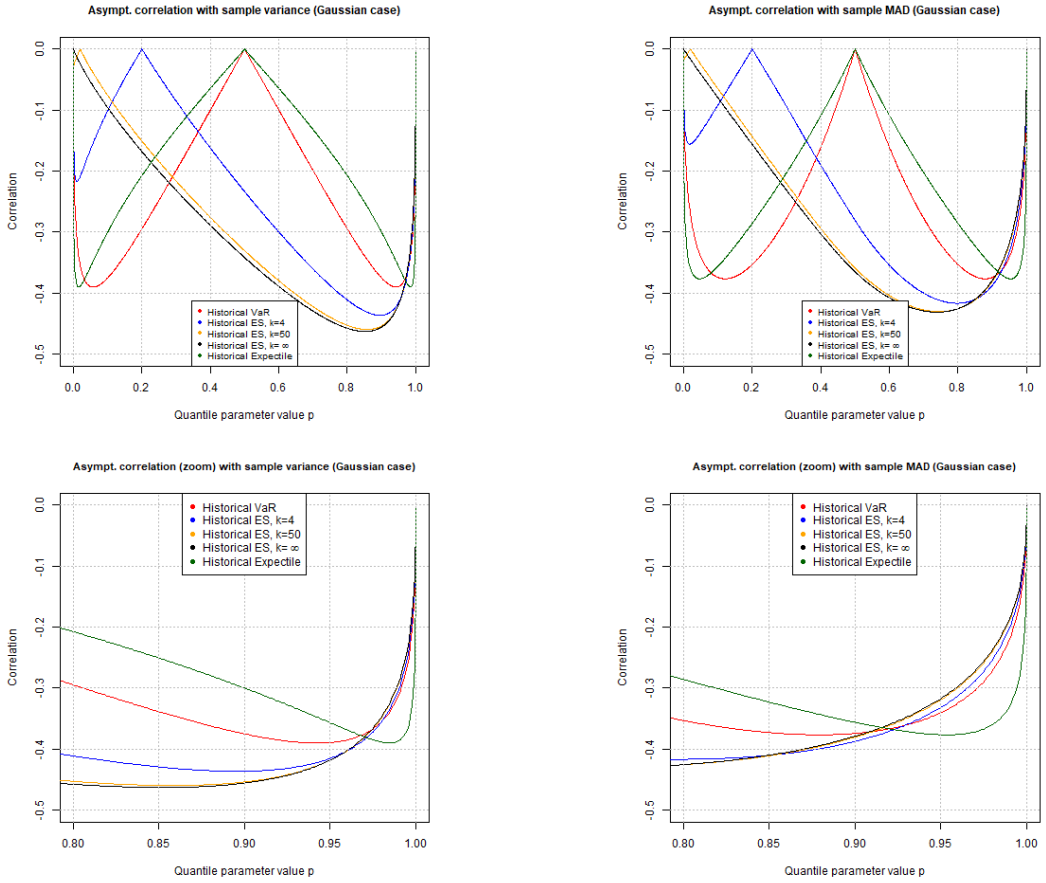


FIGURE 5.1: Pro-cyclicality as defined in (5.10), considering on each plot three different risk measures (VaR, ES, evaluated in 3 possible ways, and expectile). In the left column the measure of dispersion is the sample variance, in the right column the sample MAD. Case of an underlying Gaussian distribution.

Student-t Distribution We start by considering the case $\nu = 5$ in Figure 5.2 since we need $\nu > 4$ for (M_2) to hold. As the behaviour changes with ν, in a second step, we look in Figure 5.3 at the correlations as a function of ν by comparing the cases $\nu = 3, 4, 5, 10, 40$ with the Gaussian limiting case.

Looking first at the correlation with the sample variance in Figure 5.2 (first column), we see, generally speaking, the same trends as in the Gaussian case. However there are three articulate exceptions to that: For $p \geq 0.5$, the correlation with the ES is always higher than with VaR, and with VaR, always higher than with the expectile (in the Gaussian case there was a high threshold for p where this behaviour was inverted). Second, the correlation values with the expectile do not tend to 0 for p tending to 1, but rather seem to converge to a non-zero value. Third, the correlation with $\widetilde{\mathrm{ES}}_{n,50}$ does not apprpoximate the correlation $\widetilde{\mathrm{ES}}_{n,\infty}$ as well as in the Gaussian case.

For the correlation with the MAD (second column of Figure 5.2), we can say as well that the same trends as in the corresponding Gaussian case are visible. But we only share one exception with the case

of the sample variance: The correlation of the expectile tends for p tending to $0,1$ to a non-zero value too.

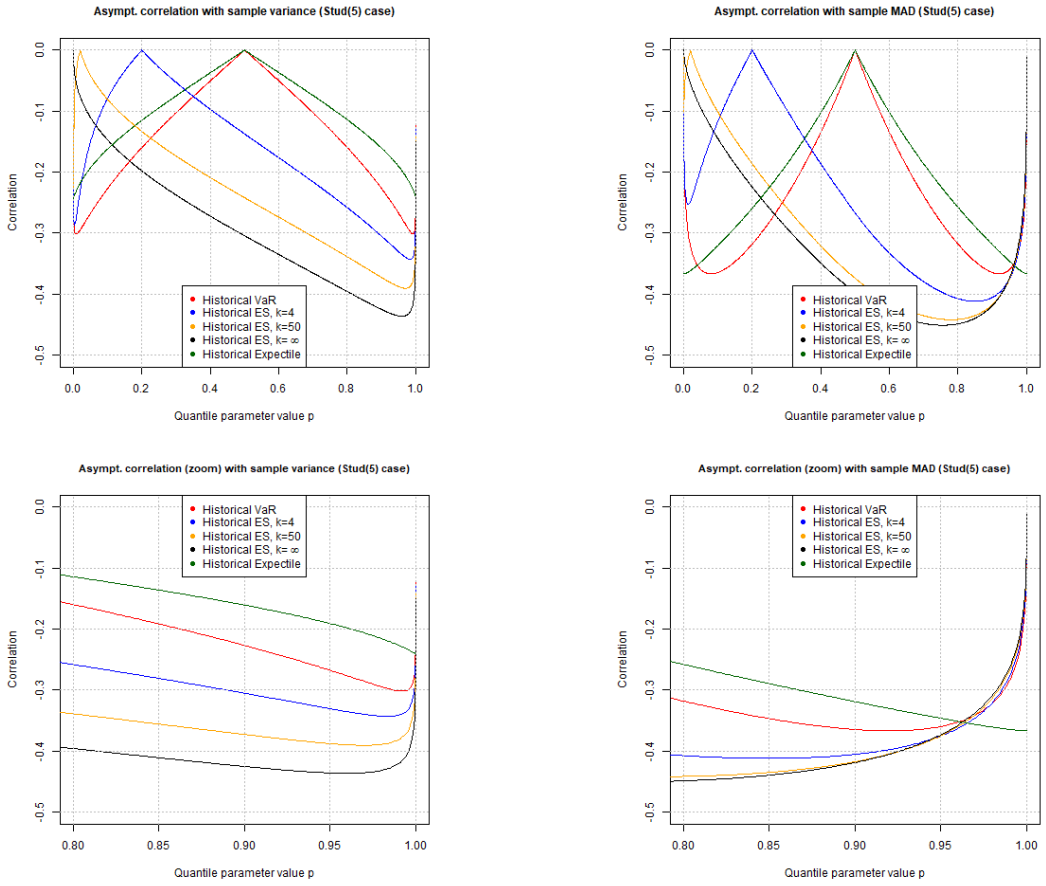


FIGURE 5.2: Pro-cyclicality as defined in (5.10), considering on each plot three different risk measures (VaR, ES, evaluated in 3 possible ways, and expectile). On each row in each plot a different measure of dispersion is considered (from left to right: sample variance, sample MAD. Case of a Student distribution with 5 degrees of freedom.

As mentioned, we also want to study the convergence of the Student correlation to the Gaussian case with respect to the degrees of freedom ν. Thus, we look in Figure 5.3 at the correlation for each pair of risk and dispersion measure separately, but showing the cases $\nu = 3, 4, 5, 10, 40$ and ∞ (Gaussian case) in the same plot.

First, we look at the case with the VaR (first row). For the sample variance (left plot), we see that the convergence, for values near $p = 0.5$ is quicker as for the other intermediate values; near the boundaries it seems to behave as near $p = 0.5$ but this is difficult to assess from the plot. Further, as we already know for the VaR, the behaviour is symmetric around the $p = 0.5$-axis. We observe a similar behaviour with the sample MAD (right plot). But we see that the convergence of the correlation for the variance is slower than for the sample MAD. Further, the convergence with the sample MAD is smoother than with the sample variance. E.g. the shape and values from $\nu = 5$ to $\nu = 10$ change more with the sample variance than with the sample MAD.

Let us now turn to the ES in the second row. Again, we start with the left plot, i.e. the convergence with the sample variance as measure of dispersion. The behaviour of the correlation changes twice. For rather low values of p, the correlation is highest (in absolute terms) for small degrees of freedom, then for

intermediate values of p this is inverted, and again for very high values of p, we have the same behaviour as for low values of p. The speed of convergence varies also with p. In contrast to this, the convergence with the MAD is very uniform. The lower the degree of freedom, the higher the correlation (in absolute terms). The quickest convergence is for values of p between 0.6 and 0.8. As we already know, the behaviour of the ES is not symmetric. To the contrary, the convergence for values of p between 0.1 and 0.3 is even the slowest. The expectile (third row) shows the same characteristics as with the VaR, apart from the fact that the convergence for boundary values of p is the slowest for all values of p.

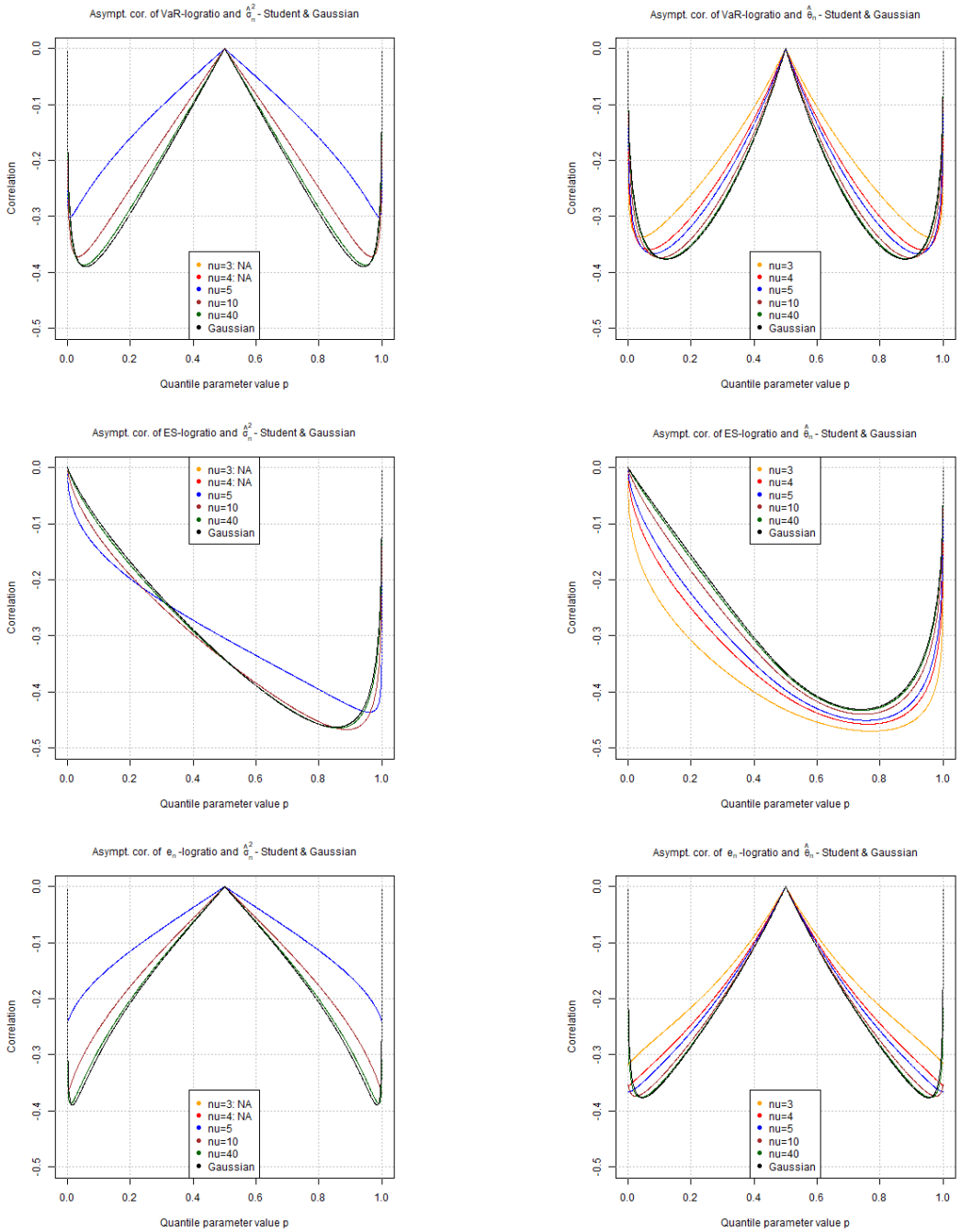


FIGURE 5.3: Pro-cyclicality as defined in (5.10), comparing the case of a Student distribution -with $\nu = 3, 4, 5, 10$ and 40 degrees of freedom as well as the Gaussian distribution. From top to bottom: sample quantile, ES, expectile; and from left to right: sample variance, sample MAD.

Implications of the pro-cyclicality for the choice of risk measure Let us finish 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 figures we have seen that the pro-cyclicality behaviour depends on the choice of underlying risk measure, dispersion measure and also the distribution. Thus, there is not one simple general tendency

to attach to the pro-cyclicality behaviour. Instead, the detailed situation has to be taken into account.

Let us exemplify this in the Gaussian case. *If* one is interested in choosing a risk measure which accentuates the pro-cyclical effect most, from the figures we have seen that the expectile would be the measure of choice - but only for high thresholds. In turn, this exact threshold depends on the corresponding measure of dispersion one is using to measure the pro-cyclcality. For the sample variance the expectile had the highest degree of pro-cyclicality for $p > 0.97$, whereas with the sample MAD this already holds for $p > 0.93$. Below these threshold values the expectile has the lowest degree of pro-cyclicality compared to the other risk measures. Thus, being aware of this threshold value is very important as it might reverse the conclusions! Also, specifically for the expectile, its behaviour is different for heavier tailed distributions. As mentioned, not making it possible, to deduce general tendencies.

When being confronted by the choice of ES or VaR (as these risk measures are more common in practice), one can say that one has, in general, more pro-cyclicality with the ES. But then again, this statement has to be quantified. This is the case for higher, but non-extreme thresholds p. Also, we saw that for heavier tailed distributions this difference was bigger. To the contrary we have seen that in the extreme tails VaR exhibits even slightly more pro-cyclicality than the ES (albeit of the same order).

Thus, to better highlight the effect of pro-cyclicality, the ES is most suited. It has a higher degree of pro-cyclicality than the VaR and in contrast to the expectile its pro-cyclicality behaviour is more consistent. It does not change as drastically (depending on the choice of distribution or measure of dispersion) as the expectile.

5.4.2 Pro-cyclicality analysis on real data (reprise)

In this last part, we want to use the thereotical results on pro-cyclicality to address the empirical claims in Chapter 2, namely that the pro-cyclicality observed is partly from an intrinsic effect of using historical estimation and partly due to the clustering and return-to-the-mean behaviour of volatility, as modeled with a GARCH(1, 1). Thus, it seems logical to use the theoretical results on the pro-cyclicality of augmented GARCH(p,q) processes, Theorem 5.9, to compute the theoretical value for a GARCH(1, 1) process and compare it with the value in the real data.

But there are some fallacies to that. First, for this family of models we do not have closed form solutions of the correlation of the asymptotic distribution. Further, it is known that, for GARCH processes, the convergence to its asymptotic distribution is slow (as e.g. [94] argue for the autocovariance/autocorrelation process). This means that, contrary to the iid case (recall the simulation study in Section 3.6.2), the asymptotic values are not a good approximation for small n. E.g. this can be also seen from the fact that we showed that $\zeta_{n/2,i,t}(p)$ is asymptotically uncorrelated from $\zeta_{n/2,i,t+n/2}(p)$, but this does not hold for the finite sample size considered in the empirical study of Chapter 2.

Thus, we proceed differently in this case. Instead of analysing the theoretical correlation for a GARCH model, we consider the residuals of a GARCH(1, 1) fitted to the data and analyse the pro-cyclicality of this residual process.

Pro-cyclicality Analysis of Residuals To start with, recall the GARCH(1,1) model as in (2.12):

$$\begin{aligned} X_{t+1} &= \varepsilon_t \sigma_t, \\ \text{with } \sigma_t^2 &= \omega + \alpha X_t^2 + \beta \sigma_{t-1}^2 \text{ and } \omega > 0, \alpha \geq 0, \beta \geq 0, \end{aligned}$$

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

For each of the 11 indices we consider the empirical residuals $\hat{\varepsilon}_t := X_{t+1}/\hat{\sigma}_t$. Using the GARCH parameters fitted in Section 2.4.2, we initialize $\hat{\sigma}_t$ by using one year of data (as 'burn-in' sample). Then, to assess the pro-cyclicality of the residuals, we compute the sample correlation between the log-ratio of sample quantiles and the sample MAD as in Section 2.3.2 - but here, on the time-series of residuals $\hat{\varepsilon}_t$

(and not the real data itself!). In theory, this time series of residuals should be iid distributed with mean 0 and variance 1. Hence, using the results of Theorem 5.4, we can exactly assess this pro-cyclicality (i.e. the correlation in the asymptotic ditribution of the SQP-logratio and the MAD) of iid models.

To compare the sample correlation (based on a finite sample) with the theoretical asymptotic value of the correlation, we provide the corresponding confidence intervals for the sample Pearson linear correlation coefficient (as done in the iid simulation study, see Section 3.6.2 for details). We iterate that those confidence interval values have to be considered with care. They are based on assuming to compute a sample correlation on a bivariate normal sample. But the bivariate normality of the log-ratios of sample quantiles with the sample MAD holds only asymptotically. Hence, it is not clear if, for the sample size considered, we can assume bivariate normality (this could be tested). Here, as in the empirical study of Chapter 2, we are computing the sample correlation on a sample of size ~ 300. From the simulation results (available in the online-appendix, [34], Appendix B.2), we can see that, for such a 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 confidence intervals, and how the sample correlation based on the real data (as computed in Table 2.4) behaves in comparison.

Before considering the plots of the pro-cyclicality of the residuals, we should first assess the mean and standard deviation of the residual process as well as the autocorrelation plots of the absolute residuals for each index to see if the iid-assumption can be considered as reasonable. The results can be found in the online-appendix, [34]; as they are positive, we can proceed with our analysis.

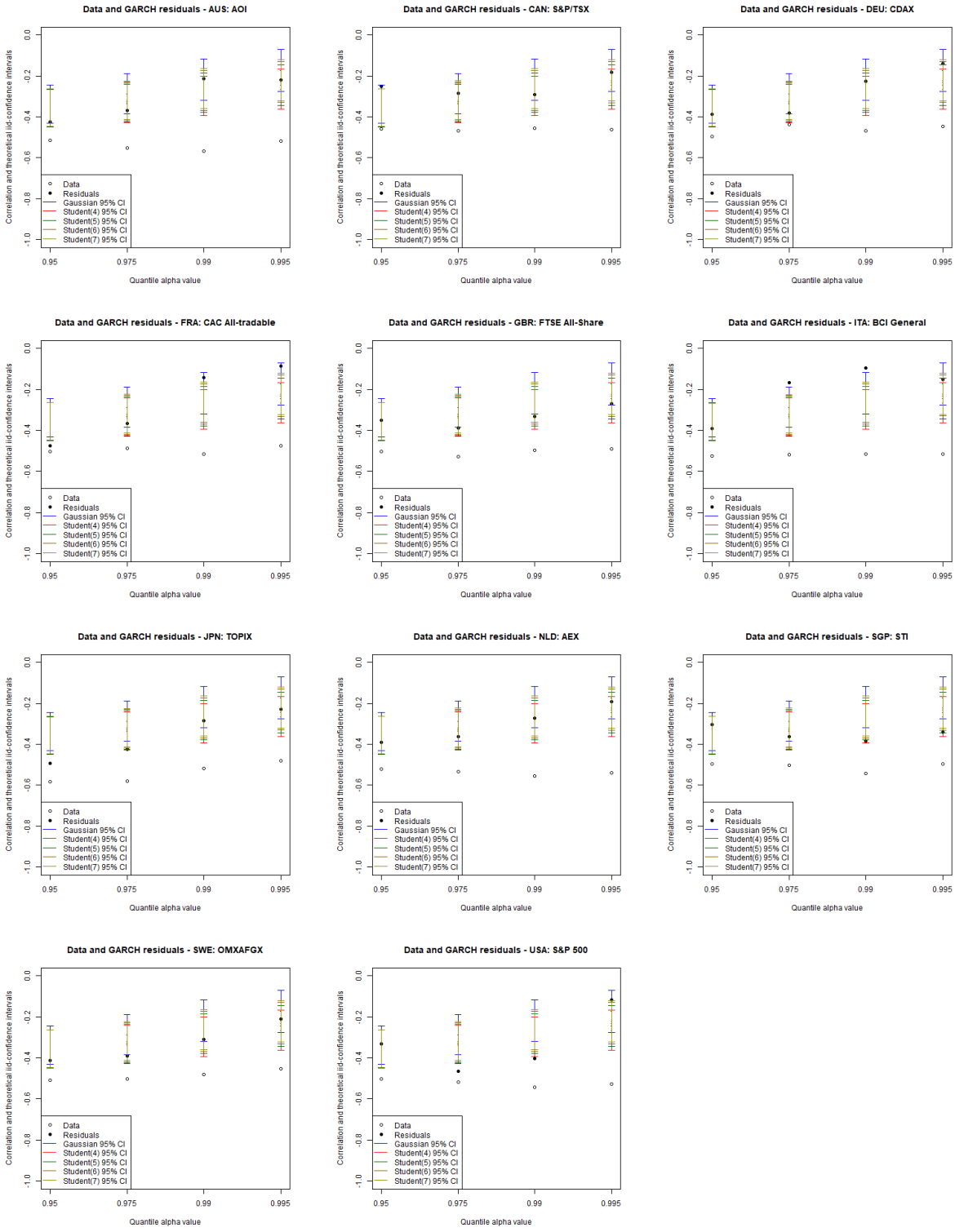


FIGURE 5.4: Comparison of pro-cyclicality in the data (blank circle) with the pro-cyclicality of the GARCH$(1,1)$-residuals (filled circle) for each index separately. Each plot contains the correlation for the four different α values. For each of them, corresponding theoretical confidence intervals (for the sample correlation) assuming a specific underlying distribution (Gaussian or Student with different degrees of freedom) are plotted.

In Figure 5.4 we have one plot for each of the 11 indices. In each plot, we compare for each threshold $\alpha = 0.95, 0.975, 0.99, 0.995$, the measured pro-cyclicality (i.e. the sample correlation between the log-ratio of sample quantiles 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 iid distribution are given - considering as alternatives a Gaussian or Student distribution, the latter with varying degrees of freedom, $\nu = 4, ..., 7$. In 38 out of 44 cases (86%), the sample correlation of the residuals falls in the 95% confidence interval of the sample correlation of an iid distribution. But in none of the cases, the sample correlation of the real data falls in these confidence intervals. Thus, we claim that the pro-cyclical behaviour of the residuals seems to be as the pro-cyclical behaviour of iid random variables. This finally means that stripping-off the GARCH features of the real data by considering its residuals, we are left with a pro-cyclicality behaviour like for iid data. Hence, the claim of Chapter 2 has been backed. Namely, that the correlation in the real data is due to two factors: One, the inherent pro-cyclicality due to the use of historical estimation as modeled with iid rv's, and a second one due to the GARCH effects, i.e. return-to-the-mean and clustering of volatility.

5.5 Conclusion

The goal in this chapter was to link the empirical evidence presented in Chapter 2 with the theoretical results of Chapters 3 and 4: In the empirical study, we considered the sample correlation of a log-ratio of sample quantiles with the sample MAD. But the theoretical results presented in Chapters 3 and 4 treated the (correlation of the) asymptotic distribution between a quantile estimator and a measure of dispersion estimator. Here, we want to assess the pro-cyclicality as measured in Chapter 2.

For this, we first needed to define the pro-cyclicality in an asymptotic sense. Also, we extended the setting beyond the VaR as risk measure, including the ES and expectile (to be able to compare the pro-cyclicality accross different risk measures). As measure of dispersion, as done for the most part in this thesis, we used the r-th absolute central sample moment.

We then started by tackling the pro-cyclicality in iid models. A bivariate CLT between the VaR and expectile estimators could be deduced from the results in Chapter 3, but for the ES we needed to establish such a result. Subsequently, we considered the pro-cyclicality in the iid setting. While the answer seemed intuitively clear (the risk measure estimators are computed on disjoint iid samples, thus are uncorrelated), we treated this formally: We considered sequences which are equal to 0 for half of the sample, and follow the underlying distribution on the other half. In this way, the estimators built on these sequences were uncorrelated. To compute the desired bivariate asymptotic distribution, we then applied a CLT for independently but non-identically distributed sequences. Note that to conclude the pro-cyclicality in an iid setting, we needed no extra conditions compared to the bivariate CLT's between the respective risk and measure of dispersion estimators.

Subsequently, we treated the case of augmented GARCH(p,q) processes, establishing analogous results to the iid case. As additional conditions, we introduced the strong mixing with geometric rate of the underlying process, as well as slightly stronger moment conditions $\Big((M_{r+\delta})$ instead of $(M_r)\Big)$: These conditions were needed, on the one hand for using a Bahadur representation of the ES to establish the FCLT with the r-th absolute moment condition, and on the other hand to prove the pro-cyclicality. As in this case the estimators computed on disjoint samples were not any more uncorrelated a priori, we needed the strong mixing with geometric rate to show that we can bound these covariances. We showed that, asymptotically, they are uncorrelated (i.e. asymptotically we recover structurally the same behaviour as in the iid case).

For both types of models considered, we showed the same results, giving a positive answer to the question **Q5** posed in the introduction: Yes, we can mathematically prove the pro-cyclicality (measured by the negative correlation in the asymptotic distribution of the log-ratio of risk measure estimators with

the r-th absolute central sample moment). Further, our results showed that it will be always present, no matter what the choice of model, risk measure (estimator) or measure of dispersion estimator.

As application of these results, we were interested in comparing the pro-cylicality behaviour for different choices of risk and dispersion measure and underlying models (c.f. question **Q6**). We considered the iid model, as we are able to derive closed form solutions in this case. We compared the pro-cyclicality of VaR, ES and expectile with the sample MAD or sample variance, when considering a Gaussian and Student-t distribution with different degrees of freedom.

The last question we wanted to answer, **Q7**, was what we could deduce from these theoretical findings on pro-cyclicality for our empirically observed pro-cyclicality in Chapter 2. As we did not have closed form solutions for the asymptotics in the GARCH$(1,1)$ case (and the asymptotics do not approximate well the finite sample behaviour), we could not use its theoretical pro-cyclicality results directly. Instead, we considered an alternative approach: We assessed the pro-cyclicality of the residual process of the GARCH$(1,1)$ fitted to the data of Chapter 2. We showed that in most of the cases (86%, i.e. 38 out of 44 cases), the pro-cyclicality of the residuals fell into the 95% confidence bands of the theoretical pro-cyclicality value for Gaussian and Student iid models. In contrast, the pro-cyclicality value of the real data (and not the residuals) did not fall in any of the cases into these confidence bands. We saw this as an alternative and additional way to support the claim that the pro-cyclicality observed on real data is to one part intrinsically due to the way risk is measured historically and to another part due to the volatility effects as modeled by a GARCH$(1,1)$, i.e. the return-to-the-mean and clustering of volatility.

Takeaways

- We formalize the notion of pro-cyclicality of risk measure estimators to a theoretical (asymptotic) setting.
- We show, in this setting, that no matter the choice of risk measure (VaR, ES, expectile) or dispersion measure (r-th absolute central sample moment), one has pro-cyclicality (see Theorem 5.4 for iid and for augmented GARCH(p,q) models, see Theorem 5.9)
- We establish the joint asymptotic normality between the ES estimator (of [38]) and the r-th absolute central sample moment (see Proposition 5.2 for iid and Proposition 5.6 for augmented GARCH(p,q) models)
- As application, we derive explicit expressions for the pro-cyclicality for Gaussian and Student-t iid samples and compare the pro-cyclicality of VaR, ES and expectile when considering the sample variance or MAD as measure of dispersion; see Section 5.4.1.
- We show that the pro-cyclicality behaviour of the residual time series of the GARCH$(1,1)$ process fitted on the real data from Chapter 2, shows a similar behaviour to the pro-cyclicality in iid samples. This way, we provide additional evidence for the claim that the pro-cyclicality observed in the data is partly due to the fact that risk is measured historically and partly due to the clustering and return-to-the mean of volatility; see Section 5.4.2.

Related contributions in the appendix

- In Appendix D.1, we provide (and derive) the explicit formulas for the correlation between the log-ratio of risk measures and the sample variance and sample MAD for a Gaussian and Student-t distribution from an iid sample (corresponding to the figures shown in Section 5.4.1).
- We provide the plots of the ACF of the absolute GARCH$(1,1)$ residuals for each of the 11 indices considered and evaluation of the sample mean and variance of the residuals, in the online-appendix, Appendix D.2.

Chapter 6

Conclusion

6.1 Summary

The empirical and theoretical contributions in this thesis have evolved out of the analysis of pro-cyclicality of historical risk estimation. Namely, the well-known effect that during crisis times future risk is over- and during quiet times underestimated. In this section, we first give an overview over the main results. Then, we comment on possible technical extensions, as, in other settings than ours, there might be a need for different conditions or results. Finally, we present further research perspectives in Section 6.2.

Results

In a first step, we proposed in Chapter 2 a methodology to measure and quantify the effect of pro-cyclicality. We introduced a ratio that quantifies the difference between the historically predicted risk and the estimated realized future risk. We conditioned this indicator of the quality of risk estimation on the current market state (as measured by the realized volatility). Measuring the degree of negative correlation between (the logarithm of) this ratio and the realized volatility, we were able to quantify the pro-cyclicality. Choosing the VaR as risk measure and the MAD as volatility estimator, we found a negative correlation on data from 11 stock indices. A similar behaviour was observed on simulations from a GARCH(1,1) model, but also from an iid model (even though to a lower extent). By comparing the pro-cyclicality in the data with those in these two models, we attributed the pro-cyclicality to two factors: Partially to an intrinsic effect of the historical estimation (as seen in the iid model) and partially to the return-to-the-mean and clustering of volatility (as modeled by a GARCH(1,1)).

In Chapters 3 and 4, we then proved joint asymptotics between quantile and measure of dispersion estimators. Chapter 3 considered the iid case. We proved the joint bivariate normality of two quantile estimators (sample quantile and location-scale quantile estimator) each with both, the sample MedianAD and the r-th absolute central sample moment. As a pre-requisit for the joint asymptotics with the latter, we established an iid-sum representation for the r-th absolute central sample moment (which is new for $r > 2$). We chose two different but consistent quantile estimators to compare their differing behaviour of the correlation in the asymptotic distributions. Moreover, we considered different measure of dispersion estimators as the associated moment and smoothness conditions on the underlying distribution varied between them. While the sample variance needed a finite fourth moment of the underlying distribution, for the sample MAD second moments sufficed and the sample MedianAD had no moment conditions at all. Additionally, in a simulation study we observed a good finite sample performance.

In Chapter 4, we extended the results of Chapter 3 to GARCH processes. Instead of only focussing on the GARCH$(1,1)$, as in the empirical study, we considered the class of augmented GARCH(p,q) for broader mathematical generality. We proved a bivariate FCLT between the sample quantile and the r-th absolute central sample moment. For this to hold, we first established an asymptotic expansion of the r-th absolute central sample moment for ergodic, stationary, short-memory processes (which, again, is new for $r > 2$).

Chapters 3 and 4 are to be seen as a contribution on its own, complementing the statistical literature that has mainly considered asymptotics between quantile and measure of *location* estimators. At the same time, those results layed the foundation to consider the pro-cyclicality from a theoretical point of view in Chapter 5.

In Chapter 5, we formalized the concept of pro-cyclicality as an asymptotic counterpart to Chapter 2: We identified it with the amount of negative correlation in the asymptotic distribution of the log-ratio of risk estimators with the r-th absolute central sample moment. We also enlarged the framework of the previous chapters, considering not only the VaR but also the ES and expectile as risk measures. As a side product, we established joint asymptotics between the ES and the r-th absolute central sample moment. Subsequently, we proved for these different choices of estimators, the joint bivariate asymptotics between the log-ratio of risk measure estimators and measure of dispersion estimators (for an underlying iid sample or an augmented GARCH(p,q) process). Considering this log-ratio of risk measure estimators meant to establish limit results for independent but not identically distributed (for the iid case) and non-stationary processes (for the augmented GARCH(p,q) case). These asymptotic results implied that, no matter the choice of risk measure (estimator), measure of dispersion estimator or underlying model considered, pro-cyclicality always exists.

In the iid case, we computed the explicit expressions of the pro-cyclicality considering different risk and dispersion measure estimators (for Gaussian and Student-t distributions) and compared their degree of pro-cyclicality. Finally, we closed the loop by coming back to the empirical findings in Chapter 2. Using the theoretical results on pro-cyclicality of this chapter, we provided additional evidence to the claims in Chapter 2: The pro-cyclicality observed in the data is partly due to the use of historical risk measure estimation (i.e. has an intrinsic part), and partly due to the GARCH(1,1) volatility effects (i.e. the return-to-the-mean and clustering of volatility).

Technical extensions

Let us comment separately on the technical extensions for the two theoretical parts of the thesis for further inevstigation by interested researchers. First, we present some possible extensions to the pro-cyclicality theorems of Chapter 5. Then, we consider the bivariate asymptotics between quantile and measure of dispersion estimators as in Chapters 3 and 4.

Pro-cyclicality theorems of Chapter 5

We point out two possible extensions. The first aims at sharpening the conditions on the pro-cyclicality results (Theorem 5.4 and 5.9). One could probably get rid of the term $\delta > 0$ in the $(M_{r+\delta})$ moment condition. For that, on the one hand, one would need to make the strong mixing condition in Theorem 5.9 superfluous and assume the L_2-NED only (as the latter holds for the class of augmented GARCH(p,q) processes when assuming (M_r)). On the other hand, one would need to establish the ES representation of [38] with a condition of L_2-NED instead of strong mixing. This could follow the lines of the corresponding results in the case of the VaR (see [121]). Then, using a non-stationary CLT that does not impose strong mixing (as [99]), could lead to the desired conclusion.

The second extension could consist of broadening the limit results to arrays (in the spirit of the Lindeberg-Feller Theorem in the iid case, and following [51], [106] for the class of augmented GARCH(p,q) processes): Instead of considering sample sizes of $n/2$ (each at time t and $t+n/2$), one would consider sample sizes k_n and l_n such that $k_n + l_n = n$, $\forall n$. From an applied point of view, this would reflect the discussion of using longer sample sizes for estimation in the empirical study (e.g. 3 years instead of 1 year to estimate the future risk, while still using a 1 year sample for the realized future risk).

Asymptotic theorems of Chapters 3 and 4

We comment on three possible extensions for the bivariate asymptotics between quantile and measure of dispersion estimators presented in Chapters 3 and 4.

It might be of interest to be able to assess the convergence behaviour of, e.g., the sample correlations instead of having results on the limiting distribution. For this, we would need the stronger results of L_2-convergence for the estimators instead of convergence in distribution. The challenge lies in the remainder terms of the representations of the estimators: The only result known to us, is the L_2-convergence of the remainder term of the sample quantile in the iid case (see [50]). An extension for the dependent case seems quite involved: For the proofs of stochastic convergence of the remainder, one assumes without loss of generality that the rv's follow a standard uniform distribution. As the behavior of the moments might change by such a transformation, a priori, this approach does not go through for L_2-convergence.

The second extension might concern the issue that financial data are known to have long-range dependence, but we considered the family of augmented GARCH(p,q) processes which has short memory. Recall that we were not looking in Chapter 2 for using a process which fits as accurately as possible the data, but one which models the volatility clustering. Still, one might be interested in corresponding results for long-memory processes. A Bahadur representation of the sample quantile for long-range dependent processes exists ([75], [124]), and is possible for the sample MAD (see Remark 3.4 in [116]). Still, it leaves the joint distribution as an open question. In general, we cannot directly apply Slutsky's theorem, as the sum of two stable random variables with different stability parameter α is not stable in general (see e.g. [101]).

Thirdly, the joint asymptotic results we presented, hold for a sample quantile with level $p \in (0,1)$. We made some remarks about the numerical behaviour in the cases p tends to 0 or 1, and linked this to the literature on the independence of sample maximum and sample mean (or sample variance) in the iid case (see [39], [92]). To accomplish extensions to all measures of dispersion considered in the iid case, one could use the results of [123]. In the dependent case, this might be tackled using either the results and techniques in [4] (for m-dependent processes), or relying on [78] (for strongly mixing processes).

All those possible extensions we discussed briefly, are only hints which we thought about as a starting point. Of course, they would need some more investigation by any interested researcher.

6.2 Further research perspectives

In the same way as the thesis was an interplay between empirical and theoretical work, the presented perspectives contain both: Proposals for further empirical investigations on the one hand, as well as theoretical questions on the other hand (that can be seen in relation to the empirical work, or on their own). We start by detailing our first ideas on constructing a counter-cyclical risk measure, which is an ongoing work. Subsequently, we lay out some extensions of the empirical study, which we think would be appealing from a practicioners point of view.

Looking for a counter-cyclical risk measure

In this thesis, we have assessed the pro-cyclicality from an empirical and theoretical perspective answering the questions **Q1**-**Q7** set in the introduction. An additional question, which one may have as follow-up, is how we can overcome the effect of pro-cyclicality itself, **Q8**. For us, it was actually the inverse. This question, and therefore the search for a counter-cyclicaly risk measure, was the main motivation when starting working on this topic. But we realized that we first had to understand the mechanism of the pro-cyclicality before then trying to circumvent it.

Thus, we first aimed at quantifying this effect and determining its main intrinsic factors. Such an analysis should then make it possible, as natural continuation, to design new dynamic risk measures.

They should be simple enough to be of practical use, but also able to properly respond to the various market states in order to avoid pro-cyclicality. This is our ongoing work. In the following, we present two lines of thought, which we want to explore further.

The first one is a natural choice that follows directly from the introduction of the SQP in Chapter 2. In the general definition of the SQP, see (2.4), we can choose freely the random measure μ. Thus, by exploring different choices of μ, we may assess empirically the change in the measured pro-cyclicality. A first example we considered in [28] has been the choice $\mu(s) = |L_s|^p$, for $p = 0, 0.5, 1, 2$, to study the impact of the tails. While the case $p = 0$ recovers the case of the VaR, a choice of $p > 0$ puts more weights on the tail values of the distribution. Following the theoretical and empirical results in this thesis, we believe that we need to include the volatility dynamics (or deviations from its average) into the random measure μ. Finding such an appropriate choice for μ is thus our current goal. Still, whatever the measure, the SQP will always remain a sample quantile, by definition. So playing on μ is not enough to remedy the pro-cyclicality.

Thus, we are investigating in parallel a second approach, completely independent of the SQP framework, which might be used separately or complementary. It is based on the empirical and theoretical observations of the negative correlation between the (logarithm of the) look-forward ratio and the realized volatility estimator. Informally establishing a corresponding regression framework (for which a rigorous and realistic model needs to be set up), we may write:

$$\log\left|\frac{q_{n,t+1y}(p)}{q_{n,t}(p)}\right| = \alpha + \gamma\hat{\theta}_{n,t} + \varepsilon_t,$$

where (ε_t) are the standard errors and we use the sample MAD $\hat{\theta}_{n,t}$ as realized volatility estimator. Thus, for a given value of realized volatility, we may correct our estimated future risk with the help of the regression estimates to have a look-forward ratio near to 1.

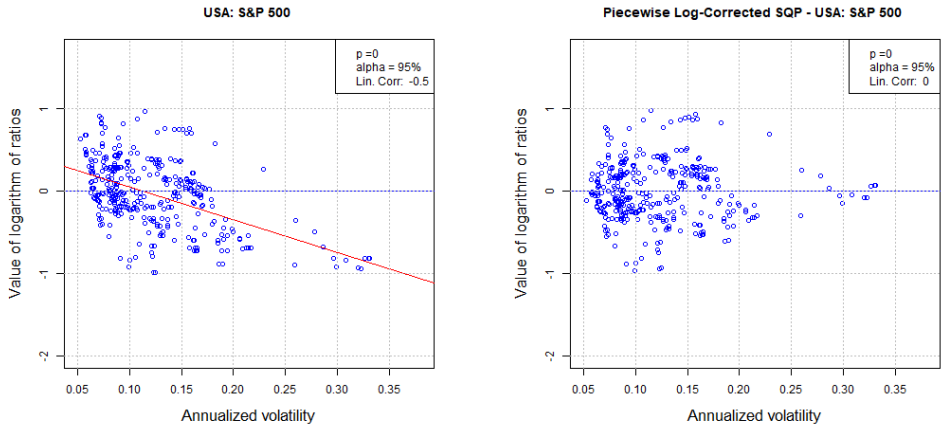


FIGURE 6.1: SQP ratios as a function of volatility (MAD) on real data with $\alpha = 0.95$: On the left original ratios, on the right the ones corrected by the regression model.

First empirical evaluations as in Figure 6.1 (with the *logarithm* of the look-forward ratio), show that the correlation between the corrected log-ratio and realized volatility exactly equals 0 (right plot). Nevertheless, a correction of the log-ratio by the regression model does only work well if the ratios are on average sufficiently close to the linear regression line. But we can see in the left plot of Figure 6.1 that this is not always the case. That is why we can still observe big over- and underestimation even after the correction (that could also be measured by the average SQP-ratio and RMSE). Thus, the correction by

linear regression in this form does not solve the problem, except for the high values of volatility. Further investigation is needed.

Finally, once we will have built a counter-cyclical risk measure, we subsequently will need to establish its corresponding standard statistical and mathematical framework: We need to verify the properties of such a risk measure, as coherence, elicitability, backtesting procedures, ..., and, for its practical use, to prove the consistency and asymptotic normality of its estimator.

Extensions of the empirical study

There are three extensions to the empirical study, which could increase even more its relevance from a practicioners perspective.

Apart from using plain historical simulation, also two of its variants are used in practice. One is filtered historical simulation (FHS) and goes back to [13], [14]: For a given volatility model, it relies on producing, given the information up to time t, a sample path up to time $t+T$. It uses a volatility model for predicting future volatilities, but relies on resampling the residuals (of the sample up to time t) as future innovations. Creating various sample paths, the VaR at time $t+T$ is assessed through its empirical distribution at time $t+T$. Popular volatility models for FHS are either a GARCH$(1,1)$ or an IGARCH$(1,1)$; for the latter see [55]. Using the former model would be similar (but not identical) to the GARCH simulations we did in Chapter 3 (we did not consider the IGARCH in our study). The second approach is the weighted historical simulation which weights the observations of the historical sample before computing its empirical quantile. Two popular variants are either using geometrically decaying weights, [26], or rescaling by a ratio of volatilities, [79]. Hence, the goal here would be to assess the pro-cyclicality in these variants of historical simulation.

Further, we have observed an asymmetric behaviour of the log-ratio with realized volatility on real data. This asymmetry is also present in the simulations with the GARCH but not with an iid model. Thus, the conclusion lies close that it should be related to the volatility process. GARCH models that take into account asymmetry are, inter alia, AGARCH ([48]), GJR-GARCH ([69]) or EGARCH ([98]). While we already know, from the results in Chapter 5, that these models remain pro-cyclical, evaluating its asymmetric behaviour might give additional insights on the reasons for pro-cyclicality.

A last point to note is that the portfolios of a bank or insurance do not equal a stock market index. Thus, given its availability, it would be good to repeat an empirical analysis using such kind of data.

Bibliography

[1] Acerbi, C. and Tasche, D. "On the coherence of expected shortfall". *Journal of Banking & Finance* 26 (7) (2002), pp. 1487–1503.

[2] Akahori, J. "Some Formulae for a New Type of Path-Dependent Option". *Ann. Appl. Probab.* 5 (2) (1995), pp. 383–388.

[3] Alfons, A. and Templ, M. "Estimation of Social Exclusion Indicators from Complex Surveys: The R Package laeken". *Journal of Statistical Software* 54 (15) (2013), pp. 1–25. URL: http://www.jstatsoft.org/v54/i15/.

[4] Anderson, C. and Turkman, K. "The joint limiting distribution of sums and maxima of stationary sequences". *Journal of Applied Probability* 28 (1) (1991), pp. 33–44.

[5] Ardia, D., Bluteau, K., Boudt, K., Catania, L., and Trottier, D.-A. "Markov-switching GARCH models in R: The MSGARCH package". *Journal of Statistical Software, Forthcoming; available at https://ssrn.com/abstract=2845809* (2016).

[6] Arismendi, J. "Multivariate truncated moments". *Journal of Multivariate Analysis* 117 (2013), pp. 41–75.

[7] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. "Coherent measures of risks". *Mathematical Finance* 9 (1999), pp. 203–228.

[8] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. "Thinking coherently". *Risk* 10 (1997), pp. 68–71.

[9] Athanasoglou, P., Daniilidis, I., and Delis, M. "Bank procyclicality and output: Issues and policies". *Journal of Economics and Business* 72 (2014), pp. 58–83.

[10] Aue, A., Berkes, I., and Horváth, L. "Strong approximation for the sums of squares of augmented GARCH sequences". *Bernoulli* 12 (4) (2006), pp. 583–608.

[11] Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. "Break detection in the covariance structure of multivariate time series models". *The Annals of Statistics* 37 (6B) (2009), pp. 4046–4087.

[12] Bahadur, R. "A note on quantiles in large samples". *The Annals of Mathematical Statistics* 37 (3) (1966), pp. 577–580.

[13] Barone-Adessi, G, Burgoin, F., and Giannopoulos, K. "VaR without correlations for nonlinear portfolios". *Risk* 11 (1998), pp. 100–104.

[14] Barone-Adessi, G, Giannopoulos, K., and Vosper, L. "VaR without correlations for nonlinear portfolios". *Journal of Future Markets* 19 (1999), pp. 583–602.

[15] Basel Committee on Banking Supervision. *Amendment to the capital accord to incorporate market risks*. Bank for International Settlements, 1996.

[16] Basel Committee on Banking Supervision. *Minimum capital requirements for market risk*. 2019.

[17] Bec, F. and Gollier, C. "Term Structure and Cyclicity of Value-at-Risk: Consequences for the Solvency Capital Requirement". *CESifo Working Paper No. 2596, available on www.ifo.de/w/47cp8fWqi* (2009).

[18] Bellini, F. and Di Bernardino, E. "Risk management with expectiles". *The European Journal of Finance* 23 (6) (2017), pp. 487–506.

[19] Bellini, F., Klar, B., Müller, A., and Gianin, E. R. "Generalized quantiles as risk measures". *Insurance: Mathematics and Economics* 54 (2014), pp. 41–48.

[20] Bera, A., Galvao, A., Wang, L., and Xiao, Z. "A new characterization of the normal distribution and test for normality". *Econometric Theory* 32 (5) (2016), pp. 1216–1252.

[21] Berkes, I., Hörmann, S., and Horváth, L. "The functional central limit theorem for a family of GARCH observations with applications". *Statistics & Probability Letters* 78 (16) (2008), pp. 2725–2730.

[22] Billingsley, P. *Convergence of probability measures*. 1st. John Wiley & Sons, 1968.

[23] Bollerslev, T. "Generalized autoregressive conditional heteroskedasticity". *Journal of Econometrics* 31 (3) (1986), pp. 307–327.

[24] Bollerslev, T. "Glossary to arch (garch)". *CREATES Research Paper* 49 (2008).

[25] Bos, C. and Janus, P. "A Quantile-based Realized Measure of Variation: New Tests for Outlying Observations in Financial Data". *Tinbergen Institute Discussion Paper 13-155/III* (2013).

[26] Boudoukh, J., Richardson, M., and Whitelaw, R. "The best of both worlds". *Risk* 11 (5) (1998), pp. 64–67.

[27] Boussama, F. "Ergodicité, mélange et estimation dans les modeles GARCH". PhD thesis. Université 7 Paris, 1998.

[28] Bräutigam, M., Dacorogna, M., and Kratz, M. "Predicting risk with risk measures: an empirical study". *ESSEC Working Paper 1803* (2018).

[29] Bräutigam, M., Dacorogna, M., and Kratz, M. "Pro-Cyclicality of Traditional Risk Measurements: Quantifying and Highlighting Factors at its Source". *arXiv:1903.03969* (2019).

[30] Bräutigam, M. and Kratz, M. "Bivariate FCLT for the Sample Quantile and Measures of Dispersion for Augmented GARCH(p,q) processes". *arXiv:1906.09332* (2019).

[31] Bräutigam, M. and Kratz, M. "On the Dependence between Functions of Quantile and Dispersion Estimators". *arXiv:1904.11871* (2019).

[32] Bräutigam, M. and Kratz, M. "On the Dependence between Quantiles and Dispersion Estimators". *ESSEC Working Paper 1807* (2018).

[33] Bräutigam, M. and Kratz, M. "The Impact of the Choice of Risk and Dispersion Measure on Procyclicality". *arXiv:2001.00529* (2020).

[34] Brautigam, M. *Online Appendix: Pro-cyclicality of Risk Measurements - Empirical Quantification and Theoretical Confirmation*. Online Appendix of Doctoral Thesis. 2019. URL: crear.essec.edu/procyclicality/thesis.

[35] Breiman, L. *Probability, volume 7 of classics in applied mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[36] Carrasco, M. and Chen, X. "Mixing and moment properties of various GARCH and stochastic volatility models". *Econometric Theory* 18 (1) (2002), pp. 17–39.

[37] Chen, J. "On Exactitude in Financial Regulation: Value-at-Risk, Expected Shortfall, and Expectiles". *Risks* 6 (2) (2018), pp. 1–29.

[38] Chen, S. "Nonparametric estimation of expected shortfall". *Journal of Financial Econometrics* 6 (1) (2008), pp. 87–107.

[39] Chow, T. and Teugels, J. "The sum and the maximum of iid random variables". In: *Proceedings of the 2nd Prague Symposium on Asymptotic Statistics*. 1978, pp. 81–92.

[40] Christoffersen, P. F. "Evaluating interval forecasts". *International Economic Review* 39 (1998), pp. 841–862.

[41] Dacorogna, M., Gencay, R., Muller, U., Pictet, O., and Olsen, R. *An introduction to High-Frequency Finance*. Academic Press, 2001.

[42] Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4. 2019. URL: https://CRAN.R-project.org/package=xtable.

[43] Daníelsson, J. "The Emperor has no clothes: Limits to risk modelling". *Journal of Banking and Finance* 26 (2002), pp. 1273–1296.

[44] DasGupta, A. and Haff, L. "Asymptotic values and expansions for the correlation between different measures of spread". *Journal of Statistical Planning and Inference* 136 (7) (2006), pp. 2197–2212.

[45] David, H. "Early sample measures of variability". *Statistical Science* (1998), pp. 368–377.

[46] Davidson, J. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford, 1994.

[47] De Rossi, G. and Harvey, A. "Quantiles, expectiles and splines". *Journal of Econometrics* 152 (2) (2009), pp. 179–185.

[48] Ding, Z., Granger, C., and Engle, R. "A long memory property of stock market returns and a new model". *Journal of Empirical Finance* 1 (1) (1993), pp. 83–106.

[49] Duan, J. "Augmented GARCH (p, q) process and its diffusion limit". *Journal of Econometrics* 79 (1) (1997), pp. 97–127.

[50] Duttweiler, D. "The mean-square error of Bahadur's order-statistic approximation". *The Annals of Statistics* (1973), pp. 446–453.

[51] Ekström, M. "A general central limit theorem for strong mixing sequences". *Statistics & Probability Letters* 94 (2014), pp. 236–238.

[52] Embrechts, P. and Samorodnitsky, G. "Sample Quantiles of heavy tailed stochastic processes". *Stochastic Processes and their Applications* 59 (2) (1995), pp. 217–233.

[53] Emmer, S., Kratz, M., and Tasche, D. "What is the best risk measure in practice? A comparison of standard risk measures". *Journal of Risk* 18 (2) (2015), pp. 31–60.

[54] Engle, R. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". *Econometrica: Journal of the Econometric Society* 50 (4) (1982), pp. 987–1007.

[55] Engle, R. and Bollerslev, T. "Modelling the persistence of conditional variances". *Econometric Reviews* 5 (1) (1986), pp. 1–50.

[56] Engle, R. and Manganelli, S. "CAViaR: Conditional autoregressive value at risk by regression quantiles". *Journal of Business & Economic Statistics* 22 (4) (2004), pp. 367–381.

[57] Engle, R. and Ng, V. "Measuring and testing the impact of news on volatility". *The Journal of Finance* 48 (5) (1993), pp. 1749–1778.

[58] European Banking Authority. "Results from the 2018 market risk benchmarking exercise". *EBA Report* (2019).

[59] European Union. "Directive 2009/138/EC (Solvency II)". *Official Journal of European Union* 335 (2009).

[60] Falk, M and Reiss, R.-D. "Independence of order statistics". *The Annals of Probability* (1988), pp. 854–862.

[61] Falk, M. "Asymptotic independence of median and MAD". *Statistics & Probability Letters* 34 (4) (1997), pp. 341–345.

[62] Ferguson, T. "Asymptotic joint distribution of sample mean and a sample quantile". *unpublished:* `http://www.math.ucla.edu/~tom/papers/unpublished/meanmed.pdf` (1999). [Online; accessed 25-April-2019].

[63] Fisher, R. "On the probable error of a coefficient of correlation deduced from a small sample". *Metron* 1 (1921), pp. 3–32.

[64] Francq, C., Horvath, L., and Zakoïan, J.-M. "Merits and drawbacks of variance targeting in GARCH models". *Journal of Financial Econometrics* 9 (4) (2011), pp. 619–656.

[65] Francq, C. and Zakoian, J.-M. *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons, 2019.

[66] Geweke, J. "Modeling the persistence of conditional variances: a comment". *Econometric Reviews* 5 (1986), pp. 57–61.

[67] Ghalanos, A. *rugarch: Univariate GARCH models.* R package version 1.4-1. 2019.

[68] Ghosh, J. "A new proof of the Bahadur representation of quantiles and an application". *The Annals of Mathematical Statistics* (1971), pp. 1957–1961.

[69] Glosten, L., Jagannathan, R., and Runkle, D. "On the relation between the expected value and the volatility of the nominal excess return on stocks". *The Journal of Finance* 48 (5) (1993), pp. 1779–1801.

[70] Gorard, S. "Revisiting a 90-year-old debate: the advantages of the mean deviation". *British Journal of Educational Studies* 53 (4) (2005), pp. 417–430.

[71] Hall, P. and Welsh, A. "Limit theorems for the median deviation". *Annals of the Institute of Statistical Mathematics* 37 (1) (1985), pp. 27–36.

[72] Hampel, F. "The influence curve and its role in robust estimation". *Journal of the American Statistical Association* 69 (346) (1974), pp. 383–393.

[73] Higgins, M. and Bera, A. "A class of nonlinear ARCH models". *International Economic Review* 33 (1) (1992), pp. 137–158.

[74] Ho, H., Lin, T.-I., Chen, H.-Y., and Wang, W.-L. "Some results on the truncated multivariate t distribution". *Journal of Statistical Planning and Inference* 142 (1) (2012), pp. 25–40.

[75] Ho, H.-C. and Hsing, T. "On the asymptotic expansion of the empirical process of long-memory moving averages". *The Annals of Statistics* 24 (3) (1996), pp. 992–1024.

[76] Hofert, M., Hornik, K., and McNeil, A. J. *qrmtools: Tools for Quantitative Risk Management*. R package version 0.0-10. 2018. URL: `https://CRAN.R-project.org/package=qrmtools`.

[77] Hörmann, S. "Augmented GARCH sequences: Dependence structure and asymptotics". *Bernoulli* 14 (2) (2008), pp. 543–561.

[78] Hsing, T. "A note on the asymptotic independence of the sum and maximum of strongly mixing stationary random variables". *The Annals of Probability* (1995), pp. 938–947.

[79] Hull, J. and White, A. "Incorporating volatility updating into the historical simulation method for value-at-risk". *Journal of Risk* 1 (1) (1998), pp. 5–19.

[80] Ibragimov, I. A. "Some limit theorems for stationary processes". *Theory of Probability & Its Applications* 7 (4) (1962), pp. 349–382.

[81] Jamalizadeh, A., Khosravi, M., and Balakrishnan, N. "Recurrence relations for distributions of a skew-t and a linear combination of order statistics from a bivariate-t". *Computational Statistics & Data Analysis* 53 (4) (2009), pp. 847–852.

[82] Jamalizadeh, A., Pourmousa, R., and Balakrishnan, N. "Truncated and limited skew-normal and skew-t distributions: properties and an illustration". *Communications in Statistics-Theory and Methods* 38 (16-17) (2009), pp. 2653–2668.

[83] James, D. and Hornik, K. *chron: Chronological Objects which Can Handle Dates and Times*. R package version 2.3-53. S original by David James, R port by Kurt Hornik. 2018. URL: https://CRAN.R-project.org/package=chron.

[84] Kim, H.-J. "Moments of truncated Student-t distribution". *Journal of the Korean Statistical Society* 37 (1) (2008), pp. 81–87.

[85] Kratz, M., Lok, Y., and McNeil, A. "Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall". *Journal of Banking and Finance* 88 (2018), pp. 393–407.

[86] Kruskal, W. "Ordinal measures of association". *Journal of the American Statistical Association* 53 (284) (1958), pp. 814–861.

[87] Kuan, C.-M., Yeh, J.-H., and Hsu, Y.-C. "Assessing value at risk with care, the conditional autoregressive expectile models". *Journal of Econometrics* 150 (2) (2009), pp. 261–270.

[88] Kulik, R. "Optimal rates in the Bahadur-Kiefer representation for GARCH sequences". *arXiv: math/0605283* (2006).

[89] Lee, O. "Functional central limit theorems for augmented GARCH (p, q) and FIGARCH processes". *Journal of the Korean Statistical Society* 43 (3) (2014), pp. 393–401.

[90] Lehmann, E. *Elements of Large-Sample Theory*. Springer Science & Business Media, 1999.

[91] Lin, P.-E., Wu, K.-T., and Ahmad, I. "Asymptotic joint distribution of sample quantiles and sample mean with applications". *Communications in Statistics-Theory and Methods* 9 (1) (1980), pp. 51–60.

[92] Loynes, R. "The variance and the range of i.I.D. random variables". *Communications in Statistics - Theory and Methods* 19 (4) (1990), pp. 1419–1432.

[93] Mazumder, S. and Serfling, R. "Bahadur representations for the median absolute deviation and its modifications". *Statistics & Probability Letters* 79 (16) (2009), pp. 1774–1783.

[94] Mikosch, T. and Stărică, C. "Limit theory for the sample autocorrelations and extremes of a GARCH (1,1) process". *The Annals of Statistics* 28 (5) (2000), pp. 1427–1451.

[95] Milhøj, A. "A multiplicative parameterization of arch models". *Working Paper* (1987).

[96] Miura, R. "A note on look-back options based on order statistics". *Hitotsubashi J. Commerce Management* 27 (1992), pp. 15–28.

[97] Morgan, J. "Introduction to riskmetrics". *New York: JP Morgan* (1994).

[98] Nelson, D. "Conditional heteroskedasticity in asset returns: A new approach". *Econometrica: Journal of the Econometric Society* 59 (2) (1991), pp. 347–370.

[99] Neumann, M. "A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics". *ESAIM: Probability and Statistics* 17 (2013), pp. 120–134.

[100] Newey, W. and Powell, J. "Asymmetric least squares estimation and testing". *Econometrica: Journal of the Econometric Society* (1987), pp. 819–847.

[101] Nolan, J. P. *Stable Distributions - Models for Heavy Tailed Data*. In progress, Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html. Boston: Birkhauser, 2018.

[102] Pantula, S. "Modeling the persistence of conditional variances: a comment". *Econometric Reviews* 5 (1986), pp. 79–97.

[103] Pérignon, C. and Smith, D. R. "The level and quality of Value-at-Risk disclosure by commercial banks". *Journal of Banking & Finance* 34 (2) (2010), pp. 362–377.

[104] Pham-Gia, T and Hung, T. "The mean and median absolute deviations". *Mathematical and Computer Modelling* 34 (7-8) (2001), pp. 921–936.

[105] Philipp, W. "A functional law of the iterated logarithm for empirical distribution functions of weakly dependent random variables". *The Annals of Probability* (1977), pp. 319–350.

[106] Politis, D., Romano, J., and Wolf, M. "Subsampling for heteroskedastic time series". *Journal of Econometrics* 81 (2) (1997), pp. 281–317.

[107] Pritsker, M. "The hidden dangers of historical simulation". *Journal of Banking & Finance* 30 (2) (2006), pp. 561–582.

[108] Pyke, R. "Spacings". *Journal of the Royal Statistical Society. Series B (Methodological)* (1965), pp. 395–449.

[109] Quagliariello, M. "Does macroeconomy affect bank stability? A review of the empirical evidence". *Journal of Banking Regulation* 9 (2) (2008), pp. 102–115.

[110] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: https://www.R-project.org/.

[111] Rockafellar, R and Uryasev, S. "Conditional value-at-risk for general loss distributions". *Journal of Banking & Finance* 26 (7) (2002), pp. 1443–1471.

[112] Rodriguez, R. "Correlation". In: *Encyclopedia of Statistical Sciences, 2nd edition*. Ed. by S. Kotz, N. Balakrishnan, C. Read, B. Vidakovic, and N. Johnson. Wiley, New York, 1982, pp. 1375–1385.

[113] Roussas, G. and Ioannides, D. "Moment inequalities for mixing sequences of random variables". *Stochastic Analysis and Applications* 5 (1) (1987), pp. 60–120.

[114] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA, 2016. URL: http://www.rstudio.com/.

[115] Schwert, G. "Why does stock market volatility change over time?" *The Journal of Finance* 44 (5) (1989), pp. 1115–1153.

[116] Segers, J. "On the asymptotic distribution of the mean absolute deviation about the mean". *arXiv:1406.4151* (2014).

[117] Serfling, R. and Mazumder, S. "Exponential probability inequality and convergence results for the median absolute deviation and its modifications". *Statistics & Probability Letters* 79 (16) (2009), pp. 1767–1773.

[118] Taylor, S. *Modelling financial time series*. Wiley, New York, 1986.

[119] Trapletti, A. and Hornik, K. *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-47. 2019. URL: https://CRAN.R-project.org/package=tseries.

[120] Vaart, A. Van der. *Asymptotic statistics*. Cambridge University Press, 1998.

[121] Wendler, M. "Bahadur representation for U-quantiles of dependent data". *Journal of Multivariate Analysis* 102 (6) (2011), pp. 1064–1079.

[122] Wickham, H. and Bryan, J. *readxl: Read Excel Files*. R package version 1.3.1. 2019. URL: https://CRAN.R-project.org/package=readxl.

[123] Withers, C. and Nadarajah, S. "Expansions for the joint distribution of the sample maximum and sample estimate". *Sankhyā: The Indian Journal of Statistics, Series A (2008-)* (2008), pp. 109–123.

[124] Wu, W. "On the Bahadur representation of sample quantiles for dependent sequences". *The Annals of Statistics* 33 (4) (2005), pp. 1934–1963.

[125] Wuertz, D., Setz, T., and Chalabi, Y. *fBasics: Rmetrics - Markets and Basic Statistics*. R package version 3042.89. 2017. URL: https://CRAN.R-project.org/package=fBasics.

[126] Wuertz, D., Setz, T., Chalabi, Y., Boudt, C., Chausse, P., and Miklovac, M. *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*. R package version 3042.83.1. 2019. URL: https://CRAN.R-project.org/package=fGarch.

[127] Yao, Q. and Tong, H. "Asymmetric least squares regression estimation: a nonparametric approach". *Journal of Nonparametric Statistics* 6 (2-3) (1996), pp. 273–292.

[128] Zakoian, J.-M. "Threshold heteroskedastic models". *Journal of Economic Dynamics and Control* 18 (5) (1994), pp. 931–955.

[129] Zumbach, G. *Correlations of the realized volatilities with the centered volatility increment*. http://www.finanscopics.com/figuresPage.php?figCode=corr_vol_r_VsDV0. [Online; accessed 25-April-2019]. 2012.

[130] Zumbach, G. "The pitfalls in fitting GARCH (1, 1) processes". In: *Advances in Quantitative Asset Management*. Springer, 2000, pp. 179–200.

[131] Zumbach, G. *Discrete time series, processes, and applications in finance*. Springer Science & Business Media, 2012.

Appendix A

Supplements to Chapter 2

This Appendix A contains a small selection of additional material related to the empirical study presented in Chapter 2. A more extensive collection of material related to the empirical study is available as online-appendix, see [34]: An empirical study using weekly data, using std instead of MAD, with ES instead of VaR and with a LM-ARCH model instead of a GARCH$(1,1)$.

Here we present two things: In Appendix A.1 further plots and tables of the empirical study in Chapter 2, and in Appendix A.4 corresponding plos and tables for the empirical study with ES. In both cases we show plots for all indices of the logarithm of SQP ratios versus annualized volatility as well as the linear Pearson, Spearman and Kendall rank correlation in a table to assess their dependence.

This material is a uncommented collection of the corresponding results.

A.1 Empirical study of Chapter 2

A.1.1 SQP on a rolling window

See the online-appendix in [34].

A.1.2 SQPratio average

See the online-appendix in [34].

A.1.3 SQPratio RMSE

See the online-appendix in [34].

A.1.4 Comparing std and MAD (1y and 3y)

See the online-appendix in [34].

A.1.5 SQP ratios and annualized volatility (MAD)

See the online-appendix in [34].

A.1.6 SQP ratios vs annualized volatility (MAD)

We present for each index the plots of the logarithm of the SQP ratios against the annualized volatility (MAD), for $\alpha = 0.95$. For the case $\alpha = 0.99$ and the plots of the SQP ratio without the logarithm, see the online-appendix in [34].

Real data

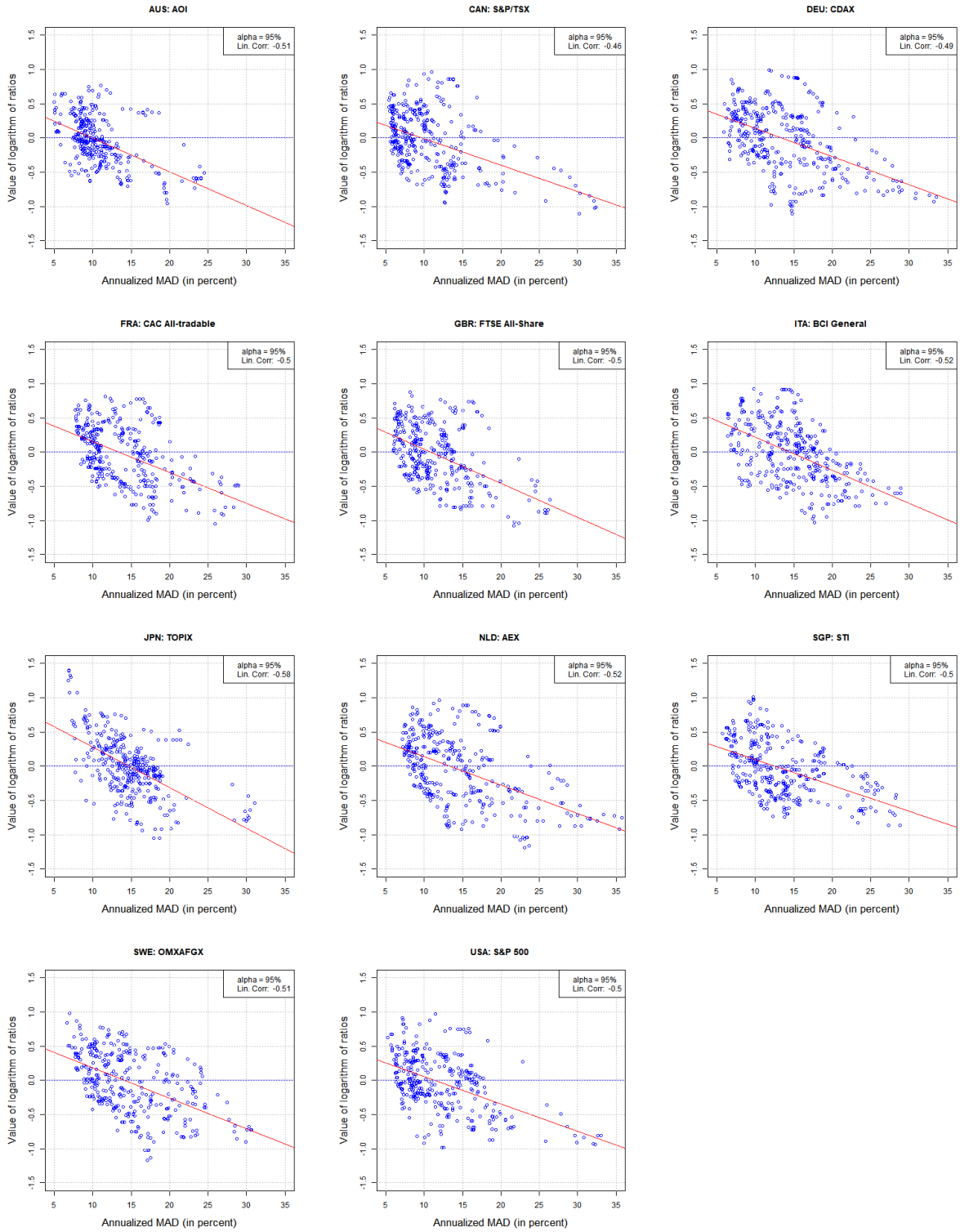


FIGURE A.1: The logarithm of SQP ratios as a function of annualized volatility (on real data) for $\alpha = 95\%$. From left to right, top to bottom: AUS, CAN, DEU, FRA, GBR, ITA, JPN, NLD, SGP, SWE, USA.

GARCH (Student innovations)

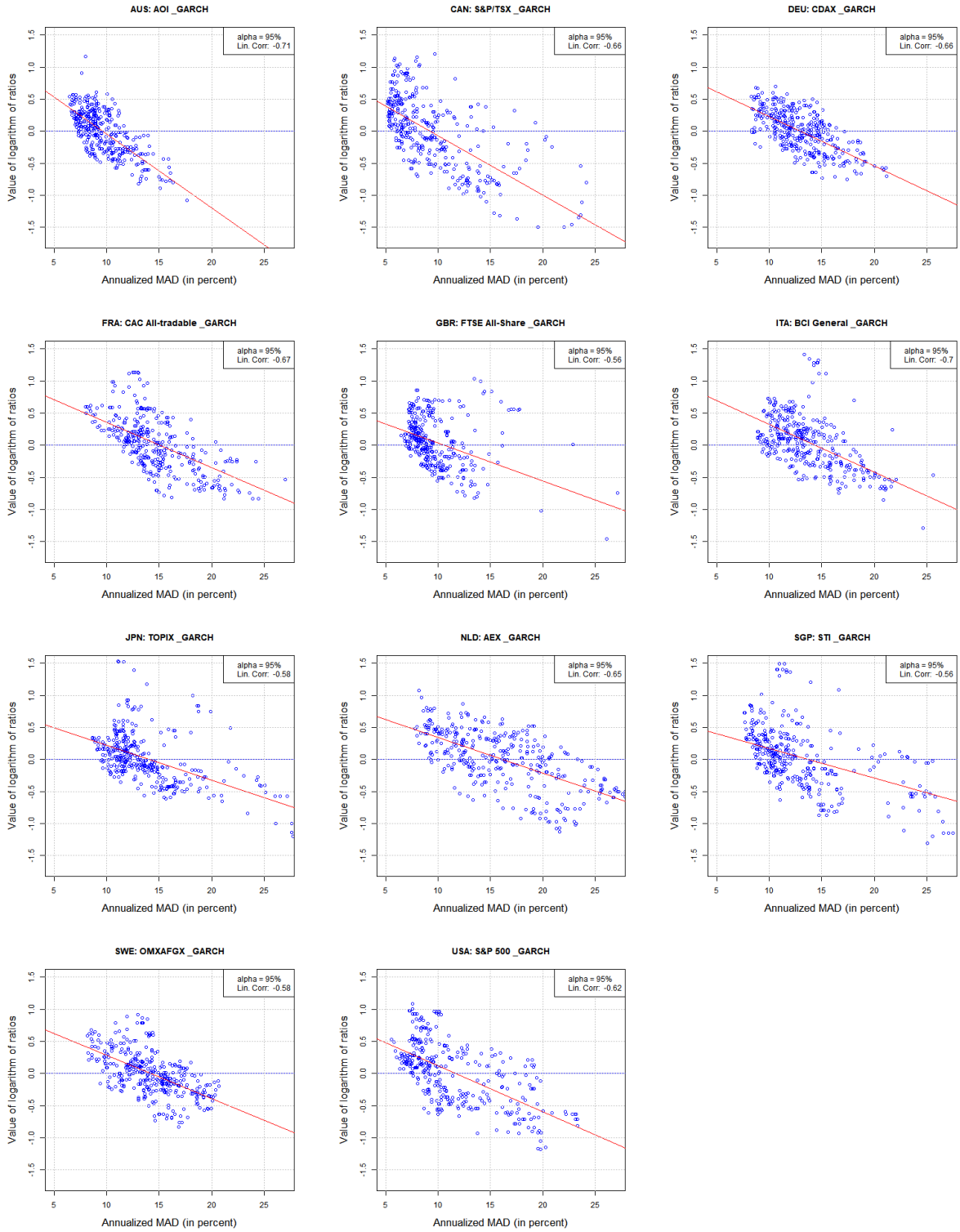


FIGURE A.2: The logarithm of SQP ratios as a function of annualized volatility (on sample paths from GARCH(1,1) simulations with Student-t innovations) for $\alpha = 95\%$. From left to right, top to bottom: AUS, CAN, DEU, FRA, GBR, ITA, JPN, NLD, SGP, SWE, USA.

GARCH (Gaussian innovations)

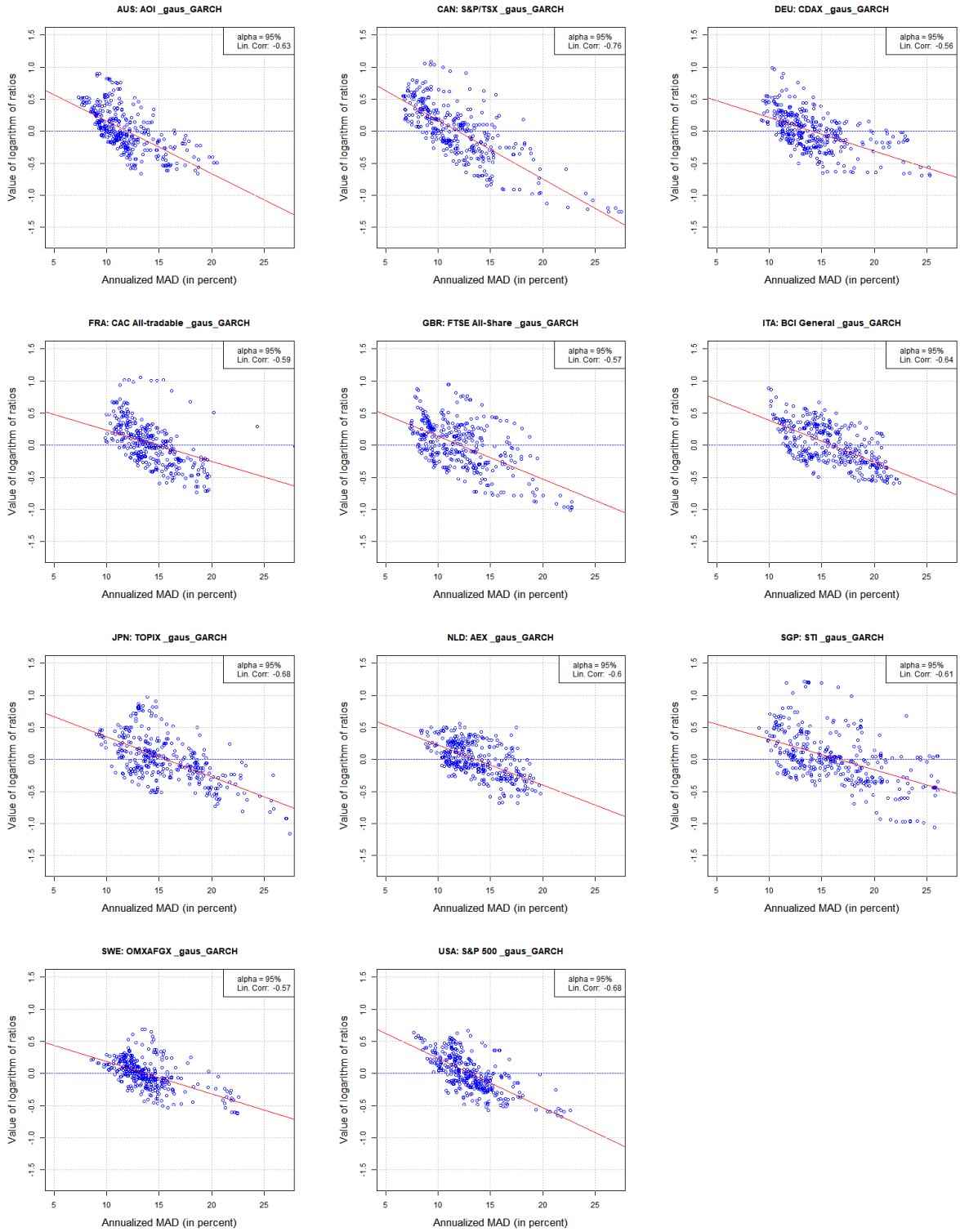


FIGURE A.3: The logarithm of SQP ratios as a function of annualized volatility (on sample paths from GARCH(1,1) simulations with Gaussian innovations) for $\alpha = 95\%$. From left to right, top to bottom: AUS, CAN, DEU, FRA, GBR, ITA, JPN, NLD, SGP, SWE, USA.

A.1.7 Pearson and rank correlations

Real Data

TABLE A.1: Pearson correlation $\rho(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (on real data), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.51	-0.46	-0.50	-0.49	-0.52	-0.58	-0.52	-0.50	-0.51	-0.50	-0.50	**-0.51** $\pm$ 0.03
97.5%	-0.55	-0.47	-0.49	-0.44	-0.52	-0.58	-0.53	-0.50	-0.50	-0.53	-0.52	**-0.51** $\pm$ 0.04
99%	-0.57	-0.46	-0.51	-0.47	-0.52	-0.52	-0.55	-0.54	-0.48	-0.50	-0.54	**-0.51** $\pm$ 0.04
99.5%	-0.52	-0.46	-0.47	-0.45	-0.51	-0.48	-0.54	-0.50	-0.45	-0.49	-0.53	**-0.49** $\pm$ 0.03

TABLE A.2: Spearman Rank correlation $\rho_S(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (on real data), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.49	-0.38	-0.48	-0.46	-0.52	-0.54	-0.49	-0.51	-0.51	-0.45	-0.42	**-0.48** $\pm$ 0.05
97.5%	-0.48	-0.43	-0.48	-0.45	-0.52	-0.53	-0.50	-0.52	-0.53	-0.48	-0.46	**-0.49** $\pm$ 0.03
99%	-0.50	-0.44	-0.49	-0.48	-0.54	-0.48	-0.52	-0.52	-0.52	-0.43	-0.48	**-0.49** $\pm$ 0.03
99.5%	-0.47	-0.43	-0.46	-0.47	-0.53	-0.48	-0.53	-0.47	-0.49	-0.43	-0.47	**-0.48** $\pm$ 0.03

TABLE A.3: Kendall rank correlation $\rho_K(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (on real data), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.35	-0.26	-0.34	-0.32	-0.36	-0.39	-0.34	-0.35	-0.35	-0.32	-0.30	**-0.34** $\pm$ 0.03
97.5%	-0.35	-0.30	-0.34	-0.31	-0.37	-0.38	-0.35	-0.35	-0.37	-0.34	-0.33	**-0.35** $\pm$ 0.03
99%	-0.37	-0.31	-0.35	-0.33	-0.37	-0.35	-0.36	-0.36	-0.36	-0.30	-0.34	**-0.35** $\pm$ 0.02
99.5%	-0.34	-0.31	-0.32	-0.32	-0.36	-0.34	-0.37	-0.32	-0.34	-0.31	-0.33	**-0.33** $\pm$ 0.02

GARCH (Student Innovations)

TABLE A.4: Pearson correlation $\rho(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Student-t innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.63	-0.62	-0.62	-0.62	-0.62	-0.64	-0.63	-0.63	-0.63	-0.63	-0.62	**-0.63** $\pm$ 0.01
97.5%	-0.63	-0.61	-0.61	-0.61	-0.61	-0.63	-0.62	-0.62	-0.62	-0.62	-0.61	**-0.62** $\pm$ 0.01
99%	-0.60	-0.58	-0.58	-0.59	-0.58	-0.60	-0.59	-0.60	-0.59	-0.59	-0.59	**-0.59** $\pm$ 0.01
99.5%	-0.58	-0.57	-0.56	-0.57	-0.56	-0.58	-0.57	-0.58	-0.57	-0.57	-0.57	**-0.57** $\pm$ 0.01

TABLE A.5: Spearman Rank correlation $\rho_S(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Student-t innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.61	-0.61	-0.60	-0.61	-0.60	-0.61	-0.61	-0.61	-0.61	-0.61	-0.61	**-0.61** $\pm$ 0
97.5%	-0.60	-0.60	-0.59	-0.60	-0.59	-0.60	-0.61	-0.61	-0.60	-0.60	-0.60	**-0.6** $\pm$ 0
99%	-0.57	-0.58	-0.57	-0.58	-0.56	-0.57	-0.58	-0.58	-0.57	-0.57	-0.58	**-0.57** $\pm$ 0.01
99.5%	-0.55	-0.56	-0.55	-0.56	-0.55	-0.55	-0.57	-0.56	-0.55	-0.56	-0.56	**-0.56** $\pm$ 0.01

TABLE A.6: Kendall rank correlation $\rho_K(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Student-t innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.44	-0.44	-0.44	-0.44	-0.44	-0.44	-0.45	-0.45	-0.44	-0.44	-0.44	**-0.44** $\pm$ 0
97.5%	-0.43	-0.43	-0.43	-0.43	-0.43	-0.44	-0.44	-0.44	-0.43	-0.43	-0.44	**-0.43** $\pm$ 0
99%	-0.41	-0.41	-0.40	-0.41	-0.40	-0.41	-0.42	-0.42	-0.41	-0.41	-0.41	**-0.41** $\pm$ 0
99.5%	-0.39	-0.40	-0.39	-0.40	-0.39	-0.40	-0.40	-0.40	-0.39	-0.40	-0.40	**-0.4** $\pm$ 0.01

GARCH (Gaussian Innovation)

TABLE A.7: Pearson correlation $\rho(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Gaussian innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.64	-0.62	-0.62	-0.62	-0.62	-0.64	-0.62	-0.63	-0.63	-0.63	-0.62	**-0.63** $\pm$ 0.01
97.5%	-0.63	-0.60	-0.61	-0.61	-0.60	-0.62	-0.61	-0.62	-0.62	-0.61	-0.61	**-0.61** $\pm$ 0.01
99%	-0.60	-0.58	-0.58	-0.58	-0.58	-0.59	-0.58	-0.59	-0.59	-0.58	-0.58	**-0.58** $\pm$ 0.01
99.5%	-0.58	-0.56	-0.56	-0.56	-0.56	-0.57	-0.57	-0.57	-0.57	-0.57	-0.56	**-0.57** $\pm$ 0.01

TABLE A.8: Spearman Rank correlation $\rho_S(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Gaussian innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.62	-0.60	-0.60	-0.60	-0.60	-0.61	-0.60	-0.61	-0.61	-0.61	-0.60	**-0.61** $\pm$ 0.01
97.5%	-0.60	-0.59	-0.59	-0.59	-0.59	-0.60	-0.59	-0.60	-0.60	-0.59	-0.59	**-0.59** $\pm$ 0
99%	-0.57	-0.56	-0.56	-0.56	-0.56	-0.56	-0.57	-0.57	-0.57	-0.57	-0.57	**-0.56** $\pm$ 0
99.5%	-0.55	-0.55	-0.54	-0.55	-0.54	-0.54	-0.55	-0.56	-0.55	-0.55	-0.55	**-0.55** $\pm$ 0

TABLE A.9: Kendall rank correlation $\rho_K(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the SQP ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Gaussian innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.45	-0.44	-0.43	-0.44	-0.43	-0.44	-0.44	-0.44	-0.44	-0.44	-0.44	**-0.44** $\pm$ 0
97.5%	-0.43	-0.43	-0.42	-0.43	-0.42	-0.43	-0.43	-0.43	-0.43	-0.43	-0.43	**-0.43** $\pm$ 0
99%	-0.41	-0.40	-0.40	-0.40	-0.40	-0.40	-0.41	-0.41	-0.40	-0.41	-0.40	**-0.4** $\pm$ 0
99.5%	-0.39	-0.39	-0.38	-0.39	-0.38	-0.39	-0.39	-0.40	-0.39	-0.39	-0.39	**-0.39** $\pm$ 0

A.1.8 Volatility binning

See the online-appendix in [34].

A.2 Empirical study using weekly data

A.2.1 Pearson and rank correlations

See the online-appendix in [34].

A.3 Empirical study with std

A.3.1 SQP ratios and annualized volatility (std)

See the online-appendix in [34].

A.3.2 SQP ratios vs annualized volatility (std)

See the online-appendix in [34].

A.3.3 Pearson and rank correlations

See the online-appendix in [34].

A.3.4 Volatility binning

See the online-appendix in [34].

A.4 Empirical study with ES

A.4.1 ES on a rolling window

See the online-appendix in [34].

A.4.2 ES Ratios and annualized volatility (MAD)

See the online-appendix in [34].

A.4.3 ES ratios vs annualized volatility (MAD)

We present for each index the plots of the logarithm of the SQP ratios against the annualized volatility (MAD), for $\alpha = 0.95$. For the case $\alpha = 0.99$ and the plots of the SQP ratio without the logarithm, see the online-appendix in [34].

Real data

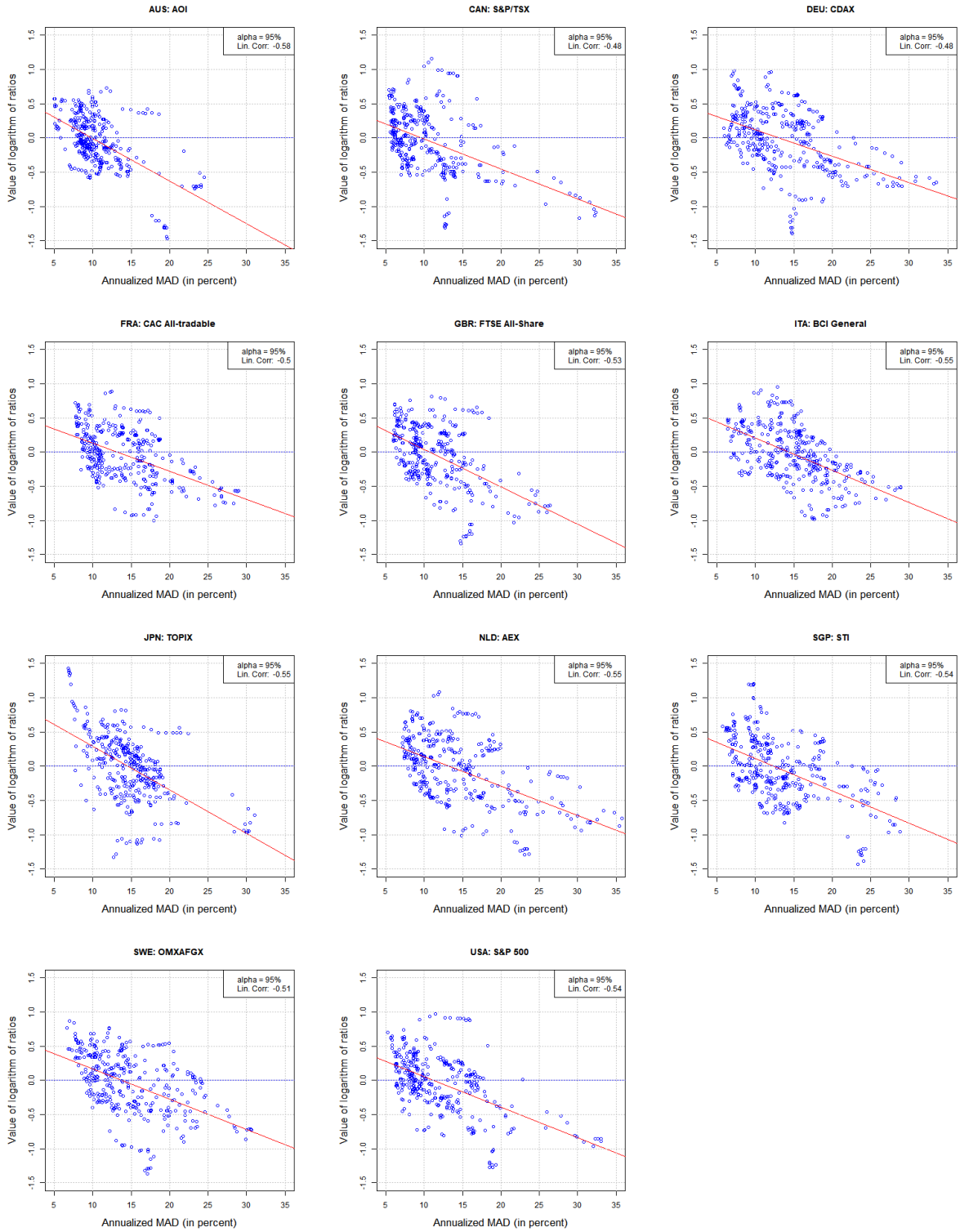


FIGURE A.4: The logarithm of ES ratios as a function of annualized volatility (on real data) for $\alpha = 95\%$. From left to right, top to bottom: AUS, CAN, DEU, FRA, GBR, ITA, JPN, NLD, SGP, SWE, USA.

GARCH (Gaussian innovations)

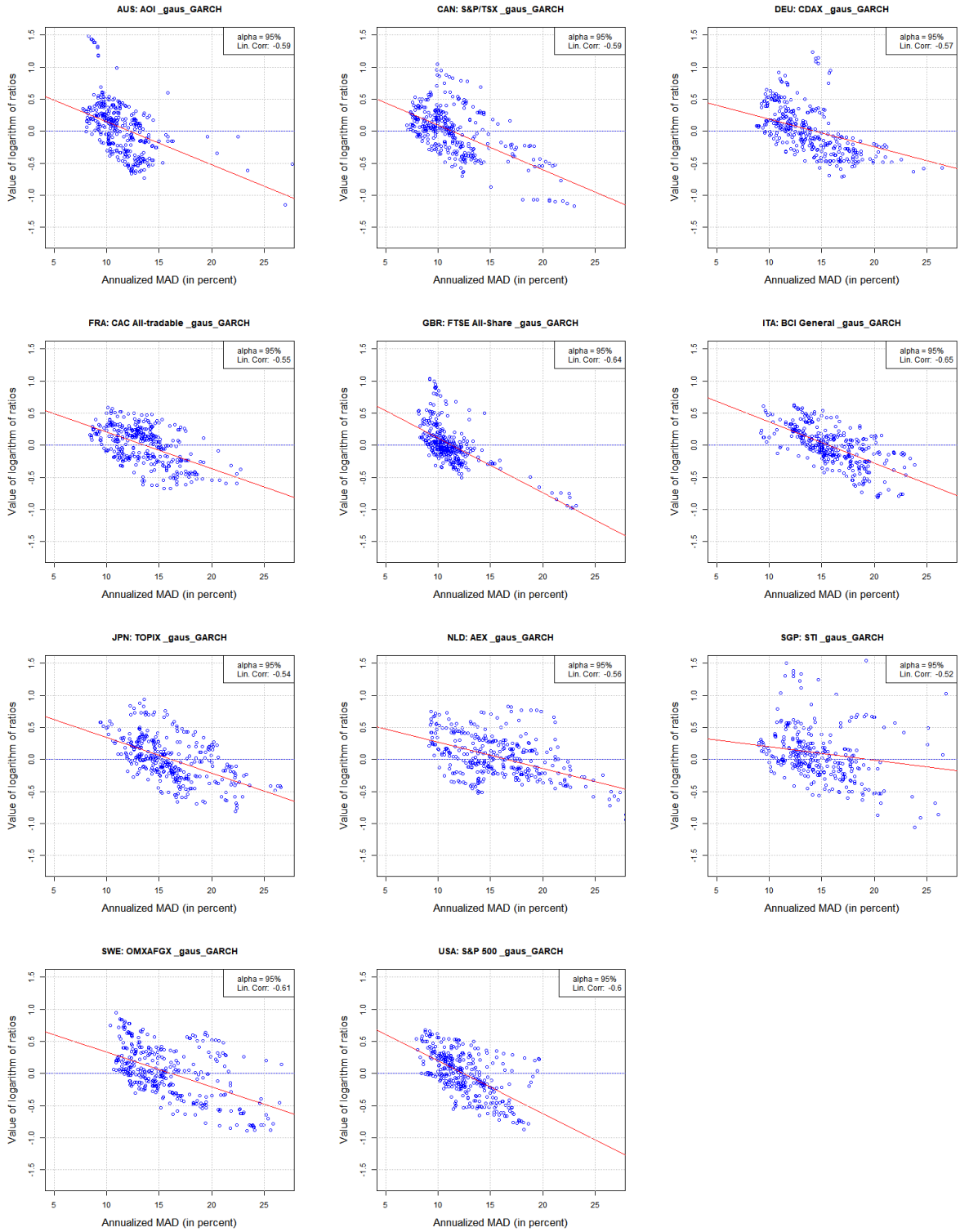


FIGURE A.5: The logarithm of ES ratios as a function of annualized volatility (on sample paths from GARCH(1,1) simulations with Gaussian innovations) for $\alpha = 95\%$. From left to right, top to bottom: AUS, CAN, DEU, FRA, GBR, ITA, JPN, NLD, SGP, SWE, USA.

A.4.4 Pearson and rank correlations

Real data

TABLE A.10: Pearson correlation $\rho(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the ES ratios and the volatility (on real data), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.58	-0.48	-0.50	-0.48	-0.55	-0.55	-0.55	-0.54	-0.51	-0.53	-0.54	**-0.53** $\pm$ 0.03
97.5%	-0.56	-0.47	-0.49	-0.46	-0.55	-0.52	-0.55	-0.53	-0.49	-0.52	-0.54	**-0.52** $\pm$ 0.03
99%	-0.53	-0.45	-0.46	-0.41	-0.53	-0.44	-0.53	-0.51	-0.44	-0.51	-0.51	**-0.48** $\pm$ 0.04
99.5%	-0.50	-0.44	-0.43	-0.37	-0.50	-0.41	-0.51	-0.48	-0.41	-0.50	-0.49	**-0.46** $\pm$ 0.05

TABLE A.11: Spearman Rank correlation $\rho_S(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the ES ratios and the volatility (on real data), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.50	-0.44	-0.48	-0.47	-0.57	-0.51	-0.51	-0.53	-0.54	-0.46	-0.48	**-0.50** $\pm$ 0.04
97.5%	-0.46	-0.44	-0.45	-0.46	-0.58	-0.47	-0.51	-0.51	-0.53	-0.46	-0.49	**-0.49** $\pm$ 0.04
99%	-0.42	-0.43	-0.42	-0.43	-0.58	-0.43	-0.49	-0.46	-0.48	-0.43	-0.47	**-0.46** $\pm$ 0.05
99.5%	-0.39	-0.42	-0.39	-0.40	-0.55	-0.40	-0.46	-0.41	-0.45	-0.41	-0.44	**-0.43** $\pm$ 0.05

TABLE A.12: Kendall rank correlation $\rho_K(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the ES ratios and the volatility (on real data), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.36	-0.31	-0.34	-0.32	-0.40	-0.37	-0.36	-0.36	-0.37	-0.33	-0.35	**-0.35** $\pm$ 0.02
97.5%	-0.33	-0.32	-0.32	-0.32	-0.40	-0.34	-0.36	-0.35	-0.37	-0.33	-0.35	**-0.34** $\pm$ 0.03
99%	-0.29	-0.31	-0.29	-0.30	-0.39	-0.30	-0.33	-0.33	-0.33	-0.30	-0.33	**-0.32** $\pm$ 0.03
99.5%	-0.27	-0.30	-0.27	-0.29	-0.37	-0.28	-0.31	-0.29	-0.31	-0.29	-0.30	**-0.30** $\pm$ 0.03

GARCH (Gaussian innovations)

TABLE A.13: Pearson correlation $\rho(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the ES ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Gaussian innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.63	-0.60	-0.60	-0.61	-0.64	-0.63	-0.59	-0.60	-0.66	-0.63	-0.62	**-0.62** $\pm$ 0.02
97.5%	-0.61	-0.59	-0.58	-0.60	-0.62	-0.61	-0.58	-0.58	-0.64	-0.61	-0.60	**-0.60** $\pm$ 0.02
99%	-0.58	-0.56	-0.56	-0.57	-0.59	-0.57	-0.55	-0.56	-0.61	-0.58	-0.58	**-0.57** $\pm$ 0.02
99.5%	-0.56	-0.55	-0.55	-0.56	-0.58	-0.55	-0.54	-0.55	-0.59	-0.57	-0.56	**-0.56** $\pm$ 0.02

TABLE A.14: Spearman Rank correlation $\rho_S(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the ES ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Gaussian innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.58	-0.59	-0.59	-0.62	-0.61	-0.61	-0.58	-0.55	-0.64	-0.61	-0.61	**-0.60** $\pm$ 0.03
97.5%	-0.55	-0.58	-0.57	-0.61	-0.60	-0.59	-0.56	-0.54	-0.62	-0.59	-0.60	**-0.58** $\pm$ 0.03
99%	-0.53	-0.56	-0.55	-0.59	-0.57	-0.55	-0.54	-0.51	-0.59	-0.57	-0.57	**-0.56** $\pm$ 0.03
99.5%	-0.51	-0.55	-0.54	-0.58	-0.56	-0.54	-0.52	-0.50	-0.58	-0.55	-0.55	**-0.54** $\pm$ 0.03

TABLE A.15: Kendall rank correlation $\rho_K(\log(R_{\alpha,T}(t), V_{1,T}(t))$ between the log of the ES ratios and the volatility (averages of 1000 repetitions from GARCH(1,1) simulations with Gaussian innovations), for each index and for $T = T_s = 1$ year, over the whole historical sample, and for two thresholds (95% and 99%). In the last column, we present the average over all indices $\pm$ the standard deviation over the 11 displayed values.

α	AUS	CAN	FRA	DEU	ITA	JPN	NLD	SGP	SWE	GBR	USA	**AVG ($\pm$ sd)**
95%	-0.42	-0.42	-0.42	-0.45	-0.45	-0.44	-0.41	-0.40	-0.47	-0.44	-0.44	**-0.43** $\pm$ 0.02
97.5%	-0.40	-0.41	-0.41	-0.44	-0.43	-0.43	-0.40	-0.39	-0.45	-0.43	-0.43	**-0.42** $\pm$ 0.02
99%	-0.38	-0.40	-0.39	-0.42	-0.41	-0.40	-0.38	-0.37	-0.43	-0.41	-0.41	**-0.40** $\pm$ 0.02
99.5%	-0.36	-0.39	-0.39	-0.41	-0.40	-0.39	-0.36	-0.36	-0.41	-0.40	-0.40	**-0.39** $\pm$ 0.02

A.5 Empirical study with LM-ARCH

A.5.1 Model explained

See the online-appendix in [34].

A.5.2 SQP ratios and annualized volatility (MAD)

See the online-appendix in [34].

A.5.3 SQP ratios vs annualized volatility (MAD)

See the online-appendix in [34].

A.5.4 Pearson and rank correlations

See the online-appendix in [34].

A.6 Empirical Study using longer samples

See the online-appendix in [34].

Appendix B

Supplements to Chapter 3

This Appendix B contains a part of the additional material related to the theoretical results presented in Chapter 3. A more extensive collection of material is available as online-appendix, see [34]: The full simulation study assessing the finite sample performance of the asymptotic results as corresponding to Section 3.6.2. Further, building upon this, a simulation study for rank correlation.

The part from the online-appendix we present here is Appendix B.1. Therein we show and derive the covariance and correlation expressions of the asymptotic distribution between the quantile estimator and a measure of dispersion estimator, in the case of an underlying Gaussian or Student-t distribution. These values were needed for the comparative plots in Section 3.6.1 and to determine the theoretical values in the simulation study in Section 3.6.2.

B.1 Correlation of asymptotic distribution - Explicit computations

We consider the covariance (or correlation) in the asymptotic distribution of a quantile estimator (either sample quantile or location-scale quantile estimator) and three measure of dispersion estimators (sample MAD, sample variance or sample MedianAD). We establish the corresponding results for an underlying Gaussian or Student-t distribution.

We introduce the standard notation $\Phi, \varphi, \Phi^{-1}(p)$ for the cdf, pdf and quantile of order p of the standard normal distribution. Further let us denote by $Y \sim t(0, \frac{\nu-2}{\nu}, \nu)$ a rv Y which follows the Student-t distribution with ν degrees of freedom, mean 0 and variance 1. In contrast to this, denote by $\tilde{Y} \sim t(0,1,\nu)$ a rv $\tilde{Y}$ which follows the standard Student-t distribution with ν degrees of freedom (i.e. mean 0 but variance $\nu/(\nu-2)$. We will usually give expressions for the Student distribution in terms of $\tilde{Y}$ as this is the standard form for the Student distribution.

Let (X_n, Y_n) be two series of random variables, converging jointly to (X, Y), respectively. We then write, by abuse of notation but to ease the presentation, $\lim_{n\to\infty} \mathrm{Cor}(X_n, Y_n)$ for the correlation in the asymptotic distribution of (X_n, Y_n), i.e. $\mathrm{Cor}(X, Y)$. And consequently $\lim_{n\to\infty} \mathrm{Cov}(X_n, Y_n)$ for the covariance in the asymptotic distribution of (X_n, Y_n), i.e. $\mathrm{Cov}(X, Y)$.

B.1.1 ... with the sample quantile

Let us consider first, in Example B.1, the sample variance as measure of dispersion.

Example B.1 *(i) In the case of a Gaussian distribution with mean μ and variance σ^2:*

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\, q_n(p), \sqrt{n}\, \hat{\sigma}_n^2) = \sigma^3 \Phi^{-1}(p), \tag{B.1}$$

$$\lim_{n\to\infty} \mathrm{Cor}(q_n(p), \hat{\sigma}_n^2) = \frac{\varphi(\Phi^{-1}(p))\Phi^{-1}(p)}{\sqrt{2p(1-p)}}. \tag{B.2}$$

(ii) In the case of a Student distribution with mean μ, variance σ^2 and $\nu > 4$ degrees of freedom,

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\, q_n(p), \sqrt{n}\, \hat{\sigma}_n^2) = \sigma^3 \sqrt{\frac{\nu-2}{\nu}}\, q_{\tilde{Y}}(p) \left(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu}\right) \tag{B.3}$$

$$\lim_{n\to\infty} \mathrm{Cor}(q_n(p), \hat{\sigma}_n^2) = \frac{f_{\tilde{Y}}(q_{\tilde{Y}}(p))\, q_{\tilde{Y}}(p) \left(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu}\right)}{\sqrt{\frac{\nu-1}{\nu-4}\, 2\, p(1-p)}} \tag{B.4}$$

As expected, letting $\nu \to \infty$ gives back the results for the Gaussian distribution given in (i).

Likewise, we consider the sample MAD as measure of dispersion in example B.2.

Example B.2 *(i) In the case of a Gaussian distribution we have, for $p \geq 0.5$,*

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n} q_n(p), \sqrt{n}\hat{\theta}_n) = \frac{\sigma^2}{\varphi(\Phi^{-1}(p))} \left(\varphi(\Phi^{-1}(p)) - 1 - p\sqrt{\frac{2}{\pi}}\right) \tag{B.5}$$

$$\lim_{n\to\infty} \mathrm{Cor}(q_n(p), \hat{\theta}_n) = \frac{\varphi(\Phi^{-1}(p)) - (1-p)\sqrt{2/\pi}}{\sqrt{p(1-p)}\,\sqrt{1-2/\pi}}. \tag{B.6}$$

(ii) For the case of a Student distribution with $\nu > 2$ degrees of freedom, for $p \geq 0.5$,

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n} q_n(p), \sqrt{n}\hat{\theta}_n) = \frac{\sigma^2(\nu-2)}{f_{\tilde{Y}}(q_{\tilde{Y}}(p))} \left(\frac{f_{\tilde{Y}}(q_{\tilde{Y}}(p))}{\nu-1} \left(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu}\right) - \frac{\Gamma(\frac{\nu-1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})} (1-p)\right) \tag{B.7}$$

$$\lim_{n\to\infty} \mathrm{Cor}(q_n(p), \hat{\theta}_n) = \frac{\frac{\sqrt{\nu(\nu-2)}}{\nu-1} f_{\tilde{Y}}(q_{\tilde{Y}}(p)) \left(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu}\right) - (1-p)\sqrt{\frac{\nu-2}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}}{\sqrt{p(1-p)}\sqrt{1 - \frac{\nu-2}{\pi}\frac{\Gamma^2(\frac{\nu-1}{2})}{\Gamma^2(\nu/2)}}}. \tag{B.8}$$

As expected, letting $\nu \to \infty$ provides the results for the Gaussian distribution given in (i). This might be not as obvious as in Example B.1, and one might need to recall the asymptotic property of the Gamma function $\lim_{n\to\infty} \frac{\Gamma(n+\alpha)}{\Gamma(n) n^\alpha} = 1$ that we need to use here with $n = \nu\alpha$ and $\alpha = 1/2$.

Recall from Proposition 3.8, the expression for the covariance of the asymptotic distribution. We remarked that it simplifies for symmetric location-scale distribution. In the case of a Gaussian or Student-t distribution, we can show that, for $p \in (0,1)$, we have

Example B.3

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\, q_n(p), \sqrt{n}\, \hat{\xi}_n) = \frac{\sigma^2}{4}\left(1 - p - 2\max\left(3/4 - \max\left(1/4, p\right), 0\right)\right) \tag{B.9}$$

$$\times \begin{cases} \frac{1}{\varphi(\Phi^{-1}(p))\varphi(\Phi^{-1}(3/4))} & \text{if } X \sim \mathcal{N}(\mu, \sigma^2) \\ f_{\tilde{Y}}(q_{\tilde{Y}}(p)) f_{\tilde{Y}}(q_{\tilde{Y}}(3/4)) \frac{\nu}{\nu-2} & \text{if } X \sim t(\mu, (\nu-2)\sigma^2/\nu, \nu) \end{cases} \tag{B.10}$$

$$\lim_{n\to\infty} \mathrm{Cor}(q_n(p), \hat{\xi}_n) = \frac{1 - p - 2\max\left(3/4 - \max\left(1/4, p\right), 0\right)}{\sqrt{p(1-p)}}, \tag{B.11}$$

B.1.2 ... with the location-scale quantile

In analogy to the cases considered in Section 3.6.1, let us assume for the location-scale quantile estimator too, that the mean μ is known. We, again, check the expressions for a Gaussian and Student-t distribution, separately for the different measures of dispersion. In the case with the sample variance, we get

Example B.4 *(i) For the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$:*

$$\lim_{n\to\infty} \mathrm{Cov}\left(\sqrt{n}\, q_{n,\hat{\sigma}}(p), \sqrt{n}\, \hat{\sigma}_n^2\right) = \sigma^3 \Phi^{-1}(p) \tag{B.12}$$

$$\lim_{n\to\infty} \mathrm{Cor}\left(q_{n,\hat{\sigma}}(p), \hat{\sigma}_n^2\right) = \mathrm{sgn}(p - 1/2) \tag{B.13}$$

(ii) For the Student distribution $t(\mu, (\nu-2)\sigma^2/\nu, \nu)$ with $\nu > 4$,

$$\lim_{n\to\infty} \mathrm{Cov}\left(\sqrt{n}\, q_{n,\hat{\sigma}}(p), \sqrt{n}\, \hat{\sigma}_n^2\right) = \sigma^3 q_{\tilde{Y}}(p) \frac{\nu - 1}{\nu - 4} \sqrt{\frac{\nu - 2}{\nu}} \tag{B.14}$$

$$\lim_{n\to\infty} \mathrm{Cor}\left(q_{n,\hat{\sigma}}(p), \hat{\sigma}_n^2\right) = \mathrm{sgn}(p - 1/2) \tag{B.15}$$

While the correlations are already the same - up to a sign - for the Gaussian and Student distributions, we can also check that taking $\nu \to \infty$ in (ii) gives back the result for the asymptotic covariance in the Gaussian case.

Consequently, when considering the sample MAD, it holds

Example B.5 *For $p \geq 0.5$, we have:*
(i) For the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$:

$$\lim_{n\to\infty} \mathrm{Cov}\left(\sqrt{n}\, q_{n,\hat{\sigma}}(p), \sqrt{n}\, \hat{\theta}_n\right) = \frac{\sigma^2 \Phi^{-1}(p)}{\sqrt{2\pi}} \tag{B.16}$$

$$\lim_{n\to\infty} \mathrm{Cor}\left(q_{n,\hat{\sigma}}(p), \hat{\theta}_n\right) = \frac{\mathrm{sgn}(p - 1/2)}{\sqrt{\pi - 2}} \tag{B.17}$$

(ii) For the Student distribution $t(\mu, (\nu-2)\sigma^2/\nu, \nu)$ with $\nu > 4$ degrees of freedom:

$$\lim_{n\to\infty} \mathrm{Cov}\left(\sqrt{n}\, q_{n,\hat{\sigma}}(p), \sqrt{n}\, \hat{\theta}_n\right) = \frac{\sigma^2(\nu-1)(\nu-2)\Gamma(\frac{\nu-1}{2}) q_{\tilde{Y}}(p)}{2(\nu-3)\sqrt{\nu\pi}\,\Gamma(\nu/2)} \tag{B.18}$$

$$\lim_{n\to\infty} \mathrm{Cor}\left(q_{n,\hat{\sigma}}(p), \hat{\theta}_n\right) = \frac{\mathrm{sgn}(p - 1/2)\sqrt{(\nu-1)(\nu-2)}}{(\nu-3)\sqrt{\frac{\pi\Gamma^2(\nu/2)}{\Gamma^2(\frac{\nu-1}{2})} - (\nu-2)\frac{2}{\nu-4}}} \tag{B.19}$$

Example B.6 *(i) For the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, we have:*

$$\lim_{n\to\infty} \mathrm{Cov}\left(\sqrt{n}\, q_{n,\hat{\sigma}}(p), \sqrt{n}\, \hat{\xi}_n\right) = \frac{\sigma^2 \Phi^{-1}(3/4)}{2}\, \Phi^{-1}(p) \tag{B.20}$$

$$\lim_{n\to\infty} \mathrm{Cor}\left(q_{n,\hat{\sigma}}(p), \hat{\xi}_n\right) = 2\sqrt{2}\Phi^{-1}\left(\frac{3}{4}\right) \varphi\left(\Phi^{-1}\left(\frac{3}{4}\right)\right) \mathrm{sgn}\left(p - \frac{1}{2}\right) \tag{B.21}$$

(ii) For the Student distribution $t(\mu, \sigma^2(\nu-2)/\nu, \nu)$ with $\nu > 4$,

$$\lim_{n\to\infty} \mathrm{Cov}\left(\sqrt{n}\, q_{n,\hat{\sigma}}(p), \sqrt{n}\, \hat{\xi}_n\right) = q_{\tilde{Y}}(p) \frac{\sigma^2 q_{\tilde{Y}}(3/4)\left(1 + \frac{q^2_{\tilde{Y}}(3/4)}{\nu}\right)}{2\frac{\nu}{\nu-2}} \tag{B.22}$$

$$\lim_{n\to\infty} \mathrm{Cor}\left(q_{n,\hat{\sigma}}(p), \hat{\xi}_n\right) = \mathrm{sgn}(p - 1/2) \times \frac{2\sqrt{2(\nu-4)}\, f_{\tilde{Y}}(q_{\tilde{Y}}(3/4))\, q_{\tilde{Y}}(3/4)\left(1 + \frac{q^2_{\tilde{Y}}(3/4)}{\nu}\right)}{\sqrt{\nu - 1}} \tag{B.23}$$

B.1.3 Sample quantile - calculations

We think it is worthwhile to provide insight into the calculations of the three examples. To ease the derivation based on Theorem 3.1, note the following three properties (for a rv X with mean μ and variance σ^2 following a location-scale distribution and its normalized version Y with mean 0 and variance 1):

$$\begin{aligned}
\tau_r(|X-\mu|, p) &= \sigma^r \,\mathrm{Cov}(|Y|^r, \mathbb{1}_{(Y > q_Y(p))}) = \sigma^r \left(\mathbb{E}[|Y|^r \mathbb{1}_{(Y>q_Y(p))}] - \mathbb{E}[|Y|^r](1-p)\right) \\
&= \sigma^r \left(-\mathbb{E}[|Y|^r \mathbb{1}_{(Y \le q_Y(p))} + p\,\mathbb{E}[|Y|^r]\right) \\
&= \sigma^r p(\mathbb{E}[|Y|^r] - \mathbb{E}[|Y|^r | Y \le q_Y(p)])
\end{aligned} \tag{B.24}$$

and analogously one can deduce

$$\tau_1(p) = \mathrm{Cov}(X, 1_{(X>q_X(p))}) = \sigma\,\mathrm{Cov}(Y, 1_{(Y>q_Y(p))}) = -\sigma\,\mathbb{E}[Y | Y \le q_Y(p)] \tag{B.25}$$

and finally, note

$$f_X(q_X(p)) = \frac{1}{\sigma} f_Y(q_Y(p)) = \frac{1}{\sigma}\sqrt{\frac{\nu}{\nu-2}} f_{\tilde{Y}}(q_{\tilde{Y}}(p)) \tag{B.26}$$

Recalling, (3.6) and (3.7), we can now compute the examples (with the help of (B.24), (B.25) and (B.26)).

Proof of Example B.1.

(i) Looking at the Gaussian distribution, we can compute, using partial integration,

$$\mathbb{E}[Y^2 | Y \le q_Y(p)] = \frac{1}{p}\int_{-\infty}^{q_Y(p)} y^2 \varphi(y) dy = \frac{1}{p}(-\varphi(\Phi^{-1}(p))\Phi^{-1}(p) + p) \tag{B.27}$$

and, as $f_Y(q_Y(p)) = \varphi(\Phi^{-1}(p))$ in this case, equation (3.7) for $r = 2$, using (B.24) and (B.26) becomes

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\hat{\sigma}_n^2, \sqrt{n} q_n(p)) = \sigma^3 \Phi^{-1}(p),$$

and accordingly, as $\mathbb{E}[Y^4] = 3$ in the case of a standard Gaussian distribution, equation (3.8) becomes $\lim_{n\to\infty} \mathrm{Cor}(\hat{\sigma}_n^2, q_n(p)) = \dfrac{\varphi(\Phi^{-1}(p))\Phi^{-1}(p)}{\sqrt{2p(1-p)}}$.

(ii) Note that the standard Student t-distribution has variance $\nu/(\nu-2)$. Thus, while we would like to work with the random variable Y that is standardised to have variance 1, we will stick to the standard Student random variable $\tilde{Y}$, which is more favourable from an implementation point of view. Thus, when considering equations (3.7),(3.8), we need to be careful about this. Analogously to the Gaussian distribution, we look at

$$\mathbb{E}[Y^2|Y \leq q_Y(p)] = \frac{1}{p}\int_{-\infty}^{q_Y(p)} y^2 f_Y(y)dy = \frac{1}{p}\frac{\nu-2}{\nu}\int_{-\infty}^{q_{\tilde{Y}}(p)} z^2 f_{\tilde{Y}}(z)dz, \tag{B.28}$$

which can be seen as the second moment of a truncated Student distribution. Results for this are readily available. Unfortunately, the provided results are sometimes wrong (see e.g. [84] for the univariate case, [6] for the multivariate case or [82] when using a truncated skew-Student distribution) So we turn to a correct derivation, found in [74], where we have that, for real-valued truncation points $-\infty < a < b < \infty, \nu > 2$

$$\frac{1}{F_{\tilde{Y}}(b) - F_{\tilde{Y}}(a)}\int_a^b z^2 f_{\tilde{Y}}(x)dx = \frac{\nu(\nu-1)}{\nu-2} \times \frac{F_{\tilde{Y},\nu-2}\left(b\sqrt{\frac{\nu-2}{\nu}}\right) - F_{\tilde{Y},\nu-2}\left(a\sqrt{\frac{\nu-2}{\nu}}\right)}{F_{\tilde{Y}}(b) - F_{\tilde{Y}}(a)} - \nu \tag{B.29}$$

where $F_{\tilde{Y}}$ denotes the standard Student cdf with ν degrees of freedom and $F_{\tilde{Y},k}$ the standard Student cdf with k degrees of freedom. Note that the condition $-\infty < a < b < \infty$ in [74] is actually not necessary for their proof, i.e. we can have $-\infty \leq a < b \leq \infty$.

The above expression can be modified (which yields a simplification for our calculations) using a recurrence relation for the cummulative distribution function of the Student distribution (as presented e.g. in [81] in the more general case of a skew-t distribution), here for $\nu > 2, a \in \mathbb{R}$:

$$F_{\tilde{Y},\nu-2}\left(a\sqrt{\frac{\nu-2}{\nu}}\right) = F_{\tilde{Y}}(a) - \frac{\Gamma(\frac{\nu-1}{2})\nu^{\frac{\nu-2}{2}}}{2\sqrt{\pi}\Gamma(\frac{\nu}{2})}a(\nu+a^2)^{-\frac{\nu-1}{2}},$$

which for our purposes, is simplified in a form where we get the density involved:

$$\begin{aligned} F_{\tilde{Y},\nu-2}\left(a\sqrt{\frac{\nu-2}{\nu}}\right) &= F_{\tilde{Y}}(a) - \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{a^2}{\nu}\right)^{-\frac{\nu+1}{2}}\frac{(1+\frac{a^2}{\nu})a}{\nu-1} \\ &= F_{\tilde{Y}}(a) - \frac{1}{\nu-1}f_{\tilde{Y}}(a)\left(1+\frac{a^2}{\nu}\right)a. \end{aligned} \tag{B.30}$$

Combining this latter equation and (B.29), taking $a = -\infty$, $b = q_{\tilde{Y}}(p)$, we obtain:

$$\begin{aligned} \int_{-\infty}^{q_{\tilde{Y}}(p)} y^2 f_{\tilde{Y}}(y)dy &= \frac{\nu(\nu-1)}{\nu-2}\left(F_{\tilde{Y}}(q_{\tilde{Y}}(p)) - f_\nu(q_{\tilde{Y}}(p))\frac{(1+\frac{q^2_{\tilde{Y}}(p))}{\nu})q_{\tilde{Y}}(p)}{\nu+1}\right) - \nu\, p \\ &= \frac{\nu}{\nu-2}\Big(p\, - f_{\tilde{Y}}(q_{\tilde{Y}}(p))q_{\tilde{Y}}(p)(1+q^2_{\tilde{Y}}(p)/\nu)\Big), \end{aligned} \tag{B.31}$$

or, equivalently,

$$\int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(y)dy = \frac{\nu}{\nu-2} - \frac{\nu}{\nu-2}\left(p\, - f_{\tilde{Y}}(q_{\tilde{Y}}(p))q_{\tilde{Y}}(p)(1+q^2_{\tilde{Y}}(p)/\nu)\right). \tag{B.32}$$

Hence, recalling (B.28), we can conclude that

$$\mathbb{E}[Y^2|Y \leq q_Y(p)] = 1 - \frac{1}{p} f_{\tilde{Y}}(F_{\tilde{Y}}(p)) q_{\tilde{Y}}(p)(1 + q_{\tilde{Y}}^2(p)/\nu), \tag{B.33}$$

and, as $f_Y(q_Y(p)) = \sqrt{\frac{\nu}{\nu-2}} f_{\tilde{Y}}(q_{\tilde{Y}}(p))$ in this case, we get from equation (3.7) for $r = 2$, using (B.24) and (B.26) that

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}\hat{\sigma}_n^2, \sqrt{n} q_n(p)) = \frac{\sigma^3 q_{\tilde{Y}}(p)(1 + q_{\tilde{Y}}^2(p)/\nu)}{\sqrt{\frac{\nu}{\nu-2}}}.$$

For the correlation, one needs to recall that in the case of the Student distribution

$$\mu_4 = \mathbb{E}[(\sigma Y)^4] = \sigma^4 \left(\frac{\nu-2}{\nu}\right)^2 \mathbb{E}[\tilde{Y}^4] = \sigma^4 \left(\frac{\nu-2}{\nu}\right)^2 \frac{3\nu^2}{(\nu-2)(\nu-4)} = 3\sigma^4 \frac{\nu-2}{\nu-4}, \tag{B.34}$$

i.e. $\mathbb{E}[Y^4] = 3\frac{\nu-2}{\nu-4}$, which gives us, by (3.8), $\lim_{n\to\infty} \mathrm{Cor}(\hat{\sigma}_n^2, q_n(p)) = \frac{f_{\tilde{Y}}(q_{\tilde{Y}}(p)) q_{\tilde{Y}}(p)(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu})}{\sqrt{2\frac{\nu-1}{\nu-4} p(1-p)}}$.

□

Proof of Example B.2. For the case with the sample MAD, the main task is to find an explicit expression for either $\mathbb{E}[|Y||Y \leq q_Y(p)]$ or $\mathbb{E}[|Y||Y > q_Y(p)]$. The easiest way to compute this is to restrict oneself to the case of $p < 0.5$ in the former or $p \geq 0.5$ in the latter. As we know that the correlation (and hence also the covariance) is point-symmetric in the cases considered, we can then deduce the case of $p \geq 0.5$ or $p < 0.5$ respectively. As in applications we are usually interested in high values p of the quantile, we show the expressions for $p > 0.5$:

As by construction, $Y \geq 0$ a-s. for $p \geq 0.5$, we have

$$\mathbb{E}[|Y||Y > q_Y(p)] = \mathbb{E}[Y|Y > q_Y(p)] = \frac{\mathbb{E}[Y \mathbb{1}_{(Y > q_Y(p))}]}{1-p},$$

which simply is the truncated first moment of the corresponding distribution.

(i) In the case of the Gaussian distribution, $\theta = \mathbb{E}|X - \mu| = \sigma\sqrt{\frac{2}{\pi}}$. Hence $\mathbb{E}\,|Y| = \theta/\sigma = \sqrt{\frac{2}{\pi}}$ and thus

$$\mathrm{Var}(|Y|) = 1 - \frac{\theta^2}{\sigma^2} = 1 - \frac{2}{\pi} \tag{B.35}$$

For $p \geq 0.5$, we have by direct calculation that

$$\mathbb{E}[|Y|\mathbb{1}_{(Y > q_Y(p))}] = \mathbb{E}[Y \mathbb{1}_{(Y > q_Y(p))}] = \varphi(\Phi^{-1}(p)). \tag{B.36}$$

Hence, we have

$$\lim_{n\to\infty} \mathrm{Cov}(\sqrt{n} q_n, \sqrt{n}\hat{\theta}_n) = \frac{\sigma^2 \left(\varphi(\Phi^{-1}(p)) - (1-p)\sqrt{\frac{2}{\pi}}\right)}{\varphi(\Phi^{-1}(p))}$$

$$\lim_{n\to\infty} \mathrm{Cor}(q_n, \hat{\theta}_n) = \frac{\varphi(\Phi^{-1}(p)) - (1-p)\sqrt{\frac{2}{\pi}}}{\sqrt{p(1-p)}\sqrt{1 - \frac{2}{\pi}}}.$$

(ii) For the case of a Student distribution, we need to pay attention as $\theta = \mathbb{E}|X-\mu| = \sigma\,\mathbb{E}|Y| = \frac{\sigma}{\sqrt{\frac{\nu}{\nu-2}}}\,\mathbb{E}|\tilde{Y}|$. Recall that $E|\tilde{Y}| = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}$, hence we have $\mathbb{E}|Y| = \frac{\theta}{\sigma} = \sqrt{\frac{\nu}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}\sqrt{\frac{\nu-2}{\nu}}$ and thus

$$\mathrm{Var}(|Y|) = 1 - \frac{\theta^2}{\sigma^2} = 1 - \frac{\nu-2}{\pi}\frac{\Gamma^2(\frac{\nu-1}{2})}{\Gamma^2(\frac{\nu}{2})} \tag{B.37}$$

Finally, using the formula for the truncated moments from [74], we get

$$\begin{aligned}
\mathbb{E}[\tilde{Y}\mathbb{1}_{(\tilde{Y}>q_{\tilde{Y}}(p))}] &= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\nu\pi}}\frac{\nu}{\nu-1}\left(1+\frac{q^2_{\tilde{Y}}(p)}{\nu}\right)^{-(\nu-1)/2} \\
&= \frac{\nu}{\nu-1}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\nu/2)\sqrt{\nu\pi}}\left(1+\frac{q^2_{\tilde{Y}}(p)}{\nu}\right)\left(1+\frac{q^2_{\tilde{Y}}(p)}{\nu}\right)^{-(\nu+1)/2} \\
&= \frac{\nu}{\nu-1} f_{\tilde{Y}}(q_{\tilde{Y}}(p))\,(1+q^2_{\tilde{Y}}(p)/\nu).
\end{aligned} \tag{B.38}$$

Hence, as $\mathbb{E}[Y\mathbb{1}_{(Y>F_Y^{-1}(p))}] = \frac{1}{\sqrt{\frac{\nu}{\nu-2}}}\mathbb{E}[\tilde{Y}\mathbb{1}_{(\tilde{Y}>q_{\tilde{Y}}(p))}]$, we conclude

$$\lim_{n\to\infty}\mathrm{Cov}(\sqrt{n}q_n, \sqrt{n}\hat{\theta}_n) = \frac{\sigma^2\left(\frac{\sqrt{\nu(\nu-2)}}{\nu-1}f_{\tilde{Y}}(q_{\tilde{Y}}(p))(1+q^2_{\tilde{Y}}(p)/\nu) - (1-p)\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}\right)}{f_{\tilde{Y}}(q_{\tilde{Y}}(p))\sqrt{\nu/(\nu-2)}}$$

$$\text{and}\quad \lim_{n\to\infty}\mathrm{Cor}(q_n, \hat{\theta}_n) = \frac{\frac{\sqrt{\nu(\nu-2)}}{\nu-1}f_{\tilde{Y}}(q_{\tilde{Y}}(p))(1+q^2_{\tilde{Y}}(p)/\nu) - (1-p)\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}}{\sqrt{p(1-p)}\sqrt{1-\frac{\nu-2}{\pi}\frac{\Gamma^2(\frac{\nu-1}{2})}{\Gamma^2(\nu/2)}}}. \qquad \square$$

Proof of Example B.3. As the Gaussian and Student distribution are both symmetric, the correlation does not depend on the underlying distribution and is directly computable. Equally we deduce the covariances by directly plugging-in. In the Student case, just recall that $Y = \frac{\tilde{Y}}{\sqrt{\nu/(\nu-2)}}$ hence $f_Y(q_Y(p)) = \sqrt{\frac{\nu}{\nu-2}}f_{\tilde{Y}}(q_{\tilde{Y}}(p))$. $\square$

B.1.4 Location scale quantile - calculations

As we did not present the limit theorems for the location-scale quantile with known μ in Chapter 3 (although we could derive them from the general case), we cannot present the calculations here. Nevertheless, the interested reader can find the derivations in Appendix B.2 of our working paper [32].

B.2 Full simulation study with Pearson correlation

See the online-appendix in [34].

B.3 Simulation study for rank correlations

B.3.1 Known asymptotics for sample estimators

See the online-appendix in [34].

B.3.2 Simulation study for rank correlation

See the online-appendix in [34].

Appendix C

Supplements to Chapter 4

C.1 Generalization of Table 4.2

See the online-appendix in [34].

Appendix D

Supplements to Chapter 5

D.1 Explicit formulas for risk measure pro-cyclicality (iid case)

Explicit formulas corresponding to examples in Section 5.4.1

In (the plots of) Section 5.4.1 we considered the degree of pro-cyclicality for different risk measure (estimators), measure of dispersion estimators and different iid distributions.

Recall that as measure of dispersion estimators we considered the sample variance or sample MAD, and looked at the following risk measure estimator $\widehat{\mathrm{VaR}}_n(p)$, $\widetilde{\mathrm{ES}}_{n,k}(p)$ for $k = 4, 50$ and $k = \infty$ as well as $e_n(p)$. As underlying distributions we considered either a Gaussian or a Student-t distribution. We measured the pro-cyclicality as the degree of correlation in the asymptotic distribution between the log-ratio of risk measure estimators and the measure of dispersion estimator.

In Table D.1 we present the expressions for an underlying Gaussian distribution and then in Table D.2 for a Student distribution with ν degrees of freedom.

To show how we obtain the expressions in Tables D.1 and D.2, we only need to focus on the quantities with $\widetilde{\mathrm{ES}}_{n,\infty}$.

Indeed, for the correlations including the sample VaR, there is nothing to do as they are simply the correlation of the asymptotic distribution of the sample quantile (with the corresponding measure of dispersion) - we computed this already, see Appendix B.1.

The same remarks hold for the expectile estimator, as it is the sample quantile at level $\kappa^{-1}(p)$ with $\kappa(\alpha)$ being defined in (5.6), which simplifies for location-scale distributions, as follows:

$$\kappa(\alpha) = \frac{\alpha q_Y(\alpha) - \int_{-\infty}^{q_Y(\alpha)} y dF_Y(y)}{-2\int_{-\infty}^{q_Y(\alpha)} y dF_Y(y) - (1-2\alpha)q_Y(\alpha)}.$$

This gives us, in the case of the Gaussian distribution (recall the first truncated moment, (B.36)),

$$\kappa_{norm}(p) = \frac{p\Phi^{-1}(p) + \varphi(\Phi^{-1}(p))}{2\varphi(\Phi^{-1}(p)) - (1-2p)\Phi^{-1}(p)}.$$

For the Student distribution (assumed to be with mean 0, and recalling the first truncated moment, (B.38)), we obtain

$$\kappa_{stud}(p) = \frac{pq_{\tilde{Y}}(p) + \frac{\nu}{\nu-1} f_{\tilde{Y}}(q_{\tilde{Y}}(p))(1 + q^2_{\tilde{Y}}(p)/\nu)}{2\frac{\nu}{\nu-1} f_{\tilde{Y}}(q_{\tilde{Y}}(p))(1 + q^2_{\tilde{Y}}(p)/\nu) - (1-2p)q_{\tilde{Y}}(p)}.$$

The ES estimator $\widetilde{\mathrm{ES}}_{n,k}$ is a sum of sample quantiles, thus this is also dealt with, by Appendix B.1. We are only left with $\widetilde{\mathrm{ES}}_{n,\infty}(p)$ which should be understood as $\widetilde{\mathrm{ES}}_{n,\infty}(p) = \frac{1}{1-p}\int_p^1 q_n(u)du$.

Because of the more compact integral representation of the asymptotic correlation we keep in the tables the correlation with $\widetilde{\mathrm{ES}}_{n,\infty}(p)$ as in [32].

The explicit solutions of this integral representation are very lengthy and can be found after presenting the two tables. First, in Table D.1 the correlations of the asymptotic distributions for a Gaussian distribution.

TABLE D.1: Correlations in the joint asymptotic distribution between the log-ratios of each, three risk measure estimators, and the two measures of dispersion estimator in the case of a Gaussian distribution

Correlation	Sample Variance	Sample MAD
$\mathrm{VaR}_n(p)$	$\frac{-1}{\sqrt{2}}\frac{\varphi(\Phi^{-1}(p))\left\lvert\Phi^{-1}(p)\right\rvert}{\sqrt{2p(1-p)}}$ (D.1)	$\frac{-1}{\sqrt{2}}\frac{\left\lvert\varphi(\Phi^{-1}(p))-(1-p)\sqrt{2/\pi}\right\rvert}{\sqrt{p(1-p)}\sqrt{1-2/\pi}}$ (D.2)
$\widetilde{\mathrm{ES}}_{n,\infty}(p)$	$\frac{-1}{\sqrt{2}}\frac{\left\lvert\int_p^1\Phi^{-1}(u)du\right\rvert}{2\sqrt{\int_p^1\int_v^1\frac{v(1-u)}{\varphi(\Phi^{-1}(u))\varphi(\Phi^{-1}(v))}dudv}}$ (D.3)	$\frac{-1}{\sqrt{2}}\frac{\left\lvert 1-p-\int_p^1\frac{1-u}{\varphi(\Phi^{-1}(u))\sqrt{2/\pi}}du\right\rvert}{2\sqrt{\left(\frac{1}{2}-\frac{1}{\pi}\right)\int_p^1\int_v^1\frac{v(1-u)}{\varphi(\Phi^{-1}(v))\varphi(\Phi^{-1}(u))}dudv}}$ (D.4)
$e_n(p)$	$\frac{-1}{\sqrt{2}}\frac{\varphi(\Phi^{-1}(\kappa^{-1}(p)))\left\lvert\Phi^{-1}(\kappa^{-1}(p))\right\rvert}{\sqrt{2\kappa^{-1}(p)(1-\kappa^{-1}(p))}}$ (D.5)	$\frac{-1}{\sqrt{2}}\frac{\left\lvert\varphi(\Phi^{-1}(\kappa^{-1}(p)))-(1-\kappa^{-1}(p))\sqrt{2/\pi}\right\rvert}{\sqrt{\kappa^{-1}(p)(1-\kappa^{-1}(p))}\sqrt{1-2/\pi}}$ (D.6)

The corresponding correlation, now for an underlying Student distribution with ν degrees of freedom, are summarised in Table D.2. The expressions look more complex than in the case with the Gaussian distribution. Still, we recover the Gaussian expressions for $\nu \to \infty$. For this, recall that $\Gamma(.)$ is the Gamma function, i.e.

$$\Gamma(x) := \begin{cases} (x-1)! & \text{for integers } x > 0, \\ \sqrt{\pi}\frac{(2x-2)!!}{2^{\frac{2x-1}{2}}} & \text{for half-integers } x, \text{ i.e. odd integer-multiples of } \frac{1}{2}, \end{cases}$$

where ! and !! denote the factorial and double-factorial function, respectively. Further, one might need to recall the asymptotic property of the Gamma function $\lim_{n\to\infty}\frac{\Gamma(n+\alpha)}{\Gamma(n)n^{\alpha}} = 1$ that we need to use here with $n = \nu\alpha$ and $\alpha = 1/2$.

TABLE D.2: Correlations in the joint asymptotic distribution between the log-ratios of each, three risk measure estimators, and the two measures of dispersion estimator in the case of a Student distribution with ν degrees of freedom

Correlation	Sample Variance	Sample MAD
$\mathrm{VaR}_n(p)$	$\frac{-1}{\sqrt{2}} \frac{f_{\tilde{Y}}(q_{\tilde{Y}}(p))\,\lvert q_{\tilde{Y}}(p)\rvert \left(1+\frac{q^2_{\tilde{Y}}(p)}{\nu}\right)}{\sqrt{\frac{\nu-1}{\nu-4}\,2\,p(1-p)}}$ (D.7)	$\frac{-1}{\sqrt{2}} \frac{\left\lvert \frac{\sqrt{\nu(\nu-2)}}{\nu-1} f_{\tilde{Y}}(q_{\tilde{Y}}(p)) \left(1+\frac{q^2_{\tilde{Y}}(p)}{\nu}\right) - (1-p)\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}\right\rvert}{\sqrt{p(1-p)}\sqrt{1-\frac{\nu-2}{\pi}\frac{\Gamma^2(\frac{\nu-1}{2})}{\Gamma^2(\nu/2)}}}$ (D.8)
$\widetilde{ES}_{n,\infty}(p)$	$\frac{-1}{\sqrt{2}} \frac{\left\lvert \int_p^1 q_{\tilde{Y}}(u)\left(1+\frac{q^2_{\tilde{Y}}(u)}{\nu}\right)du\right\rvert}{2\sqrt{\frac{\nu-1}{\nu-4}\int_p^1\int_v^1 \frac{v(1-u)}{f_{\tilde{Y}}(q_{\tilde{Y}}(v))f_{\tilde{Y}}(q_{\tilde{Y}}(u))}dudv}}$ (D.9)	$\frac{-1}{\sqrt{2}} \frac{\left\lvert \int_p^1 \sqrt{\nu-2}\left(\frac{\sqrt{\nu}}{\nu-1}\left(1+\frac{q^2_{\tilde{Y}}(u)}{\nu}\right) - \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{(1-u)}{\sqrt{\pi}f_{\tilde{Y}}(q_{\tilde{Y}}(u))}\right)\right\rvert}{\sqrt{2\int_p^1\int_v^1 \frac{v(1-u)}{f_{\tilde{Y}}(q_{\tilde{Y}}(v))f_{\tilde{Y}}(q_{\tilde{Y}}(u))}dudv}\sqrt{1-\frac{\nu-2}{\pi}\frac{\Gamma((\nu-1)/2)^2}{\Gamma(\nu/2)^2}}}$ (D.10)
$e_n(p)$	$\frac{-1}{\sqrt{2}} \frac{f_{\tilde{Y}}(q_{\tilde{Y}}(\kappa^{-1}(p)))\;\lvert q_{\tilde{Y}}(\kappa^{-1}(p))\rvert \left(1+\frac{q^2_{\tilde{Y}}(\kappa^{-1}(p))}{\nu}\right)}{\sqrt{\frac{\nu-1}{\nu-4}\,2\,\kappa^{-1}(p)(1-\kappa^{-1}(p))}}$ (D.11)	$\frac{-1}{\sqrt{2}} \frac{\left\lvert \frac{\sqrt{\nu(\nu-2)}}{\nu-1} f_{\tilde{Y}}(q_{\tilde{Y}}(\kappa^{-1}(p)))\left(1+\frac{q^2_{\tilde{Y}}(\kappa^{-1}(p))}{\nu}\right) - (1-\kappa^{-1}(p))\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\nu/2)}\right\rvert}{\sqrt{\kappa^{-1}(p)(1-\kappa^{-1}(p))}\sqrt{1-\frac{\nu-2}{\pi}\frac{\Gamma^2(\frac{\nu-1}{2})}{\Gamma^2(\nu/2)}}}$ (D.12)

Derivations: ES with Sample Variance or sample MAD

First, realise that for location-scale distributions we can rewrite the correlation of the asymptotic distribution of $\widetilde{\mathrm{ES}}_{n,\infty}$ and either the sample mean or sample variance (i.e. just considering $r = 1, 2$.) as

$$\lim_{n\to\infty} \mathrm{Cor}(\widetilde{\mathrm{ES}}_n(p), \hat{m}(X,n,r)) = \frac{\int_p^1 \frac{\tau_r(|Y|,u)+(2-r)(2F_Y(0)-1)\tau_1(Y,u)}{f_Y(q_Y(u))}du}{\sqrt{2(\int_p^1\int_v^1 \frac{v(1-u)}{f_Y(q_Y(v))f_Y(q_Y(u))}dudv)}\sqrt{\mathrm{Var}(|Y|^r+(2-r)(2F_Y(0)-1)Y)}}. \tag{D.13}$$

In a next step, we look separately at the sample variance ($r = 2$) and sample MAD ($r = 1$), each for the Gaussian and Student distribution separately.

▷ *ES with Sample Variance - Gaussian distribution.* Recall from equation (B.1) the asymptotic covariance between sample quantile and sample variance in the case of the Gaussian distribution (in this case of a standard normal distribution with $\sigma^2 = 1$). Thus, the numerator of (D.13) simplifies as $\int_p^1 \Phi^{-1}(u)du$, which we already solved by change of variables before (truncated first moment, (B.36))

$$\int_p^1 \Phi^{-1}(u)du = \int_{\Phi^{-1}(v)}^{\infty} y\varphi(y)dy = \varphi(\Phi^{-1}(p))$$

More work is needed for the computation of the double-integral in the denominator of (D.13). We first consider the inner integral, $\int_v^1 \frac{1-u}{\varphi(\Phi^{-1}(u))}du$. Again, change of variable and partial integration give the following:

$$\begin{aligned}\int_v^1 \frac{1-u}{\varphi(\Phi^{-1}(u))}du &= \int_{\Phi^{-1}(v)}^{+\infty}(1-\Phi(y))\,dy = y(1-\Phi(y))|_{\Phi^{-1}(v)}^{\infty} + \int_{\Phi^{-1}(v)}^{\infty} y\varphi(y)dy \\ &= -(1-v)\Phi^{-1}(v) + \varphi(\Phi^{-1}(v)).\end{aligned} \tag{D.14}$$

Thus, plugging this in, the double integral transforms to

$$\int_p^1 \int_v^1 \frac{v(1-u)}{f_Y(q_Y(v))f_Y(q_Y(u))}dudv = -\int_p^1 v\frac{\Phi^{-1}(v)}{\varphi(\Phi^{-1}(v))}dv + \int_p^1 v^2\frac{\Phi^{-1}(v)}{\varphi(\Phi^{-1}(v))} + \frac{1}{2}v^2|_p^1. \quad \text{(D.15)}$$

We compute the two integrals of (D.15) one after the other, again using change of variables and partial integration (with $\varphi'(y) = -y\varphi(y)$), obtaining:

$$\begin{aligned}\int_p^1 v\frac{\Phi^{-1}(v)}{\varphi(\Phi^{-1}(v))}dv &= \int_{\Phi^{-1}(p)}^{\infty} y\,\Phi(y)dy = \frac{1}{2}y^2\Phi(y)|_{\Phi^{-1}(p)}^{\infty} - \int_{\Phi^{-1}(p)}^{\infty}\frac{y^2}{2}\varphi(y)dy \\ &= \frac{1}{2}y^2\Phi(y)|_{\Phi^{-1}(p)}^{\infty} - \frac{1}{2}(y(-\varphi(y))|_{\Phi^{-1}(p)}^{\infty} + \frac{1}{2}\int_{\Phi^{-1}(p)}^{+\infty}(-\varphi(y))dy \\ &= \frac{1}{2}y^2\Phi(y)|_{\Phi^{-1}(p)}^{\infty} - \frac{1}{2}\Phi^{-1}(p)\varphi(\Phi^{-1}(p)) - \frac{1}{2}(1-p)\end{aligned}$$

and by the same techniques it follows for the second integral

$$\begin{aligned}\int_p^1 v^2\frac{\Phi^{-1}(v)}{\varphi(\Phi^{-1}(v))}dv &= \int_{\Phi^{-1}(p)}^{\infty} y\,\Phi^2(y)dy = \frac{y^2}{2}\Phi^2(y)|_{\Phi^{-1}(p)}^{\infty} - \int_{\Phi^{-1}(p)}^{\infty} y^2\varphi(y)\Phi(y)dy \\ &= \frac{1}{2}y^2\Phi^2(y)|_{\Phi^{-1}(p)}^{\infty} + y\Phi(y)\varphi(y)|_{\Phi^{-1}(p)}^{\infty} - \int_{\Phi^{-1}(p)}^{\infty}\varphi(y)\left(\Phi(y) + y\varphi(y)\right)dy \\ &= \frac{1}{2}y^2\Phi^2(y)|_{\Phi^{-1}(p)}^{\infty} + y\varphi(y)\Phi(y)|_{\Phi^{-1}(p)}^{\infty} - \frac{1}{2}\Phi^2(y)|_{\Phi^{-1}(p)}^{\infty} + \frac{1}{2}\varphi^2(y)|_{\Phi^{-1}(p)}^{\infty}.\end{aligned}$$

Hence, putting this together gives

$$\text{(D.15)} = \frac{1}{2}\Phi^{-1}(p)^2(p-p^2) + \Phi^{-1}(p)\varphi(\Phi^{-1}(p))(\frac{1}{2}-p) + \frac{1}{2}(1-p) - \frac{1}{2}\varphi^2(\Phi^{-1}(p)) \quad \text{(D.16)}$$

and overall, for (D.13) in the case of correlation with the sample variance (as $\mathbb{E}[Y^4] = 3$ for the Gaussian distribution, i.e. $\text{Var}(Y^2) = 2$),

$$\lim_{n\to\infty}\text{Cor}(\widetilde{\text{ES}}_n(p), \sigma_n^2) = \frac{\varphi(\Phi^{-1}(p))}{\sqrt{(\Phi^{-1}(p))^2(p-p^2) - \varphi^2(\Phi^{-1}(p)) + \Phi^{-1}(p)\varphi(\Phi^{-1}(p))(1-2p) + 1 - p}\sqrt{2}}.$$

▷ *ES with Sample Variance - Student distribution.* The case of the Student distribution works analogously, but needs in some details more care. Recall from equation (B.3) the asymptotic covariance between sample quantile and sample variance in the case of the Student distribution with $\nu > 3$, thus the numerator of (D.13) equals to:

$$\int_p^1 \sqrt{\frac{\nu-2}{\nu}}\,q_{\tilde{Y}}(u)\left(1 + q_{\tilde{Y}}^2(u)/\nu\right)du = \sqrt{\frac{\nu-2}{\nu}}\left(\int_{q_{\tilde{Y}}(v)}^{\infty} y f_{\tilde{Y}}(y)dy + \frac{1}{\nu}\int_{q_{\tilde{Y}}(v)}^{\infty} y^3 f_{\tilde{Y}}(y)dy\right) \quad \text{(D.17)}$$

where the equality follows by change of variables. This first truncated moment was computed already in equation (B.38). For the third truncated moment we, again, use and simplify the formula provided in [74], which gives us:

$$E[\tilde{Y}^3\mathbb{1}_{(\tilde{Y}>q_{\tilde{Y}}(p))}] = \int_{q_{\tilde{Y}}(p)}^{\infty} y^3 f_{\tilde{Y}}(y)dy = \frac{\nu^2}{(\nu-1)(\nu-3)}f_{\tilde{Y}}(q_{\tilde{Y}}(p))\left(2 + q_{\tilde{Y}}^2(p)\frac{\nu-1}{\nu}\right)\left(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu}\right),$$

giving us for (D.17) (and hence the numerator of (D.13)):

$$(\text{D.17}) = \sqrt{\frac{\nu-2}{\nu}} \int_p^1 q_{\tilde{Y}}(u)(1+q_{\tilde{Y}}^2(u)/\nu)du = \frac{\sqrt{\nu(\nu-2)}}{\nu-3}\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}(q_{\tilde{Y}}(p)).$$

We now turn to the double-integral in the denominator of (D.13), expressed as $\frac{\nu-2}{\nu}\int_p^1\int_v^1 \frac{v(1-u)}{f_{\tilde{Y}}(q_{\tilde{Y}}(v))f_{\tilde{Y}}(q_{\tilde{Y}}(u))}dudv$. The inner integral follows one-to-one from the Gaussian case above (as we already know the truncated first moment, see (B.38)):

$$\int_v^1 \frac{1-u}{f_{\tilde{Y}}(q_{\tilde{Y}}(u))}du = \int_{q_{\tilde{Y}}(v)}^{+\infty}(1-F_{\tilde{Y}}(y))\,dy = -(1-v)q_{\tilde{Y}}(v) + \frac{\nu}{\nu-1}f_{\tilde{Y}}(q_{\tilde{Y}}(v)))(1+q_{\tilde{Y}}^2(v)/\nu). \tag{D.18}$$

Plugging this in the double integral, transforms to

$$\int_p^1\int_v^1 \frac{v(1-u)}{f_{\tilde{Y}}(q_{\tilde{Y}}(v))f_{\tilde{Y}}(q_{\tilde{Y}}(u))}dudv = \int_p^1 v(v-1)\frac{q_{\tilde{Y}}(v)}{f_{\tilde{Y}}(q_{\tilde{Y}}(v))}dv + \frac{\nu}{\nu-1}\int_p^1 v\left(1+q_{\tilde{Y}}^2(v)/\nu\right)dv. \tag{D.19}$$

But, using change of variables, then partial integration and then the knowledge of the truncated second moment (B.32), we obtain:

$$\begin{aligned}
&\int_p^1 (v^2-v)\frac{q_{\tilde{Y}}(v)}{f_{\tilde{Y}}(q_{\tilde{Y}}(v))}dv = \int_{q_{\tilde{Y}}(p)}^{\infty}(F_{\tilde{Y}}^2(y)-F_{\tilde{Y}}(y))\,y\,dy \\
&= \frac{1}{2}y^2F_{\tilde{Y}}(y)(F_{\tilde{Y}}(y)-1)|_{q_{\tilde{Y}}(p)}^{\infty} - \int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(y)F_{\tilde{Y}}(y)dy + \frac{1}{2}\int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(y)dy \\
&= \frac{1}{2}\left(\frac{\nu}{\nu-2} + p\left(q_{\tilde{Y}}^2(p) - \frac{\nu}{\nu-2}\right) - q_{\tilde{Y}}^2(p)\,p^2 + \frac{\nu}{\nu-2}f_{\tilde{Y}}(q_{\tilde{Y}}(p))q_{\tilde{Y}}(p)(1+q_{\tilde{Y}}^2(p)/\nu)\right) \\
&- \int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(y)F_{\tilde{Y}}(y)dy \\
&\text{and} \quad \int_p^1 vq_{\tilde{Y}}^2(v)dv = \int_{q_{\tilde{Y}}(p)}^{\infty} F_{\tilde{Y}}(y)y^2 f_{\tilde{Y}}(y)dy,
\end{aligned}$$

hence, (D.19) becomes:

$$\frac{1}{2}\left(\frac{\nu(2\nu-3)}{(\nu-2)(\nu-1)} + p\left(q_{\tilde{Y}}^2(p) - \frac{\nu}{\nu-2}\right) - p^2\left(q_{\tilde{Y}}^2(p) + \frac{\nu}{\nu-1}\right) + \frac{\nu}{\nu-2}f_{\tilde{Y}}(q_{\tilde{Y}}(p))q_{\tilde{Y}}(p)\left(1+\frac{q_{\tilde{Y}}^2(p)}{\nu}\right)\right) - \frac{\nu-2}{\nu-1}\int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(y)F_{\tilde{Y}}(y)dy. \tag{D.20}$$

So, we are left with the integral of (D.20). Again, using partial integration with the fact that the antiderivative equals $\int y f_{\tilde{Y}}(y)dy = \frac{-\nu}{\nu-1} f_{\tilde{Y}}(y)(1+y^2/\nu)$, we obtain

$$\begin{aligned}\int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(y) F_{\tilde{Y}}(y)dy &= \frac{-\nu}{\nu-1} f_{\tilde{Y}}(y)(1+y^2/\nu) y F_{\tilde{Y}}(y)\Big|_{q_{\tilde{Y}}(p)}^{\infty} \\ &\quad + \frac{\nu}{\nu-1}\int_{q_{\tilde{Y}}(p)}^{\infty} f_{\tilde{Y}}(y)(1+y^2/\nu)(F_{\tilde{Y}}(y) + y f_{\tilde{Y}}(y))dy \\ &= \frac{\nu}{\nu-1}\left\{ p\, q_{\tilde{Y}}(p)\left(1+q_{\tilde{Y}}^2(p)/\nu\right) f_{\tilde{Y}}(q_{\tilde{Y}}(p)) + \int_{q_{\tilde{Y}}(p)}^{\infty}(1+y^2/\nu)F_{\tilde{Y}}(y) f_{\tilde{Y}}(y)dy \right. \\ &\left. + \int_{q_{\tilde{Y}}(p)}^{\infty} y(1+y^2/\nu) f_{\tilde{Y}}^2(y)dy \right\}. \end{aligned} \tag{D.21}$$

Since we have $\left(1+\frac{y^2}{\nu}\right) f_{\tilde{Y}}(y) = \frac{\nu-1}{\sqrt{\nu(\nu-2)}} f_{\tilde{Y},\nu-2}\left(y\sqrt{\frac{\nu-2}{\nu}}\right)$ (using the notation $f_{\tilde{Y},k}$ introduced after (B.29)), we can write, proceeding again with partial integration,

$$\begin{aligned}\int_{q_{\tilde{Y}}(p)}^{\infty}(1+y^2/\nu) f_{\tilde{Y}}(y) F_{\tilde{Y}}(y)dy &= \frac{\nu-1}{\sqrt{\nu(\nu-2)}}\int_{q_{\tilde{Y}}(p)}^{\infty} f_{\tilde{Y},\nu-2}\left(y\sqrt{\frac{\nu-2}{\nu}}\right) F_{\tilde{Y}}(y)dy \\ &= \frac{\nu-1}{\nu-2}\left\{ F_{\tilde{Y},\nu-2}\left(y\sqrt{\frac{\nu-2}{\nu}}\right) F_{\tilde{Y}}(y)\Big|_{q_{\tilde{Y}}(p)}^{\infty} - \int_{q_{\tilde{Y}}(p)}^{\infty} f_{\tilde{Y}}(y) F_{\tilde{Y},\nu-2}\left(y\sqrt{\frac{\nu-2}{\nu}}\right) dy \right\}. \end{aligned} \tag{D.22}$$

Now we use the recurrence relation (B.30) for the distribution functions, to evaluate the integral in (D.22), namely

$$\begin{aligned}\int_{q_{\tilde{Y}}(p)}^{\infty} f_{\tilde{Y}}(y) F_{\tilde{Y},\nu-2}\left(y\sqrt{\frac{\nu-2}{\nu}}\right) dy &= \int_{q_{\tilde{Y}}(p)}^{\infty} f_{\tilde{Y}}(y)\left(F_{\tilde{Y}}(y) - \frac{1}{\nu-1} y\,(1+y^2/\nu) f_{\tilde{Y}}(y)\right) dy \\ &= \frac{1}{2} F_{\tilde{Y}}^2(y)\Big|_{q_{\tilde{Y}}(p)}^{\infty} - \frac{1}{\nu-1}\int_{q_{\tilde{Y}}(p)}^{\infty} y\,(1+y^2/\nu) f_{\tilde{Y}}^2(y)dy. \end{aligned} \tag{D.23}$$

But, by partial integration and with the expression of the antiderivative $\int y f_{\tilde{Y}}(y)$ from above, we can write the integral of (D.23) as

$$\int_{q_{\tilde{Y}}(p)}^{\infty} y(1+y^2/\nu) f_{\tilde{Y}}^2(y)dy = -\frac{\nu}{\nu-1}(1+y^2/\nu)^2 f_{\tilde{Y}}^2(y)\Big|_{q_{\tilde{Y}}(p)}^{\infty} - \int_{q_{\tilde{Y}}(p)}^{\infty} y(1+y^2/\nu) f_{\tilde{Y}}^2(y)dy,$$

which is equivalent to

$$\int_{q_{\tilde{Y}}(p)}^{\infty} y(1+y^2/\nu) f_{\tilde{Y}}^2(y)dy = \frac{1}{2}\frac{\nu}{\nu-1}\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p)). \tag{D.24}$$

Hence we obtain for (D.23)

$$\int_{q_{\tilde{Y}}(p)}^{\infty} f_{\tilde{Y}}(y) F_{\tilde{Y},\nu-2}\left(y\sqrt{\frac{\nu-2}{\nu}}\right) dy = \frac{1}{2}(1-p^2) - \frac{1}{2}\frac{\nu}{(\nu-1)^2}\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p)),$$

so that (D.22) becomes (using again the recurrence relation (B.30))

$$\begin{aligned}
&\int_{q_{\tilde{Y}}(p)}^{\infty}(1+y^2/\nu)f_{\tilde{Y}}(y)F_{\tilde{Y}}(y)dy \\
&= \frac{1}{2}\frac{\nu-1}{\nu-2}\left(1+p^2-2p\,F_{\tilde{Y},\nu-2}\left(q_{\tilde{Y}}(p)\sqrt{\frac{\nu-2}{\nu}}\right)+\frac{\nu}{(\nu-1)^2}\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p))\right) \\
&= \frac{1}{2}\frac{\nu-1}{\nu-2}\left(1-p^2+\frac{2}{\nu-1}\,p\,q_{\tilde{Y}}(p)\left(1+q_{\tilde{Y}}^2(p)/\nu\right)f_{\tilde{Y}}(q_{\tilde{Y}}(p))+\frac{\nu}{(\nu-1)^2}\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p))\right)
\end{aligned} \tag{D.25}$$

Combining (D.25) with (D.24) in (D.21) provides:

$$\begin{aligned}
&\int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(y)F_{\tilde{Y}}(y)dy \\
&= \frac{\nu}{\nu-1}\,p\,q_{\tilde{Y}}(p)\left(1+q_{\tilde{Y}}^2(p)/\nu\right)f_{\tilde{Y}}(q_{\tilde{Y}}(p))+\frac{1}{2}\frac{\nu}{\nu-2}\left(1-p^2+\frac{2p}{\nu-1}q_{\tilde{Y}}(p)\left(1+q_{\tilde{Y}}^2(p)/\nu\right)f_{\tilde{Y}}(q_{\tilde{Y}}(p))\right) \\
&\quad+\frac{1}{2}\frac{\nu^2}{(\nu-1)(\nu-2)}\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p)).
\end{aligned} \tag{D.26}$$

Now we are ready to look at the final expression for the double integral as shown in (D.19), using the results in equations (D.20) and (D.26).

$$\begin{aligned}
&\int_p^1\int_v^1 \frac{v(1-u)}{f_{\tilde{Y}}(q_{\tilde{Y}}(v))f_{\tilde{Y}}(q_{\tilde{Y}}(u))}dudv \\
&= \frac{1}{2}\frac{\nu}{\nu-2}\left\{1+\,q_{\tilde{Y}}(p)\,\left(1+q_{\tilde{Y}}^2(p)/\nu\right)\,f_{\tilde{Y}}(q_{\tilde{Y}}(p))-\frac{\nu(\nu-2)}{(\nu-1)^2}\,\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p))\right. \\
&\left.+p\left(\frac{\nu-2}{\nu}q_{\tilde{Y}}^2(p)-1-\frac{2(\nu-2)}{\nu-1}\,q_{\tilde{Y}}(p)\,\left(1+q_{\tilde{Y}}^2(p)/\nu\right)\,f_{\tilde{Y}}(q_{\tilde{Y}}(p))\right)-\frac{\nu-2}{\nu}p^2\,q_{\tilde{Y}}^2(p)\right\}
\end{aligned} \tag{D.27}$$

Recalling, $\mathbb{E}[Y^4]=3\frac{\nu-2}{\nu-4}$, see (B.34) for the Student distribution (and hence $\mathrm{Var}(Y^2)=2\frac{\nu-1}{\nu-4}$), we get overall $\lim_{n\to\infty}\mathrm{Cor}(\widetilde{\mathrm{ES}}_n(p),\hat{\sigma}_n^2)=$

$$\frac{\frac{\sqrt{\nu(\nu-2)}}{\nu-3}f_{\tilde{Y}}(q_{\tilde{Y}}(p))(1+q_{\tilde{Y}}^2(p)/\nu)^2}{\sqrt{1+\,q_{\tilde{Y}}(p)\,\left(1+q_{\tilde{Y}}^2(p)/\nu\right)\,f_{\tilde{Y}}(q_{\tilde{Y}}(p))-\frac{\nu(\nu-2)}{(\nu-1)^2}\,\left(1+q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p))+p\left(\frac{\nu-2}{\nu}q_{\tilde{Y}}^2(p)-1-\frac{2(\nu-2)}{\nu-1}\,q_{\tilde{Y}}(p)\,\left(1+q_{\tilde{Y}}^2(p)/\nu\right)\,f_{\tilde{Y}}(q_{\tilde{Y}}(p))\right)-\frac{\nu-2}{\nu}p^2\,q_{\tilde{Y}}^2(p)}\sqrt{2\frac{\nu-1}{\nu-4}}}$$

and we get back the expression for the Gaussian distribution for $\nu\to\infty$.

▷ *ES with Sample MAD - Gaussian distribution.* Recall that we computed asymptotic correlation and covariance between the sample quantile and the sample MAD only for $p\geq 0.5$, as the case $p<0.5$ can be deduced using the point-symmetry around $p=0.5$. For ES, this argument converts to a symmetry around the $p=0.5$-axis when integrating over the asymptotic covariance, i.e. for any $p\in(0,1)$ we can write

$$\int_p^1 \lim_{n\to\infty}\mathrm{Cov}(\sqrt{n}q_n(u),\sqrt{n}\hat{\theta}_n)du=\int_{1-p}^1 \lim_{n\to\infty}\mathrm{Cov}(\sqrt{n}q_n(u),\sqrt{n}\hat{\theta}_n)du.$$

Clearly, this symmetry does not hold for the double-integral in the denominator of (D.13). Thus, the asymptotic correlation with the sample ES, in contrast to the one with the sample VaR, is not symmetric around $p=0.5$. Nevertheless, for the ease of presentation, we will only consider the case $p\geq 0.5$. The only quantity we need to compute of (D.13) is the integral over the covariance, which comes back, via

(B.5), to evaluate $\int_p^1 \left(1 - \frac{1-u}{\varphi(\Phi^{-1}(u))}\sqrt{\frac{2}{\pi}}\right) du$. Using (D.14), we obtain

$$\int_p^1 \left(1 - \frac{1-u}{\varphi(\Phi^{-1}(u))}\sqrt{2/\pi}\right) du = (1-p)\left(1 + \Phi^{-1}(p)\sqrt{2/\pi}\right) - \varphi(\Phi^{-1}(p))\sqrt{2/\pi}.$$

Recalling the asymptotic variance of the sample MAD, (B.35), and the solution of the double integral, (D.16), in the Gaussian case, this gives overall:

$$\lim_{n\to\infty} \mathrm{Cor}(\widetilde{\mathrm{ES}}_n(p), \hat{\theta}_n) = \frac{(1-p)\left(1+\Phi^{-1}(p)\sqrt{2/\pi}\right) - \varphi(\Phi^{-1}(p))\sqrt{2/\pi}}{\sqrt{\Phi^{-1}(p)^2(p-p^2) - \varphi(\Phi^{-1}(p))^2 + \Phi^{-1}(p)\varphi(\Phi^{-1}(p))(1-2p) + 1 - p}\sqrt{1-2/\pi}}$$

▷ *ES with Sample MAD - Student distribution.* The remarks made in the Gaussian case hold also for the Student distribution. We proceed analogously. Again, we first compute the integral in the numerator of the asympyotic covariance given in (B.7) (with $\sigma^2 = 1$):

$$\int_p^1 \lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}q_n(u), \sqrt{n}\hat{\theta}_n(u))du = (\nu-2)\int_p^1 \left(\frac{1}{\nu-1}\left(1 + \frac{q_{\tilde{Y}}^2(u)}{\nu}\right) - \frac{\Gamma(\frac{\nu-1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})}\frac{1-u}{f_{\tilde{Y}}(q_{\tilde{Y}}(u))}\right) du \tag{D.28}$$

Since by change of variables $\int_p^1 q_{\tilde{Y}}^2(u)\,du = \int_{q_{\tilde{Y}}(p)}^{\infty} y^2 f_{\tilde{Y}}(u)du$, using the knowledge of the second truncated moment, (B.32), for the first integral of (D.28) and (D.18) for the second integral, provides

$$\begin{aligned}
&\int_p^1 \lim_{n\to\infty} \mathrm{Cov}(\sqrt{n}q_n(u), \sqrt{n}\hat{\theta}_n(u))du \\
&= \frac{1}{\nu-1}\left\{(\nu-2)(1-p) + 1 - p + f_{\tilde{Y}}(q_{\tilde{Y}}(p))q_{\tilde{Y}}(p)\left(1 + q_{\tilde{Y}}^2(p)/\nu\right)\right\} \\
&+ \frac{(\nu-2)\Gamma(\frac{\nu-1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left((1-p)q_{\tilde{Y}}(p) - \frac{\nu}{\nu-1}f_{\tilde{Y}}(q_{\tilde{Y}}(p))(1+q_{\tilde{Y}}^2(p)/\nu)\right) \\
&= 1 + q_{\tilde{Y}}(p)\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{\nu-2}{\sqrt{\pi\nu}} + \frac{1}{\nu-1}\left[\left(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu}\right) f_{\tilde{Y}}(q_{\tilde{Y}}(p))\left(q_{\tilde{Y}}(p) - \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{(\nu-2)\sqrt{\nu}}{\sqrt{\pi}}\right)\right] - p\left(1 + q_{\tilde{Y}}(p)\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{\nu-2}{\sqrt{\pi\nu}}\right)
\end{aligned}$$

which gives us back the Gaussian case for $\nu \to \infty$, recalling an asymptotic property of the Gamma function $\lim_{n\to\infty} \frac{\Gamma(n+\alpha)}{\Gamma(n)n^\alpha} = 1$ that we need to use here with $n = \nu\alpha$ and $\alpha = 1/2$. Recalling the asymptotic variance of the sample MAD in the Student case, (B.37), and the solution of the double integral, (D.27), this gives overall $\lim_{n\to\infty} \mathrm{Cor}(\widehat{\mathrm{ES}}_n(p), \hat{\theta}_n) = \frac{1}{\sqrt{1 - \frac{\nu-2}{\pi}\frac{\Gamma((\nu-1)/2))^2}{\Gamma(\nu/2)^2}}} \times$

$$\frac{1 + q_{\tilde{Y}}(p)\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{\nu-2}{\sqrt{\pi\nu}} + \frac{1}{\nu-1}\left[\left(1 + \frac{q_{\tilde{Y}}^2(p)}{\nu}\right) f_{\tilde{Y}}(q_{\tilde{Y}}(p))\left(q_{\tilde{Y}}(p) - \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{(\nu-2)\sqrt{\nu}}{\sqrt{\pi}}\right)\right] - p\left(1 + q_{\tilde{Y}}(p)\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\frac{\nu-2}{\sqrt{\pi\nu}}\right)}{\sqrt{1 + q_{\tilde{Y}}(p)\left(1 + q_{\tilde{Y}}^2(p)/\nu\right) f_{\tilde{Y}}(q_{\tilde{Y}}(p)) - \frac{\nu(\nu-2)}{(\nu-1)^2}\left(1 + q_{\tilde{Y}}^2(p)/\nu\right)^2 f_{\tilde{Y}}^2(q_{\tilde{Y}}(p)) + p\left(\frac{\nu-2}{\nu}q_{\tilde{Y}}^2(p) - 1 - \frac{2(\nu-2)}{\nu-1}q_{\tilde{Y}}(p)\left(1 + q_{\tilde{Y}}^2(p)/\nu\right) f_{\tilde{Y}}(q_{\tilde{Y}}(p))\right) - \frac{\nu-2}{\nu}p^2 q_{\tilde{Y}}^2(p)}}$$

and we get back the expression for the Gaussian distribution for $\nu \to \infty$.

D.2 Pro-cyclicality analysis on residuals

See the online-appendix in [34].

Appendix E

GARCH Optimization

The full version of this study can be found in the online-appendix in [34]. Here we present a part of it.

E.1 Intention

In Chapter 2 we have implemented a method by Zumbach, as presented in [130], to estimate the parameters of the GARCH(1,1) process using the statistical software R. Following the claims and results in [130], it offers a robust optimization method (through reparametrization). This robustness is not guaranteed if the reparametrization is not used. In this Appendix we want to assess, from a practical point of view how some other frequently used alternatives for GARCH fitting perform when using the statistical software R.

This is especially of interest, as it is known that some estimation procedures by R packages fail every now and then when estimating GARCH parameters. Although there does not seem to be a formal large-scale investigation on that issue, one example reporting this, is this blog entry complaining about the huge estimation uncertainty when using the fGARCH or rugarch package (`https://ntguardian.wordpress.com/2017/11/02/problems-estimating-garch-parameters-r/`).

The first version of this small study was already realized in 2017 and we subsequently expanded and updated it in 2018. Also in 2018, we made Prof. Marius Hofert, the maintainer and one of the authors of the R package 'qrmtools' aware of the method of Zumbach and shared the R code. By now, you can find a GARCH(1,1) parameter estimation method using the method of Zumbach implemented in the 'qrmtools' package (acknowledging the contribution), see [76].

As packages of statistical software are frequently updated, we rerun the study in October 2019 to be sure that our conclusions and observations still hold. All in all, we consider eight different methods of GARCH(1,1) fitting from four different R packages :

- The method used in this thesis, namely our implementation of the reparametrization approach by [130].
- The popular R package 'fGarch', [126] using version 3042.83.1.
- An alternative R package 'rugarch', [67] in version 1.4-1. Therein we consider two different options in the estimation procedure, either with so called variance targeting or without. We denote this as 'rugarch var' or 'rugarch', respectively.
- The fitting function in the R package 'qrmtools', [76], version 0.0-10. As for the 'rugarch' package, we consider here two options too, either with variance targeting or without. We abbreviate it as 'qrm var' and 'qrm', respectively.
- As fourth alternative package, we consider the GARCH estimation in 'tseries' by [119], version 0.10-47.

- Last but not least, we also include the Zumbach estimation method implemented in the 'qrmtools' package. While the estimation should give the same results as in our use of the Zumbach method, it is also of interest to compare the time needed for the estimation procedure. We label it 'qrm Zum.'.

E.1.1 Limitations of our study

Sill, this is far from being a representative study. Thus, we want to mention the limitations (which could serve as a basis of extending this work to a proper representative large-scale study).

- This study does not include all available R packages for GARCH estimation. See e.g. the ones mentioned in [5] including the reference itself.
- Also, for the GARCH packages tested, it is not guaranteed that we checked all the options available in the optimization functions as for example different solver/optimization functions one can choose from. Usually we rely on the default options, only for the pacakges 'rugarch' and 'qrmtools' we modified one option, which allows us to activate the volatility targeting.
- We test the parameter estimation quality only on simulated GARCH(1,1) realizations from two sets of parameters, both with Gaussian innovations. For a proper study it should be extended to the whole range of parameters $\alpha + \beta < 1$ and also considere other innovations, like a Student distribution with varying degrees of freedom.
- This study (and the implementation of the Zumbach fitting procedure) focuses only on GARCH(1,1). No extensions to general GARCH(p,q) processes are considered. Still, in [130] it is explained that the procedure could be extended for GARCH(p,q).
- Our implementation of the Zumbach approach uses only a specific optimization (nlminb) - others were not tested thoroughly (although testing a few times with BFGS optimization gave worse results).
- This is explicitly an empirical study. For theoretical discussion on optimization techniques we refer to the paper of Zumbach, [130] and as more recent references, [64], [65].

Still, despite all those limitations, we believe that this small study can give insights into GARCH fitting with R packages.

E.1.2 Content

This study contains two different parts. The first one, Section E.2 is centered around real data, and can be found in the full version of this study in the online-appendix in [34]. The second one, Section E.3, is an analysis on simulated data from a GARCH(1,1) model (for which we know the underlying true parameters) which we present here.

On simulated data

For simulated data from a GARCH(1,1) process, knowing the true underlying parameters, we can test the ability of the different packages of correctly estimating these parameters.

Therefore, we create separately two sets of simulated sample paths (1000 sample paths for each of them). The parameters of the first one equal the estimates we obtained with the Zumbach method on the S&P 500 data. The second one was chosen arbitrarily, as an example of low α and β values.

For each of the two sets, we compute the parameter estimates with all different procedures. For each estimation method, we do this using different sample sizes (to see how the parameter estimates vary with increasing sample size). In all of those cases we take the average over all realizations and plot the average values together with the empirical quantiles in a plot for each method.

This way we can easily compare the different methods and their estimation uncertainty.

E.2 Results on real data

E.2.1 Stability of the estimates throughout repetition

See the online-appendix in [34].

E.2.2 Comparing the estimates of the three methods

See the online-appendix in [34].

E.2.3 Sensitivity to changes in the data set

See the online-appendix in [34].

E.2.4 Simulation results: Annualized volatility (over whole sample)

See the online-appendix in [34].

E.2.5 Simulation results: Annualized volatility (rolling window)

See the online-appendix in [34].

E.3 Results on simulated data

As this is not an extensive study over all parameter combinations we simply provide figures of the parameter estimates (as a function of the sample size) for each method. In each we show the parameter estimate average (over 1000 repetitions), as well as the empirical 2.5% and 97.5% quantiles. With a dotted line we depict the true parameter value. Further note, that the average was not always computed over 1000 repetitions as for some estimation procedures the convergence failed in some realizations.

Further, as a performance indicator we also compare the computation times needed to estimate the parameters in these 1000 repetitions.

Anticipating the results, we observe the following: For the first sample (which represented GARCH parameter values as they are familiar for financial data), all estimation procedures converge to the true value in the long run, see Figures E.1, E.2, E.3. The confidence intervals vary depending on the estimator but are mostly similar. Also, the confidence intervals are very big for small and moderate sample sizes, say $n \leq 1000$.

On the second sample (which we chose such that $\alpha + \beta$ are low), we observe that some techniques completely misspecify the parameter, (at least up to our sample size considered) not converging at all: rugarch, rugarch var, qrm failed, while Zumbach, fGARCH, qrm var, tseries and qrm Zum. correctly estimated the parameters, see Figures E.4, E.5, E.6. Similarly as for sample 1, the confidence intervals are usually quite big, here even up to sample sizes of $n \leq 2500$.

With respect to computational speed (irrespective of the quality of estimation), we have clear winners. From tables E.1 and E.2 we see that the tseries package as well as the qrm package using the technique by Zumbach, are clearly faster than the other (usually by a factor of 3 to 6).

E.3.1 Sample 1 (S&P 500)

	Methods							
Estimates for	Zumbach	fGarch	rugarch	rugarch var	qrm	qrm var	tseries	qrm Zum.
min	6.76	11.72	24.65	93.99	26.96	12.11	2.81	3.79

TABLE E.1: Computation time in minutes for estimating the GARCH parameters in 1000 repetitions from sample 1.

Estimation Quality: Omega value

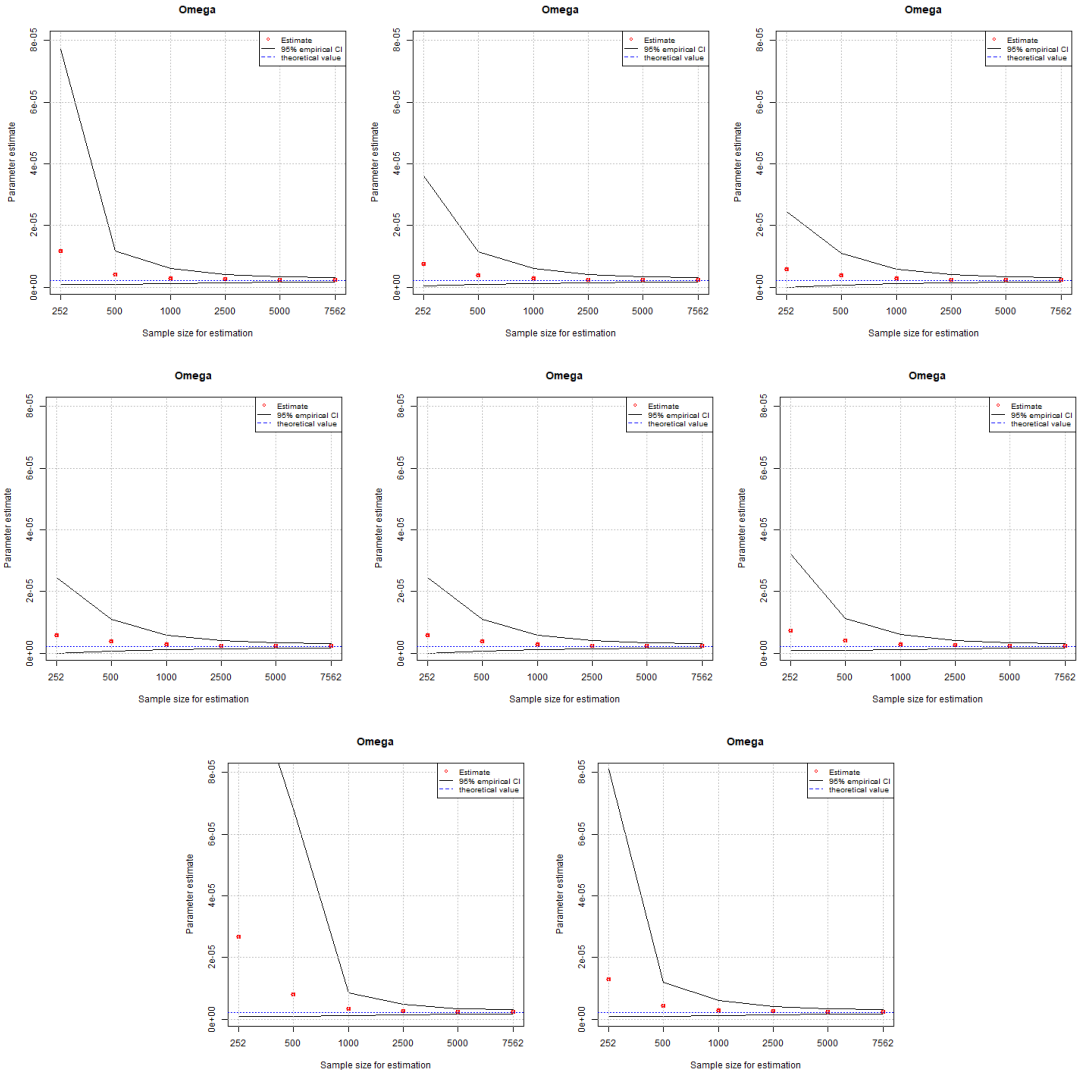


FIGURE E.1: Parameter estimates of omega for sample 1 with the different methods. We see the mean of 1000 repetitions, and its corresponding empirical 2.5% and 97.5% quantiles. From left to right, top to bottom, the estimation methods are: Zumbach, fGARCH, rugarch, rugarch var, qrm, qrm var, tseries, qrm Zum.

Estimation Quality: Alpha value

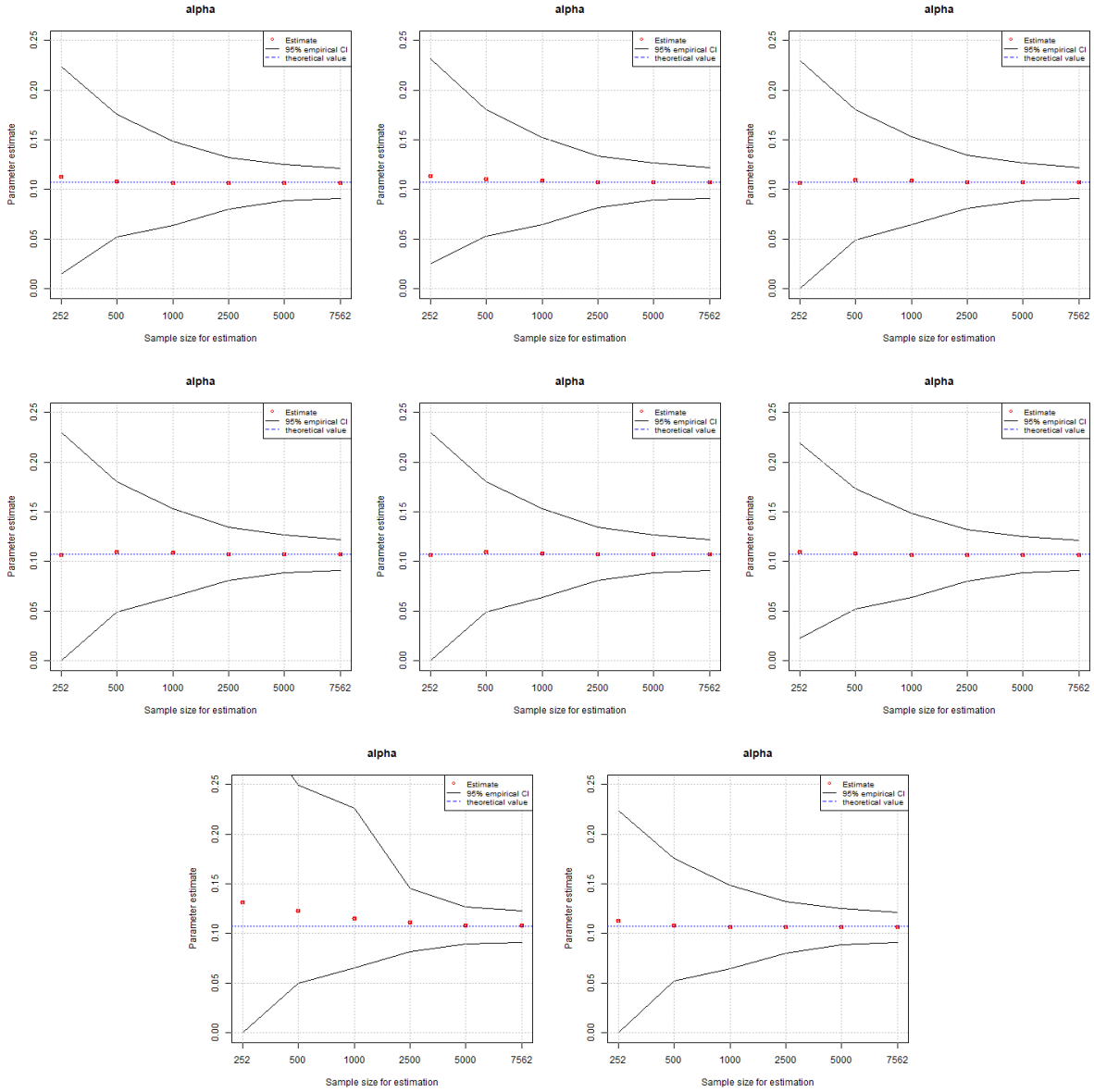


FIGURE E.2: Parameter estimates of alpha for sample 1 with the different methods. We see the mean of 1000 repetitions, and its corresponding empirical 2.5% and 97.5% quantiles. From left to right, top to bottom, the estimation methods are: Zumbach, fGARCH, rugarch, rugarch var, qrm, qrm var, tseries, qrm Zum.

Estimation Quality: Beta value

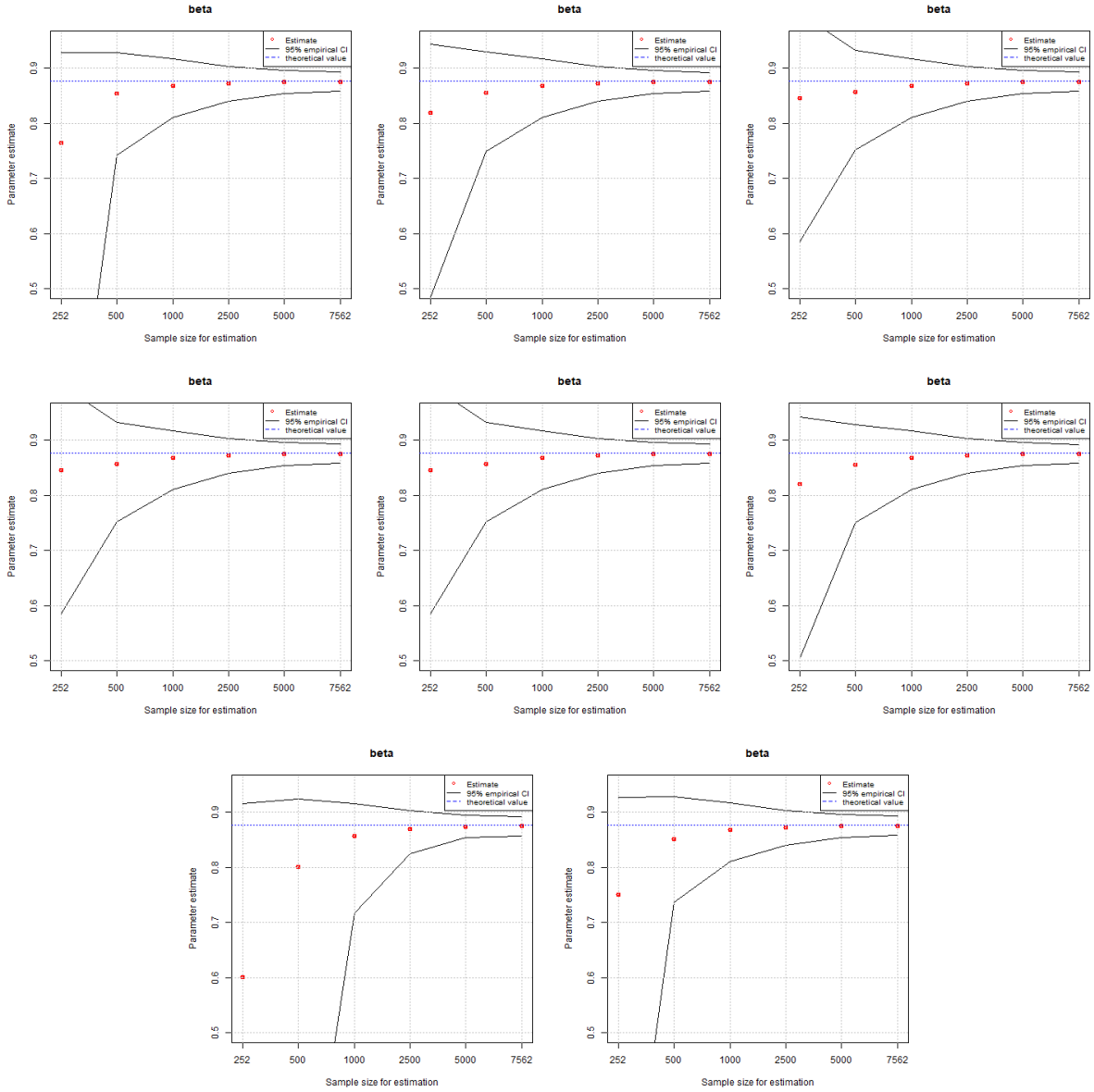


FIGURE E.3: Parameter estimates of beta for sample 1 with the different methods. We see the mean of 1000 repetitions, and its corresponding empirical 2.5% and 97.5% quantiles. From left to right, top to bottom, the estimation methods are: Zumbach, fGARCH, rugarch, rugarch var, qrm, qrm var, tseries, qrm Zum.

E.3.2 Sample 2 (low $\alpha + \beta$)

	Methods							
Estimates for	Zumbach	fGarch	rugarch	rugarch var	qrm	qrm var	tseries	qrm Zum.
min	9.75	9.92	18.52	34.24	16.74	10.28	1.75	2.44

TABLE E.2: Computation time in minutes for estimating the GARCH parameters in 1000 repetitions from sample 2.

Estimation Quality: Omega value

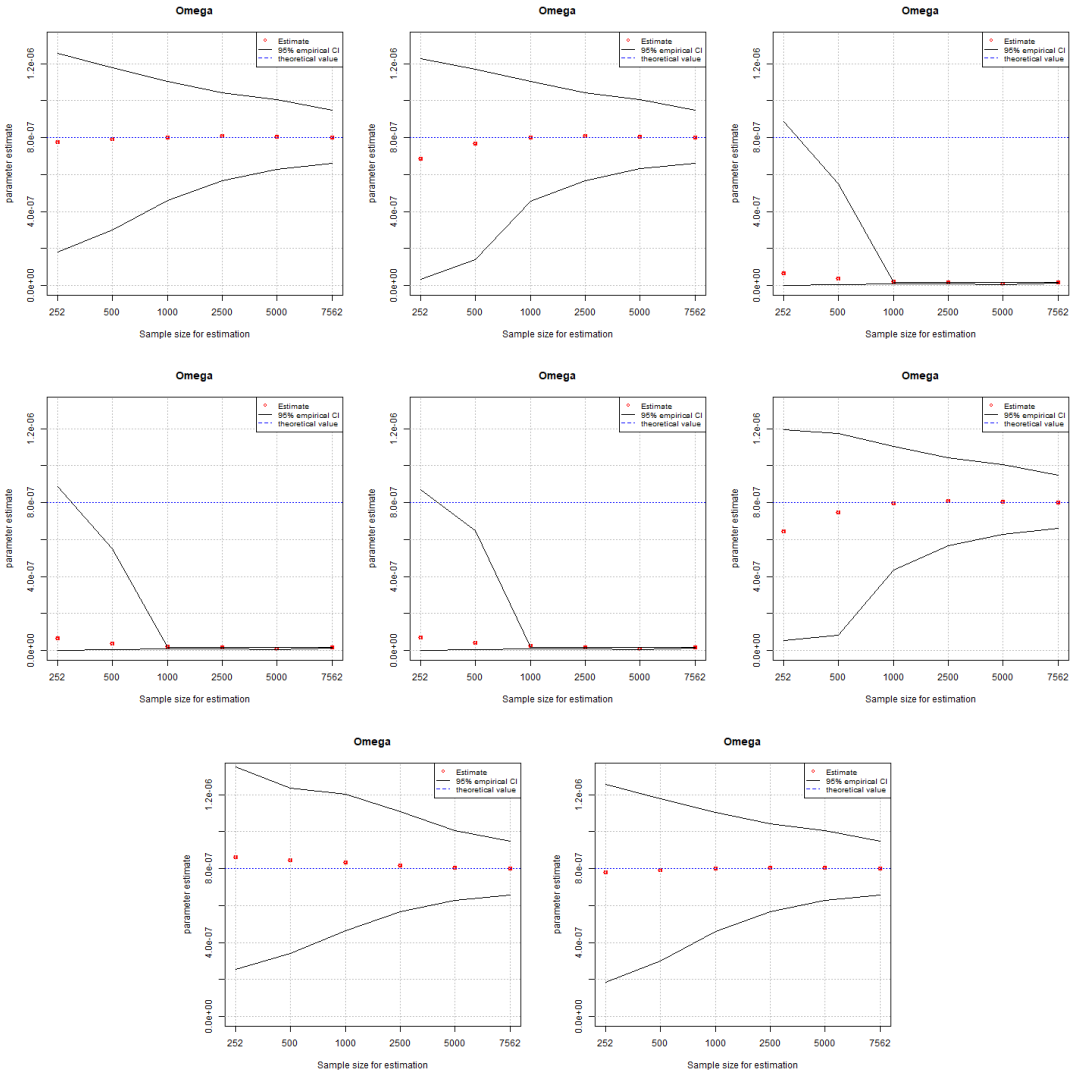


FIGURE E.4: Parameter estimates of omega in sample 2 with the different methods. We see the mean of 1000 repetitions, and its corresponding empirical 2.5% and 97.5% quantiles. From left to right, top to bottom, the estimation methods are: Zumbach, fGARCH, rugarch, rugarch var, qrm, qrm var, tseries, qrm Zum.

Estimation Quality: Alpha value

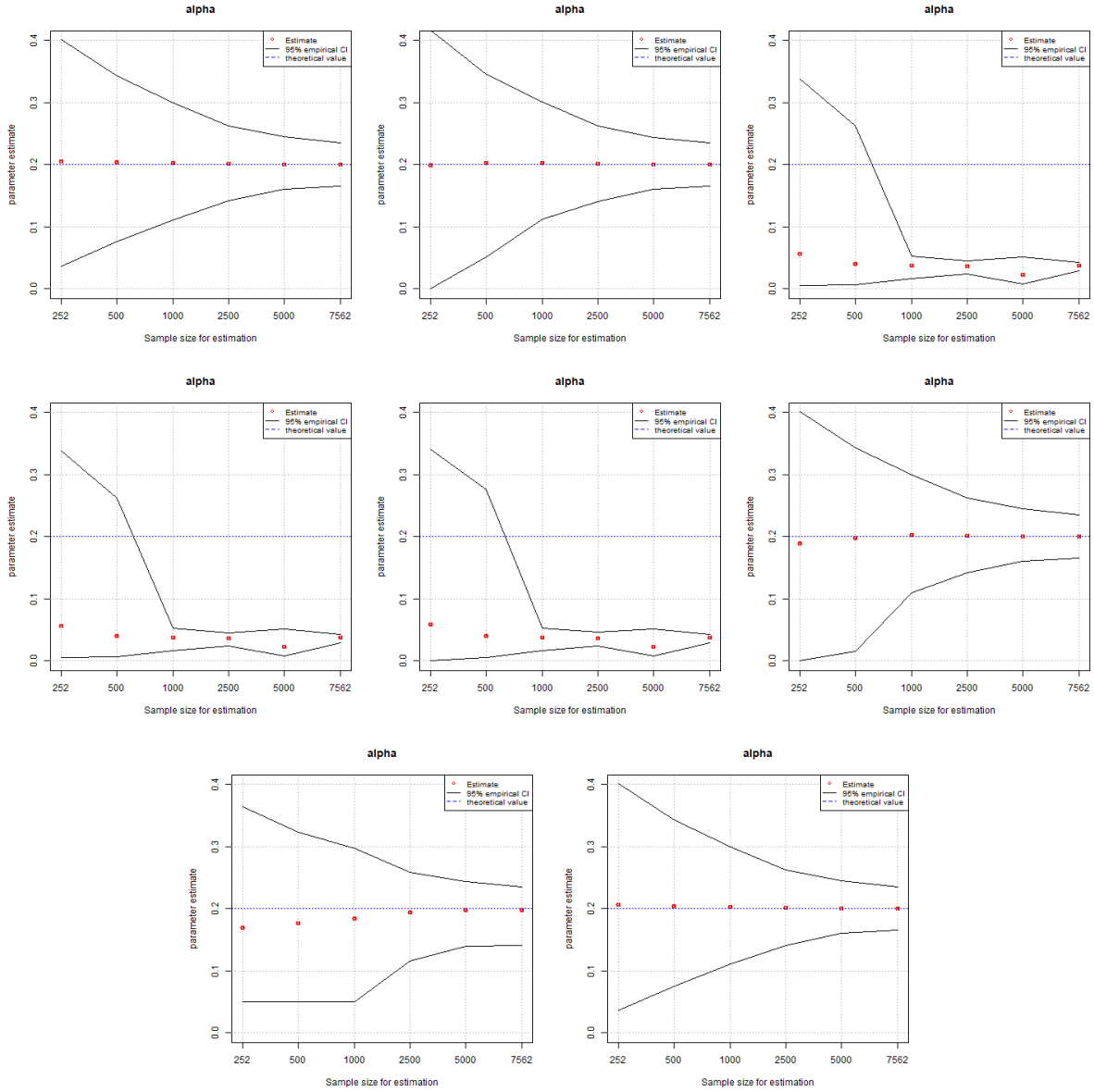


FIGURE E.5: Parameter estimates of alpha in sample 2 with the different methods. We see the mean of 1000 repetitions, and its corresponding empirical 2.5% and 97.5% quantiles. From left to right, top to bottom, the estimation methods are: Zumbach, fGARCH, rugarch, rugarch var, qrm, qrm var, tseries, qrm Zum.

Estimation Quality: Beta value

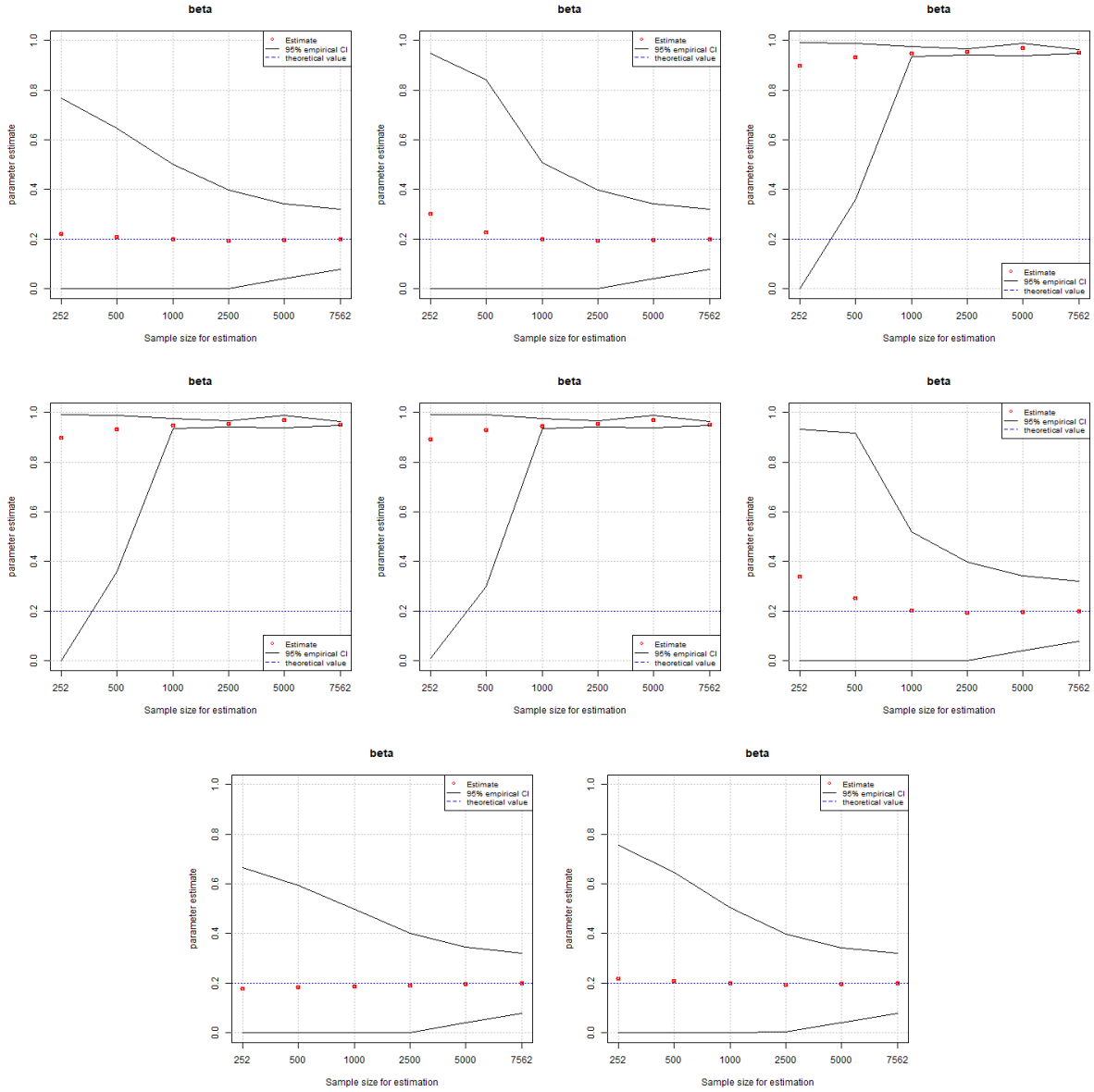


FIGURE E.6: Parameter estimates of beta in sample 2 with the different methods. We see the mean of 1000 repetitions, and its corresponding empirical 2.5% and 97.5% quantiles. From left to right, top to bottom, the estimation methods are: Zumbach, fGARCH, rugarch, rugarch var, qrm, qrm var, tseries, qrm Zum.

E.4 Overall summary

On real data we observed the following (although not presented here, but only in the online-appendix, [34]):

- In general all methods behave similarly (and sensitive) to changes in the data set

- All methods do not capture well the extreme movements of the annualized volatility (measured on a rolling window) in the data, i.e. standard deviation and/or range of the volatility are not replicated well
- We witnessed a case (the japanese index TOPIX) where the packages fGarch, rugarch, qrm, tseries completely fail in replicating the annualized volatility behaviour of the data but Zumbach, rugarch var, qrm var succeed.

On simulated we observed the following:

- When estimating the parameters on 1000 replications of a simulated dataset (i.e. with known parameters) we saw that some methods might completely fail (rugarch, rugarch var, qrm). From the working methods, the confidence intervals of fGarch and tseries seem to be the widest. Comparing the smaller confidence intervals of Zumbach and qrm var, both cases exist: Sometimes Zumbach is more precise, sometimes qrm var.
- The computation time was clearly superior with tseries and qrm Zum.

Thus, combining the observations on the real data and simulated data, we find that either the Zumbach method or qrm var seems to be the most reliable. Given the fact that the Zumbach method implemented in the qrmtools package (what we labeled as 'qrm Zum.'), is more than five times quicker than qrm var. This seems the best choice.

Disclaimer: This is not a representative analysis but only the result from observations on three different indices.

Appendix F

Information on computational resources

The empirical studies and simulations were implemented in the programming language R (version 3.6.1), [110], and using RStudio Desktop (version 1.1.423) as integrated development environment, [114].

None of the code material has been published online, but is available on request. The only exception being, as remarked in Appendix E, that we shared our R code implementation for the GARCH optimization method of Zumbach, [130], with Prof. Marius Hofert, the maintainer and one of the authors of the R package 'qrmtools'. By now, you can find a GARCH(1,1) parameter estimation method using the method of Zumbach implemented in the 'qrmtools' package (acknowledging my contribution), see [76].

In the following we list and acknowledge the additional R packages we used for our computations:

- The package 'laeken', [3], to compute weighted sample quantiles more quickly.
- The packages 'chron', [83], and 'readxl', [122], to import and correctly format the financial data for our purposes.
- The package 'xtable', [42], to output the simulation results conveniently in formatted tables.
- The package 'fBasics', [125], to compute row-wise standard deviations of a matrix.