NCoVaR Granger causality

Cees Diks[a,b], Marcin Wolski[c,*]

[a] *Center for Nonlinear Dynamics in Economics and Finance (CeNDEF), University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands*
[b] *Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands*
[c] *European Investment Bank, 98-100 Boulevard Konrad Adenauer, 2950 Luxembourg, Luxembourg*

August 2018

Abstract

We propose a new methodology to assess risk transmission effects between individual companies or sectors and to measure their contributions to systemic risk. We extend the Conditional Value-at-Risk (CoVaR) approach, introducing explicit nonparametric CoVaR (NCoVaR) measures of cross-sectional dependence and Granger causality. By showing that the natural estimators are U-statistics, we construct formal nonparametric tests for independence and NCoVaR Granger non-causality. Numerical simulations confirm that in common situations the nonparametric tests have better size and power properties than their parametric counterparts. The methodology is illustrated empirically by assessing risk transmissions between sovereigns and banking sectors in the euro area.

Keywords: Financial Risk, Nonparametric Tests, General Granger Causality

JEL Codes: C1, G1, G20, G32

*Corresponding author
Email addresses: C.G.H.Diks@uva.nl (Cees Diks), M.Wolski@eib.org (Marcin Wolski)

1. Introduction

The global financial crisis and the subsequent sovereign debt crisis shed new light on the complexity within the financial sector. The linkages and risk exposures between various institutions and sectors proved to be of great importance in transmitting distress across the financial system. Additionally, during systemic events the spread of malaise is accelerated through indirect channels, such as price effects and liquidity spirals (Brunnermeier, 2009). In effect, market values of various financial assets tend to move closer together during financial distress, drifting away from their fundamentals and demonstrating dependence in the form of tail co-movements (Adrian and Brunnermeier, 2016).

A commonly-used econometric approach to assess the level of such co-risk shifts is Conditional Value-at-Risk (CoVaR), introduced by Adrian and Brunnermeier (2016). It extends the concept of Value-at-Risk (VaR), which determines the maximum loss on returns within the γ-percentile confidence interval (Kupiec, 2002). CoVaR assesses VaR_γ of one institution conditional on distress in the other.[1]

The estimation of CoVaR usually involves (linear) quantile regression techniques and/or GARCH-type models. As noted by Rothe (2010) and Jeong et al. (2012), a shortcoming of taking a parametric approach lies in its susceptibility to model misspecification. Clearly, standard linear and/or (G)ARCH type parametric approaches may overlook deviations from linearity and/or normality, and hence might lead to inaccurate co-risk estimation results.

Motivated by this, we develop a nonparametric measure for tail co-movement, which is robust to the above-mentioned shortcomings but it is still intuitively comparable to CoVaR. Our measure is dubbed NCoVaR, which stands for 'nonparametric CoVaR'. Below we first extend the original instantaneous CoVaR setup to the nonparametric setting and then introduce Granger causal effects by extending this notion to NCoVaR Granger causality, a form of nonparametric (general) Granger causality in distribution. To complete the picture we

[1] In this paper we use the term 'institution' as describing returns on a specific asset, company, or sector.

also introduce a parametric Granger causality measure by extending CoVaR parametrically. The family of CoVaR measures, highlighting the ones developed in this paper, is summarized in Table 1.

Table 1: Taxonomy of CoVaR measures. Entries in italics highlight the contributions of this paper.

method	type of dependence	
	cross-sectional/instantaneous	causal/dynamic
parametric	CoVaR	*CoVaR Granger causality*
nonparametric	*NCoVaR*	*NCoVaR Granger causality*

By using U-statistic representations we derive asymptotic normality of the nonparametric estimators and demonstrate numerically that NCoVaR and NCoVaR Granger causality offer superior performance in the presence of dynamic spillover effects, compared to their parametric counterparts. In the same spirit, Jeong et al. (2012) proposed a nonparametric test for causality in quantile, extending the work of Hong et al. (2009) to nonparametric Value-at-Risk estimation. These methodologies are complementary to ours in the sense that whereas these authors focus on the effect of one variable on the conditional quantiles of another, we study conditional tail probability effects.

Finally, we apply the proposed methodology to assess the bank-sovereign risk transmission between selected Euro Area (EA) countries. We confirm the differences between the core- and vulnerable-EA dynamics, identified earlier by Paries et al. (2014), Ohnsorge et al. (2014) and EIB (2016). We furthermore argue that NCoVaR is a more conservative methodology, grasping key fundamental causal relationships, and to a higher extent it captures the information spanned by extra confounding variables.

A major motivation for considering causality lies in its possible applications to networks and contagion analysis (see e.g. Chinazzi and Fagiolo, 2013). Looking at any pair of institutions, the possible mutual risk transmission effects do not have to be bilaterally equal (as they are implicitly assumed to be in a linear Gaussian setting). For instance, a lender has a different kind of risk exposure to a creditor than vice versa. NCoVaR Granger causality cap-

tures that phenomenon explicitly, allowing for a more detailed analysis of network spillover effects, cascades and shock propagation. The general type of Granger causality employed in this study, i.e. a nonparametric version of the concept originally proposed by Granger (1969), is intuitive and requires minimal model restrictions. A parametric Granger causality measure has already been successfully applied as a network mapping tool in financial analysis (Gao and Ren, 2013). Such a general notion of causality appears to be particularly relevant in the exact characterization of feedback loops and propagation mechanisms.

The use of nonparametric methods makes NCoVaR perform well in nonlinear and/or non-Gaussian settings. In fact, the existence of nonlinearities is widely recognized in the financial literature. For instance, von Borstel et al. (2016) find that the sovereign debt crisis changed the composition of the pass-through, adjusting for indirect effects from lower sovereign risk premia in the EA. Kitamura et al. (2016) find that banks with high shares of relationship lending appear to be characterized by nonlinear pass-through effects. This is somehow in line with a more general finding of Huang et al. (2010) and He and Krishnamurthy (2012), who suggest that banking risk is a nonlinear function of asset exposure.

The motivation behind the empirical part stems from the strong adverse effects from the bank-sovereign feedback loops, i.e. the interdependence of the banking sectors and corresponding sovereigns, on the real economy and taxpayers. The majority of econometric approaches in these fields focus on co-risk measures, where the risk of one sector is assessed in relation to the risk of the other one. The intuition behind these models lies in negative externalities. As argued by Adrian and Brunnermeier (2016), such externalities are a consequence of asset exposures, excessive risk taking and leverage. Given, for instance, that the banking sector is facing a liquidity shock, it liquidates its assets, including sovereign securities, at fire-sale prices as given, affecting the borrowing constraints of the sovereign. On the other hand, sovereign characteristics are often perceived as a country-wide benchmark in credit risk assessment. It is rarely the case that a financial entity 'pierces the sovereign ceiling' in a credit rating context. Furthermore, sovereigns are often the implicit guarantors

of the financial system (Acharya et al., 2014).

Angelini et al. (2014) look at the correlation between sovereign and bank CDS premia, and compare them with the feedback loops between sovereign and corporate sectors. They point out that the sovereign risk spillovers are not limited to the financial sector, but in a similar fashion they affect the corporate risk. Consequently, they argue that country-wide risk appears to be the major underlying risk factor behind the bank-sovereign relation. Furthermore, through a prism of a correlation network analysis, Ohnsorge et al. (2014) report that sovereign-bank relations may amplify shock propagation. They suggest that there is a difference in sovereign-bank relations between so-called safe-havens and other countries, where safe-havens are defined as having two triple-A ratings and a negative beta. In line with Ohnsorge et al. (2014), our empirical evidence confirms different bank-sovereign risk transmissions between such defined country groups, on the sample of EA countries.

This paper is organized as follows. In Section 2 we explain the methodology of NCoVaR and NCoVaR Granger causality. We evaluate the asymptotic properties of the test statistic and we confirm them numerically in Section 3. In Section 4 we test our approach on the EA sample. Section 5 concludes.

2. Methodology

In this section we introduce the main mechanics of the parametric and nonparametric risk transmission measures. For convenience we begin by highlighting the main features of the standard CoVaR methodology. We then introduce a nonparametric extension of CoVaR to finally complement both CoVaR methods with their Granger causality analogues determined by conditional rather than unconditional dependence.

2.1. CoVaR

To start with the basics, the unconditional Value-at-Risk (VaR_γ) of a continuous random variable Y representing losses of an institution, j, say, is defined as the γ^{th} quantile of Y for

a given probability $\gamma \in (0,1)$, that is

$$\mathrm{P}\left(Y \leq \mathrm{VaR}_{\gamma}^{Y}\right) = \gamma. \tag{1}$$

To make the presentation transparent and consistent with our further argumentation, let us denote the losses of another institution, i, say, by X. The conditional Value-at-Risk, or CoVaR, proposed by Adrian and Brunnermeier (2016), measures the effect of an event $\mathbb{C}(X)$ occurring in institution i on VaR_{γ} of institution j. To put it formally, one can rewrite CoVaR as

$$\mathrm{P}\left(Y \leq \mathrm{CoVaR}_{\gamma}^{Y|\mathbb{C}(X)} | \mathbb{C}(X)\right) = \gamma. \tag{2}$$

To capture risk transmission effects, also referred to as tail dependence, Adrian and Brunnermeier (2016) introduced the measure ΔCoVaR, which measures the change in CoVaR of an institution j when the conditioning event of institution i changes. ΔCoVaR measures the effect of a shift in X from the median to the tail quantile (from a safe to a risky state) of institution i on the performance of institution j. Formally, one can define

$$\Delta\mathrm{CoVaR}_{\gamma}^{Y|X} = \mathrm{CoVaR}_{\gamma}^{Y|X=\mathrm{VaR}_{\gamma}^{X}} - \mathrm{CoVaR}_{\gamma}^{Y|X=\mathrm{VaR}_{0.5}^{X}}, \tag{3}$$

measuring the shift in the conditional Y-quantile, in response of a change in X from the median to the γ-th quantile.

2.2. Nonparametric CoVaR (NCoVaR)

We first intrduce the nonparametric extension of ΔCoVaR, called ΔNCoVaR. It is conceptually closely related to CoVaR, but we focus on conditional tail event probabilities rather than conditional quantiles, and on nonparametric estimation rather than parametric.

Let A denote a set of extreme events for Y, such as Y being near a given unconditional tail quantile $y_{\gamma} \equiv \mathrm{VaR}_{\gamma}^{Y}$, and C and D sets of events where X is either near a given unconditional tail quantile (x_{γ}) of X or near its unconditional median ($x_{0.5}$). To make this explicit, we

consider the intervals

$$\begin{aligned} A &= [y_\gamma - \mu, y_\gamma + \mu], \\ C &= [x_\gamma - \mu, x_\gamma + \mu], \\ D &= [x_{0.5} - \mu, x_{0.5} + \mu], \end{aligned}$$

for some positive parameter μ. Then we can quantify the effect of a change in X from region C to D on the tail event probability of Y by defining

$$\Delta\text{NCoVaR} = \text{P}(Y \in A | X \in C) - \text{P}(Y \in A | X \in D).$$

The advantage of focusing on probabilities of events rather than on quantiles is that, as we will show next, this allows for natural nonparametric estimation using U-statistics, for which asymptotic theory is readily available. From the definition of ΔNCoVaR, we obtain

$$\begin{aligned} \Delta\text{NCoVaR} &= \text{P}\left(Y \in A | X \in C\right) - \text{P}\left(Y \in A | X \in D\right) \\ &= \frac{\text{P}\left(Y \in A, X \in C\right)}{\text{P}\left(X \in C\right)} - \frac{\text{P}\left(Y \in A, X \in D\right)}{\text{P}\left(X \in D\right)}. \end{aligned}$$

The null hypothesis of no (instantaneous) NCoVaR relation of the variable X on the tail probability of Y, implies that $\Delta\text{NCoVaR} = 0$ for all events A, C and D for which $P(X \in C) > 0$ and $P(X \in D) > 0$. Upon multiplication by $\text{P}\left(X \in D\right)\text{P}\left(X \in C\right)$, H_0 can be seen to imply

$$q \equiv \text{P}\left(Y \in A, X \in C\right)\text{P}\left(X \in D\right) - \text{P}\left(Y \in A, X \in D\right)\text{P}\left(X \in C\right) = 0,$$

which then is, in fact, also defined in cases where $P(X \in C)$ and/or $P(X \in D)$ happen to be zero. Note that equivalently, we may write H_0 as

$$q = \text{E}\left[I_{A\times C}(Y_1, X_1) I_D(X_2) - I_{A\times D}(X_1, Y_1) I_C(X_2)\right] = 0,$$

for two vectors (X_1, Y_1) and (X_2, Y_2), drawn independently from the joint distribution of

(X, Y).

Given a sample of the process $\{(X_t, Y_t)\}$, $t = 1, \ldots, n$, a natural frequency count-based estimate of q is

$$q_n = \frac{1}{n(n-1)} \sum_{\ell \neq k} \sum \left[I_{A \times C}(Y_k, X_k) I_D(X_\ell) - I_{A \times D}(Y_k, X_k) I_C(X_\ell) \right].$$

To develop asymptotics, define $W_t = (X_t, Y_t)$ and write the estimator as a weighted average of a symmetric kernel function (as a U-statistic)

$$q_n = \frac{1}{n(n-1)} \sum_{k \neq \ell} \sum \mathcal{K}(W_k, W_\ell),$$

where $\mathcal{K}(W_k, W_\ell)$ is the symmetric (w.r.t. swapping W_k and W_ℓ) kernel function

$$\mathcal{K}(W_k, W_\ell) = \frac{1}{2} \left[I_{A \times C}(Y_k, X_k) I_D(X_\ell) - I_{A \times D}(Y_k, X_k) I_C(X_\ell) + \ell \leftrightarrow k \right],$$

where $\ell \leftrightarrow k$ stands for similar terms with ℓ and k swapped. This shows that q_n is in fact a U-statistic estimator of $q = E(\mathcal{K}(W_i, W_j))$, (where W_i and W_j are drawn independently from the stationary distribution of W). Although in a time series process the observed vectors W_t, $t = 1, \ldots, n$ are not independent, as long as the time series $\{W_t\}$ is stationary and satisfies some rather mild mixing conditions, the asymptotic theory of U-statistics still apply, provided that a HAC estimator of variance is used (Denker and Keller, 1983, 1986). This leads to the following theorem.

Theorem 1. *Consider a sample $\{W_t\}_{t=1}^n$ from bivariate random process $\{W_t\} \equiv \{(X_t, Y_t)\}$ with $t \in \mathbb{Z}$ that is strictly stationary and β-mixing with exponential decay rate. Then for the kernel function $\mathcal{K}(\cdot, \cdot)$ as defined above, for fixed A, C and D,*

$$\sqrt{n} \frac{q_n - q}{S_n} \xrightarrow{d} N(0, 1),$$

where S_n^2 is a heteroskedasticity and covariance consistent (HAC) estimator of the asymptotic variance of $\sqrt{n}(q_n - q)$.

The proof of Theorem 1 is provided in Appendix A.

The present definition of NCoVaR focuses on extreme losses equal to, or μ-near, a given tail quantile, rather than above this quantile, which is the case considered more often in the literature. The background of this choice is historical, as we were initially thinking of NCoVaR in terms of densities near given tail quantiles. If desired, the test can be readily adjusted by changing the events A (and/or C and D) to the regions above a certain quantile rather than μ-close to them, without affecting the validity of the proof of asymptotic normality. To anticipate on this possibility we performed numerical size and power comparisons (shown in Appendix E due to space considerations) for such an alternative setup. For commonly-used quantiles, such as $\gamma = 0.95$ or $\gamma = 0.99$, the results indicate that the test as introduced here actually offers comparable or better power, depending on the simulation setup.

2.3. CoVaR Granger causality

In Granger causality testing, the goal is to find evidence against the null hypothesis of Granger non-causality, defined as follows.

Definition 1 (Granger non-causality (bivariate))**.** *For a strictly stationary bivariate time series process $\{(X_t, Y_t)\}$, $t \in \mathbb{Z}$, $\{X_t\}$ does not Granger cause $\{Y_t\}$ if, for all $k \geq 1$,*

$$(Y_{t+1}, \ldots, Y_{t+k})|(\mathcal{F}_{X,t}, \mathcal{F}_{Y,t}) \sim (Y_{t+1}, \ldots, Y_{t+k})|\mathcal{F}_{Y,t},$$

where $\mathcal{F}_{X,t}$ and $\mathcal{F}_{Y,t}$ are information sets spanned by X_s, $s \leq t$ and Y_s, $s \leq t$, respectively.

For now, we focus on $k = 1$ lag. Following Diks and Panchenko (2006) and Diks and Wolski (2016) one can represent the null hypothesis of Granger non-causality in terms of equality of conditional probabilities. Throughout we assume that the process $\{(X_t, Y_t), t \in \mathbb{Z}\}$ is

strictly stationary and β-mixing with exponential decay rate. For the ease of notation we introduce the lead variable $Z_t = Y_{t+1}$. In this notation the null hypothesis is a statement about the invariant distribution evaluated at conditional quantile levels of the 3-dimensional vector $W_t = (X_t, Y_t, Z_t)$. For clarity, since the null hypothesis concerns the invariant distribution of W_t, in formulating the null hypothesis we often drop the time index and refer simply to the distribution of the random variable $W = (X, Y, Z)$.

Under the null hypothesis of Granger non-causality X and Z are conditionally independent given $Y = y^*$, where y^* is a given unconditional quantile of Y. Consequently, the Granger causality setting adds an additional conditioning event relative to NCoVaR. We can either address this using quantile regression, leading to an extension of CoVaR to CoVaR Granger causality, or nonparametrically, giving rise to NCoVaR Granger causality.

CoVaR Granger causality is based on quantile regression, but we now consider quantiles of Z, being the future Y-variable, instead of Y. In that context the null hypothesis H_0 implies that the conditional quantiles of Z are independent of X given Y, or, the conditional quantiles of $Z|X = x$, $Y = y$ and $Z|Y = y$ are the same for all (x, y) in the support of (X, Y).

Note that unlike CoVaR, CoVaR Granger causality is directional, i.e. the effect of X on future values of Y need not be the same as that of Y on future X-values. As such it is a measure of Granger causality from one variable to another. This stands in great contrast with linear CoVaR, which essentially depends on the instantaneous correlation between two variables, and hence is inherently symmetrical.

2.4. NCoVaR Granger causality

The null hypothesis is Granger non-causality, i.e. for $(X, Y, Z) \sim (X_t, Y_t, Y_{t+1})$ under the null hypothesis H_0 vectors X and Z are conditionally independent given Y. The same as with NCoVaR, we take a nonparametric approach, starting by considering events where Z

is in of A, and X in C and/or D, where now

$$\begin{aligned} A &= [z_\gamma - \mu, z_\gamma + \mu] \\ C &= [x_\gamma - \mu, x_\gamma + \mu] \\ D &= [x_{0.5} - \mu, x_{0.5} + \mu]. \end{aligned}$$

The null hypothesis implies (see Appendix B)

$$\frac{P(Z \in A, X \in C|Y = y)}{P(X \in C|Y = y)} = \frac{P(Z \in A, X \in D|Y = y)}{P(X \in D|Y = y)}, \qquad \forall A, C, D, y.$$

The analogy with NCoVaR above now suggests testing the implication of the null hypothesis

$$\begin{aligned} H_0' : \qquad & P(Z \in A, X \in C|Y = y_*)P(X \in D|Y = y_*) \\ & -P(z \in A, X \in D|Y = y_*)P(X \in C|Y = y_*) = 0, \end{aligned}$$

for a given past Y_t value y_* (e.g. some unconditional Y-quantile) and given A, C and D. For instance $P(X \in D|Y = y_*)$ is now estimated by counting the frequency of events $X \in D$, among the vectors close to y_*.

A natural estimator for $P(X \in D|Y = y_*)$ is the Nadaraya-Watson nonparametric regression function estimator

$$\widehat{P(X \in D|Y} = y_*) = \frac{\frac{1}{n}\sum_{k=1}^{n} I_D(Z_k)K_h(y_* - Y_k)}{\frac{1}{n}\sum_{k=1}^{n} K_h(y_* - Y_k)}, \tag{4}$$

where we take $K_h(\cdot)$ to be a density estimation kernel. Although the theory holds more general, in the simulations and applications presented herein we focus on the Gaussian kernel

$$K_h(s) = \frac{1}{\sqrt{2\pi}h} \exp(-s^2/(2h^2)),$$

and its associated higher-order kernels (see e.g. Hansen, 2009). Note that the denominator in Eq. (4), which is preventing us from writing it as a U-statistic, is just a kernel density estimate

of $f_Y(y_*)$. We can get rid of the denominator and obtain simple U-statistics estimates[2] if we multiply the probabilities by $f_Y(y_*)$. Therefore, for a given unconditional quantile y_* of Y we define

$$\begin{aligned} q_* &= f_Y^2(y_*)\left(P(Z \in A, X \in C|Y = y_*)P(X \in D|Y = y_*)\right. \\ &\quad \left. -P(z \in A, X \in D|Y = y_*)P(X \in C|Y = y_*)\right). \end{aligned}$$

By construction, $q_* = 0$ under H_0. Now the term

$$f_Y(y_*)P(X \in D|Y = y_*)$$

e.g. can be simply estimated as

$$\frac{1}{n}\sum_{k=1}^{n} I_D(X_k)K_h(y_* - Y_k).$$

The corresponding U-statistic kernel used for estimation of q_* is

$$\begin{aligned} \mathcal{K}(W_k, W_\ell; h) &= \tfrac{1}{2}\left[I_A(Z_k)I_C(X_k)K_h(y_* - Y_k)I_D(X_\ell)K_h(y_* - Y_\ell)\right. \\ &\quad -I_A(Z_k)I_D(X_k)K_h(y_* - Y_k)I_C(X_\ell)K_h(y_* - Y_\ell) \\ &\quad \left. +k \leftrightarrow \ell\right], \end{aligned} \tag{5}$$

where '$k \leftrightarrow \ell$' represents the same terms with k and ℓ swapped and $K_h(w) = h^{-1}K(w/h)$ a scaled version of the kernel function $K(w)$, satisfying

$$\int |K(w)|dw < \infty, \qquad \int K(w)dw = 1 \qquad \text{and} \qquad |wK(w)| \to 0 \quad \text{as} \quad |w| \to \infty. \tag{6}$$

The next theorem states that the U-statistics estimator of q_*, given by

$$q_{*,n} = \frac{1}{n(n-1)}\sum\sum_{\ell \neq k} \mathcal{K}(W_k, W_\ell; h),$$

[2]Note that we slightly abuse language here. Strictly speaking, we should call the resulting estimators sample averages of kernel functions rather than U-statistics, since they are only asymptotically unbiased.

is asymptotically normally distributed provided that the bandwidth $h = h_n$ tends to zero with n at an appropriate rate.

Theorem 2. *Consider a strictly stationary bivariate random process* $\{(X_t, Y_t)\}$ *with* $t \in \mathbb{Z}$. *Then for given the events A, C and D, and Y-quantile* y_*, *for a kernel density bandwidth parameter tending to zero at an appropriate rate*

$$\sqrt{n}\frac{q_{*,n} - q_*}{S'_n} \xrightarrow{d} N(0,1),$$

where S'^2_n *is a consistent (HAC) estimator of the asymptotic variance of* $\sqrt{n}(q_{*,n} - q_*)$

The proof of Theorem 2 is provided in Appendix C. Here by an 'appropriate' rate for the bandwidth to tend to zero, we mean that in the derivation of asymptotic normality we let the bandwidth tend to zero as $h_n = cn^{-\beta}$, with an appropriate fixed rate β. This rate should satisfy a number of conditions depending on the density estimation kernel order and the dimension of W. For a second-order ($\alpha = 2$) density estimation kernel, and $d_Y = 1$ (first Markov-order bivariate), we find that $\beta \in (\frac{1}{4}, \frac{1}{2})$ with MSE-optimal rate $\beta = \frac{1}{3}$. See Appendix C for details.

2.5. Optimal bandwidth selection

By minimising the MSE of $q_{*,n}$ we can, in addition to the MSE optimal bandwidth rate β, also find the asymptotically optimal value of c in the sequence $h_n = cn^{-\beta}$. Unfortunately, this optimal value of c will be not independent of the data generating process assumed. Nevertheless it will be instructive to find the optimal values of c for a number of commonly considered processes, to get an estimate the order of magnitude of the MSE optimal value of c in these cases. Recall that asymptotically c is not relevant to achieve asymptotic normality, and even β isn't, as long as it is in the appropriate range, but for finite sample sizes it is preferable to use values for β and c that are motivated by optimality arguments, even if for somewhat stylised processes that are only approximations to the real unknown DGP.

Naturally, it is impossible to consider all possible DGPs. Moreover, not for all assumed DGPs it is possible to numerically calculate MSE optimal bandwidths. We here restrict attention to VAR and ARCH models, as these represent common DGPs allowing for linear dependence in the first and second conditional moment, respectively.

An alternative approach would be to apply cross-validation techniques to calculate optimal data-driven bandwidths (Li and Racine, 2008, 2013). Due to stochasticity of minimization algorithms in small samples, such methods can lead to power losses in the context of hypothesis testing. Nevertheless, we consider them as an interesting avenue for future research.

Vector AR (VAR) model. As shown in Appendix D, in the context of a specific linear VAR model with normal innovations we find an optimal bandwidth for, for instance $\gamma = 0.95$ and $\mu = 0.8$, equal to

$$h_{\text{opt}}^{\text{AR(1)}} = 3.07 \times n^{-1/3},$$

after standardizing the data to zero mean and unit variance and choosing the dependence parameter a equal to 0.4.

ARCH model. In the ARCH model context the bias is of a lower order in the kernel bandwidth, so the optimal bandwidth also has a different rate. We obtain, again for dependence parameter value $a = 0.4$, $\gamma = 0.95$ and $\mu = 0.8$,

$$h_{\text{opt}}^{\text{ARCH(1)}} = 2.14 \times n^{-1/5},$$

after standardizing the data to zero mean and unit variance. For more details and for other values of μ and γ we refer to Appendix D.

Although the optimal bandwidths presented here were obtained for very specific processes, they might help in practice to give an idea of the order of magnitude of the optimal

bandwidth in empirical applications. Specifically, we propose to use $h_{\text{opt}}^{\text{AR(1)}}$ in case there may still be VAR dependence in the time series, and $h_{\text{opt}}^{\text{ARCH(1)}}$ in cases where VAR structure has been removed by filtering using a VAR model.

2.6. Extensions to higher order processes and/or confounding variables

Most of the remarks above concern the bivariate case, with Markov order $k = 1$ lag and density estimation kernel order $\alpha = 2$. Adding more conditional variables, such as lagged Y-variables or possibly confounding variables puts additional restrictions on the feasible bandwidth rates (β-values). Mathematically, extra lagged Y-values or additional variables can be added to the conditioning variable Y, in $W = (X, Y, Z)$, when testing whether X and Z are conditionally independent given Y. The increase of the dimension of Y places additional conditions on the feasible rates β at which h_n can tend to zero without letting the bias or the variance of $q_{*,n}$ dominate asymptotically. Specifically, α and d_Y put the restrictions $\frac{1}{2\alpha} < \beta < \frac{1}{2d_Y}$, on β. Note that for the usual second order density estimation kernels, which have $\alpha = 2$, there are already no feasible rates β as soon as we increase the dimension d_Y of Y from 1 to 2, e.g. by adding a single extra lag of Y_t or the first lag of a possible extra, confounding, variable, V_t, say.

When addressing this issue, one has a choice between Data Sharpening (DS) on one hand, and the use of higher order kernels on the other. Both these methods reduce the order of the bias, opening up some room for the bandwidth to tend to zero slower, and hence reduce the variance. However, there is a huge practical difference between these methods, in that higher order kernels require only a single bandwidth to be considered, while in DS there is one bandwidth for the data sharpening step and another for the estimation step, and these need to be carefully adjusted to each other. For instance, as noted by Diks and Wolski (2016) the DS bandwidth should go to zero at a slower rate than the density estimation bandwidth, to make sure that the gradient of the density is estimated consistently. Therefore, in this study we decided to use higher-order kernels rather than DS when Y is multivariate. In the case described above, with $d_Y = 2$, using a 4th-order ($\alpha = 4$) kernel provides a range of

feasible β-values, $\beta \in (\frac{1}{8}, \frac{1}{4})$.

Formally, if we denote the vector of extra conditioning variables by V, one can represent the null hypothesis of the multivariate NCoVaR as

$$\mathrm{P}(Y \in A | X \in C, V = v_*) = \mathrm{P}(Y \in A | X \in D, V = v_*),$$

and for multivariate NCoVaR Granger causality as

$$\frac{P(Z \in A, X \in C | Y = y_*, V = v_*)}{P(X \in C | Y = y_*, V = v_*)} = \frac{P(Z \in A, X \in D | Y = y_*, V = v_*)}{P(X \in D | Y = y_*, V = v_*)}.$$

The latter equation illustrates explicitly that adding extra control variables to condition on is mathematically equivalent to increasing the dimension of the conditioning variable Y.

3. Size/power simulations

To compare the parametric and nonparametric methods, we measure their ability to detect given dependency structures between variables. Our strategy is to simulate processes with stylized transmission channels and to verify statistical power of both methodologies. Since CoVaR and NCoVaR focus on instantaneous dependence, whereas CoVaR Granger causality and NCoVaR Granger causality considers Granger-type dependence, for the former we simulate processes with simultaneous and for the latter with lagged dependence. We then benchmark the results against standard and lagged parametric CoVaR specifications, respectively.

As argued by Adrian and Brunnermeier (2016), ΔCoVaR can be estimated as $\Delta\text{CoVaR} =$

$\hat{\beta}^i_\gamma(\text{VaR}^i_\gamma - \text{VaR}^i_{0.5})$, where $\hat{\beta}^i_\gamma$ comes from the quantile regression[3]

$$\hat{Y_\gamma}^{j|X} = \hat{\alpha}^i_\gamma + \hat{\beta}^i_\gamma X, \tag{7}$$

where $\hat{Y_\gamma}^{j|X}$ is the predicted value for γ-quantile of institution j conditional on a return realization X of institution i. In fact, variable $\hat{\beta}^i_\gamma$ captures the tail dependence between the institutions and is the core variable of interest for our further investigation. Under standard distributional assumptions, the estimated coefficient follows a Student's t-distribution with $n-2$ degrees of freedom. The statistical significance of $\hat{\beta}^i_\gamma$ is therefore a direct measure to assess the size and power of the parametric approach. Furthermore, to correct for possible heteroskedasticity we estimate the t-statistics for $\hat{\beta}^i_\gamma$ with robust standard errors.

We carry out a one-sided t-test where under the null $H_0 : \hat{\beta}^i_\gamma \leq 0$ and under the alternative $H_a : \hat{\beta}^i_\gamma > 0$, to make the test size directly comparable with our further investigation.

To construct CoVaR estimates for lagged dependency structure, we adjust the lag composition of Eq. (7) to match Definition 1. Besides, the testing framework follows the same principles as described above.

We consider two groups of processes, highlighting the dependence in the first and second conditional moments of the random variables. These processes have been widely used as simulation benchmarks and therefore constitute a natural testing environment (see Li and Racine (2007) or Diks and Wolski (2016)). One could also focus on a combination of both types of dependencies simultaneously, as for instance Rothe (2010). However, as we demonstrate later, the sensitivity of parametric and nonparametric methods is vastly different across these two groups of processes. We therefore consider them separately.

[3]An alternative approach would be to estimate ΔCoVaR via a multivariate GARCH model. Under the Gaussian case, the main difference between two techniques is that whereas in the quantile regression case the estimate is proportional to the overall correlation between variables, the multivariate GARCH estimate is proportional to the instantaneous correlation. Consequently, neither method can capture the dependency structure which is not related to correlation, which builds an argument for the nonparametric extensions proposed herein.

3.1. Dependence in the first conditional moment

To give an example of a data-generating process with causality in mean, consider the linear bivariate Vector Autoregressive model of order 1 (VAR(1)) given by

$$\begin{aligned} x_t &= a x_{t-1} + \sqrt{1-a^2}\varepsilon_{1,t}, \\ y_t &= a x_{t-\tau} + \sqrt{1-a^2}\varepsilon_{2,t}, \end{aligned} \tag{8}$$

where $a \in (0,1)$ is a tuning parameter and $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ represent i.i.d. zero-mean innovations. We restrict the parameter a to be within the unit interval to guarantee that the process is stationary. The process is designed so that the causality runs from X to Y, which constitutes a foundation for power assessment. For the size assessment we use the same process, but switch the causality from Y to X, so that the null hypothesis of non-causality holds.

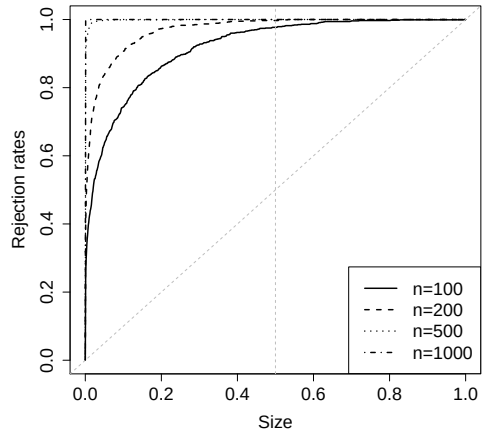
The subscripts denote the time dimension. Here we make an important distinction between instantaneous and lagged dependence spillovers or, by a slight abuse of terminology, 'instantaneous' and 'standard' Granger causality. By adjusting the dependence lag structure, we match the timing of the data generating process to (N)CoVaR (Granger causality) measures, providing an appropriate testing framework for each. In particular, we set $\tau = 0$ for 'instantaneous' Granger causality to assess the size and power of CoVaR and NCoVaR. We investigate the properties of the CoVaR and NCoVaR Granger causality-based tests on the process with lagged dependence, i.e. $\tau = 1$.

For the simulations we set $a = 0.4$ and run 1000 independent realizations of the process in Eq. (8), after a burn-in period of 100 time steps. Sample size-dependent bandwidths are set at the calculated MSE-optimal value and the fixed-range parameter is set at $\mu = 0.8$, as this seems to provide consistently good size and power properties (for comparison see Appendix F).

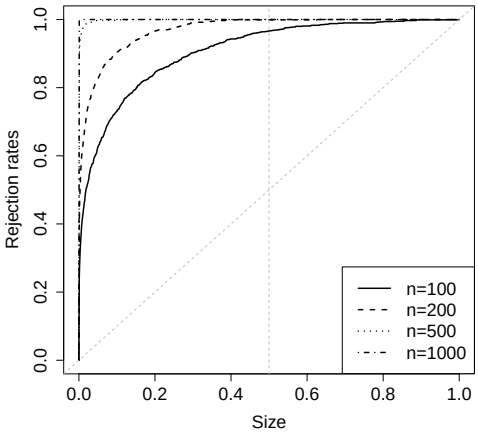
The results for the former are presented in Fig. 1, whereas for the latter in Fig. 2. Estimations focus on the risky quantile of $\gamma = 0.95$. The results for $\gamma = 0.99$ and for above-quantile specification are given in Appendix E. However, since for smaller samples standard

errors cannot be estimated consistently, the results comprise larger samples. The detailed results for selected nominal size levels are additionally summarized in Table 2.

CoVaR

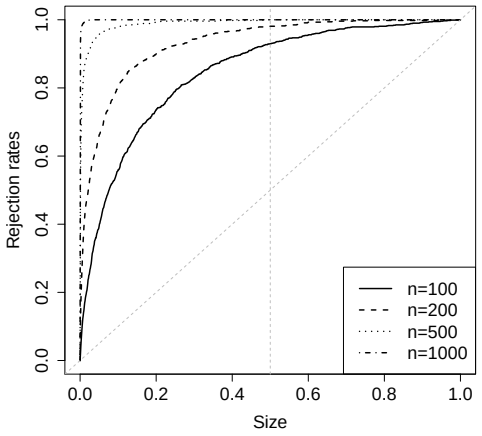


(a) Size-size plot

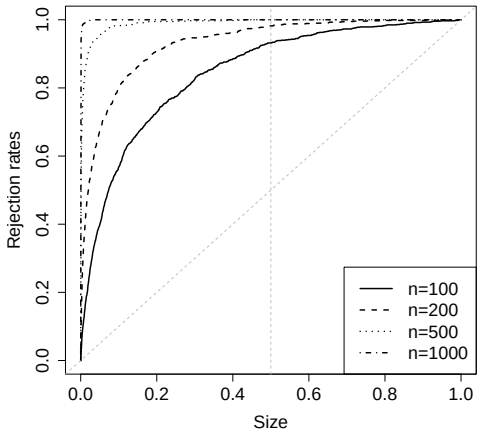


(b) Size-power plot

NCoVaR



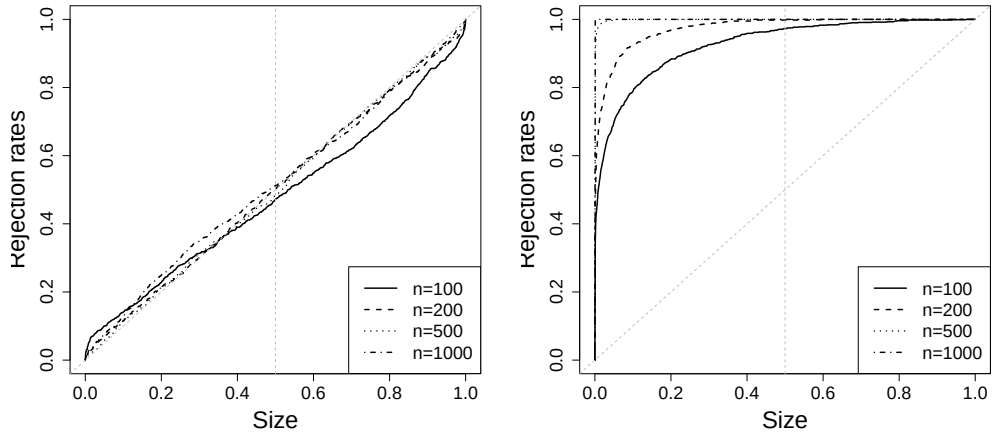
(c) Size-size plot



(d) Size-power plot

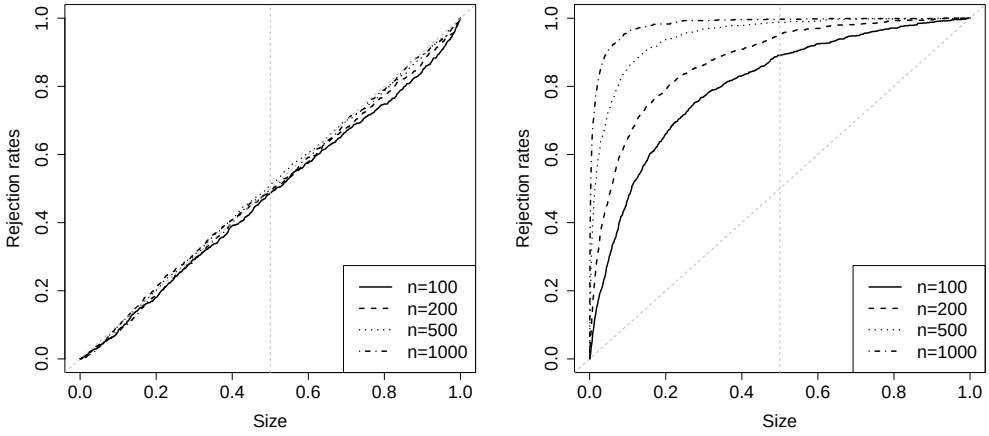
Figure 1: CoVaR and NCoVaR performance in VAR class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (8) with instantaneous dependence ($\tau = 0$). The quantile at which the risk is defined is set to $\gamma = 0.95$. The fixed-range parameter is set to $\mu = 0.8$. The results are aggregated over 1000 simulations.

CoVaR Granger causality



(a) Size-size plot (b) Size-power plot

NCoVaR Granger causality



(c) Size-size plot (d) Size-power plot

Figure 2: CoVaR and NCoVaR Granger causality performance in VAR class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (8) with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $\gamma = 0.95$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

Table 2: Performance summary of CoVaR and NCoVaR methodologies in VAR class of models for selected nominal size levels (α). Actual rejection rates under $X \to Y$ causality (power) and $Y \to X$ causality (size) for different sample sizes (n), generated under process in Eq. (8) with instantaneous dependence ($\tau = 0$) and with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $\gamma = 0.95$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

		Instantaneous dependence				Granger causality			
α	n	size		power		size		power	
		CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR
0.01	100	0.395	0.147	0.462	0.142	0.053	0.007	0.511	0.087
	200	0.621	0.391	0.625	0.374	0.024	0.004	0.681	0.186
	500	0.976	0.850	0.972	0.831	0.016	0.002	0.979	0.432
	1000	1.000	0.991	1.000	0.990	0.030	0.003	1.000	0.702
0.05	100	0.610	0.394	0.633	0.397	0.100	0.037	0.698	0.289
	200	0.837	0.661	0.825	0.675	0.070	0.035	0.857	0.473
	500	0.998	0.957	0.997	0.953	0.058	0.034	0.999	0.730
	1000	1.000	1.000	1.000	1.000	0.074	0.042	1.000	0.908
0.1	100	0.734	0.560	0.741	0.567	0.141	0.081	0.789	0.468
	200	0.907	0.803	0.904	0.803	0.115	0.098	0.923	0.648
	500	0.999	0.979	0.999	0.983	0.114	0.095	1.000	0.849
	1000	1.000	1.000	1.000	1.000	0.130	0.078	1.000	0.958

There are three main observations that can be made based on this experiment. Firstly, the instantaneous CoVaR and NCoVaR measures seem to be bi-directional, whereas the Granger causality-based specifications are sensitive to the direction of causality. Regarding the former, we note that Adrian and Brunnermeier (2016) consider such a property a virtue, as the methodology captures the co-risk effects between variables. Our numerical exercise confirms this feature for both CoVaR and NCoVaR measures.

Secondly, regarding the Granger causality measures, it seems that parametric estima-

tion modestly over-rejects under the null, i.e. the methodology finds evidence for CoVaR Granger causality too often when it is actually absent. This over-rejection is confirmed for conventional nominal size levels and doesn't seem to diminish with increasing sample size (see Table 2). On the contrary, NCoVaR Granger causality displays much more conservative size properties.

Thirdly, for the VAR(1) process we find a considerable power gain for CoVaR relative to NCoVaR in small samples for the $\gamma = 0.95$ quantile. The differences evaporate as the sample size increases and the instantaneous NCoVaR converges faster in power than the Granger causality setup. There are several possible reasons for these results. As confirmed by Rothe (2010), parametric models are characterized by higher efficiency (in terms of MSE) for correctly specified models. Our simulation setup assumes the simplest process dynamics, which corresponds to the model specifications. Secondly, the strong power of CoVaR can be partially driven by its over-rejection bias, at least to some extent. Thirdly, the slower convergence for NCoVaR Granger causality can be attributed to the 'curse of dimensionality', i.e. less precise estimates in higher dimensions. There are several possible remedies to this problem, including data sharpening, principal components, projection pursuit or informative components analysis (Scott, 1992; Hall and Minnotte, 2002). This topic is, however, beyond the scope of this paper and we leave it for further investigation.

We also find the interesting pattern that for the same sample size estimation at more risky quantiles, i.e. $\gamma = 0.99$, yields superior statistical power for both CoVaR and NCoVaR, with marginal effects on the size (see Table E2 in Appendix E). Points around more aggressive quantiles contain clearer evidence against the null hypothesis. Nevertheless, due to data scarcity for smaller data sets estimation at less conservative quantiles may be preferred.

3.2. Dependence in the second conditional moment

The second-moment dependence is analysed through a prism of a class of (Generalized) Autoregressive Conditional Heteroskedasticity ((G)ARCH) models. In particular, we focus on a stylized bi-variate (G)ARCH process with (possibly lagged) volatility spillovers from

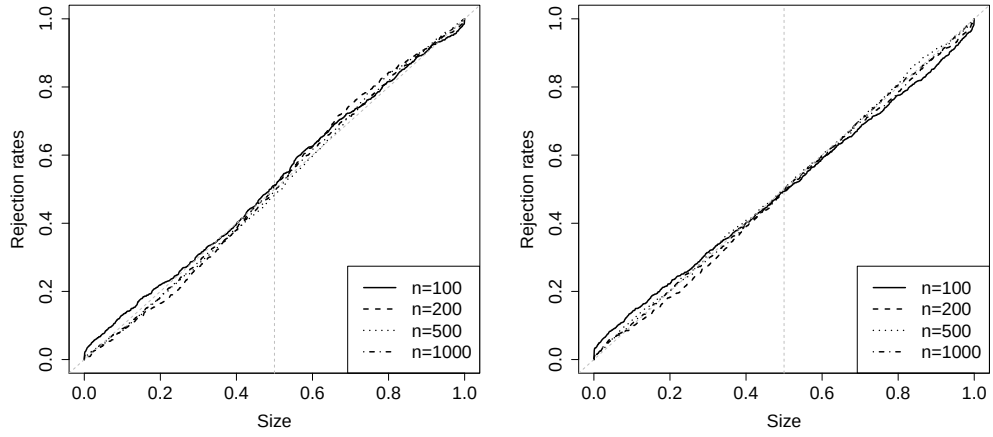
$\{X_t\}$ onto $\{Y_t\}$ of the form

$$\begin{aligned} X_t &\sim N(0,1), \\ Y_t &\sim N(0, 1 + aX_{t-\tau}^2), \end{aligned} \tag{9}$$

where $a > 0$ is again tuning parameter and $N(0, \sigma^2)$ denotes the zero-centred normal distribution with variance σ^2.

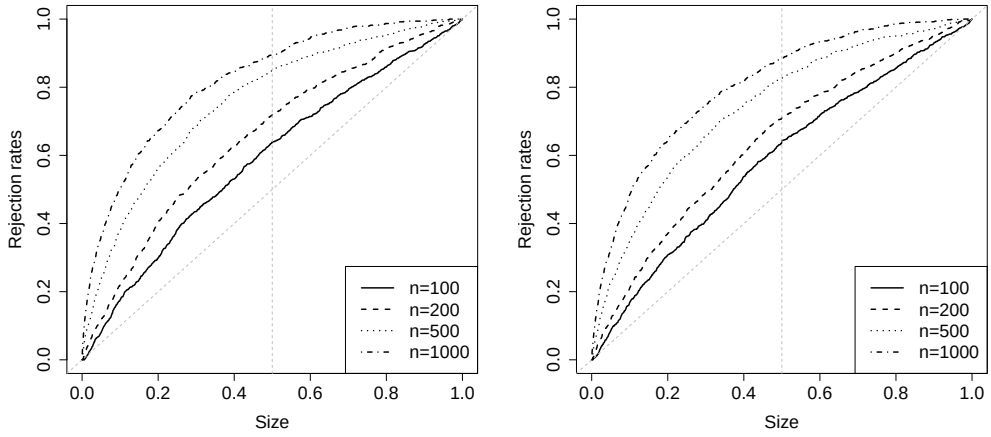
The nomenclature and testing procedure are the same as for the VAR(1) process described in Section 3.1. In the simulations we set $a = 0.4$ and we focus on a risky quantile of $\gamma = 0.95$. (Results for $\gamma = 0.99$ are provided in Appendix E.) The size-size and size-power plots are shown in Figs 3 and 4. The corresponding detailed results for selected nominal size levels can be found in Table 3.

CoVaR



(a) Size-size plot (b) Size-power plot

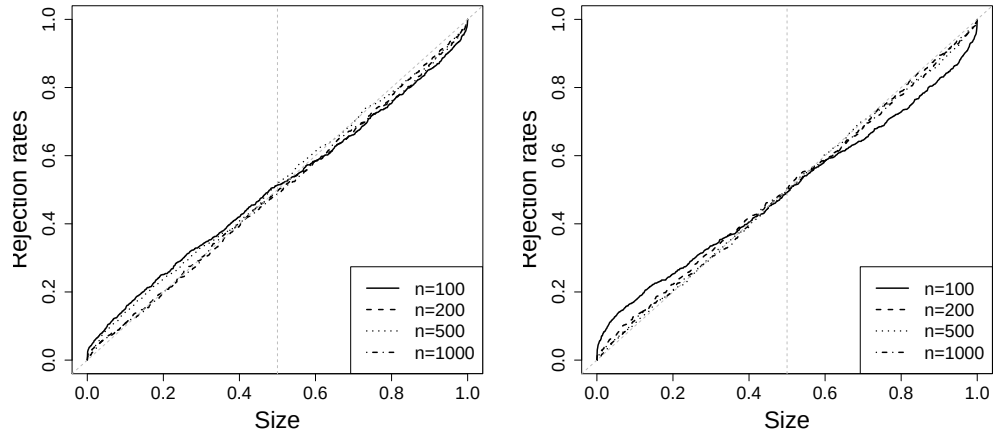
NCoVaR



(c) Size-size plot (d) Size-power plot

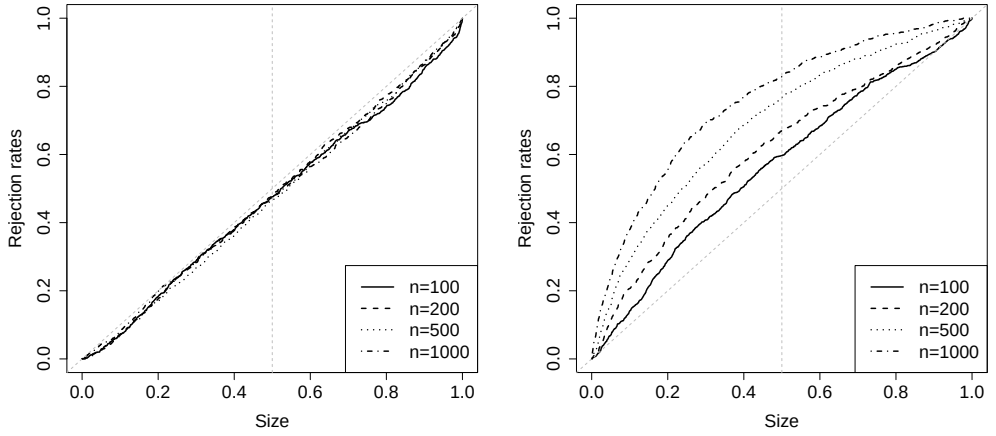
Figure 3: CoVaR and NCoVaR performance in (G)ARCH class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (9) with instantaneous dependence ($\tau = 0$). The quantile at which the risk is defined is set to $\gamma = 0.95$. The fixed-range parameter is set to $\mu = 0.8$. The results are aggregated over 1000 simulations.

CoVaR Granger causality



(a) Size-size plot (b) Size-power plot

NCoVaR Granger causality



(c) Size-size plot (d) Size-power plot

Figure 4: CoVaR and NCoVaR Granger causality performance in (G)ARCH class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (9) with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $\gamma = 0.95$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

Table 3: Performance summary of CoVaR and NCoVaR methodologies in (G)ARCH class of models for selected nominal size levels (α). Actual rejection rates under $X \to Y$ causality (power) and $Y \to X$ causality (size) for different sample sizes (n), generated under process in Eq. (9) with instantaneous dependence ($\tau = 0$) and with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $\gamma = 0.95$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

		Instantaneous dependence				Granger causality			
α	n	size		power		size		power	
		CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR
0.01	100	0.023	0.006	0.042	0.008	0.042	0.002	0.061	0.008
	200	0.018	0.027	0.020	0.024	0.026	0.004	0.025	0.017
	500	0.018	0.078	0.017	0.066	0.029	0.000	0.025	0.022
	1000	0.009	0.146	0.018	0.120	0.015	0.005	0.011	0.065
0.05	100	0.061	0.080	0.088	0.086	0.096	0.028	0.129	0.070
	200	0.043	0.118	0.062	0.116	0.066	0.026	0.073	0.109
	500	0.045	0.235	0.062	0.229	0.086	0.029	0.063	0.158
	1000	0.045	0.363	0.063	0.324	0.066	0.043	0.060	0.240
0.1	100	0.114	0.176	0.139	0.169	0.150	0.070	0.174	0.141
	200	0.084	0.220	0.096	0.211	0.106	0.072	0.129	0.207
	500	0.084	0.381	0.115	0.351	0.141	0.071	0.107	0.290
	1000	0.091	0.505	0.105	0.485	0.108	0.083	0.117	0.380

The experiments on the (G)ARCH process confirm the first two findings reported in Section 3.1, i.e. the bi-directionality of the CoVaR and NCoVaR measures, and the over-rejection bias of CoVaR Granger causality, although the latter seems to somehow contained. Regarding the power results, parametric CoVaR estimation is unable to detect the volatility spillovers generated by process in Eq. (9), irrespective of the sample size. On the contrary, NCoVaR and NCoVaR Granger causality still capture this type of dependence, however, the power is considerably subdued compared to the VAR experiments.

The poor performance of CoVaR measures in (G)ARCH environment can be explained by its methodological design. As argued by Mainik and Schaanning (2014), CoVaR is a correlation-driven measure. Having pointed this out, it misses any type of dependence in the higher moments of the conditional variable distributions. To put it pragmatically, $\hat{\beta}^{i}_{\gamma}$ estimates capture the average linear quantile effects between the variables of interest. In the case of (G)ARCH (or volatility spillovers) processes, the left-tail effects are offset by the right-tail equivalents, on average, which escalates the standard errors and reduces the statistical power of the method. Under such circumstances, NCoVaR provides a robust estimation alternative.

4. Empirical illustration

To demonstrate the performance of NCoVaR and NCoVaR Granger causality we choose the Euro Area (EA) financial environment. In particular, we investigate the so-called feedback loops (after Ohnsorge et al. (2014)), between sovereigns and banks in selected EA Member States. Feedback loops are of particular importance for policy makers and regulators as they serve as a shock transmission channel during distress times. Banks, as important sovereign debt holders, are directly exposed to debt valuation and sovereign risk. Should a sovereign be under distress banks might be required to increase capital buffers or take a haircut. On the other side, sovereigns are implicit guarantors of the banking sector and they took a huge hit on their debt accounts during the financial and subsequent sovereign debt crises. In essence, the prices of both instruments showed a high degree of co-movement in the recent history across different parts of the Europe (Caporin et al., 2012).

The data used in the empirical analysis covers seven countries, i.e. two so-called core EA Member States: Germany and France, and five vulnerable EA Member States: Spain, Portugal, Italy, Ireland and Greece. The Sovereign Price Index (SPI) is calculated from the price-yield relation of a 1-year zero-coupon bond, on the basis of a generic 1-year sovereign bond yield for each country. The Banking Price Index (BPI) is taken as the FTSE banking

price index for each country. SPI come from Bloomberg and BPI come from Datastream. The missing observations are interpolated using linear interpolation technique but the results fully hold when excluding the missing values. We focus on daily observations between January 1994 and September 2016, however, due to data availability the precise ranges differ across countries. The exact coverage together with basic summary statistics are depicted in Table 4.

Table 4: Summary statistics of the data used in the empirical analysis. Banking prices are taken from the FTSE Banking Price Index (BPI) and the Sovereign Price Index (SPI) is calculated from 1-year zero-coupon bonds of corresponding countries. Source: Bloomberg and Datastream.

Sector	Country	Obs.	Sample period	Mean	St. Dev.	Min.	Max.
BPI	Germany	5932	Jan-1994/Sep-2016	125.9	62.3	23.6	267.4
	France	5932	Jan-1994/Sep-2016	208.3	93.8	69.9	506.7
	Spain	5932	Jan-1994/Oct-2016	359.9	146.2	109.6	704.1
	Portugal	4244	Apr-1998/Aug-2014	151.5	94.3	3.1	373.7
	Italy	5932	Jan-1994/Oct-2016	158.6	86.1	32.0	352.6
	Ireland	5932	Jan-1994/Oct-2016	670.5	617.7	7.9	2279.0
	Greece	4803	May-1998/Oct-2016	622.3	495.3	0.3	1713.8
SPI	Germany	5932	Jan-1994/Sep-2016	976.6	17.0	942.7	1007.0
	France	4803	May-1998/Sep-2016	980.1	15.9	949.7	1006.2
	Spain	4803	May-1998/Oct-2016	975.9	13.9	939.5	1002.4
	Portugal	4803	May-1998/Oct-2016	968.5	25.4	818.0	1001.3
	Italy	4694	Oct-1998/Oct-2016	976.2	14.0	924.5	1002.1
	Ireland	4803	May-1998/Oct-2016	971.2	18.6	855.0	1004.7
	Greece	4215	Aug-2000/Oct-2016	936.0	101.4	222.8	989.5

The data delivers interesting first-touch evidence on the nature of sovereign and banking sectors in the sample. Firstly, BPIs of Germany, France, Portugal and Italy are have, on average, lower valuations and volatility than the ones in Spain, Ireland and Greece. In fact, Germany is characterized by the lowest average in-sample price and standard deviation, with

Ireland having the highest values on both statistics.

Looking at the SPIs, the average debt valuations look, on average, similar, with the exception of Greece, with bonds traded at more than double-the-average discounts. Interestingly, Greece was the only country in the sample, for which the yields never marked the negative territory, and were the most volatile.

In the empirical analysis, to guarantee the stationarity of the time series, we look at the log returns of respective variables. We also standardize the data magnitude by the standard normal transformation. The fixed-bandwidth are set to $\mu = 0.8$ and the size-dependent window is chosen as indicated for the VAR(1) process. (The results are largely robust to different parametrization.)

The goal of the exercise is to quantify the sovereign-bank feedback loops on the sample countries and to compare the NCoVaR and NCoVaR Granger causality estimates against their parametric CoVaR equivalents. The main results for $\gamma = 0.95$ are depicted in Table 5.[4]

[4]The results for $\gamma = 0.99$ are largely in line but their statistical significance is weaker. For transparency reasons we do not report them in this paper but they are available from the authors upon request.

Table 5: Bank-sovereign feedback loops in selected euro area countries. BPI and SPI denote the Banking Price Index and Sovereign Price Index, respectively. Columns CoVaR and NCoVaR denote the instantaneous specifications, whereas columns CoVaR Gc and NCoVaR Gc correspond to Granger causality setups. ***, **, ** denote 1%, 5% and 10% significance levels. For nonparametric NCoVaR and NCoVaR Granger causality tests we set $\mu = 0.8$. Risky quantiles are estimated at $\gamma = 0.95$.

			CoVaR		NCoVaR		CoVaR Gc		NCoVaR Gc	
	X	Y	$X \to Y$	$Y \to X$	$X \to Y$	$Y \to X$	$X \to Y$	$Y \to X$	$X \to Y$	$Y \to X$
Germany	BPI	SPI						***		
France	BPI	SPI						***		
Spain	BPI	SPI	***	**	***	***	**	***		**
Italy	BPI	SPI	***	***		**	***	***		*
Portugal	BPI	SPI	***	***			***	***		
Ireland	BPI	SPI					***	***		
Greece	BPI	SPI	***				**	***		

It can be readily observed that the NCoVaR estimates are somehow more conservative than the CoVaR estimates. This finding holds for both quantile specifications as well as across the variables and countries. In fact, this evidence is in line with our numerical conclusion that the linear CoVaR framework over-rejects under the null. Looking at the results, the size of overrejection is quite substantial. They strongly suggest CoVaR Grange causality spillovers from sovereign onto banks in all sample countries, with moderate support from NCoVaR Granger causality only in the case of Spain and Italy. Consequently, we consider the CoVaR Granger causality results to be inconclusive.

Looking at the instantaneous NCoVaR results, we find evidence for feedback loops in Spain and risk spillovers from sovereigns onto banks in Italy. NCoVaR Granger causality results appear to confirm the directional dependence from sovereigns onto banks in both countries, however, the spillovers from banks onto sovereigns in Spain disappear.

The parametric CoVaR results largely support the findings of simultaneous NCoVaR, suggesting also further bi-directional effects in Portugal and bank-to-sovereign spillovers in

Greece. Overall, with the exception of Ireland, the exercise confirms the differences in bank-sovereign feedback loops between vulnerable and core EA countries (Ohnsorge et al., 2014).

4.1. Extra controls

As a robustness check we tests whether the nonlinear dependencies between banks and sovereigns found above can be explained by potential common-factor effects. As pointed out in Section 2.6, Theorems 1 and 2 require higher-order kernel smoothing and slower convergence rate of the bandwidth. In our example, we take the 4th-order Gaussian kernel, allowing to include two extra conditioning variables, which makes the bandwidth rates equal to $n^{-1/6}$ for NCoVaR and $n^{-1/7}$ for NCoVaR Granger causality. We take the bandwidth constants consistently with the higher-order kernels and under the no dependence against extra co-variates. As common-factor benchmarks, we take the daily changes of the USD/EUR exchange rate and the STOXX Europe 600 equity index. Both time series cover the entire time span of the main variables of interest so that the number of observations in each of the countries fully corresponds to Table 4. The conditioning densities are evaluated around the median of the conditioning variables, so that the results are indicative of the bank-sovereign dependencies in the absence of substantial shocks in the extra control variables, which in our example represent currency and equity markets.

The results are given in Table 6. The framework assumes $\gamma = 0.95$ but again the more conservative quantiles confirm the main findings. It can be observed that both linear CoVaR and CoVaR Granger causality vastly resembles the structure observed in the basic specification in Table 5. However, the statistical significance of dependence between variables weakens in the nonparametric results after controlling for the confounding variables.

Table 6: Bank-sovereign feedback loops in selected euro area countries correcting for common-factor effects. BPI and SPI denote the Banking Price Index and Sovereign Price Index, respectively. Columns CoVaR and NCoVaR denote the instantaneous specifications, whereas columns CoVaR Gc and NCoVaR Gc correspond to Granger causality setups. ***, **, * denote 1%, 5% and 10% significance levels. For nonparametric NCoVaR and NCoVaR Granger causality tests we set $\mu = 0.8$. Risky quantiles are estimated at $\gamma = 0.95$.

			CoVaR		NCoVaR		CoVaR Gc		NCoVaR Gc	
	X	Y	$X \to Y$	$Y \to X$	$X \to Y$	$Y \to X$	$X \to Y$	$Y \to X$	$X \to Y$	$Y \to X$
Germany	BPI	SPI						**		
France	BPI	SPI						**		
Spain	BPI	SPI	***	***			**	***	*	
Italy	BPI	SPI	***	***	*	*	**			
Portugal	BPI	SPI	***	***			***	***		
Ireland	BPI	SPI					***	***		
Greece	BPI	SPI	***				**	***		

The cross-sectional NCoVaR measure does suggest some evidence for a bank-sovereign feedback loop in Italy. Yet, it seems that the dynamics behind the Spanish feedback loop discovered in Table 5 is fully captured by the information present in the extra variables. Similarly, NCoVaR Granger causality detects weak evidence for a bank-onto-sovereign risk spillovers in Spain when conditioning for confounding variables.

The results indicate a clear difference how the parametric and nonparametric setups incorporate extra information from the confounding variables. It seems that the former remains intact whereas the latter is more agile. It may be that the extra information is present at higher moments of distribution of the confounding variables, which may be difficult to be discovered by linear frameworks. The exact nature of this phenomenon is, however, beyond the scope of this paper.

5. Conclusions and discussion

NCoVaR and NCoVaR Granger causality build a new methodological framework to assess co-risk relations, designed to capture the possible nonlinear effects. We derive the regular asymptotic properties of the NCoVaR tests and we confirm them numerically. Importantly, because of the use of nonparametric methods, the framework is able to capture risk dependencies even in highly nonlinear environments, mimicking for instance volatility spillovers, which the standard CoVaR methodology is unable to capture. Moreover, we demonstrate that the CoVaR Granger causality measure is vulnerable to a false positive error, which in the risk-management context may lead to overpriced hedge instruments.

We apply our methodology to assess the bank-sovereign co-risk relations in the Euro Area (EA). Our findings suggest substantial differences between core and vulnerable EA countries, as often highlighted in the literature (Ohnsorge et al., 2014; Angelini et al., 2014). The findings are preserved when conditioning for common-factor effects, which include currency and equity markets' dynamics. In particular, our findings suggest substantial instantaneous and lagged co-movement between bank and sovereign asset returns in Spain, Italy, Portugal and Greece, with negligible effects in Germany, France and Ireland.

The NCoVaR framework can be of great use for macroprudential policy makers. Our extensive numerical and empirical studies suggest that NCoVaR tests provide more conservative estimates, compared to their parametric equivalents. In other words, standard CoVaR estimates may overprice the co-risk relevance between given entities or asset classes, exacerbating their risk premia, and possibly leading to inefficient allocation of macroprudential attention.

The novel methodology reveals some intriguing phenomena on the nonlinear nature of the co-risk relations. A tempting idea is to investigate the underlying structures analytically in models of the aggregate economy. Such settings would allow to capture not only the risk contribution of relevant sectors but also measure the dynamics of aggregate disturbances. One may also apply NCoVaR as a mapping tool and bring the risk analysis to the network

level.

References

Acharya, V., I. Drechsler, and P. Schnabl (2014). A Pyrrhic victory? Bank bailouts and sovereign credit risk. *The Journal of Finance 69*(6), 2689–2739.

Adrian, T. and M. K. Brunnermeier (2016, July). CoVaR. *American Economic Review 106*(7), 1705–41.

Angelini, P., G. Grande, and F. Panetta (2014, January). The negative feedback loop between banks and sovereigns. Questioni di Economia e Finanza (Occasional Papers) 213, Bank of Italy, Economic Research and International Relations Area.

Brunnermeier, M. (2009). Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic Perspectives 23*(1), 77–100.

Caporin, M., L. Pelizzon, F. Ravazzolo, and R. Rigobon (2012). Measuring sovereign contagion in Europe. Working Paper 05, Norges Bank.

Chinazzi, M. and G. Fagiolo (2013). Systemic risk, contagion, and financial networks: A survey. LEM Working Paper 2013/08, Sant'Anna School of Advanced Studies, Pisa.

Denker, M. and G. Keller (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 64*, 505–522.

Denker, M. and G. Keller (1986). Rigorous statistical procedures for data from dynamical systems. *Journal of Statistical Physics 44*, 67–93.

Diks, C. and V. Panchenko (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control 30*, 1647–1669.

Diks, C. and M. Wolski (2016). Nonlinear Granger causality: Guidelines for multivariate analysis. *Journal of Applied Econometrics 31*(7), 1333–1351.

EIB (2016). *Investment and investment finance in Europe*. Luxembourg: European Investment Bank.

Gao, B. and R.-e. Ren (2013). The topology of a causal network for the Chinese financial system. *Physica A: Statistical Mechanics and its Applications 392*(13), 2965–2976.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*(3), 424–438.

Hall, P. and M. C. Minnotte (2002). High order data sharpening for density estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 64*(1), 141–157.

Hansen, B. (2009). Lecture notes on nonparametrics. Working paper, University of Wisconsin – Madison.

He, Z. and A. Krishnamurthy (2012). A macroeconomic framework for quantifying systemic risk. Working Paper 233, National Bank of Belgium.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics 19*, 293–325.

Hong, Y., Y. Liu, and S. Wang (2009). Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics 150*(2), 271–287.

Huang, X., H. Zhou, and H. Zhu (2010). Systemic risk contributions. Working Paper 60, Bank for International Settlements.

Jeong, K., W. K. Härdle, and S. Song (2012). Nonparametric test for causality in quantile. *Econometric Theory 28*, 861–887.

Kitamura, T., I. Muto, and I. Takei (2016). Loan interest rate pass-through and changes after the financial crisis: Japan's evidence. *Journal of the Japanese and International Economies 42*, 10 – 30.

Kupiec, P. H. (2002). Stress testing in a value at risk framework. In M. A. H. Dempster (Ed.), *Risk Management: Value at Risk and Beyond*, pp. 76–99. Cambridge: Cambridge University Press.

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton and Oxford: Princeton University Press.

Li, Q. and J. S. Racine (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics 26*(4), 423–434.

Li, Q. and J. S. Racine (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *Journal of Business & Economic Statistics 31*(1), 57–65.

Mainik, G. and E. Schaanning (2014). On dependence consistency of CoVaR and some other systemic risk measures. *Statistics & Risk Modeling 31*(1), 49–77.

Ohnsorge, F. L., M. Wolski, and S. Y. Zhang (2014). Safe havens, feedback loops, and shock propagation in global asset prices. IMF Working Paper 14/81, International Monetary Fund.

Paries, M., D. Moccero, E. Krylova, and C. Marchini (2014). The retail bank interest rate pass-through: The case of the euro area during the financial and sovereign debt crisis. *Occasional Paper Series of the European Central Bank* (no 154).

Powell, J. L. and T. M. Stoker (1996). Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics 75*, 219–316.

Rothe, C. (2010). Nonparametric estimation of distributional policy effects. *Journal of Econometrics 155*(1), 56 – 70.

Scott, D. W. (1992). *Multivariate Density Estimation (Wiley Series in Probability and Mathematical Statistics)*. New York: Wiley-Interscience.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

von Borstel, J., S. Eickmeier, and L. Krippner (2016). The interest rate pass-through in the euro area during the sovereign debt crisis. *Journal of International Money and Finance 68*, 386 – 402.

Appendix A. NCoVaR asymptotics

First we derive asymptotic normality under the assumption that the W_t, $t = 1, \ldots, n$, are i.i.d. and at the end extend to the case with dependence. For the instantaneous NCoVaR case, our test is based on

$$\begin{aligned} q &= P(X \in C)P(X \in D)\left(P(Y \in A|X \in C) - P(Y \in A|X \in D)\right) \\ &= P(Y \in A, X \in C)P(X \in D) - P(Y \in A, X \in D)P(X \in C), \end{aligned}$$

where

$$\begin{aligned} A &= [y_\gamma - \mu, y_\gamma + \mu], \\ C &= [x_\gamma - \mu, x_\gamma - \mu], \\ D &= [x_{0.5} - \mu, x_{0.5} + \mu]. \end{aligned}$$

Given γ and μ, q can be estimated unbiasedly using a U-statistic of degree $r = 2$, with U-statistic kernel

$$K(W_k, W_\ell) = I_A(Y_k)I_C(X_k)I_D(X_\ell) - I_A(Y_k)I_D(X_k)I_C(X_\ell) + k \leftrightarrow \ell. \tag{A.1}$$

The resulting estimator is unbiased and asymptotically normal by the theory of U-statistics developed by Hoeffding (1948). Following Serfling (1980) and van der Vaart (1998), the variance of a U-statistic U_n is, up to leading orders, given by

$$\mathrm{Var}(U_n) \simeq \sum_{s=1}^{r} \frac{r!^2}{k!(r-k)!} \frac{1}{n^s \zeta_s}, \tag{A.2}$$

where in our case ($r = 2$)

$$\zeta_1 = \mathrm{Cov}(\mathcal{K}(W_1, W_2), \mathcal{K}(W_1, W_2') = \mathrm{Var}(\mathcal{K}_1(W_i))$$

and

$$\zeta_2 = \mathrm{Cov}(\mathcal{K}(W_1, W_2), \mathcal{K}(W_1, W_2)) = \mathrm{Var}(\mathcal{K}(W_1, W_2)).$$

Discarding higher order terms in $1/n$ gives

$$\begin{aligned}\mathrm{Var}(q_{*,n}) &= \frac{4}{n}\zeta_1 + \text{h.o.t} \\ &= \frac{4}{n}\mathrm{Var}(\mathcal{K}_1(W_1)) + \text{h.o.t.}\end{aligned}$$

where 'h.o.t.' stands for higher order terms.

One finds

$$\sqrt{n}\frac{q_n - q}{S_n} \xrightarrow{d} N(0,1) \tag{A.3}$$

where S_n^2 is a consistent estimator of the asymptotic variance $4\mathrm{Var}\left[\mathcal{K}_1(W_t)\right]$. In the time series setting, under the assumption that the processes are stationary and weakly dependent, the long-run variance of $\sqrt{n}(q_n - q)$ is given by

$$\sigma^2 = 4\left(\mathrm{Var}\left(\mathcal{K}_1(W_t)\right) + 2\sum_{\ell=1}^{\infty}\mathrm{Cov}\left[\mathcal{K}_1(W_t), \mathcal{K}_1(W_{t+\ell})\right]\right).$$

The term in brackets can then be estimated consistently using a HAC estimator for the long-run variance of $2\mathcal{K}_1(W_t)$ (Denker and Keller 1983; 1986).

Appendix B. Logical equivalences and implications of Granger non-causality

The null hypothesis is Granger non-causality from $\{X_t\}$ to $\{Y_t\}$, that is H_0: $\{X_t\} \not\to \{Y_t\}$, for one lag. Let $(X, Y, Z) \sim (X_t, Y_t, Y_{t+1})$, then H_0 states that X and Z are conditionally independent given Y. In terms of densities and probabilities, the following logical

equivalences and implications hold

$$
\begin{array}{rcl}
H_0: \quad f_{X,Z|Y=y}(x,z|y) & = & f_{X|Y=y}(x|y) f_{Z|Y=y}(z|y), \quad \forall x,y,z \\
& \Updownarrow & \\
f_{Z|X=x,Y=y}(z|x,y) & = & f_{Z|Y=y}(z|y) \quad \forall x,y,z \\
& \Updownarrow & \\
f_{z|X=x_1,Y=y}(z|x_1,y) & = & f_{z|X=x_2,Y=y}(z|x_2,y) \quad \forall x_1,x_2,y,z \\
& \Downarrow & \\
P(Z \in A|X=x_1,Y=y) & = & P(z \in A|X=x_2,Y=y) \quad \forall A,x_1,x_2,y \\
& \Updownarrow & \\
P(Z \in A|X \in C,Y=y) & = & P(z \in A|X \in D,Y=y) \quad \forall A,C,D,y \\
& \Updownarrow & \\
\frac{P(Z \in A, X \in C|Y=y)}{P(X \in C|Y=y)} & = & \frac{P(Z \in A, X \in D|Y=y)}{P(X \in D|Y=y)} \quad \forall A,C,D,y \\
& \Downarrow & \\
H_0': \quad \frac{P(Z \in A, X \in C|Y=y_*)}{P(X \in C|Y=y_*)} & = & \frac{P(Z \in A, X \in D|Y=y_*)}{P(X \in D|Y=y_*)} \quad \text{for some specific } A,\ C,\ D \text{ and } y=y_*.
\end{array}
\tag{B.1}
$$

In the NCoVaR Granger causality context we focus on testing the implication H_0' for a given unconditional Y-quantile y_*, and events where X and Z fall in the sets

$$
\begin{array}{rcll}
A & = & [z_\gamma - \mu, z_\gamma + \mu] & (Z \text{ near tail quantile}) \\
C & = & [x_\gamma - \mu, x_\gamma - \mu] & (X \text{ near tail quantile}) \\
D & = & [x_{0.5} - \mu, x_{0.5} + \mu] & (X \text{ near median}).
\end{array}
$$

Appendix C. NCoVaR Granger causality asymptotics

Recall that for a given unconditional quantile y_* of Y, the quantity to be estimated is

$$
\begin{array}{rcl}
q_* & = & f_Y^2(y_*) \left(P(Z \in A, X \in C|Y=y_*) P(X \in D|Y=y_*) \right. \\
& & \left. - P(Z \in A, X \in D|Y=y_*) P(X \in C|Y=y_*) \right),
\end{array}
$$

where, as motivated in Appendix B,

$$\begin{aligned} A &= [z_\gamma - \mu, z_\gamma + \mu] \\ C &= [x_\gamma - \mu, x_\gamma + \mu] \\ D &= [x_{0.5} - \mu, x_{0.5} + \mu]. \end{aligned}$$

By construction, $q_* = 0$ under H_0, and terms like $f_Y(y_*)P(X \in D|Y = y_*)$ can be simply estimated as $\frac{1}{n}\sum_{k=1}^{n} I_D(Z_k)K_h(y_* - Y_k)$.

A U-statistic type of estimator (sample-weighted average) of q is given by

$$q_{*,n} = \frac{1}{n(n-1)} \sum_{\substack{k \\ k \neq \ell}} \sum_{\ell} K(W_k, W_\ell; h),$$

where $K(\cdot, \cdot; \cdot)$ is the symmetric (with respect to interchanging k and ℓ) kernel function

$$\begin{aligned} \mathcal{K}(W_k, W_\ell; h) &= \frac{1}{2}[I_A(Z_k)I_C(X_k)K_h(y_* - Y_k)I_D(X_\ell)K_h(y_* - Y_\ell) \\ &\quad - I_A(Z_k)I_D(X_k)K_h(y_* - Y_k)I_C(X_\ell)K_h(y_* - Y_\ell) \\ &\quad + k \leftrightarrow \ell]. \end{aligned}$$

Note that although $\lim_{h\to 0} E(K(W_i, W_j; h)) = q_*$, $q_{*,n}$ is biased as estimator of q_* for positive bandwidth values and hence finite sample size n if the bandwidth tends to zero only in the limit where n tends to infinity. Since $q_{*,n}$ is not unbiased in finite samples, it formally is not a U-statistic estimator of q_*. However, for a given value of the bandwidth h it nevertheless has the same analytic form as a U-statistic, which means that it can still be interpreted as a U-statistic estimator, not of q_* but of $q(h) \equiv E(K(W_i, W_j; h))$. This property allows us to prove asymptotic normality of $q_{*,n}$ using the asymptotic theory for U-statistics, provided that we make sure that the bias of $q_{*,n}$ as an estimator of q_* is asymptotically negligible.

Below we will consider the leading terms in the mean squared error (MSE), which can

be decomposed as

$$\mathrm{MSE}(q_{*,n}) = \mathrm{bias}^2(q_{*,n}) + \mathrm{Var}(q_{*,n}),$$

where the first term is the squared finite sample bias of $q_{*,n}$ as an estimator of q (rather than $q(h)$), and the second is the finite sample variance of $q_{*,n}$, which is obtained from the theory of U-statistics. Intuitively, provided that the bias vanishes sufficiently fast with n, only the variance part is relevant asymptotically, and we are in the realm of U-statistics with bandwidth-dependent sample sizes.

In order to proceed we introduce some notation. Let $K_1(w_1; h) = E(K(w_1, W_2; h))$, $r(w_1; h) = E(K_1(w_1; h))$ and $r_0(w_1) = \lim_{h\to 0} K_1(w_1; h)$. Note that $E(r_0(W_1)) = q_*$ by construction. We follow Powell and Stoker (1996) in assuming that the underlying densities are sufficiently regular, ensuring that the point-wise (local) bias of $q_{*,n}$ (as estimator of q_*) and the variance of the U-statistics kernel (the expectation of which is $q_*(h)$) behave as

$$\mathcal{K}_1(w_i; h) - r_0(w_i) = s(w_i)h^\alpha + s^*(w_i; h),$$

for some $\alpha > 0$ and

$$\mathrm{Var}(\mathcal{K}(W_1, W_2; h)) = E((\mathcal{K}(W_1, W_2; h) - q_*(h))^2) = bh^{-\gamma} + c^*(h),$$

for some $\gamma > 0$, respectively, where the remainder terms make asymptotically negligible contributions to $\mathrm{MSE}(q_n)$, i.e. $E\left((s^*(W_i; h))^2\right) = o(h^{2\alpha})$ and $c^*(h) = o(h^{-\gamma})$. It is well known that the order (α here) of the local bias of a kernel density estimate equals the order the density estimation kernel function K. That is, $\alpha = 2$ for commonly-used second-order kernels, such as the Gaussian kernel, and $\alpha = 4, 6, \ldots$ for higher-order kernels. Straightforward calculations (not provided here due to space considerations, but available from the authors upon request) show that $\gamma = 2d_Y$.

Using Eq. (A.2) for the variance of a U-statistic again, but now with the bandwidth-

dependent U-statistic kernel, we have

$$\zeta_1 = \mathrm{Cov}(\mathcal{K}(W_1, W_2; h), \mathcal{K}(W_1, W_2'; h) = \mathrm{Var}(\mathcal{K}_1(W_i; h))$$

and

$$\zeta_2 = \mathrm{Cov}(\mathcal{K}(W_1, W_2; h), \mathcal{K}(W_1, W_2; h)) = \mathrm{Var}(\mathcal{K}(W_1, W_2; h)).$$

Note that

$$\begin{aligned}
\mathrm{Var}(\mathcal{K}_1(W_1; h)) &= \mathrm{Var}(r_0(W_1) + s(W_1)h^\alpha + \text{h.o.t.}) \\
&= \mathrm{Var}(r_0(W_1)) + 2\mathrm{Cov}(r_0(W_1), s(W_1))h^\alpha + \text{h.o.t.} \\
&\equiv \mathrm{Var}(r_0(W_1)) + 2C_0 h^\alpha + \text{h.o.t.} \\
&= \mathrm{Var}(r_0(W_1)) + \text{h.o.t.},
\end{aligned}$$

where 'h.o.t.' stands for higer order terms.

Using this we find

$$\begin{aligned}
\mathrm{MSE}(q_{*,n}) &= (\mathrm{bias}(q_{*,n}))^2 + \mathrm{Var}(q_{*,n}) \\
&\simeq (E(s(W_1)))^2 h^{2\alpha} + \frac{4}{n}\zeta_1 + \frac{2}{n^2}\zeta_2, \\
&\simeq (E(S(W_i)))^2 h^{2\alpha} + \frac{4}{n}\mathrm{Var}(r_0(W_i)) + \frac{2}{n^2} b h^{-\gamma},
\end{aligned}$$

up to leading order, in the sense that one or more of the terms on the right hand side will dominate asymptotically as n tends to infinity with the bandwidth $h = h_n$ decreasing to zero exponentially according to $h_n = cn^{-\beta}$.

The first and third term on the right-hand-side correspond to the leading (squared) bias and variance terms, which dominate when the bandwidth h_n tends to zero too slow or too fast with n, respectively. The second term does not depend on h_n, and in fact corresponds to the leading variance $\frac{4}{n}\zeta_1$ of $\frac{2}{n}\sum_t r_0(W_t)$, which is a regular U-statistic of degree 1 with kernel $2r_0(W)$ (in fact, $r_0(W)$ is the Hájek projection $\lim_{h\to 0}\mathcal{K}_1(W_1, h)$ of the kernel on the

data in the limit $h \to 0$). It turns out that $q_{*,n}$ is asymptotically normal with asymptotic mean q_* and asymptotic variance $\frac{4}{n}\text{Var}(r_0(W_i))$ if and only if the bandwidth-independent term $\frac{4}{n}\text{Var}(r_0(W_i))$ dominates the MSE expansion; see Powell and Stoker (1996) for details.

It follows that asymptotic normality is obtained if both

$$h^{2\alpha} \propto n^{-2\alpha\beta} \ll n^{-1} \qquad \text{and} \qquad \frac{1}{n^2}h^{-\gamma} \propto n^{\beta\gamma-2} \ll n^{-1} \qquad \text{for } n \text{ large,}$$

which holds as long as $\beta > \frac{1}{2\alpha}$ and $\beta < \frac{1}{\gamma} = \frac{1}{2d_Y}$. This means that for commonly-used second order ($\alpha = 2$) kernels there are only feasible rates (β-values) for $d_Y = 1$, being $\beta \in (\frac{1}{4}, \frac{1}{2})$. For $d_Y = 2$ and larger, larger α-values are required. Either data sharpening or higher-order kernels can be used to reduce the order of the bias term, opening up some range of β-values for which the bandwidth tends to zero sufficiently slow, so that the variance doesn't explode.

As a consequence, using a density estimation kernel of appropriate order $\alpha > d_Y$, and a constant rate $\beta \in \left(\frac{1}{2\alpha}, \frac{1}{2d_Y}\right)$ for the bandwidth sequence $h_n = cn^{-\beta}$, one obtains

$$\sqrt{n}\frac{q_{*,n} - q_*}{S'_n} \xrightarrow{d} N(0,1), \tag{C.1}$$

where S'^2_n is a consistent estimator of the asymptotic variance $4\text{Var}\left[\lim_{h\to 0} K_1(W_t, h)\right] = 4\text{Var}\left[r_0(W_t)\right]$. In the time series setting, under the assumption that the processes are stationary and weakly dependent, the long-run variance of $\sqrt{n}(q_{*,n} - q_*)$ is given by

$$\sigma'^2 = 4\left(\text{Var}\left(r_0(W_t)\right) + 2\sum_{\ell=1}^{\infty}\text{Cov}\left[r_0(W_t), r_0(W_{t+\ell})\right]\right),$$

where $r_0(W_t) = \lim_{h\to 0}(K_1(W_t, h))$. The term in brackets can then be estimated consistently using a HAC estimator for the long-run variance of $r_0(W_t)$ (Denker and Keller 1983; 1986).

Appendix D. MSE optimal bandwidth for NCoVaR Granger causality

We first focus on the basic case $d_Y = 1$, where density kernel order $\alpha = 2$ suffices, and discuss extensions to higher-variate cases later. Within the feasible window of β-values ($\beta \in (\frac{1}{4}, \frac{1}{2})$) for $d_Y = 1$ the MSE optimal rate can be found by minimising the right hand side of the expression for the MSE with respect to h. The first order condition is

$$2\alpha(E(s(W_i)))^2 h^{2\alpha-1} - \frac{2}{n^2} b\gamma h^{-\gamma-1} = 0,$$

which gives the asymptotically optimal bandwidth

$$h_n^* = \left(\frac{b\gamma}{\alpha(E(s(W_i)))^2}\right)^{1/(\gamma+2\alpha)} n^{-2/(\gamma+2\alpha)} = \left(\frac{b\gamma}{\alpha(E(s(W_i)))^2}\right)^{1/(2d_Y+2\alpha)} n^{-1/(d_Y+\alpha)}.$$

The optimal rate is $\beta_{\text{opt}} = 1/(d_Y + \alpha) = \frac{1}{3}$ if $d_Y = 1$ and $\alpha = 2$.

We wish to calculate the MSE of $q_{*,n}$ as an estimator of

$$q_* = \lim_{h\to 0} E(K(W_k, W_\ell; h)),$$

which is 0 under H_0, for W_k and W_ℓ two independent random variables, distributed as $W_t = (X_t, Y_t, Y_{t+1})$. We first derive some expressions for the expectation of the squared kernel, which can be decomposed as

$$E\left[(K(W_k, W_\ell; h))^2\right] = E\left(\text{diag. terms} + \text{off-diag. terms}\right),$$

where the diagonal terms are the squares of the four terms in the U-statistics kernel (5), which, using the independence and equal distributions of W_k and W_ℓ, can be simplified to

$$\begin{aligned}\text{diag. terms} \;=\; & (I_A(Z_k) I_C(X_k) K_h^2(y_* - Y_k) I_D(X_\ell) K_h^2(y_* - Y_\ell)) \\ & + I_A(Z_k) I_D(X_k) K_h^2(y_* - Y_k) I_C(X_\ell) K_h^2(y_* - Y_\ell))\,/2,\end{aligned}$$

while the 12 off-diagonal (cross-product) terms can be represented as

$$\begin{aligned}\text{off-diag. terms} &= \tfrac{1}{2}K_h^2(y_*-Y_k)K_h^2(y_*-Y_\ell)\,[I_A(Z_k)I_C(X_k)I_D(X_\ell)I_A(Z_k)I_D(X_k)I_C(X_\ell)\\ &\quad +I_A(Z_k)I_C(X_k)I_D(X_\ell)I_A(Z_\ell)I_C(X_\ell)I_D(X_k)\\ &\quad -I_A(Z_k)I_C(X_k)I_D(X_\ell)I_A(Z_\ell)I_D(X_\ell)I_C(X_k)\\ &\quad -I_A(Z_k)I_D(X_k)I_C(X_\ell)I_A(Z_\ell)I_C(X_\ell)I_D(X_k)\\ &\quad +I_A(Z_k)I_D(X_k)I_C(X_\ell)I_A(Z_\ell)I_D(X_\ell)I_C(X_k)\\ &\quad -I_A(Z_\ell)I_C(X_\ell)I_D(X_k)I_A(Z_\ell)I_D(X_\ell)I_C(X_k)]\,.\end{aligned}$$

Assuming that regions C and D do not overlap (which can always be achieved by taking μ sufficiently small), only the third and fourth of the six terms within brackets can be nonzero, since then $I_C(X_\ell)I_D(X_\ell)\equiv 0$[5]. Taken together with the diagonal terms

$$\begin{aligned}E\left[(K(W_k,W_\ell;h))^2\right] &= \tfrac{1}{2}E\left(K_h^2(y_*-Y_k)K_h^2(y_*-Y_\ell)I_A(Z_k)I_C(X_k)I_D(X_\ell)(1-I_A(Z_\ell))\right)\\ &\quad +\tfrac{1}{2}E\left(K_h^2(y_*-Y_k)K_h^2(y_*-Y_\ell)I_A(Z_k)I_D(X_k)I_C(X_\ell)(1-I_A(Z_\ell))\right)\\ &= \tfrac{1}{2}E\left(K_h^2(y_*-Y_k)I_A(Z_k)I_C(X_k)\right)E\left(K_h^2(y_*-Y_\ell)I_D(X_\ell)(1-I_A(Z_\ell))\right)\\ &\quad +\tfrac{1}{2}E\left(K_h^2(y_*-Y_k)I_A(Z_k)I_D(Z_k)\right)E\left(K_h^2(y_*-Y_\ell)I_C(X_\ell)(1-I_A(Z_\ell))\right).\end{aligned}$$

VAR(1) optimal bandwidth. Consider the the bivariate process defined by

$$\begin{aligned}X_t &= aY_t+\sqrt{1-a^2}\varepsilon_{X,t}\\ Y_{t+1} &= aY_t+\sqrt{1-a^2}\varepsilon_{Y,t}\end{aligned}$$

where $\varepsilon_{X,t}$ and $\varepsilon_{Y,t}$ are i.i.d. standard normal. Note that both X_t and Y_t are $N(0,1)$-distributed as well. Let $W_t=(X_t,Y_t,Z_t)=(X_t,Y_t,Y_{t+1})$, then it is easy to verify that

[5]In case C and D do overlap, the test statistic is still asymptotically normal, but the derivation of the asymptotic variance would become much more involved.

the density of W is given by

$$f_W(x,y,z;a) = \frac{1}{(2\pi)^{3/2}(1-a^2)} \mathrm{e}^{-y^2/2-(x-ay)^2/(2(1-a^2))-(z-ay)^2/(2(1-a^2))}. \tag{D.1}$$

The MSE optimal bandwidth for $a = 0.4$, $\gamma = 0.95$, $\mu = 0.8$, for example, turns out to be (obtained through extensive analytical manipulation with Mathematica)

$$h_{\mathrm{opt}}^{\mathrm{AR(1)}} = 3.07 \times n^{-1/3}.$$

ARCH(1) optimal bandwidth. To obtain an idea how the optimal bandwidth h behaves in a setting with conditional heteroskedasticity, we also considered the following distribution of (X, Y, Z).

$$Y \sim N(0,1), \quad X|Y \sim N(0, 1+aY^2), \quad \text{and} \quad Z|(X,Y) \sim N(0, 1+aY^2). \tag{D.2}$$

The MSE optimal bandwidth for $a = 0.4$, $\gamma = 0.95$, $\mu = 0.8$, for example, turns out to be (obtained using Mathematica)

$$h_{\mathrm{opt}}^{\mathrm{ARCH(1)}} = 2.14 \times n^{-1/5}.$$

Interestingly, the MSE optimal bandwidth for the ARCH process tends to zero at a slower rate than $1/(d_Y + \alpha) = 1/3$ here, since some terms that are leading in the bandwidth cancel for the ARCH process while they don't for the VAR process.

In the higher-variate case, only the fact that we use a higher dimension and/or a higher-order kernel affects the optimal bandwidth parameter c under the assumption that the additional variable is independent of the others. Adding a possible confounding variable that is independent of the other variables has no effect on the constant c in the expression for the optimal bandwidth, since their densities integrate out in the calculation of the relevant

expectations. The optimal sample size-dependent constants for different dimensions d_Y, choices of μ and γ are given in Table D1 for the VAR and ARCH processes with dependence parameter $a = 0.4$.

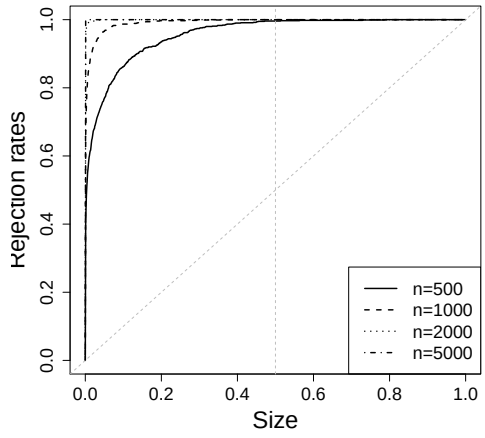
Table D1: Optimal constant c for sample size-dependent bandwidth ($h_n = cn^{-\beta}$) in NCoVaR Granger causality test for VAR and ARCH processes, as given in Eqs. (D.1) and (D.2) respectively, with $a = 0.4$. The values are given for different quantile levels (γ), fixed bandwidth parameter μ and for 2nd and 4th-order Gaussian kernels under the assumption of independent confounding variables. Rate of convergence should satisfy $1/(2\alpha) < \beta < 1/(2d_Y)$ as discussed in Section 2.6.

		$\gamma = 0.95$		$\gamma = 0.99$	
Estimation kernel	Fixed bandwidth (μ)	VAR	ARCH	VAR	ARCH
Gaussian of order $\alpha = 2$	0.1	7.95	3.42	8.74	3.23
	0.2	5.64	2.79	6.22	2.63
	0.3	4.63	2.50	5.13	2.35
	0.4	4.05	2.33	4.49	2.17
	0.5	3.67	2.23	4.07	2.06
	0.6	3.40	2.17	3.76	1.98
	0.7	3.21	2.14	3.53	1.92
	0.8	3.07	2.14	3.35	1.89
	$0.9^\dagger$	2.96	2.16	3.20	1.87
	$1.0^\dagger$	2.88	2.21	3.08	1.86
	$1.1^\dagger$	2.82	2.29	2.98	1.86
	$1.2^\dagger$	2.78	2.40	2.89	1.88
	$1.3^\dagger$	2.75	2.57	2.82	1.91
Gaussian of order $\alpha = 4$	0.1	3.30	3.50	5.64	3.30
	0.2	2.68	2.86	4.37	2.69
	0.3	2.37	2.56	3.66	2.40
	0.4	2.18	2.39	3.18	2.22
	0.5	2.05	2.28	2.84	2.11
	0.6	1.95	2.22	2.59	2.03
	0.7	1.87	2.19	2.39	1.97
	0.8	1.81	2.19	2.24	1.93
	$0.9^\dagger$	1.76	2.21	2.11	1.91
	$1.0^\dagger$	1.73	2.26	2.01	1.90
	$1.1^\dagger$	1.70	2.34	1.93	1.91
	$1.2^\dagger$	1.68	2.46	1.87	1.92
	$1.3^\dagger$	1.67	2.63	1.81	1.95

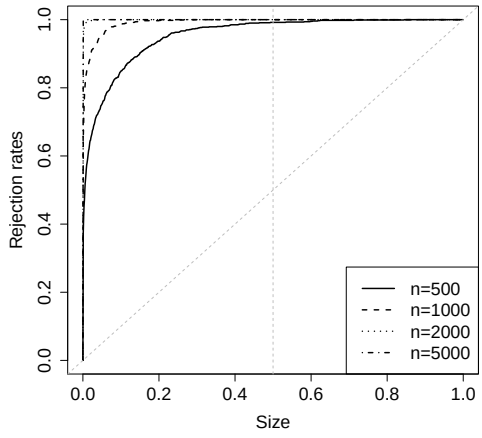
† for larger μ values the conditioning events C and D have some overlap, which is ignored in the calculations.

Appendix E. Additional results

CoVaR

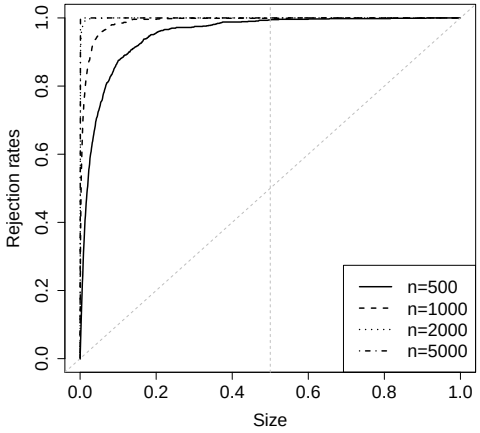


(a) Size-size plot

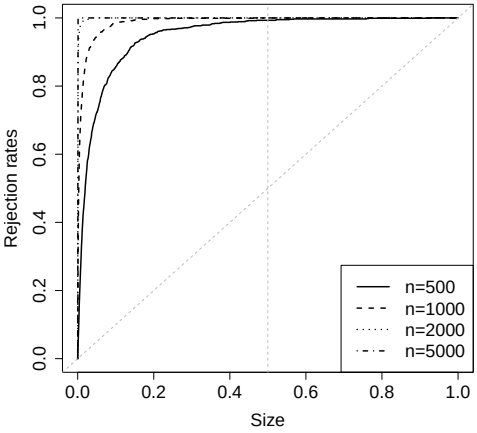


(b) Size-power plot

NCoVaR

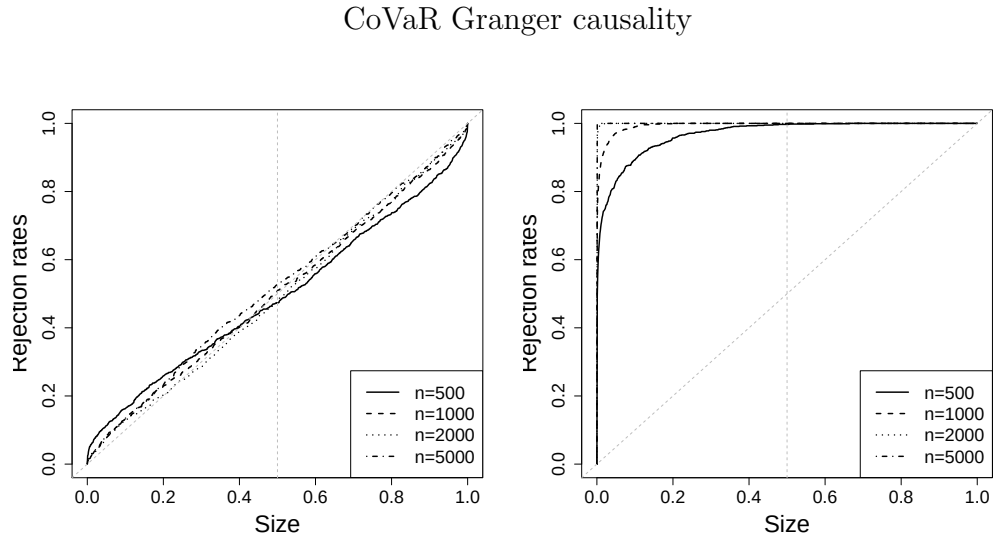


(c) Size-size plot

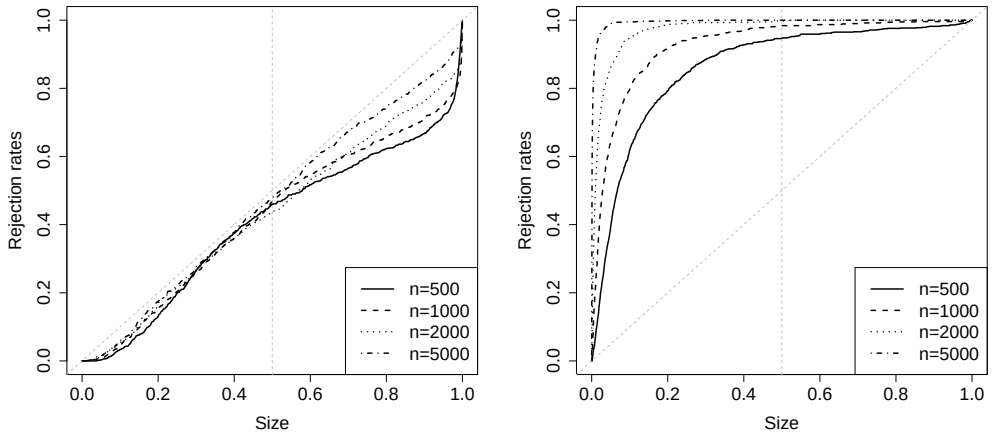


(d) Size-power plot

Figure E1: CoVaR and NCoVaR performance in VAR class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (8) with instantaneous dependence ($\tau = 0$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.



(a) Size-size plot (b) Size-power plot



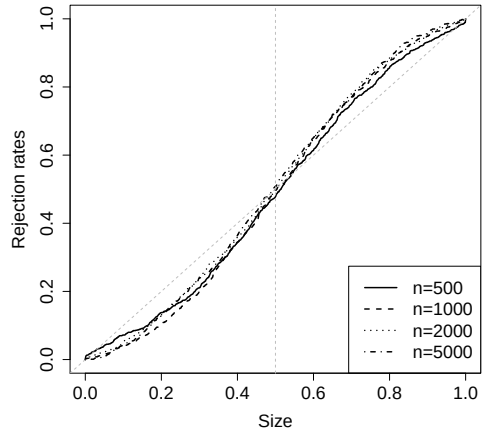
(c) Size-size plot (d) Size-power plot

Figure E2: CoVaR and NCoVaR Granger causality performance in VAR class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (8) with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

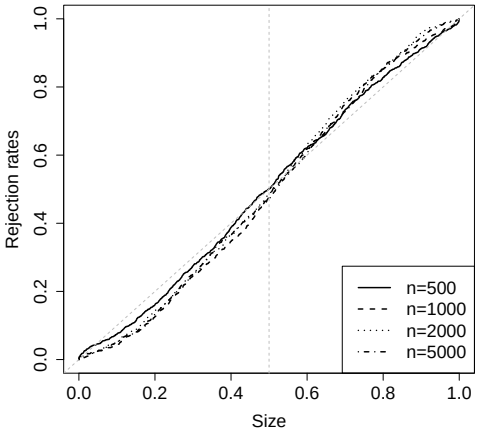
Table E2: Performance summary of CoVaR and NCoVaR methodologies in VAR class of models for selected nominal size levels (α). Actual rejection rates under $X \to Y$ causality (power) and $Y \to X$ causality (size) for different sample sizes (n), generated under process in Eq. (8) with instantaneous dependence ($\tau = 0$) and with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

		Instantaneous dependence				Granger causality			
α	n	size		power		size		power	
		CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR
0.01	500	0.596	0.334	0.574	0.331	0.057	0.000	0.695	0.070
	1000	0.865	0.753	0.849	0.744	0.027	0.000	0.879	0.216
	2000	0.993	0.986	0.993	0.983	0.028	0.000	0.995	0.510
	5000	1.000	1.000	1.000	1.000	0.023	0.001	1.000	0.899
0.05	500	0.768	0.732	0.752	0.723	0.113	0.003	0.830	0.395
	1000	0.966	0.953	0.954	0.946	0.078	0.008	0.971	0.643
	2000	1.000	0.999	1.000	1.000	0.076	0.012	1.000	0.862
	5000	1.000	1.000	1.000	1.000	0.084	0.016	1.000	0.990
0.1	500	0.862	0.874	0.849	0.856	0.163	0.034	0.893	0.619
	1000	0.987	0.983	0.987	0.985	0.131	0.047	0.990	0.798
	2000	1.000	1.000	1.000	1.000	0.130	0.053	1.000	0.949
	5000	1.000	1.000	1.000	1.000	0.135	0.057	1.000	0.995

CoVaR

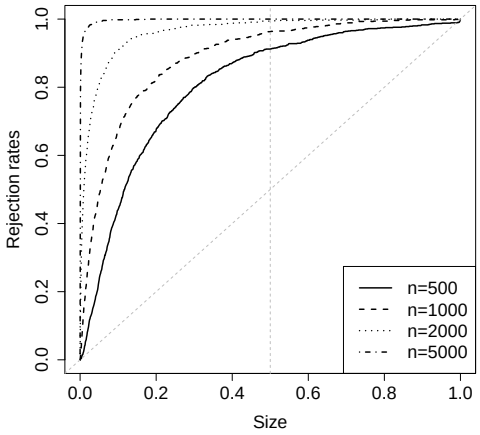


(a) Size-size plot

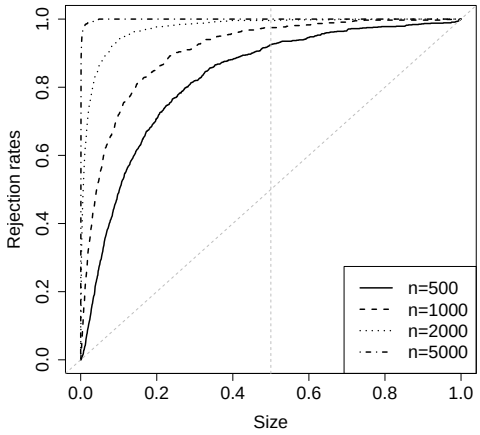


(b) Size-power plot

NCoVaR



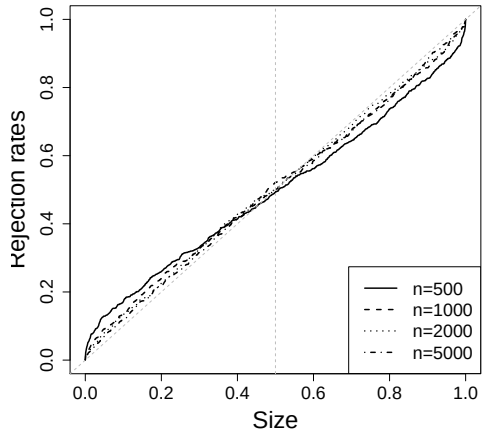
(c) Size-size plot



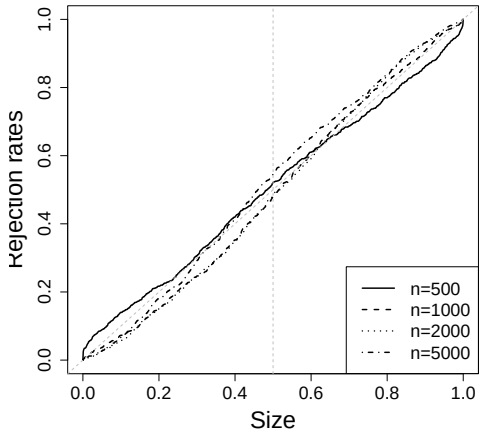
(d) Size-power plot

Figure E3: CoVaR and NCoVaR performance in (G)ARCH class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (9) with instantaneous dependence ($\tau = 0$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

CoVaR Granger causality

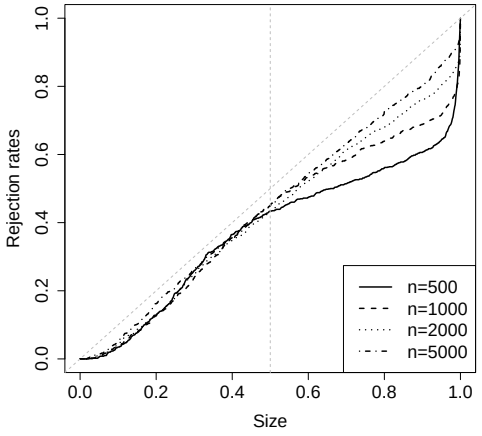


(a) Size-size plot

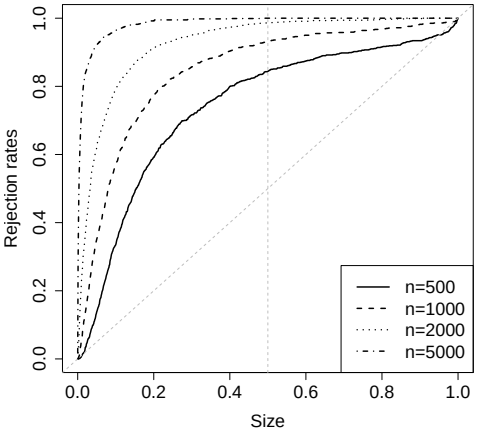


(b) Size-power plot

NCoVaR Granger causality



(c) Size-size plot



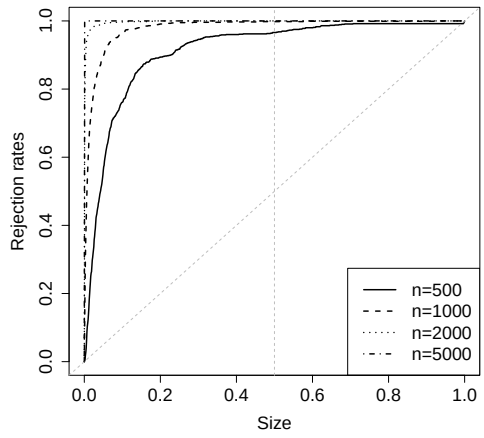
(d) Size-power plot

Figure E4: CoVaR and NCoVaR Granger causality performance in (G)ARCH class of models. Size-size and size-power plots for different sample sizes generated under process in Eq. (9) with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

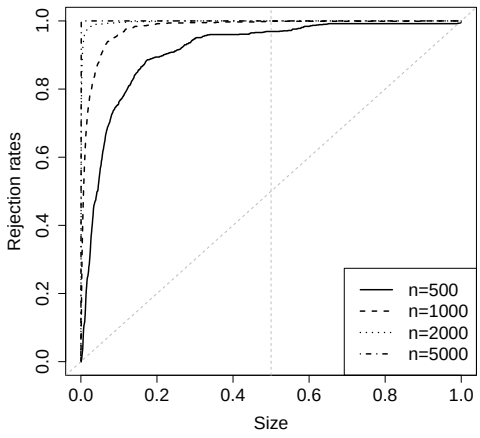
Table E3: Performance summary of CoVaR and NCoVaR methodologies in (G)ARCH class of models for selected nominal size levels (α). Actual rejection rates under $X \to Y$ causality (power) and $Y \to X$ causality (size) for different sample sizes (n), generated under process in Eq. (9) with instantaneous dependence ($\tau = 0$) and with lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

		Instantaneous dependence				Granger causality			
α	n	size		power		size		power	
		CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR	CoVaR	NCoVaR
0.01	500	0.016	0.028	0.018	0.035	0.057	0.000	0.042	0.008
	1000	0.004	0.155	0.012	0.174	0.032	0.000	0.012	0.054
	2000	0.007	0.502	0.008	0.587	0.027	0.000	0.005	0.230
	5000	0.001	0.957	0.005	0.976	0.023	0.000	0.008	0.725
0.05	500	0.041	0.234	0.046	0.279	0.129	0.006	0.089	0.138
	1000	0.016	0.484	0.025	0.537	0.091	0.008	0.037	0.351
	2000	0.023	0.806	0.027	0.867	0.080	0.010	0.027	0.621
	5000	0.010	0.992	0.025	1.000	0.069	0.014	0.026	0.922
0.1	500	0.072	0.441	0.074	0.490	0.172	0.035	0.140	0.335
	1000	0.040	0.669	0.043	0.720	0.138	0.038	0.073	0.569
	2000	0.054	0.908	0.050	0.940	0.133	0.041	0.057	0.790
	5000	0.042	0.998	0.051	1.000	0.120	0.054	0.067	0.965

NCoVaR (above quantile)

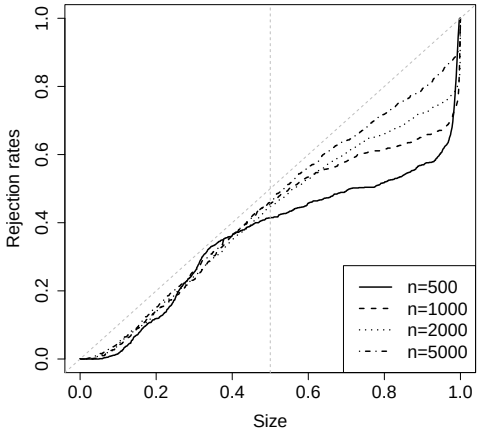


(a) Size-size plot

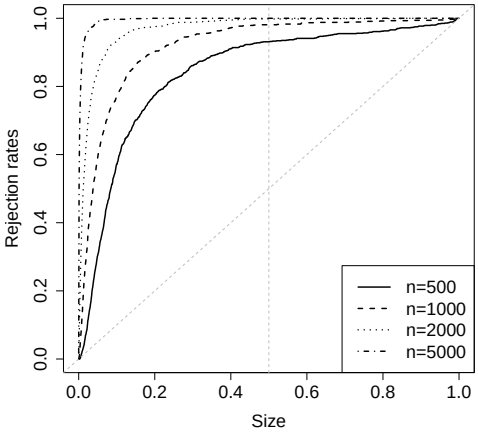


(b) Size-power plot

NCoVaR Granger causality (above quantile)



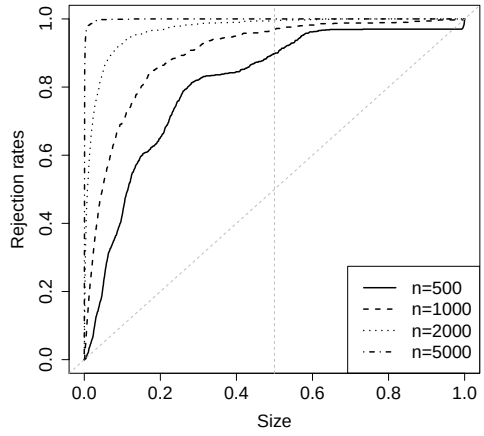
(c) Size-size plot



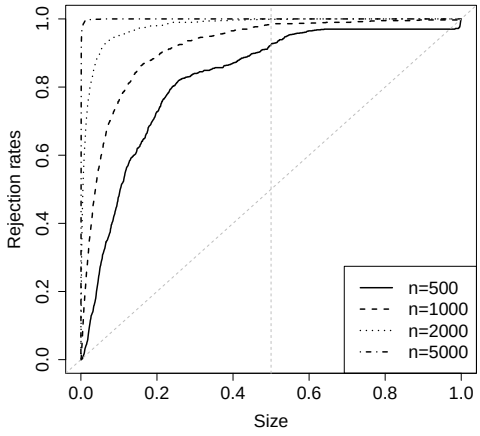
(d) Size-power plot

Figure E5: NCoVaR and NCoVaR Granger causality performance in VAR class of models in above quantile definition. Size-size and size-power plots for different sample sizes generated under process in Eq. (8) with instantaneous ($\tau = 0$) and lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

NCoVaR (above quantile)

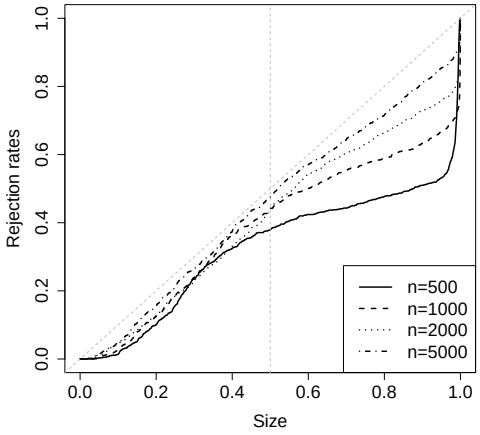


(a) Size-size plot

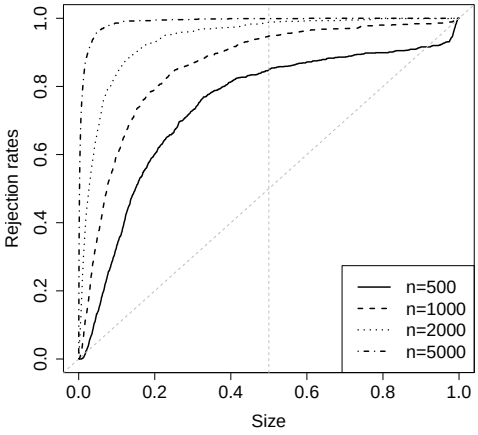


(b) Size-power plot

NGCoVaR Granger causality (above quantile)



(c) Size-size plot



(d) Size-power plot

Figure E6: NCoVaR and NCoVaR Granger causality performance in (G)ARCH class of models in above quantile definition. Size-size and size-power plots for different sample sizes generated under process in Eq. (9) with instantaneous ($\tau = 0$) and lagged dependence ($\tau = 1$). The quantile at which the risk is defined is set to $q = 0.99$. We apply the MSE-optimal bandwidth and we set the fixed-range parameter to $\mu = 0.8$. The results are aggregated over 1000 simulations.

Appendix F. Fixed bandwidth parametrization

Table F4: Relative size and power efficiency of NCoVaR and NCoVaR Granger causality measures depending on the fixed bandwidth parametrization (μ) for a sample size of $n = 500$ and risky quantile estimated at $\gamma = 0.95$.

			Fixed bandwidth choice									
Test	Process	Measure	$\mu = 0.4$	$\mu = 0.5$	$\mu = 0.6$	$\mu = 0.7$	$\mu = 0.8$	$\mu = 0.9^{\dagger}$	$\mu = 1.0^{\dagger}$	$\mu = 1.1^{\dagger}$	$\mu = 1.2^{\dagger}$	$\mu = 1.3^{\dagger}$
NCoVaR	VAR	size	1132%	1466%	1728%	1854%	1914%	1966%	1988%	1996%	2000%	2000%
	ARCH	size	290%	376%	422%	430%	458%	394%	362%	302%	218%	194%
	VAR	power	56%	73%	85%	93%	95%	98%	99%	100%	100%	100%
	ARCH	power	15%	19%	21%	22%	23%	20%	18%	15%	11%	10%
NCoVaR Gc	VAR	size	22%	34%	44%	44%	68%	72%	96%	98%	100%	104%
	ARCH	size	36%	36%	32%	42%	56%	74%	78%	110%	94%	106%
	VAR	power	25%	39%	53%	66%	72%	81%	89%	92%	96%	97%
	ARCH	power	7%	12%	14%	15%	16%	15%	13%	12%	9%	7%

Table F5: Relative size and power efficiency of NCoVaR and NCoVaR Granger causality measures depending on the fixed bandwidth parametrization (μ) for a sample size of $n = 500$ and risky quantile estimated at $q = 0.99$.

			Fixed bandwidth choice									
Test	Process	Measure	$\mu = 0.4$	$\mu = 0.5$	$\mu = 0.6$	$\mu = 0.7$	$\mu = 0.8$	$\mu = 0.9^{\dagger}$	$\mu = 1.0^{\dagger}$	$\mu = 1.1^{\dagger}$	$\mu = 1.2^{\dagger}$	$\mu = 1.3^{\dagger}$
NCoVaR	VAR	size	242%	512%	842%	1182%	1464%	1656%	1792%	1862%	1938%	1972%
	ARCH	size	62%	132%	248%	366%	468%	592%	684%	682%	698%	692%
	VAR	power	14%	26%	42%	58%	72%	84%	89%	93%	96%	99%
	ARCH	power	5%	9%	15%	21%	28%	32%	37%	37%	37%	36%
NCoVaR Gc	VAR	size	0%	0%	4%	2%	6%	12%	24%	30%	38%	52%
	ARCH	size	0%	0%	6%	4%	10%	20%	28%	36%	34%	36%
	VAR	power	3%	7%	13%	25%	36%	48%	58%	64%	74%	83%
	ARCH	power	2%	3%	6%	11%	15%	23%	27%	26%	27%	26%

Notes: Relative efficiency is calculated as the ratio between actual rejection rates and perfect rejection rates for size and power, i.e. 0.05 and 1.0, respectively. Simulations are run on two processes specified in Eqs. (8) and (9). The numbers are aggregated over 1000 runs. for larger μ values the conditioning events C and D have some overlap, which is ignored in the calculations.