Backtesting CoVaR

Author 1

Department of YYY, University of XXX

and

Author 2

Department of ZZZ, University of WWW

June 7, 2018

Abstract

We propose a new backtesting method for Delta-CoVaR. This measure of systemic risk is defined as the increase in Value of Risk of a financial institution when the financial system is in distress (i.e., when it violates its own VaR) compared to a non-distress state (when its return is at the median). By relying on a conditional version of the probability integral transform, we are able to exploit the entire sample of CoVaR predictions, rather than just the periods in which the system is in distress. A Monte Carlo study demonstrates that the new test has dramatically increased power vis-a-vis existing tests. An empirical application demonstrates the use of the test in a real-world scenario.

Keywords: Backtest; CoVaR; Systemic risk.

1 Introduction

Since the financial crisis, there has been a lot of attention for systemic risk, the risk that one institution in financial distress can contaminate the entire financial system. Federal Reserve Governor Daniel Tarullo formulated systemic importance in a 2009 testimony before the Senate Banking, Housing, and Urban Affairs Committee as follows:[1]

"Financial institutions are systemically important if the failure of the firm to meet its obligations to creditors and customers would have significant adverse consequences for the financial system and the broader economy."

The most widely used risk measure in finance was (and still is) the Value at Risk (VaR), which is defined as the maximum return loss of a financial institution within a certain probability. The VaR focuses on one financial institution in isolation and does not capture its risk contribution to the financial system. Therefore, there has been a lot of interest in risk measures which take the interaction between financial institutions into account. Systemic risk measures aim to capture the amount of risk in the financial system that can be attributed to a financial institution.

Several systemic risk measures have been developed. Billio et al. (2012) use the time variation in the number of principal components necessary to explain a certain fraction of the volatility of the return of the financial system, to determine the time variation in the level of interconnectedness between individual financial institutions. They analyse the directionality of shocks with Granger causality tests.

Huang et al. (2012) construct a systemic risk measure from a hypothetical insurance premium against distress of the financial system based on the probability of default of individual institutions and their asset correlations.

Acharya et al. (2010) define the Systemic Expected Shortfall (SES) as the Expected Shortfall of an institution given that the financial system is in crisis. (The Expected Shortfall is defined as the expected loss of an institution given that its loss is above it's VaR level.) Acharya et al. (2012) extend this definition to the expected amount of capital that a firm needs in case of a financial crisis.

This paper focuses on a systemic risk measure developed by Adrian & Brunnermeier

---

[1]Source: https://www.federalreserve.gov/newsevents/testimony/tarullo20090723a.htm

(2016). It is called the *Difference in Conditional Value at Risk* or $\Delta CoVaR$. The $\Delta CoVaR$ of an institution j is defined as the VaR of the financial system conditional on j being in distress minus the VaR of the financial system conditional on the return of j being at its median. This definition captures the systemic risk in line with the definition mentioned earlier, the amount of systemic risk that can be attributed to institution j.

A lot of further research on $\Delta CoVaR$ has been published after the first working paper of Adrian & Brunnermeier was published in 2008. Girardi & Ergün (2013), Mainik & Schaanning (2012), and Banulescu et al. (2016) provide backtesting methods to test whether an estimated $(\Delta)CoVaR$ series is specified correctly.

The general idea of these methods is to use conventional VaR backtesting methods on the subset of $CoVaR$ predictions in which the condition is satisfied (i.e. where institution j is in distress). A disadvantage of these methods is that testing only this subset (often only about 1% or 5% of the predictions) could lead to limited test power in realistic sample sizes.

In this paper, we propose a new $CoVaR$ backtesting method. Instead of using a subset in which the condition is satisfied, we predict the $CoVaR$ as a function of its condition (the return of j) and evaluate the predicted $CoVaR$ at the realisation of this condition (the observed return of j). This allows us to perform tests on the full set of $CoVaR$ predictions.

The remainder of this paper is organised as follows. Section 2 formally defines the $\Delta CoVaR$ and discusses different estimation methods. Section 3 introduces a previous $CoVaR$ backtesting method as well as our new $CoVaR$ backtesting method.

In Section 4, we use Monte Carlo simulation to assess the finite sample properties of our test. We find that our test is useful in a realistic setting and that it has better small sample properties than previous tests. In Section 5 we apply our test to real world data and compare the performance of three $CoVaR$ estimation methods. Finally, Section 6 makes some concluding remarks and some suggestions for further research.

2 CoVaR

2.1 Definition

The one-day ahead Value at Risk $\left(VaR^i_{q,t}\right)$ is defined as the q-th quantile of the return loss of institution or group of institutions i at time t, given the information available at time $t-1$:

$$\mathbb{P}\left[X^i_t \leq VaR^i_{q,t} \,\middle|\, \mathcal{F}_{t-1}\right] = q, \tag{1}$$

where X^i_t is the return loss of i and $\mathcal{F}_{t-1}$ is the information set available at time $t-1$. Often t is measured in days and q is 0.95 or 0.99. Subsequently, the Conditional Value at Risk ($CoVaR$) is defined as the Value at Risk of institution or group of institutions i conditional on some event $\mathcal{E}^j_t$ on institution or group of institutions j at time t:

$$\mathbb{P}\left[X^i_t \leq CoVaR^{i|\mathcal{E}^j_t}_{q,t} \,\middle|\, \mathcal{F}_{t-1}, \mathcal{E}^j_t\right] = q, \tag{2}$$

where $\mathcal{E}^j_t$ is often a realisation of the return loss of j.

In the literature, we find two different definitions of ΔCoVaR, which both attempt to capture the amount of systemic risk of i that can be attributed to j:

$$\Delta CoVaR^{i|j}_{q,t} = CoVaR^{i|X^j_t = VaR^j_{q,t}}_{q,t} - CoVaR^{i|X^j_t = VaR^j_{0.5,t}}_{q,t}, \tag{3}$$

$$\Delta CoVaR^{i|j}_{q,t} = CoVaR^{i|X^j_t \geq VaR^j_{q,t}}_{q,t} - CoVaR^{i|\mu^j_t - \sigma^j_t \leq X^j_t \leq \mu^j_t + \sigma^j_t}_{q,t}. \tag{4}$$

Adrian & Brunnermeier (2016) define the $\Delta CoVaR$ according to (3), which is the VaR of i given that the return loss of j (X^j_t) is at its q-th quantile (its VaR) minus the VaR of i given that the return loss of j is at its 0.5-quantile (its median).[2] We will refer to this by *Quantile $\Delta CoVaR$*.

Girardi & Ergün (2013) change this definition to (4), which is the VaR of i given that the return loss of j (X^j_t) is larger than or equal to its q-th quantile minus the VaR of i given that the return loss of j is no more than one standard deviation from its mean.[3] We will refer to this by *Tail $\Delta CoVaR$*.

---

[2]The first working paper of Adrian & Brunnermeier was released in 2008.

[3]The final $\Delta CoVaR$ definition used by Girardi & Ergün (2013) is

$$\Delta CoVaR^{i|j}_{q,t} = 100 \times \left(CoVaR^{i|X^j_t \geq VaR^j_{q,t}}_{q,t} - CoVaR^{i|b^j_t}_{q,t}\right) \middle/ CoVaR^{i|b^j_t}_{q,t}$$

where b^j_t is the base state of X^j_t: $\mu^j_t - \sigma^j_t \leq X^j_t \leq \mu^j_t + \sigma^j_t$.

One of the reasons Girardi & Ergün (2013) prefer $CoVaR$ based on an inequality condition, is monotonicity in the dependence parameter. Mainik & Schaanning (2012) argue that conditioning on an equality condition could violate this relationship. Under a joint normality assumption for example, $CoVaR_{q,t}^{i|X_t^j=VaR_{q,t}^j}$ is not monotonically increasing in the correlation parameter. However, after subtracting $CoVaR_{q,t}^{i|X_t^j=VaR_{0.5,t}^j}$, $\Delta CoVaR$ based on equality conditions and a normality assumption is monotonically increasing in the correlation parameter.

The advantage of using quantiles in the return realisation condition (instead of a fixed level of return loss) is that we also take the likelihood of such an event into account, which enables us to compare the $\Delta CoVaR$s over time.

Similar to Adrian & Brunnermeier (2016) and Girardi & Ergün (2013), and in line with the definition of systemic risk discussed in the introduction, we will focus on the $\Delta CoVaR$ of the financial system conditional on the return of a specific institution ($\Delta CoVaR^{system|j}$). Adrian & Brunnermeier (2016) argue that it is an advantage that the $\Delta CoVaR$ does not causally allocate the source of systemic risk (the correlation might originate from a common source) because it assigns a high risk value to small financial institutions that are of no significant importance on their own but are significant as part of a larger group of financial institutions with similar risks.

In the following sections, we will use the following notation: $CoVaR$ based on an equality condition (e.g. $CoVaR_{q,t}^{i|X_t^j=VaR_{q,t}^j}$) is referred to by Q-$CoVaR$ ($Quantile\ CoVaR$) and $CoVaR$ based on an inequality condition (e.g. $CoVaR_{q,t}^{i|X_t^j\geq VaR_{q,t}^j}$) is referred to by T-$CoVaR$ ($Tail\ CoVaR$).

2.2 Estimating Q-CoVaR

2.2.1 Quantile regression

Adrian & Brunnermeier (2016) use quantile regression to estimate $\Delta CoVaR$. Quantile regression is similar to ordinary linear regression. However, instead of estimating the mean of the dependent variable, it estimates a quantile of the dependent variable.

Since the Q-$CoVaR$ is defined as the q-th quantile of the return loss of i given the return loss of j, we can use quantile regression to estimate the Q-$CoVaR$. The advantage

of the quantile regression method is that it does not need any distributional assumptions. The approach is similar to VaR estimation by quantile regression, but this time x_t^j (the realisation of X_t^j) is included as an explanatory variable:

$$x_t^i = \alpha_q + \gamma_q \, M_{t-1} + \beta_q \, x_t^j + \varepsilon_{q,t} \tag{5}$$

$$\widehat{F}^{-1}_{X_t^i|X_t^j} (q) = \widehat{CoVaR}^{i|X_t^j=x_t^j}_{q,t} = \hat{\alpha}_q + \hat{\gamma}_q \, M_{t-1} + \hat{\beta}_q \, x_t^j, \tag{6}$$

where $F^{-1}_{X_t^i|X_t^j} (\cdot)$ is the inverse cumulative distribution function of X_t^i given $X_t^j = x_t^j$ and M_{t-1} contains lagged market state variables (e.g., the lagged risk-free rate or lagged market volatility). The parameters are found using the optimization method of Koenker & Bassett (1978). Note that the expression for Q-$\Delta CoVaR$ simplifies to

$$\widehat{\Delta CoVaR}^{i|j}_{q,t} = \hat{\beta}_q \left(\widehat{VaR}^j_{q,t} - \widehat{VaR}^j_{0.5,t} \right). \tag{7}$$

2.2.2   Asymmetric GARCH-DCC model with Gaussian innovations

A disadvantage of the quantile regression method is that the parameter β_q, which measures the dependency of the $CoVaR$ of i on the return loss of j, does not change over time, even when the correlation between X_t^i and X_t^j changes. Therefore, we will also discuss a second estimation method which is similar to the GARCH-DCC model used by Girardi & Ergün (2013).

We follow the derivation of Adrian & Brunnermeier (2016) to find an analytical expression of Q-$CoVaR$ based on a joint normality assumption:

$$CoVaR^{i|X_t^j=x_t^j}_{q,t} = F^{-1}_{X_t^i|X_t^j} (q) = \frac{\sigma_t^i \rho_t^{ij}}{\sigma_t^j} x_t^j + \sigma_t^i \sqrt{1 - (\rho_t^{ij})^2} \, \Phi^{-1} (q) . \tag{8}$$

where $\Phi^{-1} (\cdot)$ is the inverse standard normal CDF. The expression for Δ CoVaR simplifies to:

$$\Delta CoVaR^{i|j}_{q,t} = \Phi^{-1} (q) \, \rho_t^{ij} \sigma_t^i, \tag{9}$$

as the right-hand side in (8) cancels from taking the difference, and because under the normality assumption $VaR^j_{q,t} = F^{-1}_{X_t^j} (q) = \Phi^{-1} (q) \, \sigma_t^j$ and $\Phi^{-1} (0.5) = 0$.

In order to find $CoVaR$ estimates using (8) we need a method to find the volatility and correlation series. For this purpose, we use the Asymmetric GARCH-DCC model.

6

Girardi & Ergün (2013) use a skewed-t distribution (Hansen 1994) for the innovations to account for the asymmetry in return distributions and use a regular GARCH-DCC model to estimate the volatility and correlation series. In this paper, we will use symmetric distributions for the innovations and we will use the asymmetric GARCH-DCC model (TGARCH) to account for the leverage effect (Gloston et al. 1993). The leverage effect refers to the property that large negative returns often lead to more increased volatility than large positive returns.

The first step is to find the volatility series by estimating two univariate Threshold Generalized Autoregressive Conditional Heteroscedasticity (TGARCH) models. The TGARCH(1,1,1) model is:

$$r_t = \sigma_t \nu_t \tag{10}$$

$$\sigma_t^2 = \alpha + (\beta + \gamma \mathbb{I}(r_{t-1} < 0)) r_{t-1}^2 + \delta \sigma_{t-1}^2, \tag{11}$$

where $\{r_t\}$ is a univariate return series ($r_t = -x_t$), $\alpha > 0$, $\beta \geq 0$, $\delta \geq 0$ and $\nu_t \sim N(0,1)$ i.i.d. A value of $\gamma > 0$ would represent the leverage effect. The parameter values are found by maximum likelihood.

After we have estimated the volatility series $\{\hat{\sigma}_t^i\}$ and $\{\hat{\sigma}_t^j\}$, we obtain the correlation series $\{\hat{\rho}_t^{ij}\}$ using the Asymmetric Dynamic Conditional Correlation (DCC) model. First we obtain the standardized return vector series

$$\hat{\epsilon}_t = \begin{pmatrix} \hat{\epsilon}_t^i \\ \hat{\epsilon}_t^j \end{pmatrix} = \begin{pmatrix} r_t^i/\hat{\sigma}_t^i \\ r_t^j/\hat{\sigma}_t^j \end{pmatrix} = \begin{pmatrix} -x_t^i/\hat{\sigma}_t^i \\ -x_t^j/\hat{\sigma}_t^j \end{pmatrix}, \tag{12}$$

then we estimate by maximum likelihood

$$Q_t = (1 - \theta_1 - \theta_3)\bar{Q} + \theta_1 \epsilon_{t-1}\epsilon_{t-1}' + \theta_2(v_{t-1}v_{t-1}' - \bar{N}) + \theta_3 Q_{t-1}, \tag{13}$$

where $\theta_1 \geq 0$, $\theta_3 \geq 0$, $v_t = max(0, -\epsilon_t)$, $\bar{Q}$ is the unconditional covariance matrix of ϵ_t and $\bar{N}$ is the unconditional covariance matrix of v_t. The correlation series are obtained by

$$\hat{\rho}_t^{ij} = \frac{\hat{q}_{12,t}}{\sqrt{\hat{q}_{11,t}\hat{q}_{22,t}}}, \tag{14}$$

where $q_{kl,t}$ is the kl'th element of Q_t. The correlation matrix $\hat{\rho}_t$ has unit diagonal elements and off-diagonal elements $\hat{\rho}_t^{ij}$. The maximum likelihood specification is based on $\epsilon_t | \mathcal{F}_{t-1} \sim N(0, \rho_t)$. Maximum likelihood is equivalent to maximising $-\frac{1}{2}\sum_{t=2}^{T} (log|\hat{\rho}_t| + \hat{\epsilon}_t' \hat{\rho}_t \hat{\epsilon}_t)$ over θ_1, θ_2, and θ_3.

2.2.3 Asymmetric GARCH-DCC model with Student-t innovations

Instead of assuming a bivariate normal distribution for (X_t^i, X_t^j), we can also assume a standardized bivariate Student-t distribution for the standardized returns. This might be more appropriate as it is often argued that stock return distributions have heavier tails than the normal distribution (e.g., Bradley & Taqqu 2001).

Using quasi maximum likelihood, we find the same parameters (and therefore, the same volatility and correlation series) from the Asymmetric GARCH-DCC model estimated under a normality assumption. Once we have obtained the standardized return loss vector series $\{\hat{\epsilon}_t\}$ and the correlation matrix series $\{\hat{\rho}_t\}$ (with unit diagonals), we can find the number of degrees of freedom ν by maximising the likelihood of the standardized return vector series assuming a standardized bivariate Student-t distribution. This bivariate distribution has the following pdf:

$$f_t\left(\epsilon_t^i, \epsilon_t^j\right) = \frac{\Gamma\left((\nu+2)/2\right)}{\Gamma\left(\nu/2\right)(\nu-2)\pi\left|\rho_t\right|^{1/2}}\left(1 + \frac{1}{\nu-2}\epsilon_t'\rho_t^{-1}\epsilon_t\right)^{-(\nu+2)/2}. \tag{15}$$

From this expression, we follow the derivation of Kotz & Nadarajah (2004) to derive the pdf of the conditional distribution of X_t^i given $X_t^j = x_t^j = -\sigma_t^j\epsilon_t^j$:

$$f_t\left(x_t^i\middle|x_t^j\right) = \frac{\Gamma\left((\nu+2)/2\right)}{\sigma_t^i\sqrt{(\nu-2)\pi}\Gamma\left((\nu+1)/2\right)\left|\rho_t\right|^{1/2}} \times \frac{\left(1 + (1/(\nu-2))\left(x_t^j/\sigma_t^j\right)^2\right)^{(\nu+1)/2}}{\left(1 + (1/(\nu-2))x_t'\Sigma_t^{-1}x_t\right)^{(\nu+2)/2}}, \tag{16}$$

where Σ_t is the covariance matrix of (X_t^i, X_t^j) and $x_t = (x_t^i\ x_t^j)'$. Finally, the Q-$CoVaR$ can be found by solving

$$\int_{-\infty}^{\widehat{CoVaR}_{q,t}^{i|X_t^j=x_t^j}} \widehat{f}_t\left(x_t^i\middle|x_t^j\right) dx_t^i = q \tag{17}$$

for $\widehat{CoVaR}_{q,t}^{i|X_t^j=x_t^j}$, where $\widehat{f}_t\left(x_t^i\middle|x_t^j\right)$ is the estimated pdf of X_t^i conditional on $X_t^j = x_t^j$.

2.3 Estimating T-CoVaR

In this section we will shortly explain how the $T\text{-}CoVaR$ can be found. From the $T\text{-}CoVaR$ definition we know that

$$\mathbb{P}\left[X_t^i \geq CoVaR_{q,t}^{i|X_t^j \geq VaR_{q,t}^j} \,\middle|\, X_t^j \geq VaR_{q,t}^j, \mathcal{F}_{t-1}\right] = 1 - q \tag{18}$$

$$\implies \frac{\mathbb{P}\left[X_t^i \geq CoVaR_{q,t}^{i|X_t^j \geq VaR_{q,t}^j}, X_t^j \geq VaR_{q,t}^j \,\middle|\, \mathcal{F}_{t-1}\right]}{\mathbb{P}\left[X_t^j \geq VaR_{q,t}^j \,\middle|\, \mathcal{F}_{t-1}\right]} = 1 - q. \tag{19}$$

Next, since the denominator on the left-hand side of (19) is equal to $1 - q$, we have

$$\mathbb{P}\left[X_t^i \geq CoVaR_{q,t}^{i|X_t^j \geq VaR_{q,t}^j}, X_t^j \geq VaR_{q,t}^j \,\middle|\, \mathcal{F}_{t-1}\right] = (1 - q)^2. \tag{20}$$

Finally, after one has found $\widehat{VaR}_{q,t}^j$ and the joint distribution of X_t^i and X_t^j using some estimation method, the $T\text{-}CoVaR$ can be found, for given $\widehat{VaR}_{q,t}^j$, by solving

$$\int_{\widehat{CoVaR}_{q,t}^{i|X_t^j \geq VaR_{q,t}^j}}^{\infty} \int_{\widehat{VaR}_{q,t}^j}^{\infty} \widehat{f}_t(x_t^i, x_t^j) dx_t^j dx_t^i = (1 - q)^2 \tag{21}$$

for $\widehat{CoVaR}_{q,t}^{i|X_t^j \geq VaR_{q,t}^j}$ (where $\widehat{f}_t(x_t^i, x_t^j)$ is the estimated joint probability distribution function of (X_t^i, X_t^j) conditional on $\mathcal{F}_{t-1}$). For a distribution that is symmetric around zero (e.g. a Gaussian distribution with zero mean) the equation simplifies to:

$$\widehat{F}_t\left(-\widehat{CoVaR}_{q,t}^{i|X_t^j \geq VaR_{q,t}^j}, -\widehat{VaR}_{q,t}^j\right) = (1 - q)^2, \tag{22}$$

where $\widehat{F}_t(\cdot, \cdot)$ is the estimated cumulative distribution function of (X_t^i, X_t^j).

2.4 Backtesting T-CoVaR

A second reason why Girardi & Ergün (2013) prefer $Tail\ CoVaR$ (besides the monotonicity property of the $T\text{-}CoVaR$) is that it enables them to backtest their $CoVaR$ predictions. In other papers about backtesting $CoVaR$, a similar aproach is used (Mainik & Schaanning 2012, Banulescu et al. 2016). We will briefly explain the backtesting procedure used in

these papers, which is based on conventional VaR backtesting methods. We know that for a correctly specified T-$CoVaR$:

$$\mathbb{P}\left[X_t^i > CoVaR_{q,t}^{i|X_t^j \geq VaR_{q,t}^j} \,\big|\, \mathcal{F}_{t-1}, X_t^j \geq VaR_{q,t}^j\right] = 1 - q. \tag{23}$$

If we focus on the subset of T-$CoVaR$ predictions for which the observed $x_t^j \geq \widehat{VaR}_{q,t}^j$ (where x_t^j is the realisation of X_t^j), we can define a hit series similar to a VaR hit series:

$$I_n^i = \mathbb{I}\left(x_n^i > \widehat{CoVaR}_{qn}^{i|X_n^j \geq VaR_{qn}^j}\right), \qquad n \in \{1, \ldots, N\}, \tag{24}$$

where N is the number of VaR violations of j, x_n^i is the realisation of X_n^i, and $\mathbb{I}(\cdot)$ is the indicator function which is equal to one if the condition is true and zero otherwise. This subset has an expected cardinality of $\mathbb{E}[N] = (1-q)T$, provided that $\widehat{VaR}_{q,t}^j$ is correctly specified (where T is the cardinality of the full set of $CoVaR$ predictions).

Under the null hypothesis of a correctly specified T-$CoVaR$, it follows from (23) that $\mathbb{P}[I_n^i = 1] = (1-q)$ and $\mathbb{P}\left[I_n^i = 1 \,\big|\, I_{n-1}^i = 1\right] = \mathbb{P}\left[I_n^i = 1 \,\big|\, I_{n-1}^i = 0\right]$. The first equation corresponds to the unconditional coverage test, which tests whether the fraction of hits is not significantly different from $(1-q)$. The second equation corresponds to the independence test, which tests whether the historic information set $\mathcal{F}_{n-1}$ provides no additional information about the occurrence of the hits. A third test is the conditional coverage test, which tests both hypotheses simultaneously (Christoffersen 1998).

3 A new backtest for CoVaR

A disadvantage of the T-$CoVaR$ backtesting method explained in the previous section, is that it only uses the subset of $CoVaR$ predictions for which $x_t^j \geq \widehat{VaR}_{q,t}^j$. This means that the expected number of predictions in the test sample (hit series) is $\mathbb{E}[N] = (1-q)T$, where T is the cardinality of the full set of $CoVaR$ predictions. In this test sample, the expected number of hits is $(1-q)\mathbb{E}[N] = (1-q)^2 T$. This means that for $q = 0.95$ we have one expected test sample hit in 400 predictions in the original set. For $q = 0.99$, this is only one in 10,000, which is approximately one every forty years. Therefore, these tests have little power for T-$CoVaR$ prediction series of limited length. In this paper, we propose a backtesting procedure for Q-$CoVaR$ which uses the full set of predictions.

10

To ease the notational burden, let $F_t(\cdot \mid \cdot)$ denote the cdf of the return of asset i, conditional on $\mathcal{F}_{t-1}$ and the return on asset j; i.e.,

$$F_t(r \mid s) \equiv \mathbb{P}\left[X_t^i < -r \mid \mathcal{F}_{t-1}, X_t^j = -s\right].$$

Our test is based on the hit series

$$J_t \equiv \begin{cases} 1, & \text{if } F_t(-x_t^i \mid -x_t^j) \leq 1-q, \\ 0, & \text{otherwise,} \end{cases}$$

where x_t^i and x_t^j denote the observed relative losses. We then have the following result.

Theorem 1. *Let $\mathcal{G}_{t-1} \equiv \mathcal{F}_{t-1} \cup \sigma(x_t^j)$. Then*

$$\mathbb{E}[J_t \mid \mathcal{G}_{t-1}] = 1-q.$$

Proof. Let $R \equiv -X_t^i$, $S \equiv -X_t^j$, $Y \equiv F_t(R \mid S)$, $f_t(r \mid s) \equiv \partial/\partial r F_t(r \mid s)$, and let $F_t^{-1}(\cdot \mid \cdot)$ denote the quantile function of the return of asset i, conditional on $\mathcal{F}_{t-1}$ and the return on asset j, i.e., the functional inverse of F in its first argument. The result follows from a conditional version of the probability integral transform, viz.,

$$\mathbb{P}\left[Y < y \mid \mathcal{G}_{t-1}\right] = \mathbb{E}\left[1_{F_t(R,S)<y} \mid \mathcal{G}_{t-1}\right] = \mathbb{E}\left[1_{R<F_t^{-1}(y,S)} \mid \mathcal{G}_{t-1}\right]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{F_t^{-1}(y,s)} f_t(r \mid s)\mathrm{d}r = \int_{-\infty}^{\infty} F_t(F_t^{-1}(y,s),s) = y.$$

$\square$

The implication in Theorem 1 can be tested by forming a feasible version of J_t and testing if its sample average, conditional on x_t^j and variables in $\mathcal{F}_{t-1}$ (e.g., x_{t-1}^i, x_{t-1}^j, VaR_{t-1}^i, VaR_{t-1}^j, $CoVaR_{q,t-1}^{i|X_{t-1}^j=x_{t-1}^j}$, J_{t-1}^i, etc.), equals $1-q$. Here, we specifically choose to include J_{t-1}^i and x_t^j. Our first test is analogous to the VaR backtesting method of Engle & Manganelli (2004) and uses the linear regression

$$J_t^i = \alpha + \beta\ J_{t-1}^i + \gamma\ x_t^j + \varepsilon_t. \tag{25}$$

Two implications of Theorem 1 will be tested: $H_{0,a} : \gamma = 0$, and $H_{0,b} : \alpha = (1-q)$, $\beta = \gamma = 0$. Both hypotheses will be tested using a likelihood-ratio statistic (which are asymptotically

11

$\chi^2_{(1)}$ and $\chi^2_{(3)}$ distributed).[4] The expression for the LR-statistic is:

$$logLR = T \left(log(u_r^2) - log(u_u^2) \right),$$ (26)

where u_r and u_u are the standard errors of the restricted and unrestricted regressions, and T is the number of observations in the regressions. Our second test is analogous to the VaR backtesting method of Dumitrescu et al. (2012). It explicitly takes the binary nature of the hit series into account by testing $H_{0,a}$ and $H_{0,b}$ in the logistic regression

$$\mathbb{P}\left[J_t^i = 1 \right] = F(\alpha + \beta \; J_{t-1}^i + \gamma \; x_t^j),$$ (27)

where $F(\cdot)$ is the cumulative distribution function of the logistic distribution. The LR-statistic for this test can be expressed as follows:

$$logLR = \sum_{t=1}^{T} J_t^i \cdot log \left(\hat{\mathbb{P}}\left[J_t^i = 1 \right] \right) + \left(1 - J_t^i \right) \cdot log \left(\hat{\mathbb{P}}\left[J_t^i = 0 \right] \right),$$ (28)

where $\hat{\mathbb{P}}[\cdot]$ is the estimated probability based on the logistic regression. It is useful to contrast our test with that of Girardi & Ergün (2013) based on the hit series in (24). To restate their test in our notation, let

$$I_t \equiv \begin{cases} 1, & \text{if } \int_{-\infty}^{-VaR_q^j} F_t(-x_t^i \mid s) f_t(s) \mathrm{d}s \leq (1-q)^2 \wedge F_t(-x_t^j) \leq 1-q, \\ 0, & \text{otherwise,} \end{cases}$$ (29)

where $F_t(s)$ is the cdf of $-X_t^j$ conditional on $\mathcal{F}_{t-1}$ and $f_t(s)$ its derivative. I_n in (24) is obtained by re-indexing the hit series (??) to include only those observations for which $F_t(-x_t^j) \leq 1-q$; in a slight abuse of notation, we use the same symbol for the two series, distinguishing them only by their subscript. The implication tested by Girardi & Ergün's (2013) test is

$$\mathbb{E}[I_t \mid \mathcal{G}_{t-1}] = (1-q)N_t/t,$$

where $N_t = \sum_{\tau=1}^{t} 1_{F_\tau(-x_\tau^j) \leq 1-q}$; note that $\mathbb{E}[N_t] = t(1-q)$ if the VaR is correctly specified. It is immediately clear that the test based on I_t only considers observations (x_t^i, x_t^j) for

---

[4]To the best of our knowledge, this $CoVaR$ backtesting procedure has not been proposed before. Especially, testing if the dependence on X_t^j has been specified correctly by testing the explanatory power of x_t^j on the hit series is (to the best of our knowledge) new.

which both components lie in their respective tail, whereas the test based on J_t considers the left tail of the entire conditional distribution of X_t^i. As discussed earlier, this implies a more powerful test. Apart from such power considerations, another reason to prefer the J_t-based test over that based on I_t is that the Δ-CoVaR as defined in (3) and (4) requires a correctly specified conditional distribution for X_t^i at (at least) two quantiles of X_t^j: the VaR and the median. The test based on I_t only considers the former, whereas our new test considers all possible values.

4 Monte Carlo study

4.1 Misspecified contemporaneous dependence

In order to assess the finite sample properties of our test (the Q-$CoVaR$ test discussed in Section 3), we implement a Monte Carlo simulation. In this section, we will investigate whether our test is able to detect a distributional misspecification. In Section 4.2, a dynamic misspecification will be investigated.

4.1.1 Size

The first step is to simulate random returns that have real world return properties. To this end, we simulate returns following the Asymmetric GARCH-DCC model, which is explained in Section 2.2.2. The parameters used in the random return simulation are obtained by estimating an Asymmetric GARCH-DCC model on real world daily log-returns of the Dow Jones US Financials Index and JPMorgan Chase ranging from January 2001 to December 2014 (3500 observations).[5]

In each replication, a new bivariate return series with 3500 observations is simulated

---

[5]The estimated univariate TGARCH parameters for the Dow Jones US Financials Index are $\alpha = 0.0000015, \beta = 0.0064825, \gamma = 0.1294960$, and $\delta = 0.9217418$, and for JPMorgan Chase, $\alpha = 0.0000026, \beta = 0.0222182, \gamma = 0.1038000$, and $\delta = 0.9229578$, where the parameters correspond to (11). The positive values for γ represent the leverage effect.

The unconditional covariance matrices of the standardized returns are estimated as:

$$\bar{Q} = \mathbb{V}\left(\epsilon_t\right) = \begin{pmatrix} 1 & 0.847 \\ 0.847 & 1 \end{pmatrix}, \text{ and } \bar{N} = \mathbb{V}\left(v_t\right) = \begin{pmatrix} 0.525 & 0.444 \\ 0.444 & 0.502 \end{pmatrix}.$$

using this model. The first 2500 simulated observations are used to estimate new Asymmetric GARCH-DCC parameters. Based on these parameters, one-day ahead $Q\text{-}CoVaR$ predictions will be made for the last 1000 observations. Using these 1000 out-of-sample $CoVaR$ predictions we find the hit series and subsequently the likelihood-ratio test statistics. Since this process is quite time consuming, all the reported simulation results are based on 1000 replications (with 1000 daily $CoVaR$ predictions each).

Using a bivariate Gaussian distribution to simulate the standardized returns (12) and using the analytical $Q\text{-}CoVaR$ expression based on the joint normality assumption (8) we find the empirical size as the rejection frequency at the nominal 5% level.

In the first size simulation, we assume the parameters to be known. As can be seen from Table 1, when the parameters are assumed to be known, the empirical sizes are close to their nominal 5% level. In only two out of the eight cases the empirical size is significantly different from 5%.

When we estimate the parameters in each size simulation, the empirical sizes become more different from their nominal 5% level. In Table 2 the results of this empirical size simulation are shown. We find that for seven out of the eight test specifications, the size is significantly different from the nominal 5% level. To arrive at the asymptotic distribution, the number of observations in the out-of-sample hit series as well as the number of in-sample observations used for the Asymmetric GARCH-DCC parameter estimation should go to infinity. Since the number of test specifications for which the size is significantly different from the nominal 5% level decreases from seven to two when we assume the parameters to be known, we can conclude that one of the reasons why the test over-rejects is the parameter estimation error.

Finally, the Asymmetric DCC parameters are estimated as $\theta_1 = 0.0089828, \theta_2 = 0.0295914,$ and $\theta_3 = 0.9614814$, where the parameters correspond to (13). The data is discussed in Section 5.

	95% $CoVaR$		99% $CoVaR$	
Linear, $H_0 : \gamma = 0$	0.048	(3.78) [3.84]	0.053	(4.01) [3.84]
Logistic, $H_0 : \gamma = 0$	0.045	(3.73) [3.84]	0.050	(3.84) [3.84]
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	0.052	(7.85) [7.81]	0.069***	(9.13) [7.81]
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	0.056	(7.91) [7.81]	0.031***	(6.54) [7.81]

Table 1: Empirical sizes from the simulation with known parameters. The number of stars indicate whether the size is significantly different from 0.050 (*:=10% level, **:=5% level, ***:=1% level). The empirical 5% critical values are reported between parentheses with the corresponding asymptotic 5% critical values between brackets.

	95% $CoVaR$		99% $CoVaR$	
Linear, $H_0 : \gamma = 0$	0.062*	(4.32) [3.84]	0.064**	(4.26) [3.84]
Logistic, $H_0 : \gamma = 0$	0.062*	(4.26) [3.84]	0.059	(4.14) [3.84]
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	0.081***	(9.14) [7.81]	0.079***	(9.27) [7.81]
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	0.077***	(8.49) [7.81]	0.036**	(7.06) [7.81]

Table 2: Empirical sizes from the simulation with parameter estimation. The number of stars indicate whether the size is significantly different from 0.050 (*:=10% level, **:=5% level, ***:=1% level). The empirical 5% critical values are reported between parentheses with the corresponding asymptotic 5% critical values between brackets.

4.1.2 Power

After assessing the finite sample properties of the test when the Q-$CoVaR$ is specified correctly, we change the simulation design to see how well our test performs when the $CoVaR$ is misspecified. One option would be to estimate unconditional VaRs and use them as misspecified $CoVaR$s to investigate the power of the test. However, it is obvious that x_t^j will have explanatory power on the VaR hit series (provided that X_t^i and X_t^j are correlated). Therefore, it is more interesting to change the dependency structure between X_t^i and X_t^j in the return simulation and check whether our test is able to detect the distributional

misspecification of our Q-$CoVaR$ when we still use the the analytical $CoVaR$ expression based on the joint normality assumption (8).

In the size simulation we used a bivariate Gaussian distribution to simulate the standardized returns. Two random variables which are jointly normally distributed can be described as two variables with Gaussian marginal distributions and a Gaussian dependency structure. In our power simulation, we would like to change the dependency structure to a Student-t's while maintaining the Gaussian marginals.

This is implemented as follows. If we draw a random vector (T_1, T_2) from a bivariate Student-t distribution with ν degrees of freedom, correlation parameter ρ (the ρ_t from the Asymmetric GARCH-DCC model) and marginal CDFs $F(\cdot\,; \nu)$, then $C_1 = F(T_1; \nu)$ and $C_2 = F(T_2; \nu)$ are both uniformly distributed on $[0, 1]$. The distribution of (C_1, C_2) is called a "Student-t copula". If we transform these random variables as follows, $Z_1 = \Phi^{-1}(C_1)$ and $Z_2 = \Phi^{-1}(C_2)$, then (Z_1, Z_2) is a bivariate random variable with Gaussian marginals and a Student-t dependency structure. Finally, if we multiply them by the volatilities from the Asymmetric GARCH-DCC model, we have the simulated returns.

We assume the parameters to be unknown in the power simulation (so we estimate new parameters in each replication) just as in the first size simulation. Therefore, we use the empirical critical values from the first size simulation to find the size-corrected power. Since the predicted $CoVaR$s using (8) are misspecified in this case (due to the Student-t copula used in the simulation) we will find whether the test is able to detect this misspecification.
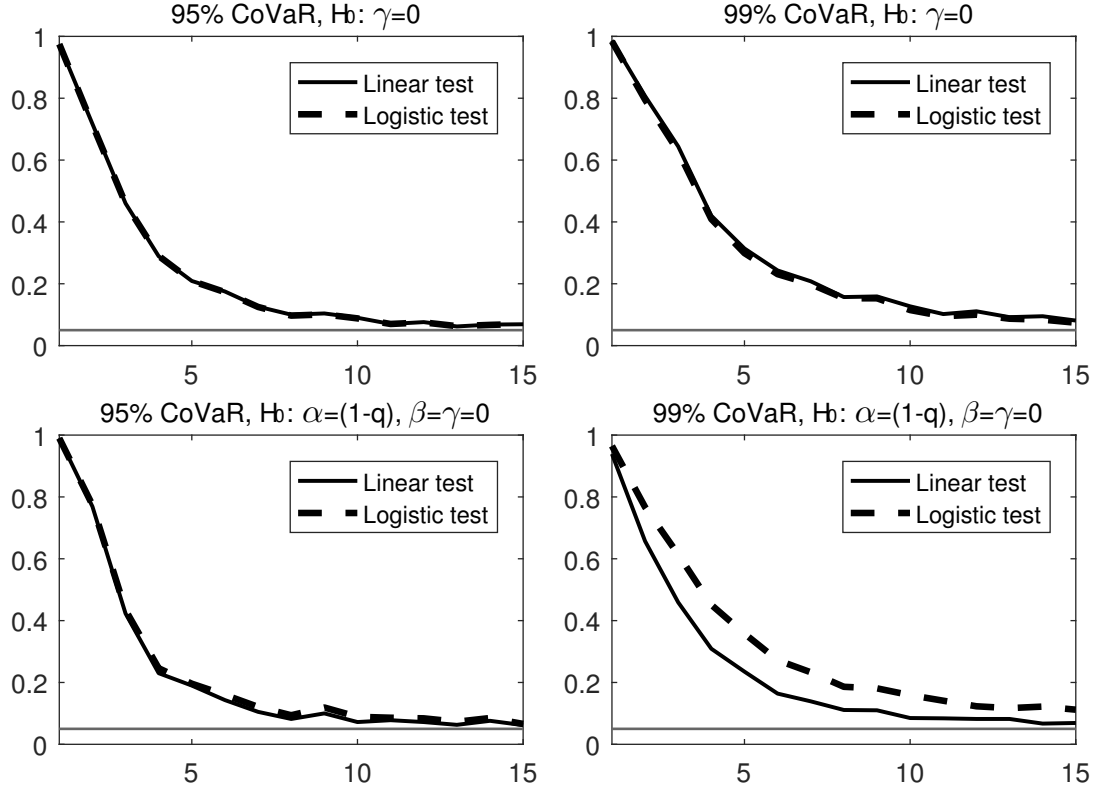
Figure 1: For each test specification, the graph shows the size-corrected power (y-axis) as a function of the number of degrees of freedom of the Student-t copula used in the simulation (x-axis). The horizontal grey line represents the nominal 5% level.

In Figure 1 the size-corrected power is shown as a function of the number of degrees of freedom of the Student-t copula used to simulate the returns. Note that as the number of degrees of freedom approaches infinity, the $CoVaR$ is correctly specified. When the number of degrees of freedom becomes small, the dependency structure becomes more different from the Gaussian dependency structure (which is assumed in the $CoVaR$ prediction), and our test detects the misspecification more often. In line with Dumitrescu et al. (2012), we find that the performance of the two tests (linear and logistic) is similar in most cases, although in some situations the logistic test outperforms the linear test (bottom-right).
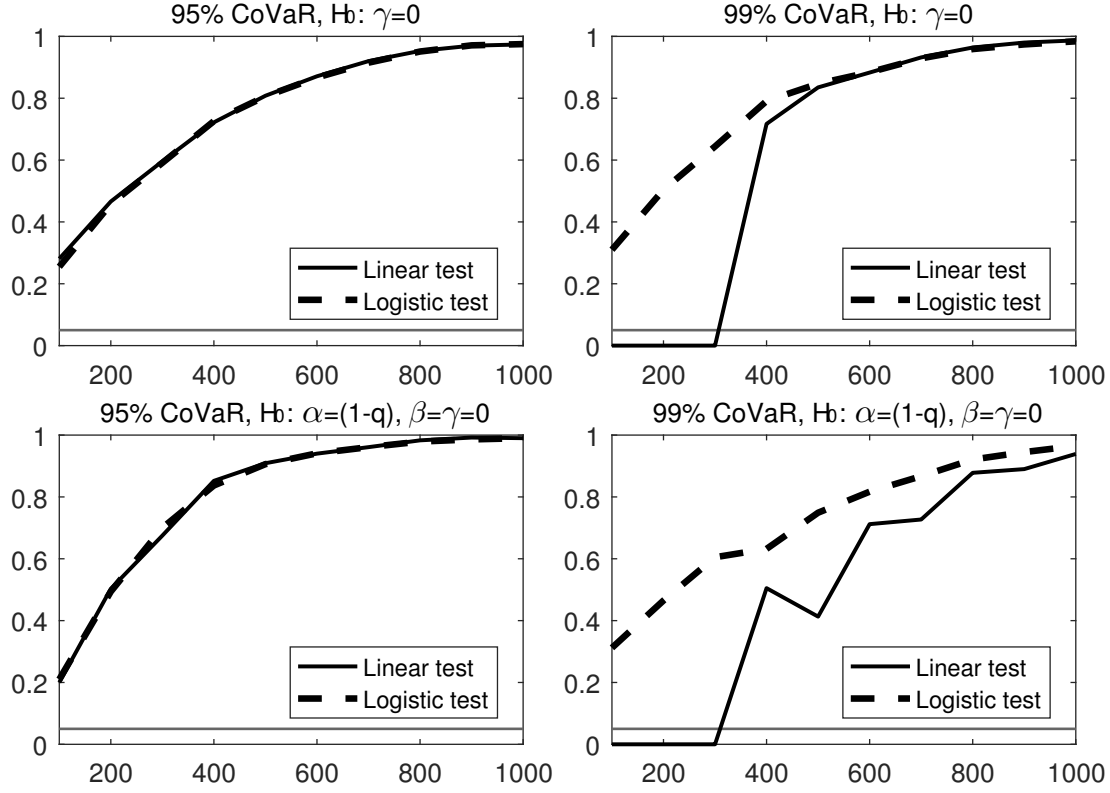
Figure 2: For each test specification, the graph shows the size-corrected power (y-axis) as a function of the number of out-of-sample $CoVaR$ predictions (x-axis).

Figure 2 shows the size-corrected power as a function of the number of out-of-sample $CoVaR$ predictions (the length of the hit series). The number of in-sample observations is still fixed at 2500 and the Student-t copula used in the simulation has one degree of freedom. For each power simulation, we use the empirical critical value from the corresponding size simulation. As expected, the size-corrected power increases when the length of the hit series increases.

The zero size-corrected power for small out-of-sample sizes on the bottom-right is caused by a perfect fit in the linear test regression in the size simulation when there are zero hits, which leads to infinite empirical critical values.

	95% $CoVaR$	99% $CoVaR$
Linear test, $H_0 : \gamma = 0$	0.97 (0.06)	0.99 (0.06)
Logistic test, $H_0 : \gamma = 0$	0.98 (0.06)	0.98 (0.06)
Linear test, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	0.99 (0.08)	0.94 (0.08)
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	0.99 (0.08)	0.97 (0.04)

Table 3: Size-corrected power at nominal 5% level. The corresponding sizes are reported between parentheses.

	1,000 predictions		10,000 predictions	
	95% $T\text{-}CoVaR$	99% $T\text{-}CoVaR$	95% $T\text{-}CoVaR$	99% $T\text{-}CoVaR$
Uncond. coverage test	0.04 (0.11)	0.06 (0.02)	0.07 (0.09)	0.05 (0.03)
Independence test	0.05 (0.08)	0.07 (0.01)	0.04 (0.11)	0.06 (0.02)
Cond. coverage test	0.06 (0.08)	0.06 (0.03)	0.07 (0.13)	0.06 (0.04)

Table 4: Size-corrected power for the T-$CoVaR$ backtesting method discussed in Section 2.4. The corresponding sizes are reported between parentheses.

In Table 3 the size-corrected power is shown with the corresponding sizes between parentheses. The number of out-of-sample predictions is 1000 and the standardized returns are simulated using a Student-t copula with one degree of freedom. Since all test specifications have a size-corrected power above 0.9, we can conclude that our test detects the misspecification very well in this setting.

The power simulation shows us that our test is able to detect a distributional misspecification in a realistic sample size. Combined with the size simulation results from the previous section, we can conclude that our test as proposed in section 3 is useful in a realistic setting when we want to test whether a $CoVaR$ specification describes the underlying distribution correctly.

Now we have shown that our test has the desired finite sample properties and that it is able to detect a distributional misspecification, we want to assess the performance of the T-$CoVaR$ backtesting method discussed in Section 2.4 in this setting. This test only tests

the subset where the inequality condition is satisfied.

In Table 4 the simulation results for Test 2 are shown. The first two columns show the results of the simulation where 1,000 out of sample T-$CoVaR$ predictions are tested. We find that this test has almost no power. Since only a small subset is tested here (so about 50 predictions for the 95% $CoVaR$ and about 10 for the 99% $CoVaR$), we increase the number of out-of-sample predictions to 10,000. The results are shown in the last two columns. It shows that the test still has almost no power.

4.2   Dynamic misspecification

In this section we will investigate whether our test is also able to detect a dynamic misspecification. We will use a Gaussian distribution both in the return simulation and in the $CoVaR$ prediction. The TGARCH specification for both volatility series will remain the same as in the previous section. However, in the $CoVaR$ prediction, we will assume a constant correlation over time. The goal of the Monte Carlo study is to find whether our test is able to detect this misspecification.

4.2.1   Size

The design of the size simulation is exactly the same as in the previous section. The only difference is that a constant correlation is assumed in the Q-$CoVaR$ prediction and used in the return simulation. In Table 5 the results of the size simulation with known parameters are shown, and in Table 6 the results of the size simulation are shown. Similar to the previous section, we find that the test over-rejects for most test specifications. In line with our previous findings, the number of test specifications for which the empirical size is significantly different from the nominal 5% level decreases and most of them become closer to the nominal 5% level.

	95% $CoVaR$		99% $CoVaR$	
Linear, $H_0 : \gamma = 0$	0.045	(3.65) [3.84]	0.068***	(4.49) [3.84]
Logistic, $H_0 : \gamma = 0$	0.045	(3.62) [3.84]	0.063*	(4.27) [3.84]
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	0.051	(7.88) [7.81]	0.072***	(9.21) [7.81]
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	0.054	(7.87) [7.81]	0.030***	(6.51) [7.81]

Table 5: Empirical sizes from the simulation with known parameters. The number of stars indicate whether the size is significantly different from 0.050 (*:=10% level, **:=5% level, ***:=1% level). The empirical 5% critical values are reported between parentheses with the corresponding asymptotic 5% critical values between brackets.

	95% $CoVaR$		99% $CoVaR$	
Linear, $H_0 : \gamma = 0$	0.055	(4.13) [3.84]	0.070***	(4.59) [3.84]
Logistic, $H_0 : \gamma = 0$	0.056	(4.10) [3.84]	0.062*	(4.27) [3.84]
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	0.068***	(8.86) [7.81]	0.089***	(9.69) [7.81]
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	0.065**	(8.58) [7.81]	0.042	(7.24) [7.81]

Table 6: Empirical sizes from the simulation with parameter estimation. The number of stars indicate whether the size is significantly different from 0.050 (*:=10% level, **:=5% level, ***:=1% level). The empirical 5% critical values are reported between parentheses with the corresponding asymptotic 5% critical values between brackets.

4.2.2   Power

In the power simulation we continue to use the Gaussian distribution and TGARCH volatilities both in the return simulation and in the Q-$CoVaR$ prediction. However, the DCC equation (13) used in the return simulation will be specified as follows:

$$Q_t = \left(1 - \theta_1 - \frac{1}{2}(1-\delta)\theta_2 - \theta_3\right)\bar{Q} + \delta\ \theta_1\epsilon_{t-1}\epsilon'_{t-1} + \delta\ \theta_2(v_{t-1}v'_{t-1} - \bar{N})$$
$$+ \left(\theta_3 + (1-\delta)\left(\theta_1 + \frac{1}{2}\theta_2\right)\right)Q_{t-1},$$
$$(30)$$

where θ_1, θ_2, and θ_3 are the empirically estimated values reported in Section 4.1.1. Note that the stationarity condition is satisfied for all values of δ, and that the unconditional expectation $\mathbb{E}[Q_t] = \bar{Q}$ for every δ. Also note that for $\delta = 0$, the correlation is constant, so for $\delta = 0$ the model is correctly specified (the size simulation). For $\delta = 1$, the real world coefficients are used, and for $\delta > 1$ the correlation will be more fluctuating than in the real world situation. Since all coefficients should be larger than or equal to zero the following condition should hold: $0 \leq \delta \leq \delta_{max} \approx 41$.

The simulation design will be the same as in the previous sections (with parameter estimation), and a constant correlation will be assumed in the $Q\text{-}CoVaR$ prediction.
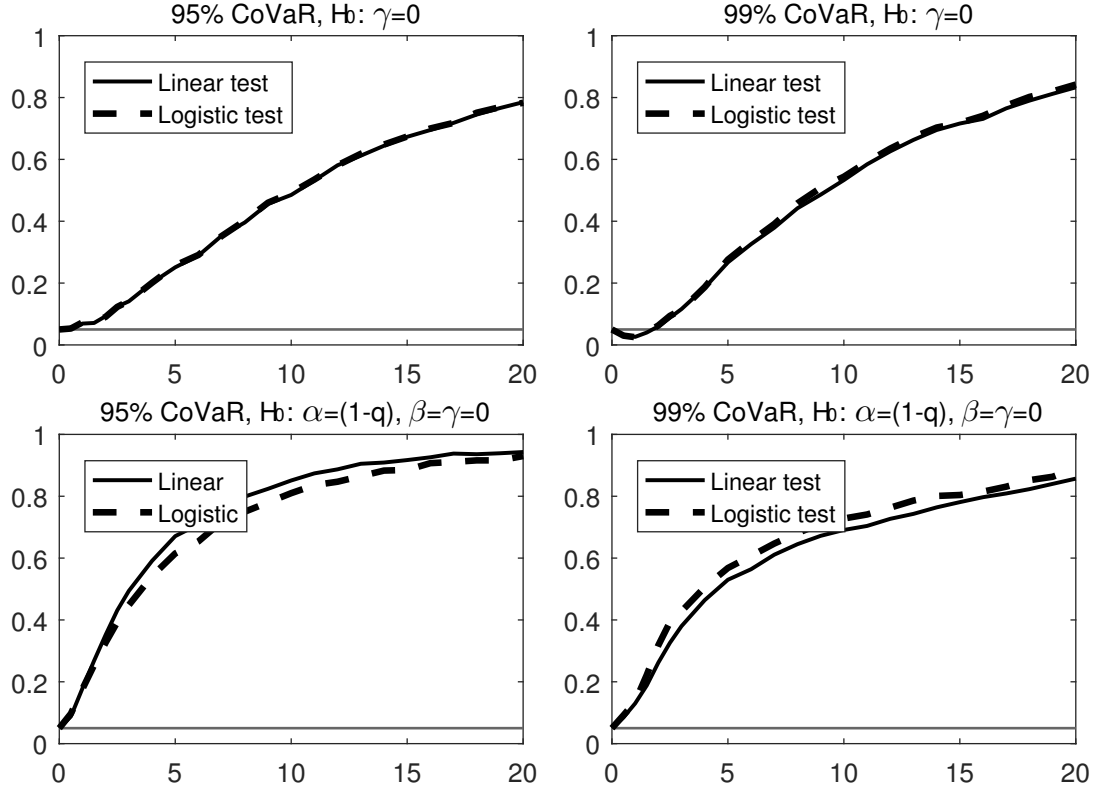

Figure 3: For each test specification, the graph shows the size-corrected power (y-axis) as a function of δ (x-axis) used in the DCC equation underlying the simulation (29).

In Figure 3 the size-corrected power is shown as a function of δ, for different test specifications. The empirical critical values from the size simulation are used. We find that in the real world situation ($\delta = 1$), where the correlation is relatively stable, our test is not performing very well. However, in more extreme cases, the test detects the dynamic
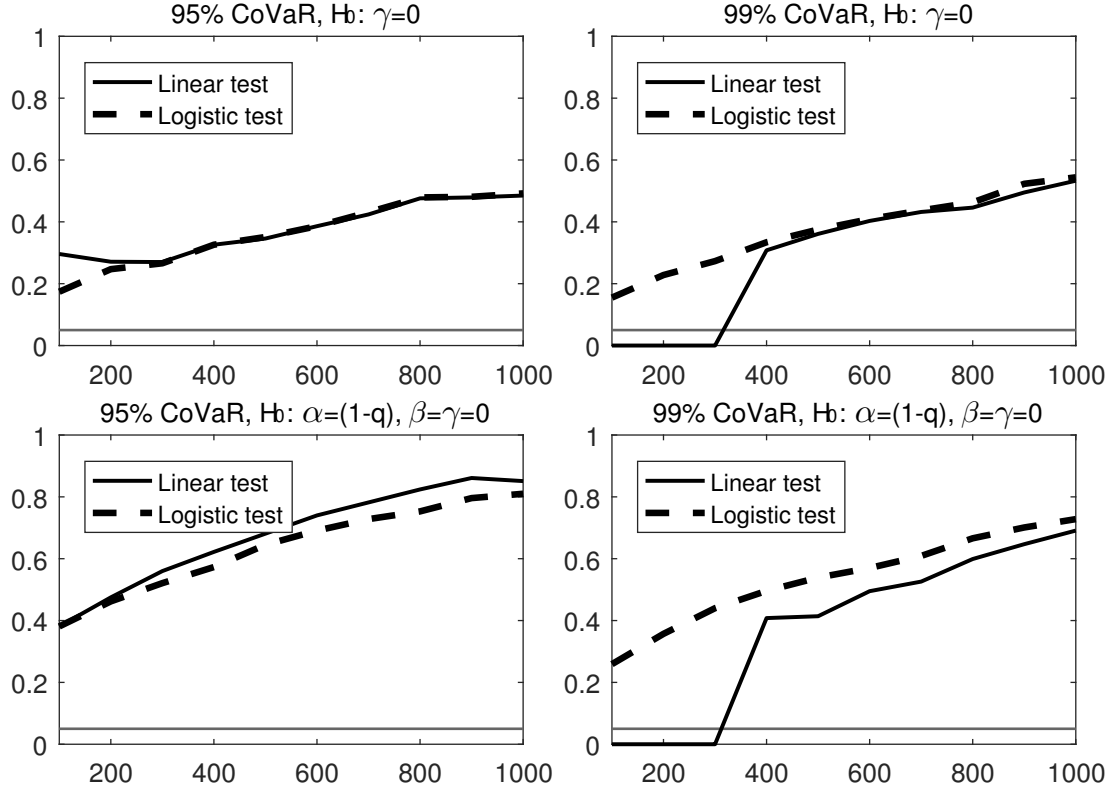
misspecification most of the times.

Figure 4: For each test specification, the graph shows the size-corrected power (y-axis) as a function of the number of out-of-sample $CoVaR$ predictions (x-axis).

In Figure 4 the size-corrected power is shown as a function of the number of out-of-sample $CoVaR$ predictions. The number of in-sample observations is fixed at 2500 and δ is fixed at 10. For each power simulation, we use the empirical critical values from the corresponding size simulation. As expected, the size-corrected power increases when the number of out-of-sample $CoVaR$ predictions increases.

23

	95% $CoVaR$	99% $CoVaR$
Linear, $H_0 : \gamma = 0$	0.49 (0.06)	0.53 (0.07)
Logistic, $H_0 : \gamma = 0$	0.49 (0.06)	0.54 (0.06)
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	0.85 (0.07)	0.69 (0.09)
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	0.81 (0.07)	0.73 (0.04)

Table 7: *Size-corrected power at nominal 5% level from the simulation with $\delta = 10$. The corresponding sizes are reported between parentheses.*

	1,000 predictions		10,000 predictions	
	95% $T\text{-}CoVaR$	99% $T\text{-}CoVaR$	95% $T\text{-}CoVaR$	99% $T\text{-}CoVaR$
Uncond. coverage test	0.05 (0.12)	0.05 (0.02)	0.07 (0.09)	0.05 (0.03)
Independence test	0.05 (0.08)	0.05 (0.02)	0.05 (0.11)	0.06 (0.02)
Cond. coverage test	0.05 (0.09)	0.05 (0.03)	0.06 (0.13)	0.06 (0.04)

Table 8: *Size-corrected power for for the $T\text{-}CoVaR$ backtesting method discussed in Section 2.4. The corresponding sizes are reported between parentheses.*

Table 7 reports the size-corrected power for $\delta = 10$. In the next simulation, we assess whether the $T\text{-}CoVaR$ backtesting method discussed in Section 2.4 is able to detect the dynamic misspecification. The same DCC equation, simulation design, and $CoVaR$ prediction method are used. In Table 8 the size-corrected power for the $T\text{-}CoVaR$ test is shown, where $\delta = 10$. Similarly to the distributional misspecification, we do not find any observable power, even for 10,000 out-of-sample predictions.

5 Empirical application

In this section we will discuss an empirical application of our test (the $Q\text{-}CoVaR$ test with x_t^j included in the test regression equations as explained in Section 3). We will use our test to compare the performance of the three estimation methods discussed in Section 2.2. As a proxy for the financial system, we use the Dow Jones US Financials Index. We

will make out-of-sample $Q\text{-}CoVaR$ predictions for the five largest US banks by market capitalisation according to Relbanks (2017), which are also the five largest components of the Dow Jones US Financials Index (Wikinvest 2008). These banks are JPMorgan Chase, Bank of America, Wells Fargo, Citigroup and Goldman Sachs.

We obtain daily log-return series ranging from January 2, 2001 to December 1, 2014, which contain 3500 observations each. The Dow Jones US Financials Index return series is obtained from Yahoo Finance and the return series of the banks are obtained from the Center for Research in Security Prices (CRSP) database.

Similar to Adrian & Brunnermeier (2016), we will use the following market state variables in the quantile regression:

- The change in the three-month Treasury bill rate (change in risk free rate)

- The change in the spread between the ten-year Treasury constant maturity rate and the three-month Treasury bill rate (change in the slope of the yield curve)

- The difference between the three-month LIBOR rate and the three-month Treasury bill rate in the secondary market (TED spread)

- The change in the spread between Moody's Seasoned Baa Corporate Bond Yield and the ten-year Treasury constant maturity rate (change in the credit spread)

- The daily Dow Jones US Financials Index log-return (financial system return)

- The 22-day rolling standard deviation of the daily Dow Jones US Financials Index log-return (financial system volatility)

The daily three-month Treasury bill rate series, the daily ten-year Treasury constant maturity rate series, the daily TED spread series, and the daily Moody's Seasoned Baa Corporate Bond Yield series are obtained from the Federal Reserve Economic Data (FRED) database. Table 9 provides summary statistics for the data.

Daily log-returns	Mean	St.dev	Min	Max
Dow Jones US Financials Index	0.00	0.02	-0.18	0.15
JPMorgan Chase	0.00	0.03	-0.23	0.22
Bank of America	0.00	0.03	-0.34	0.30
Wells Fargo	0.00	0.03	-0.27	0.28
Citigroup	0.00	0.03	-0.49	0.46
Goldman Sachs	0.00	0.02	-0.21	0.23

Daily market state variables	Mean	St.dev	Min	Max
Δ 3-month T-Bill rate	0.00	0.05	-0.81	0.74
Δ Yield curve slope	0.00	0.07	-0.51	0.74
TED spread	0.44	0.47	0.09	4.58
Δ Credit spread	0.00	0.04	-1.20	1.15
Financial system return	0.00	0.02	-0.18	0.15
Financial system volatility	0.01	0.01	0.00	0.08

Table 9: Summary statistics (3500 observations).

For each bank, we use the first 2500 observations to estimate the quantile regression and Asymmetric GARCH-DCC parameters. Based on these parameters, we make one-day ahead out-of-sample predictions for the next 10 days and predict the 95% and 99% Q-$CoVaR$. Then we re-estimate the parameters based on observation 11 - 2510 and predict the $CoVaR$ for day 2511 - 2520. This procedure is repeated until we have 1000 out-of-sample $CoVaR$ predictions for day 2501 - 3500.

Once we have obtained the 95% and 99% out-of-sample $CoVaR$ prediction series for the three estimation methods (the quantile regression, the Asymmetric GARCH-DCC model with Gaussian innovations, and the Asymmetric GARCH-DCC model with Student-t innovations) for each bank, we evaluate the predicted $CoVaR$s at the observed x_t^j, we find the hit series, and we apply our tests. The results for each bank are shown in Table $10-14$.

JPMorgan Chase	Quantile regression	Asymmetric GARCH-DCC	
95% $CoVaR$		Gaussian	Student-t
Linear, $H_0 : \gamma = 0$	3.96 (0.05) **	1.29 (0.26)	0.26 (0.61)
Logistic, $H_0 : \gamma = 0$	3.91 (0.05) **	1.27 (0.26)	0.26 (0.61)
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	4.73 (0.19)	1.50 (0.68)	0.62 (0.89)
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	4.65 (0.20)	1.47 (0.69)	0.60 (0.90)
99% $CoVaR$			
Linear, $H_0 : \gamma = 0$	8.73 (0.00) ***	0.17 (0.68)	3.68 (0.06) *
Logistic, $H_0 : \gamma = 0$	7.33 (0.01) ***	0.17 (0.68)	3.33 (0.07) *
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	9.61 (0.02) **	1.53 (0.67)	6.38 (0.09) *
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	8.51 (0.04) **	2.01 (0.57)	5.28 (0.15)

Table 10: Q-CoVaR LR test statistics with corresponding p-values between parentheses. The number of stars corresponds to the significance level (*:=10%, **:=5%, ***:=1%).

Bank of America	Quantile regression	Asymmetric GARCH-DCC	
95% $CoVaR$		Gaussian	Student-t
Linear, $H_0 : \gamma = 0$	35.57 (0.00) ***	8.10 (0.00) ***	2.93 (0.09) *
Logistic, $H_0 : \gamma = 0$	34.11 (0.00) ***	8.04 (0.00) ***	2.95 (0.09) *
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	43.15 (0.00) ***	8.87 (0.03) **	6.49 (0.09) *
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	41.64 (0.00) ***	8.74 (0.03) **	5.71 (0.13)
99% $CoVaR$			
Linear, $H_0 : \gamma = 0$	21.31 (0.00) ***	15.57 (0.00) ***	2.55 (0.11)
Logistic, $H_0 : \gamma = 0$	17.01 (0.00) ***	13.57 (0.00) ***	2.38 (0.12)
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	23.87 (0.00) ***	23.63 (0.00) ***	2.66 (0.45)
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	20.62 (0.00) ***	27.17 (0.00) ***	2.59 (0.46)

Table 11: Q-CoVaR LR test statistics with corresponding p-values between parentheses. The number of stars corresponds to the significance level (*:=10%, **:=5%, ***:=1%).

Wells Fargo	Quantile regression	Asymmetric GARCH-DCC	
95% $CoVaR$		Gaussian	Student-t
Linear, $H_0 : \gamma = 0$	16.72 (0.00) ***	6.77 (0.01) ***	0.63 (0.43)
Logistic, $H_0 : \gamma = 0$	14.14 (0.00) ***	6.21 (0.01) **	0.63 (0.43)
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	35.19 (0.00) ***	12.28 (0.01) ***	6.14 (0.11)
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	25.67 (0.00) ***	10.58 (0.01) **	5.00 (0.17)
99% $CoVaR$			
Linear, $H_0 : \gamma = 0$	2.11 (0.15)	1.25 (0.26)	0.01 (0.94)
Logistic, $H_0 : \gamma = 0$	1.71 (0.19)	1.14 (0.28)	0.01 (0.94)
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	12.88 (0.00) ***	5.70 (0.13)	20.17 (0.00) ***
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	5.19 (0.16)	4.86 (0.18)	5.42 (0.14)

Table 12: Q-CoVaR LR test statistics with corresponding p-values between parentheses. The number of stars corresponds to the significance level (*:=10%, **:=5%, ***:=1%).

Citigroup	Quantile regression	Asymmetric GARCH-DCC	
95% $CoVaR$		Gaussian	Student-t
Linear, $H_0 : \gamma = 0$	5.00 (0.03) **	0.30 (0.58)	0.09 (0.77)
Logistic, $H_0 : \gamma = 0$	4.55 (0.03) **	0.30 (0.59)	0.09 (0.77)
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	6.19 (0.10)	4.43 (0.22)	2.77 (0.43)
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	5.56 (0.13)	3.77 (0.29)	2.42 (0.49)
99% $CoVaR$			
Linear, $H_0 : \gamma = 0$	23.22 (0.00) ***	6.20 (0.01) **	3.35 (0.07) *
Logistic, $H_0 : \gamma = 0$	13.58 (0.00) ***	5.80 (0.02) **	2.96 (0.09) *
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	23.41 (0.00) ***	6.69 (0.08) *	12.33 (0.01) ***
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	13.85 (0.00) ***	6.47 (0.09) *	7.68 (0.05) *

Table 13: Q-CoVaR LR test statistics with corresponding p-values between parentheses. The number of stars corresponds to the significance level (*:=10%, **:=5%, ***:=1%).

Goldman Sachs	Quantile regression	Asymmetric GARCH-DCC	
95% $CoVaR$		Gaussian	Student-t
Linear, $H_0 : \gamma = 0$	1.17 (0.28)	1.58 (0.21)	6.67 (0.01) ***
Logistic, $H_0 : \gamma = 0$	1.07 (0.30)	1.54 (0.21)	6.33 (0.01) **
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	12.04 (0.01) ***	2.15 (0.54)	7.03 (0.07) *
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	9.33 (0.03) **	2.08 (0.56)	6.67 (0.08) *
99% $CoVaR$			
Linear, $H_0 : \gamma = 0$	4.49 (0.03) **	6.63 (0.01) **	0.22 (0.64)
Logistic, $H_0 : \gamma = 0$	3.58 (0.06) *	5.66 (0.02) **	0.22 (0.64)
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	9.50 (0.02) **	7.12 (0.07) *	0.78 (0.85)
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	6.72 (0.08) *	6.34 (0.10) *	0.77 (0.86)

Table 14: Q-CoVaR LR test statistics with corresponding p-values between parentheses. The number of stars corresponds to the significance level (*:=10%, **:=5%, ***:=1%).

Number of rejections	Quantile regression			Asymmetric GARCH-DCC					
				Gaussian			Student-t		
95% $CoVaR$	10%	5%	1%	10%	5%	1%	10%	5%	1%
Linear, $H_0 : \gamma = 0$	4	4	2	2	2	2	2	1	1
Logistic, $H_0 : \gamma = 0$	4	4	2	2	2	1	2	1	0
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	3	3	3	2	2	1	2	0	0
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	3	3	2	2	2	0	1	0	0
99% $CoVaR$									
Linear, $H_0 : \gamma = 0$	4	4	3	3	3	1	2	0	0
Logistic, $H_0 : \gamma = 0$	4	3	3	3	3	1	2	0	0
Linear, $H_0 : \alpha = (1-q), \beta = \gamma = 0$	5	5	3	3	1	1	3	2	2
Logistic, $H_0 : \alpha = F^{-1}(1-q), \beta = \gamma = 0$	4	3	2	3	1	1	1	0	0

Table 15: Number of rejections (out of the five banks) for the different estimation methods and significance levels.

Table 15 shows the number of rejections (out of the five banks) for the different estimation methods. We find that the Asymmetric GARCH-DCC model with Student-t innovations has the least rejections, and the quantile regression has the most rejections for most of the test specifications. This indicates that the Asymmetric GARCH-DCC model with Student-t innovations is the best in describing the underlying dynamics and that the quantile regression is the worst (out of the three models).
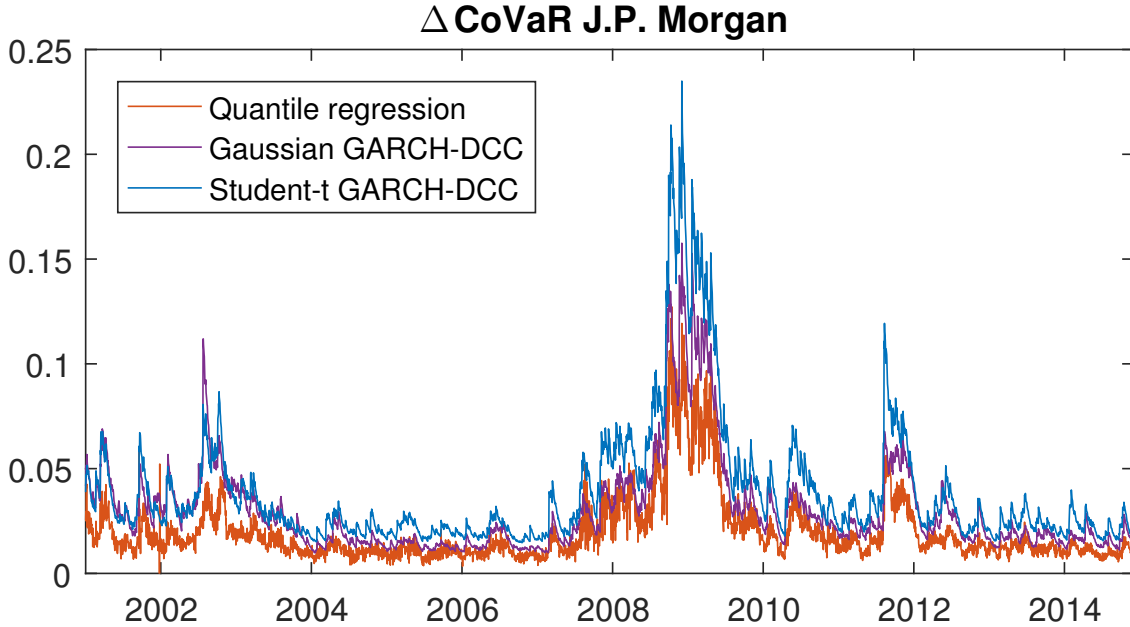

Figure 5: Estimated $\Delta CoVaR$ series for J.P. Morgan.

In Figure 5 the estimated $\Delta CoVaR$ series are shown for J.P. Morgan. The credit crisis of 2008 is clearly visible.

In the quantile regression, the dependency of the $CoVaR$ of i on the return loss of j, captured by the parameter $\hat{\beta}_q$ in (6), does not change over time. In contrast, the Asymmetric GARCH-DCC model allows for time-varying correlation between the returns of i and j. This is likely to be one of the reasons why the Asymmetric GARCH-DCC model outperforms the quantile regression. If, for example, the correlation between X_t^i and X_t^j increases over time, then $\hat{\beta}_q$ will underestimate the dependency and therefore, x_t^j will have significant explanatory power on the occurrence of hits.

Table 15 also shows that for the 95% $CoVaR$, the model using Gaussian innovations and the model using Student-t innovations perform approximately equally well. However,

for the 99% $CoVaR$, the Student-t model clearly outperforms the Gaussian model. This is in line with our expectations since it is assumed that stock return distributions have heavier tails than the Gaussian distribution, in particular in the end of the tails (e.g., Bradley & Taqqu 2001).

6   Conclusion

In this paper a new method for backtesting *Quantile CoVaR* predictions is introduced. Instead of testing the subset of *Tail CoVaR* predictions where the condition is satisfied (the approach used in previous research), we predict the Q-$CoVaR$ as a function of its condition (the return of j) and evaluate the predicted Q-$CoVaR$ at the realisation of this condition (the observed return of j). This allows us to perform tests on the full set of Q-$CoVaR$ predictions.

By definition, the observed return of j should have no additional explanatory power on the hit series of a correctly specified Q-$CoVaR$ evaluated at the at the observed return of j. We include this property in the tests, which allows us to test whether there is misspecification in the dependency structure.

The Monte Carlo simulation shows that our test is useful in a realistic setting when we want to test whether a $CoVaR$ prediction method is specified correctly. It can detect both a distributional and a dynamic misspecification. Our test also outperforms previous $CoVaR$ backtesting methods in our simulation. The test is a little over-sized which is partly caused by the parameter estimation error. In line with Dumitrescu et al. (2012), we find that the linear test and the logistic test have a similar performance.

We predicted Q-$CoVaR$ series for the five largest US banks, where we used the Dow Jones US Financials Index as a proxy for the financial system. We applied our tests to compare the performance of the quantile regression, the Asymmetric GARCH-DCC model with Gaussian innovations, and the Asymmetric GARCH-DCC model with Student-t innovations.

The results indicate that the Asymmetric GARCH-DCC model with Student-t innovations has the best ability to describe the underlying dynamics and that the quantile regression has the worst performance (out of the three estimation methods). A plausible

explanation is that in the quantile regression, the dependency coefficient is fixed over time, while the Asymmetric GARCH-DCC model allows for time-varying correlation.

We also observe that the Student-t innovations are outperforming the Gaussian innovations in particular in the end of the tail (i.e. the 99% $CoVaR$). This may be due to the fatter tails of the Student-t distribution.

Even if one prefers to use T-$CoVaR$ (based on an inequality condition) for a certain application (e.g. because of its monotonicity in the dependence parameter), our test might still be useful. One approach could be to first predict the Q-$CoVaR$ using different models and to apply our tests to find which model has the best performance. Then, using the preferred model, the T-$CoVaR$ can be predicted. This does not necessarily result in the optimal T-$CoVaR$ predictions. However, it may give better finite sample results than testing only a small subset of the T-$CoVaR$ predictions directly.

Bernardi et al. (2013) apply Bayesian updating to the quantile regression coefficients to enable them to describe the time-varying dependency. An interesting approach for future research would be to use their prediction methods and to apply our tests to the Bayesian Q-$CoVaR$ predictions.

In this paper we focussed on the conditioning direction "VaR of the financial system given the return loss of j" ($CoVaR^{system|j}$). However, our test is equally applicable to the opposite definition where the $CoVaR$ is defined as the VaR of an institution given that the financial system is in distress ($CoVaR^{i|system}$). This is also called "*Stressed Value at Risk*". Banks are required to calculate a *Stressed VaR* by the Basel II regulations. The *Stressed VaR* should be based on 10-day ahead predictions of the 99% quantile (Basel Comittee on Banking Supervision 2009b, p. 14). Therefore, it would be an interesting approach for future research to investigate the performance of our test for 10-day ahead predictions.

References

Acharya, V. V., Engle, R. & Richardson, M. (2012), 'Capital shortfall: A new approach to ranking and regulating systemic risks', *American Economic Review* **102(3)**, 59–64.

Acharya, V. V., Pedersen, L. H., Philippon, T. & Richardson, M. (2010), 'Measuring

systemic risk', *people.stern.nyu.edu/lpederse/papers/MeasuringSystemicRisk.pdf* .

Adrian, T. & Brunnermeier, M. K. (2016), 'CoVaR', *American Economic Review* **106(7)**, 1705–1741.

Banulescu, D., Hurlin, C., Leymarie, J. & Scaillet, O. (2016), 'Backtesting marginal expected shortfall and related systemic risk measures'.

Basel Comittee on Banking Supervision (2009b), 'Revisions to the Basel II market risk framework', *Bank for International Settlements* .

Bernardi, M., Gayraud, G. & Petrella, L. (2013), 'Bayesian Inference for CoVaR', *http://arxiv.org/abs/1306.2834* .

Billio, M., Getmansky, M., Lo, A. W. & Pelizzon, L. (2012), 'Econometric measures of connectedness and systemic risk in the finance and insurance sectors', *Journal of Financial Economics* **104(3)**, 535–559.

Bradley, B. O. & Taqqu, M. S. (2001), 'Financial risk and heavy tails', *Boston University* .

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39(4)**, 841–862.

Dumitrescu, E., Hurlin, C. & Pham, V. (2012), 'Backtesting Value at Risk: From dynamic quantile to dynamic binary tests', *Presses universitaires de Grenoble* **33(1)**, 79–112.

Engle, R. & Manganelli, S. (2004), 'CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles', *Journal of Business & Economic Statistics* **22**, 367–381.

Girardi, G. & Ergün, A. T. (2013), 'Systemic risk measurement: Multivariate GARCH estimation of CoVaR', *Journal of Banking & Finance* **37**, 3169–3180.

Gloston, L., Jagannathan, R. & Runkle, D. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *The Journal of Finance* **48**, 1779–1801.

Hansen, B. (1994), 'Autoregressive conditional density estimation', *International Economic Review* **35**, 705–730.

Huang, X., Zhou, H. & Zhu, H. (2012), 'Systemic risk contributions', *Journal of Financial Services Research* **42(1-2)**, 55–83.

Koenker, R. W. & Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46(1)**, 33–50.

Kotz, S. & Nadarajah, S. (2004), 'Multivariate t distributions and their applications', *Cambridge University Press* .

Mainik, G. & Schaanning, E. (2012), 'On dependence consistency of CoVaR and some other systemic risk measures', *Working paper, ETH Zurich.* .

Relbanks (2017), 'Top US banks by market capitalization', *http://www.relbanks.com/top-us-banks/market-cap* .

Wikinvest (2008), 'Dow Jones U.S.General Financial Index (DJUSFN)', *http://www.wikinvest.com/index/Dow_Jones_U.S.General_Financial_Index_(DJUSFN)* .