FUNDAÇÃO GETULIO VARGAS

EPGE

Escola de Pós-Graduação em Economia

Ensaios Econômicos

Escola de Pós-Graduação em Economia da Fundação Getulio Vargas

N° 679 ISSN 0104-8910

Evaluating Value-at-Risk models via Quantile regressions

Wagner Piazza Gaglianone, Oliver Linton, Luiz Renato Regis de Oliveira Lima

Setembro de 2008

URL: http://hdl.handle.net/10438/1718

```
     Piazza Gaglianone, Wagner
        Evaluating Value-at-Risk models via Quantile
regressions/ Wagner Piazza Gaglianone, Oliver Linton,
Luiz Renato Regis de Oliveira Lima - Rio de Janeiro :  FGV,EPGE,
2010
        (Ensaios Econômicos; 679)

        Inclui bibliografia.

                                      CDD-330
```

Evaluating Value-at-Risk models via Quantile Regression

Wagner Piazza Gaglianone[*] Luiz Renato Lima[†] Oliver Linton[‡]

29th August 2008

Abstract

This paper is concerned with evaluating value at risk estimates. It is well known that using only binary variables to do this sacrifices too much information. However, most of the specification tests (also called backtests) available in the literature, such as Christoffersen (1998) and Engle and Maganelli (2004) are based on such variables. In this paper we propose a new backtest that does not rely solely on binary variable. It is shown that the new backtest provides a sufficient condition to assess the performance of a quantile model whereas the existing ones do not. The proposed methodology allows us to identify periods of an increased risk exposure based on a quantile regression model (Koenker & Xiao, 2002). Our theoretical findings are corroborated through a Monte Carlo simulation and an empirical exercise with daily S&P500 time series.

Keywords: Value-at-Risk, Backtesting, Quantile Regression.

JEL Classification: C12, C14, C52, G11.

[*]Research Department, Central Bank of Brazil (e-mail: wagner.gaglianone@bcb.gov.br).

[†]Corresponding author. Department of Economics, University of Illinois and Graduate School of Economics, Getulio Vargas Foundation, Praia de Botafogo 190, s.1104, Rio de Janeiro, RJ 22.253-900, Brazil (e-mail: luizr@fgv.br).

[‡]Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom (e-mail: o.linton@lse.ac.uk).

1 Introduction

Recent financial disasters have emphasized the need for accurate risk measures for financial institutions. Value-at-Risk (VaR) models were developed in response to the financial disasters of the early 90s, and have become a standard measure of market risk, which is increasingly used by financial and non-financial firms as well. In fact, VaR is a statistical risk measure of potential losses, and summarizes in a single number the maximum expected loss over a target horizon, at a particular significance level. Despite several other competing risk measures proposed in the literature, VaR has effectively become a cornerstone of internal risk management systems in financial institutions, following the success of the J.P. Morgan (1996) RiskMetrics system, and nowadays form the basis of the determination of market risk capital, since the 1996 Amendment of the Basel Accord.

Another advantage of VaR is that it can be seen as a coherent risk measure for a large class of distributions, that is, it satisfies the following properties: (i) subadditivity (the risk measure of a portfolio cannot be greater than the sum of the risk measures of the smaller portfolios that comprise it); (ii) homogeneity (the risk measure is proportional to the scale of the portfolio); (iii) monotonicity (if portfolio Y dominates X, in the sense that each payoff of Y is at least as large as the corresponding payoff of X, i.e., $X \leq Y$, then X must be of lesser or equal risk) and; (iv) risk free condition (adding a risk-free instrument to a portfolio decreases the risk by the size of the investment in the risk-free instrument).[1]

A crucial question that arises in this context is how to evaluate the performance of a VaR model? According to Giacomini and Komunjer (2005), when several risk forecasts are available, it is desirable to have formal testing procedures for comparison, which do not necessarily require knowledge of the underlying model, or, if the model is known, do not restrict attention to a specific estimation procedure. The literature has proposed several tests (also known as "backtests"), such as Kupiec (1995), Christoffersen (1998) and Engle and Manganelli (2004), mainly based on binary variables, from which statistical properties are derived and further tested.

Since binary variables sacrifice too much information, the main goal of this paper is to develop a test that is not based solely on such variables. This led us to propose a quantile regression-based backtest, which provides a sufficient condition to assess the performance of a quantile model. We show that standard backtests based on binary variables (Christoffersen (1998) and Engle and Manganelli (2004)) do not provide a sufficient condition to assess the performance of a quantile

[1] Daníelsson et al. (2005) and Ibragimov and Walden (2007) show that VaR is coherent and satisfies subadditivity or the first moment of the investigated variable does not exist. In this sense, they show that VaR is subadditive for the tails of all fat distributions, provided the tails are not super fat (i.e., Cauchy distribution). This way, for a very large class of distributions of continuous random variables, one does not have to worry about subadditivity violations for a VaR risk measure.

model, but (instead) they are all implied by the quantile-based test under the null hypothesis. Our Monte Carlo simulations as well as our empirical application using the S&P500 series corroborate our theoretical findings. Moreover, the proposed test is quite simple to compute and can be carried out using software available for conventional quantile regression, and is applicable even when the VaR does not come from a conditional volatility model.

This study is organized as follows: Section 2 defines Value-at-Risk and the model we use, section 3 presents a quantile regression-based hypothesis test to evaluate VaRs. In Section 4, we briefly describe the existing backtests and establish a sufficient condition to assess a quantile model. Section 5 shows the Monte Carlo simulation comparing the size and power of the competing backtests. Section 6 provides an empirical exercise based on daily S&P500 series, and Section 7 concludes

2 The Model

A Value-at-Risk model reports the maximum loss that can be expected, at a particular significance level, over a given trading horizon. If R_t denotes return of a portfolio at time t, and $\tau^* \in (0,1)$ denotes a (pre-determined) significance level, then the respective VaR (V_t) is implicitly defined by the following expression:

$$\Pr\left[R_t \leq -V_t|\mathcal{F}_{t-1}\right] = \tau^*, \tag{1}$$

where $\mathcal{F}_{t-1}$ is the information set available at time $t-1$.[2] From the above definition, it is clear that V_t is nothing else but the τ^*th conditional quantile of R_t. In other words, V_t is the one-step ahead forecast of the τ^*th quantile of R_t based on the information available up to period $t-1$. Notice that $\Pr\left[R_t \leq -V_t|\mathcal{F}_{t-1}\right] = \tau^*$ is a left-tail VaR and that the reason for the sign convention is to have the VaR be a positive number. As for the right-tail VaR, it can be defined as

$$\Pr\left[R_t \geq V_t|\mathcal{F}_{t-1}\right] = 1-\tau^*. \tag{2}$$

Therefore, if the trader is in a short (long) position, then the risk will come to her whenever price goes up (down), which leads us to compute the VaR at the right (left) tail.

From Equation (1) or (2) it is clear that finding a VaR is essentially the same as finding the conditional quantile of R_t . Following the idea of Christoffersen et al. (2001), one can think of

[2] According to Nankervis et al. (2006), it is usual that VaR is separately computed for the left and right tails of the distribution depending on the position of the risk managers or traders: If a trader is in a short (long) position, then the risk will come to her whenever price goes up (down) which lead us to compute the VaR at the right (left) tail.

generating a VaR measure as the outcome of a quantile regression, treating volatility as a regressor.[3] In this paper, we adapt the idea of Christoffersen et al. (2001) to investigate the accuracy of a given VaR model. In particular, instead of using the conditional volatility as a regressor, we simply use the VaR measure of interest (V_t). We embed this measure in a general class of models for stock returns in which the specification that delivered V_t is nested as a special case. In this way, we can provide a test of the VaR model through a conventional hypothesis test. Specifically, we consider that there is a random coefficient model for R_t, generated in the following way:

$$
\begin{aligned}
R_t &= \alpha_0(U_t) + \alpha_1(U_t)V_t && (3)\\
&= x_t'\beta(U_t), && (4)
\end{aligned}
$$

where V_t is $\mathcal{F}_{t-1}$-measurable in the sense that it is already known at period t, $U_t \sim iid\ U(0,1)$, and $\alpha_i(U_t)$, $i = 0, 1$ are assumed to be comonotonic in U_t, with $\beta(U_t) = [\alpha_0(U_t), \alpha_1(U_t)]'$ and $x_t' = [1, V_t]$.

Given the random coefficient model (3) and the comonotonicity assumption of $\alpha_i(U_t)$, $i = 0, 1$, the τth conditional quantile of R_t can be written as

$$Q_{R_t}(\tau \mid \mathcal{F}_{t-1}) = \alpha_0(\tau) + \alpha_1(\tau)V_t \text{ ; for all } \tau \in (0,1). \tag{5}$$

Now, recall what we really want to test: $\Pr(R_t \leq V_t \mid \mathcal{F}_{t-1}) = \tau^*$, that is, V_t is indeed the $\tau^* th$ conditional quantile of R_t. Therefore, considering the conditional quantile model (5), a natural way to test for the overall performance of a VaR model is to test the null hypothesis

$$H_o : \begin{cases} \alpha_0(\tau^*) = 0 \\ \alpha_1(\tau^*) = 1 \end{cases} \tag{6}$$

against the general alternative.

The null hypothesis can be presented in a classical formulation as $H_o : W\beta(\tau^*) = r$, for the fixed significance level (quantile) $\tau = \tau^*$, where W is a 2×2 identity matrix; $\beta(\tau^*) = [\alpha_0(\tau^*), \alpha_1(\tau^*)]'$ and $r = [0, 1]$. Note that, due to the simplicity of our restrictions, the latter null hypothesis can still be reformulated as $H_o : \theta(\tau^*) = 0$, where $\theta(\tau^*) = [\alpha_0(\tau^*), (\alpha_1(\tau^*) - 1)]'$. Notice that the null hypothesis is interpretable like a Mincer and Zarnowitz (1969) type-regression framework for quantile setup.

[3] In this sense, Engle and Patton (2001) argue that a volatility model is typically used to forecast the absolute magnitude of returns, but it may also be used to predict quantiles.

3 The Test Statistic and Its Null Distribution

Let $\widehat{\theta}(\tau^*)$ be the quantile regression estimator of $\theta(\tau^*)$. The asymptotic distribution of $\widehat{\theta}(\tau^*)$ can be derived following Koenker (2005, p.74), it is normal with covariance matrix that takes the form of a Huber (1967) sandwich:

$$\sqrt{T}(\widehat{\theta}(\tau^*) - \theta(\tau^*)) \xrightarrow{d} N(0, \tau^*(1-\tau^*)H_{\tau^*}^{-1}JH_{\tau^*}^{-1}) = N(0, \Lambda_{\tau^*}), \tag{7}$$

where $J = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t'$ and $H_{\tau^*} = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t'[f_t(Q_{R_t}(\tau^*|x_t)]$ under the quantile regression model $Q_{R_t}(\tau \mid x_t) = x_t'\theta(\tau)$. The term $f_t(Q_{R_t}(\tau^*|x_t)$ represents the conditional density of R_t evaluated at the quantile τ^*. Consistent estimators of J and H_{τ^*} are computed by using, for instance, the techniques in Koenker and Machado (1999). Given that we are able to compute the covariance matrix of the estimated $\widehat{\theta}(\tau)$ coefficients, we can now construct our hypothesis test to verify the performance of the Value-at-Risk model based on quantile regressions (hereafter, VQR test).

Definition 1: Let our test statistic be defined by

$$\zeta_{VQR} = T[\widehat{\theta}(\tau^*)'(\tau^*(1-\tau^*)H_{\tau^*}^{-1}JH_{\tau^*}^{-1})^{-1}\widehat{\theta}(\tau^*)]. \tag{8}$$

In addition, consider the following assumptions:

Assumption 1: Let x_t be measurable with respect to $\mathcal{F}_{t-1}$ and $z_t \equiv \{R_t, x_t\}$ be a strictly stationary process;

Assumption 2: (Density) Let $\{R_t\}$ have conditional (on x_t) distribution functions F_t, with continuous Lebesgue densities f_t uniformly bounded away from 0 and ∞ at the points $Q_{R_t}(\tau \mid x_t) = F_t^{-1}(\tau \mid x_t)$ for all $\tau \in (0,1)$;

Assumption 3: (Design) There exist positive definite matrices J and H_τ, such that $J = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t'$ and $H_\tau = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t'[f_t(Q_{R_t}(\tau \mid x_t))]$ for all $\tau \in (0,1)$;

Assumption 4: $\max_{i=1,\ldots,T} \|x_i\| / \sqrt{T} \xrightarrow{p} 0$.

The asymptotic distribution of the VQR test statistic, under the null hypothesis that $Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1}) = V_t$, is given by Proposition 1 below, which is merely an application of Hendricks and Koenker (1992) and Koenker (2005, Theorem 4.1) for a fixed quantile τ^*.

Proposition 1 *(VQR test) Consider the quantile regression (5). Under the null hypothesis (6), if assumptions (1)-(4) hold, then, the test statistic* ζ_{VQR} *is asymptotically chi-squared distributed with two degrees of freedom.*

Proof. See appendix. ■

Remark 1: Assumption (1) together with comonotonicity of $\alpha_i(U_t)$, $i = 0, 1$ guarantee the monotonic property of the conditional quantiles. We recall the comment of Robinson (2006), in which the author argues that comonotonicity may not be sufficient to ensure monotonic conditional quantiles, in cases where x_t can assume negative values. In our case, $x_t \geq 0$. Assumption (2) relaxes the iid assumption in the sense that allows for non-identical distributions. Bounding the quantile function estimator away from 0 and ∞ is necessary to avoid technical complications. Assumptions (2)-(4) are quite standard in the quantile regression literature (e.g., Koenker and Machado (1999) and Koenker and Xiao (2002)) and familiar throughout the literature on M-estimators for regression models, and are crucial to apply the CLT of Koenker (2005, Theorem 4.1).

Remark 2: Under the null hypothesis it follows that $V_t = Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$, but under the alternative hypothesis the random nature of V_t, captured in our model by the estimated coefficients $\widehat{\theta}(\tau^*) \neq 0$, can be represented by $V_t = Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1}) + \eta_t$, where η_t represents the measurement error of the VaR on estimating the latent variable $Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$. Note that assumptions (1)-(4) are easily satisfied under the null and the alternative hypotheses. In particular, note that assumption (4) under H_1 implies that also η_t is bounded.

Remark 3: Assumptions (1)-(4) do not restrict our methodology to those cases in which V_t is constructed from a conditional volatility model. Indeed, our methodology can be applied to a broad number of situations, such as:

(i) The model used to construct V_t is known. For instance, a risk manager trying to construct a reliable VaR measure. In such a case, it is possible that: (ia) V_t is generated from a conditional volatility model, e.g., $V_t = g(\widehat{\sigma}_t^2)$, where $g(\cdot)$ is some function of the estimated conditional variance $\widehat{\sigma}_t^2$, say from a GARCH model; or (ib) V_t is directly generated, for instance, from a CAViaR model or an ARCH-quantile method (See Koenker & Zhao (1996) and Wu & Xiao (2002) for further details);

(ii) V_t is generated from an unknown model, and the only information available is $\{R_t, V_t\}$. In this case, we are still able to apply Proposition 1 as long as assumptions (1)-(4) hold. This might be the case described in Berkowitz and O'Brien (2002), in which a regulator investigates the VaR measure reported by a supervised financial institution;

(iii) V_t is generated from an unknown model, but besides $\{R_t, V_t\}$ a confidence interval of V_t is also reported. Suppose that a sequence $\{R_t, V_t, \underline{V}_t, \overline{V}_t\}$ is known, in which $\Pr\left[\underline{V}_t < V_t < \overline{V}_t \mid \mathcal{F}_{t-1}\right] = \delta$, where $[\underline{V}_t, \overline{V}_t]$ are respectively lower and upper bounds of V_t, generated (for instance) from a bootstrap procedure, with a confidence level δ (see Christoffersen and Goncalves (2005), Hartz et. al. (2006) and Pascual et al. (2006)). One could use this additional information to investigate the considered VaR by making a connection between the confidence interval of V_t and the previously mentioned measurement error η_t. The details of this route remain an issue to be further explored.

4 Other Backtests and the Sufficient Condition to Asses a Quantile Model

Recall that a correctly specified VaR model at level τ^* is nothing else than the $\tau^* th$ conditional quantile of R_t. The goal of the econometrician is to test the null hypothesis that V_t correctly approximates the conditional quantile for a specified level τ^*. In this section, we show that the existing backtests, which are based on binary variables, do not provide a sufficient condition to assess a quantile model. Such tests are, however, implied by the VQR test under the null hypothesis that $Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1}) = V_t$, meaning that they provide a necessary condition to assess a quantile model. In what follows, we briefly describe some existing backtests and compare each other with the proposed VQR test.

We define a violation sequence by the following indicator function or hit sequence:

$$H_t = \begin{cases} 1 & ; \text{ if } R_t > V_t \\ 0 & ; \text{ if } R_t \leq V_t \end{cases}, \tag{9}$$

and compute the number of violations $N = \sum_{t=1}^{T} H_t$. Based on these definitions, we now present some backtests usually mentioned in the literature to identify misspecified VaR models (see Dowd (2005) and Jorion (2007) for a detailed description):

(i) Kupiec (1995): Some of the earliest proposed VaR backtests is due to Kupiec (1995), which proposes a nonparametric test based on the proportion of exceptions. Assume a sample size of T observations and a number of violations of N. The objective of the test is to know whether $\widehat{p} \equiv N/T$ is statistically equal to $1 - \tau^*$.

$$H_o : p = E(H_t) = 1 - \tau^*. \tag{10}$$

The probability of observing N violations over a sample size of T is driven by a Binomial distribution and the null hypothesis $H_o : p = 1 - \tau^*$ can be verified through a LR test of the form (also known as the unconditional coverage test):

$$LR_{uc} = 2 \ln \left(\frac{\widehat{p}^N (1 - \widehat{p})^{T-N}}{(1 - \tau^*)^N (\tau^*)^{T-N}} \right), \tag{11}$$

which follows (under the null) a chi-squared distribution with one degree of freedom. It also should be mentioned that this test is uniformly most powerful (UMP) test for a given T. However, Kupiec (1995) finds that the power of his test is generally low in finite samples, and the test becomes more powerful only when the number of observations is very large.[4]

(ii) Christoffersen (1998): The unconditional coverage property does not give any information about the temporal dependence of violations, and the Kupiec (1995) test ignores conditioning coverage, since violations could cluster over time, which should also invalidate a VaR model. In this sense, Christoffersen (1998) extends the previous LR statistic to specify that the hit sequence should also be independent over time. The author argues that we should not be able to predict whether the VaR will be violated, since if we could predict it, then, that information could be used to construct a better risk model. The proposed test statistic is based on the mentioned hit sequence H_t, and on T_{ij} that is defined as the number of days in which a state j occurred in one day, while it was at state i the previous day. The test statistic also depends on π_i, which is defined as the probability of observing a violation, conditional on state i the previous day. It is also assumed that the hit sequence follows a first order Markov sequence with transition matrix given by

$$\Pi = \overset{\text{Previous day}}{\begin{bmatrix} 1 - \pi_0 & 1 - \pi_1 \\ \pi_0 & \pi_1 \end{bmatrix}} \begin{matrix} \text{current day (violation)} \\ \text{no violation} \end{matrix} \tag{12}$$

Note that under the null hypothesis of independence, we have that $\pi = \pi_0 = \pi_1 = (T_{01} + T_{11})/T$, and the following LR statistic can, thus, be constructed:

$$LR_{ind.} = 2 \ln \left(\frac{(1 - \pi_0)^{T_{00}} \pi_0^{T_{01}} (1 - \pi_1)^{T_{10}} \pi_1^{T_{11}}}{(1 - \pi)^{(T_{00} + T_{10})} \pi^{(T_{01} + T_{11})}} \right). \tag{13}$$

The joint test, also known as "conditional coverage test", includes unconditional coverage and independence properties, and is simply given by $LR_{cc} = LR_{uc} + LR_{ind.}$; where each component follows a chi-squared distribution with one degree of freedom, and the joint statistic LR_{cc} is asymptotically distributed as $\chi^2_{(2)}$. An interesting feature of this test is that a rejection of the conditional

[4] According to Kupiec (1995), it would require more than six violations during a one-year period (250 trading days) to conclude that the model is misspecified.

coverage may suggest the need for improvements on the VaR model, in order to eliminate the clustering behavior. On the other hand, the proposed test has a restrictive feature, since it only takes into account the autocorrelation of order 1 in the hit sequence.

(iii) Engle and Manganelli (2004): Although the test by Christoffersen (1998) can detect the presence of serial correlation in the sequence of the indicator function H_t, Engle & Manganelli (2004) pointed out that this is only a necessary condition (but not sufficient) to assess the performance of a quantile model. They provided a counter example in which the Christoffersen (1998) test has no power against misspecifications caused by quantile measurement errors (i.e., any noise introduced into the quantile estimates).

Engle & Manganelli (2004) proposed a new test that incorporates a variety of alternatives. Using the previous notation, the random variable $Hit_t = H_t - (1 - \tau^*)$ is defined by the authors, in order to construct the dynamic conditional quantile (DQ) test, which involves the following statistic:

$$DQ_{oos} = (Hit_t' X_t [X_t' X_t]^{-1} X_t' Hit_t)/(T\tau(1-\tau)), \tag{14}$$

where the vector of instruments X_t might include lags of Hit_t, V_t and its lags). This way, Engle & Manganelli (2004) test the null hypothesis that Hit_t and X_t are orthogonal. Under their null hypothesis, the proposed metric to evaluate one-step ahead forecasts follows a χ^2_q, in which $q = rank(X_t)$. Note that the DQ test can be used to evaluate the performance of any type of VaR methodology (and not only the CAViaR family, proposed in their paper).[5]

(iv) Berkowitz et al. (2006): These authors recently proposed a unified approach to a VaR assessment, based on the fact that the unconditional coverage and independence hypotheses are nothing but consequences of the martingale difference hypothesis of the Hit_t process, i.e., $E(Hit_t \mid \mathcal{F}_{t-1}) = 0$. Based on a Ljung-Box type-test, they consider the nullity of the first K autocorrelations of the Hit_t process, instead of only considering the autocorrelation of order one, as done by Christoffersen (1998).

Several other related procedures are also well documented in the literature, such as the nonparametric test of Crnkovic and Drachman (1997),the duration approach of Christoffersen and Pelletier

[5] The Conditional Autoregressive Value-at-Risk by Regression Quantiles (CAViaR) model is proposed by the authors, which define $V_t(\tau)$ as the solution to $\Pr[R_t < -V_t(\tau) \mid \mathcal{F}_{t-1}) = \tau$; and describe the (generic) specification: $V_t(\tau) = \beta_0 + \sum_{i=1}^{q} \beta_i V_{t-i}(\tau) + \sum_{j=1}^{r} \gamma_j x_{t-j}$; where $[\beta_i; \gamma_j]$ are unknown parameters to be estimated and x_t is a generic vector of time t observable variables. The CAViaR approach directly models the return quantile rather than specifying a complete data generating process. The authors define various dynamic models for V_t itself, including the adaptative model: $V_t(\tau) = V_{t-1}(\tau) + \beta\left[\mathbf{1}(R_{t-1} \leq -V_{t-1}) - \tau\right]$; symmetric absolute value: $V_t(\tau) = \beta_0 + \beta_1 V_{t-1}(\tau) + \beta_2 |R_{t-1}|$; asymmetric slope: $V_t(\tau) = \beta_0 + \beta_1 V_{t-1}(\tau) + \beta_2 R^+_{t-1} + \beta_3 R^-_{t-1}$; in which $\{R_t\}$ is the return series.

(2004), and the encompassing test of Giacomini and Komunjer (2005). However, as shown in several simulation exercises (e.g., Kupiec (1995), Pritsker (2001) and Campbell (2005)), backtests generally have low power and are, thus, prone to misclassifying inaccurate VaR estimates as "acceptably accurate".

4.1 A Sufficient Condition to Assess Quantile Models

In this section, we compare our setup with some standard backtests. Since the main concern of the backtest literature is to evaluate the VaR accuracy, we pose a relevant question in this context: What do we really want to test? Given that a VaR measure is implicitly defined by Property 1 (hereafter, P1, reproduced below), the core issue of a backtest should be to verify whether it (in fact) is true. As we next show, the quantile regression framework provides a natural way to investigate the performance of a VaR model, and the proposed VQR test consists on a sufficient condition for P1. We also show that the existing tests are all implied by the VQR test under the null hypothesis., meaning that they are only necessary conditions for P1.

Property 1: $\Pr\left[R_t \leq V_t | \mathcal{F}_{t-1}\right] = \tau^*$.

Statement 1 (S1): Null hypothesis of the VQR test: $V_t = Q_{R_t}\left(\tau^* \mid \mathcal{F}_{t-1}\right)$.

Statement 2 (S2): Berkowitz et al. (2006): $E(Hit_t \mid \mathcal{F}_{t-1}) = 0$, in which we assume that the conditional expectation is represented by a linear model.

Statement 3 (S3): Engle and Manganelli (2004): $E[X_t' Hit_t] = 0$, in which $X_t \in \mathcal{F}_{t-1}$.

Statement 4 (S4): Christoffersen (1998): $E[Hit_t' Hit_{t-1}] = 0$.

Statement 5 (S5): Kupiec (1995): $E(Hit_t) = 0$.

Proposition 2 *Consider Property 1 and statements S1-S5. If assumptions (1)–(4) hold for the regression (3), then, it follows that:*

(i) $P1 \Leftrightarrow S1$

(ii) $\{S2, S3, S4, S5\} \nRightarrow S1$

(iii) $S1 \Rightarrow \{S2, S3, S4, S5\}$

Proof. See appendix. ■

Proposition 2 shows that the VQR test breaks down the paradigm of the binary variables (hit sequence) in the backtest literature. Indeed, if $E(Hit_t \mid \mathcal{F}_{t-1}) = 0$ is linear, then the tests based on binary variables (such as the one proposed by Engle and Manganelli (2004)) do not provide a sufficient condition to assess the performance of a quantile model. We make such a comparison by using the optimality (or subgradient) condition of the quantile regression problem. Moreover, Proposition 2 also show that the existing Backtests (such as Kupiec (1995), Christoffersen (1998) and Engle and Manganelli (2004)) are implied by the VQR test. Intuitively, this happens because if $V_t = Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$ then V_t will provide a filter to transform a (possibly) serially correlated and heteroskedastic time series into a serially independent sequence of indicator functions. A small Monte Carlo simulation is conducted in Section 4 to verify these theoretical findings as well as to compare the VQR test with the existing ones in terms of power and size.

In next section, we provide an additional framework that might be useful for those interested in improving the performance of a VaR model as well as choosing the best model among competing measures.

5 Local analysis of VaR models: identifying periods of risk exposure

The conditional coverage literature is concerned with the adequacy of the VaR model, in respect to the existence of clustered violations. In this section, we will take an alternative route to analyze the conditional behavior of a VaR measure. According to Engle and Manganelli (2004), a good Value-at-Risk model should produce a sequence of unbiased and uncorrelated hits, and any noise introduced into the Value-at-Risk measure would change the conditional probability of a hit vis-à-vis the related VaR. Given that our study is entirely based on a quantile framework, besides the VQR test, we are also able to identify the exact periods in which the VaR produces an increased risk exposure in respect to its nominal level τ^*, which is quite a novelty in the literature. To do so, let us first introduce some notation:

Definition 2: $W_t \equiv \{\widetilde{\tau} \in [0,1] \mid V_t = \widehat{Q}_{R_t}(\widetilde{\tau} \mid \mathcal{F}_{t-1})\}$, representing the "fitted quantile" of the VaR measure at period t given the regression model (3).

In other words, W_t is obtained by comparing V_t with a full range of estimated conditional quantiles evaluated at $\tau \in [0,1]$. Note that W_t enables us to conduct a local analysis, whereas the proposed VQR test is designed for a global evaluation based on the whole sample. It is worth mentioning that, based on our assumptions, $Q_{R_t}(\tau \mid \mathcal{F}_{t-1})$ is monotone increasing in τ, and W_t by definition is equivalent to a quantile level, i.e., $W_t > \tau^* \Leftrightarrow Q_{R_t}(W_t \mid \mathcal{F}_{t-1}) > Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$. Also note that if V_t is a correctly specified VaR model, then W_t should be as close as possible to τ^* for all t. However, if V_t is misspecified, then it will vague away from τ^*, suggesting that V_t does not correctly approximate the $\tau^* th$ conditional quantile.

Notice that, due to the quantile regression setup, one does not need to know the true returns distribution in order to construct W_t. In practical terms, based on the series R_t, V_t one can estimate the conditional quantile functions $\widehat{Q}_{R_t}(\tau \mid \mathcal{F}_{t-1})$ for a (discrete) grid of quantiles $\tau \in [0,1]$. Then, one can construct W_t by simply comparing (in each time period t) the VaR series V_t with the set of estimated conditional quantile functions $\widehat{Q}_{R_t}(\tau \mid \mathcal{F}_{t-1})$ across all quantiles τ inside the adopted grid.

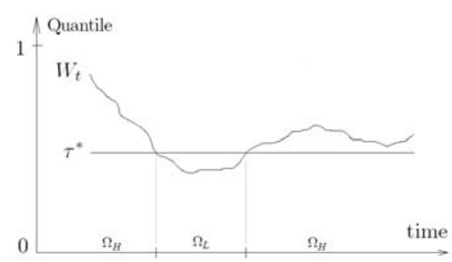
Now consider the set of all observations $\Omega = 1, \ldots, T$, in which T is the sample size, and define the following partitions of Ω:

Definition 3: $\Omega_H \equiv \{t \in \Omega \mid W_t \geqslant \tau^*\}$, representing the periods in which the VaR belongs to a quantile above the level of interest τ^* (indicating a conservative model);

Definition 4: $\Omega_L \equiv \{t \in \Omega \mid W_t < \tau^*\}$, representing the periods in which the VaR is below the nominal τ^* level and, thus, underestimate the risk in comparison to τ^*.

Since we partitioned the set of periods into two categories, i.e. $\Omega = \Omega_H + \Omega_L$, we can now properly identify the so-called periods of "risk exposure" Ω_L. Let us summarize the previous concepts through the following schematic graph:

Figure 1 - Periods of risk exposure



It should be mentioned that a VaR model that exhibits a good performance in the VQR test (i.e., in which H_o is not rejected) is expected to exhibit W_t as close as possible to τ^*, fluctuating around τ^*, in which periods of W_t below τ^* are balanced by periods above this threshold. On the other hand, a VaR model rejected by the VQR test should present a W_t series detached from τ^*, revealing the periods in which the model is conservative or underestimate risk. This additional information can be extremely useful to improve the performance of the underlying Value-at-Risk model, since the periods of risk exposure are now easily revealed.

Another important issue regarding model analysis is the choice of competing VaRs. Instead of only checking the performance of a single model, one might be interested in ranking several VaR measures (see Giacomini and Komunjer, 2005). Although this is not the main objective of this paper, we outline a simple nonparametric procedure, inspired by Lopez (1999), in which a loss function is used to measure the "conditional coverage distance" of a VaR from its nominal benchmark τ^*. According to the author, a numerical score could reflect regulatory concerns and provide a measure of relative performance to compare competing VaR models across time and institutions.

The generic loss function suggested by Lopez (1999) is given by $C(\{R_t\},\{V_t\}) = \sum_{t=1}^{T} C_t(R_t, V_t)$, where $C_t(.) = \begin{cases} f(R_t, V_t) & ; \text{ if } R_t > V_t \\ g(R_t, V_t) & ; \text{ if } R_t \leq V_t \end{cases}$. Accurate VaR estimates are expected to generate lower numerical scores. Once the f and g functions are defined, the loss function can be constructed and used to evaluate the performance of a set of VaR models. Among several different specifications, Lopez (1999) suggests $C_t(.) = \begin{cases} 1 + (R_t - V_t)^2; \text{ if } R_t > V_t \\ 0 \;\; ; \text{ if } R_t \leq V_t \end{cases}$, in which a quadratic term is used to penalize a VaR model based on the magnitude of the violation (i.e., $R_t > V_t$ in our setup). An interesting advantage of this specification is to consider the magnitude of violations, since the magnitude as well as the number of violations is a serious matter of concern to regulators and risk managers. In addition, loss functions may be more suited to discriminate among competing VaR models than deciding for the accuracy of a single VaR model.

In this paper, we adapt the previous approach to our setup in order to rank competing VaRs. Let's first define C_t as the "empirical distance" of V_t in respect to the conditional quantile function $Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$, based on the W_t series:

$$C_t \equiv |W_t - \tau^*| . \tag{15}$$

Now, define the loss function $L(V_t)$ that summarizes the distances C_t, and assigns weights $[\gamma_1, \gamma_2]$ to each distance, according to the indicator function $I_t = \begin{cases} 1 \;\; ; \text{ if } W_t > \tau^* \\ 0 \;\; ; \text{ if } W_t \leq \tau^* \end{cases}$:

$$L(V_t) \equiv \frac{1}{T} \sum_{t=1}^{T} C_t * (\gamma_1 I_t + \gamma_2 (1 - I_t)). \tag{16}$$

This *ad-hoc* loss function is easy to compute and non-parametrically provides us an empirical way to rank a set of VaRs. Thus, the choice among $i = 1, \ldots, n$ competing models could be based on the minimization of the proposed loss function: $\underset{i=1,\ldots,n}{\arg\min}[L(V_t^i)]$. In addition, note that by setting

$\gamma_1 < \gamma_2$ the econometrician could penalize more the periods of risk exposure than those periods in which the "fitted quantile" W_t is above τ^*. Therefore, an asymmetric evaluation is allowed by this framework, in which the choice of weights $[\gamma_1, \gamma_2]$ could also be driven by the risk aversion degree of the regulator (or the risk manager). Despite its simplicity, the descriptive statistic $L(.)$ might be useful to illustrate model comparison in our empirical exercise. It is worth mentioning that the backtest literature is mainly focused just on the signal I_t (as detailed in the previous section), whereas in this paper we try to go a step further by also considering the valuable information contained in the magnitude of the VaR violations.

6 Monte Carlo simulation

In this section we conduct a simulation experiment to investigate the finite sample properties of the VQR test. In particular we are interested in showing under which conditions our theoretical findings (that is, some of the results stated in Proposition 2) are observed in finite samples. To save space, we consider, besides the VQR test, only the conditional coverage test of Christoffersen (1998) and the out-of-sample DQ test of Engle and Manganelli (2004). In the DQ test we considered the instruments $X_t = [1\ V_t]'$.[6]

To investigate the size properties of tests, we assume that the VaR model estimated by the econometrician coincides with VaR model obtained using the DGP. The DGP corresponds to a GARCH(1,1) model, that is, $R_t = \sigma_t \varepsilon_t$, where $\sigma_t^2 = 0.02 + 0.06 y_{t-1}^2 + 0.94 \sigma_{t-1}^2$ and $\varepsilon_t \sim N(0,1)$. Notice that this model is simply the well known RiskMetrics model developed by JPMorgan.

We investigate (size-adjusted) power of tests by considering (local) misspecification of the model estimated by the econometrician. In other words, we change the DGP slightly, but the econometrician keeps on estimating VaR using the (misspecified) RiskMetrics. To do so, we considers a sequence of DGPs based on a GARCH(1,1) model, with coefficients: $c = 0.02$, $\alpha = 0.06 - \phi/20$; $\beta = 0.94 - \phi/2$, and a Gamma (a, b) distribution with parameters $a = 200e^{-5\phi}; b = 5$. We control the "degree of misspecification" through the parameter $\phi \in [0, 1]$, which ranges from 0 to 1 with increments of 0.1. Then, in order to replicate a realistic situation, a VaR is estimated for each DGP via a RiskMetrics model with normal distribution. Note that when $\phi = 0$ we are under the null hypothesis, but as long as we increase ϕ the alternative hypothesis is simulated. Recall that a Gamma (a, b) distribution tends to a normal distribution as long as $a \to \infty$.

For each DGP, we generate $T + 2,000$ observations, discarding the first $2,000$ observations. Then, a total amount of $i = 5,000$ replications of the $\{R_t\}_{t=1}^T$ process are considered for each

[6] Simulations including the Kupiec (1995) test are available under request.

DGP. We follow here the same computational strategy of Lima & Neri (2006), in which a hybrid solution using R and Ox environments is adopted, since the proposed simulation is extremely computational intensive. Ox is much faster than R in large computations, and also makes use of the package G@RCH 4.2 (see Laurent & Peters, 2006), which easily allows us to generate GARCH specifications. On the other hand, the R language is more interactive and user-friendly than Ox and the VQR test must in fact be conducted in R, since its package for quantile regressions (quantreg) is more complete and updated than the Ox package. Therefore, we proceed as follows: an Ox code initially generates the time series R_t for each DGP, and save all the replications in the hard disk. Next, an R code computes the four considered backtests for all replications and saves the final results in a text file.

Finally, we consider samples of size $T = \{250, 500, 1000\}$ and quantile level $\tau^* = 99\%$. The choice of $\tau^* = 99\%$ was made to evaluate the performance of the VQR at the most extreme quantile. Moreover, the 99% significance level is largerly used im empirical applications for traders in short position. Table 1 presents the empirical size of 5% tests.

Table 1 - Size investigation (T=250)

	$\tau^* = 99\%$
ζ_{VQR}	0.1801
$\zeta_{\text{Christ.}}$	0.0215
ζ_{DQ}	0.0806
T=500	
ζ_{VQR}	0.1114
$\zeta_{\text{Christ.}}$	0.0325
ζ_{DQ}	0.0762
T=1000	
ζ_{VQR}	0.0950
$\zeta_{\text{Christ.}}$	0.0374
ζ_{DQ}	0.0801

Note: The values above represent the percentage of p-values below the nominal level of significance $\alpha = 5\%$.

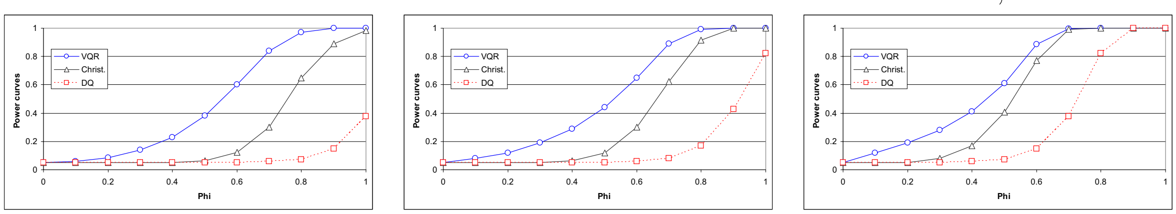
Firstly, notice that for samples as small as $T = 250$, the VQR and DQ backtests tend to over-reject the null hypothesis, whereas the Christoffersen (1998) backtest tends to under-reject it. The main reason is that, for $T = 250$ only a small number of observations is expected at the extreme quantiles, $\tau^* = 99\%$, which is a serious problem for all backtests, and also affect the QR estimation.

The increase of the sample size T can give us some flavor of the asymptotic behavior in the size investigation. Notice that an increase of the sample size produces the following effect in our simulation: As long as T increases, the estimation of the extreme quantiles becomes more precise, leading to a better estimation of the quantile density function evaluated at those quantiles. As a result, the empirical size of the VQR test approaches its nominal size (5%) as T goes to infinity.[7] In general, the empirical size of the three tests tends to the nominal size of 5%. This confirms our finding that, under the null hypothesis, the tests based on binary variables are implied by the VQR test and, therefore, are supposed to have correct size under the null hypothesis as T goes to infinity

We now turn to (size-corrected) power analysis. Backtesting involves balancing two types of errors and dealing with the tradeoff between rejecting a correct model versus accepting a misspecified one. According to Christoffersen (2003, p.186) it is very costly if a test fails to reject an incorrect model. Therefore, if one is more concerned with discarding a poorVaR model, then it is important to use a backtest that delivers high power against misspecified models. Figure 2 exhibits the size-corrected power curves for $\tau^* = 99\%$.

Figure 2 - Size-corrected Power Curves ($\tau^* = 99\%$)

$T = 250$ $\qquad$ $T = 500$ $\qquad$ $T = 1,000$



Notes: Nominal level of significance is $\alpha = 5\%$.

Recall that model misspecification is introduced through the parameter $\phi \in [0, 1]$, which ranges from 0 to 1 with increments of 0.1. If we consider small values of ϕ, say $\phi = 0.1$, 0.2 and 0.3, we will learn from Figure 2 that, even when the sample is as large as $T = 1000$, the tests based on binary variables such as Christoffersen (1998) and DQ_{oos} have almost no power against the misspecified RiskMetrics model. On the other hand, the VQR test delivers considerable power against misspecified models. This result illustrates the main finding of this paper which says that tests based on binary variable do not provide a sufficient condition to assess a quantile model. Indeed, for $\phi \neq 0$ the RiskMetrics model V_t is misspecified, and therefore the vector $\beta_0(\tau^*) = (0, 1)'$

[7] Despite the relatively large sample size when $T = 1,000$, note that for $\tau^* = 99\%$ one should expect only 10 observations of R_t above the VaR measure, which could seriously influence the performance of any backtest. However, the small sample size should not be viewed anymore as a restriction, given that nowadays it is common to deal with intra-day data, and even for daily frequency, a sample size of $T = 1,000$ only requires four years of database.

is not an optimal solution to the quantile regression problem, leading the VQR test to deliver power against such a misspecified model. The same does not happen to tests based on binary variables because they are formulated in terms orthogonality conditions.

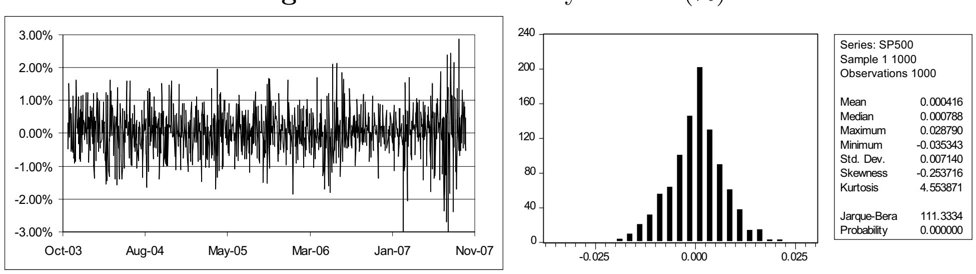
In sum, these simulation results indicate that the test based on the quantile regression (VQR test) has empirical size that converges to the nominal size as sample size increases and the same happen to other backtests based on binary variables. As far as power is concerned, the results are more favorable to the VQR test in the sense that it delivers more power than other backtests for a variety of misspecified models while other tests may fail to reject the null hypothesis in many cases where the orthogonality condition is satisfied by a misspecified model.

7 Empirical exercise

7.1 Data

In this section, we explore the empirical relevance of the theoretical results previously derived. This is done by evaluating and comparing five different VaR models, based on the VQR test and other competing procedures commonly presented in the backtest literature. To do so, we investigate the daily returns of S&P500 over the last 4 years, with an amount of $T = 1000$ observations, depicted in the following figure:

Figure 3 - S&P500 daily returns (%)



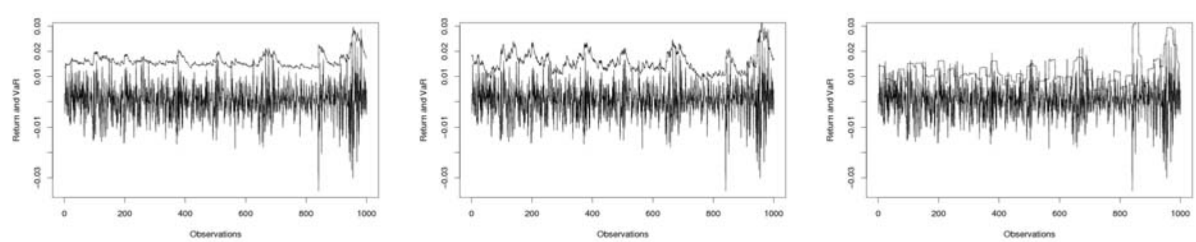
Notes: a) The sample covers the period from 23/10/2003 until 12/10/2007;

b) Source: Yahoo!Finance.

Note from the graph and the summary statistics the presence of common stylized facts about financial data (e.g., volatility clustering; mean reverting; skewed distribution; excess kurtosis ; and non-normality, see Engle and Patton (2001) for further details). The five Value-at-Risk models adopted in our evaluation procedure are the following: Rolling Window (1 and 3 months), GARCH (1,1), RiskMetrics (hereafter, RM) and CAViaR. In the first two approaches, the last 30 (and 90) days of data are used to calculate the conditional variance (σ_t^2), based on a moving average of past observations. The third and fourth approaches are nothing else than conditional volatility models based on a GARCH (1,1) model, since RiskMetrics is just an integrated GARCH(1,1) model with the autoregressive parameter set to 0.94.[8] We adopted these five models merely for illustrative purposes, since our focus here is the backtests comparison. In addition, recall that we are testing the null hypothesis that the model V_t correctly approximates the true $\tau^{*}th$ conditional quantile of the return series R_t. We are not testing the null hypothesis that V_t correctly approximates the entire distribution of R_t. Therefore, it is possible that for different τ's (target probabilities) the model V_t might do well at a target probability, but otherwise poorly (see Kuester et al., 2005). The respective VaR measures of these first four volatility models are, then, constructed by a linear function of σ_t (assuming normality). For instance, the Value-at-Risk for $\tau^{*} = 99\%$ is given by $V_t = 2.33{*}\sigma_t$. Regarding the CAViaR model, we considered the asymmetric slope model: $V_t(\tau) = \beta_0 + \beta_1 V_{t-1}(\tau) + \beta_2 R_{t-1}^{+} + \beta_3 R_{t-1}^{-}$; in which $\{R_t\}$ is the return series.

Practice generally shows that these various models lead to widely different VaR time series for the same considered return series, leading us to the crucial issue of model comparison and hypothesis testing. The Rolling Window method (also called Historical Simulation, hereafter, HS) has serious drawbacks and is expected to generate poor VaR measures, since it ignores the dynamic ordering of observations, and volatility measures look like "plateaus", due to the so-called "ghost effect". On the other hand, as shown by Christoffersen et al. (2001), the GARCH-VaR model is the only VaR measure, among several alternatives considered by the authors, which passes the Christoffersen's (1998) conditional coverage test. The JP Morgan's RiskMetrics-VaR model is chosen as a benchmark model commonly used by practitioners. Finally, Engle and Manganelli (2004) show that the "asymmetric absolute value" and "asymmetric slope" models are the best CAViaR specifications for the S&P500 data.

Figure 4 - S&P500 daily returns (R_t) and VaR (99%) V_t
GARCH(1,1), CAViaR and Rolling Window (1 month)



[8] An amount of 500 observations is used to estimate the GARCH models in a moving window scheme.

7.2 Results

Based on the quantile regression framework, we are now able to compute the VQR test for the five considered VaRs. The main results are summarized in the following table:

Table 2 - Results of the VQR test ($\tau^* = 99\%$)

$H_o : V_t = Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$

	CAViaR	GARCH	RM	HS1m	HS3m
$\widehat{\alpha}_0(\tau^*)$	0.00205 (0.00232)	-0.00594 (0.00976)	-0.00267 (0.00202)	0.00677 (0.00252)	0.00298 (0.00255)
$\widehat{\alpha}_1(\tau^*)$	0.83323 (0.19955)	1.39269 (0.63097)	1.16941 (0.08170)	0.80103 (0.22783)	0.91397 (0.19632)
ζ_{VQR}	0.81351	0.39766	15.23240	31.94366	11.87233
p-value	0.66581	0.81968	0.00049	1.15e-07	0.00264

Note: a) Standard error in parentheses.

As already expected, the rolling window models are all rejected, whereas the GARCH(1,1) and CAViaR models do not fail the VQR test, which is a result perfectly in line with the literature (e.g., Christoffersen et al. (2001) and Giacomini and Komunjer (2005)). In addition, the RiskMetrics-VaR is rejected for $\tau^* = 99\%$. It should be mentioned that violations that are clustered in time are more likely to occur in a VaR model obtained from a rolling window procedure, which increases the number of scenarios for our backtest evaluation. We now present the results of other backtests often used in the literature for VaR evaluation:

Table 3 - Backtests comparison ($\tau^* = 99\%$)

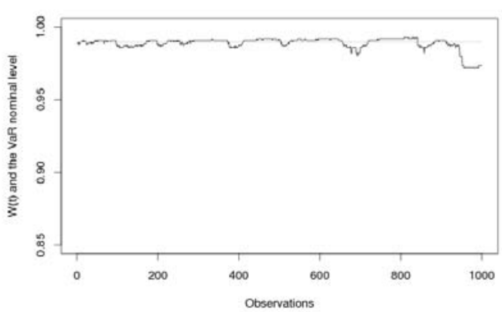
	CAViaR	GARCH	RM	HS1m	HS3m
% of hits	0.9	1.2	1.1	5.6	2.5
ζ_{Kupiec}	0.74884	0.53556	0.75198	0.00000 (**)	0.00000 (**)
$\zeta_{\text{Christ.}}$	0.87539	0.71333	0.84163	0.00000 (**)	0.00017 (**)
ζ_{DQ}	0.97173	0.94656	0.13848	0.00000 (**)	0.00000 (**)
ζ_{VQR}	0.66581	0.81968	0.00049 (**)	0.00000 (**)	0.00264 (**)

Notes: P-values are shown in the ζ's rows; (**) means rejection at 1%.

Note that the GARCH(1,1)-VaR model provides a quite good VaR measure, according to all considered backtests, despite its simplicity and the assumption of normality. Overall, the results are similar to those obtained from the VQR test, excepting the RiskMetrics model. The results of Table 3 indicate that RiskMetrics is only rejected by the VQR test, which is compatible with the previous results of the Monte Carlo simulation (see Figure 2, $T = 1,000$). In other words, our methodology is able to reject more misspecified VaR models in comparison to other backtests, which might be a major advantage of our approach. In fact, recall that the VQR test has more power against a variety of misspecified models, as described in the Monte Carlo simulations.

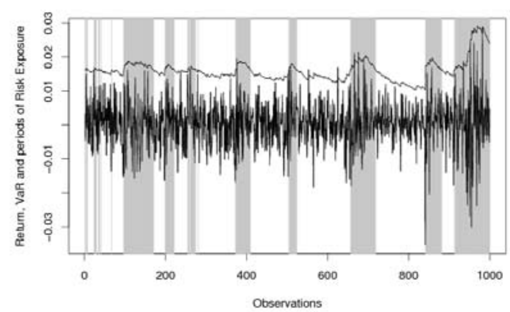
Recall that a secondary result of our methodology is the introduction of the time series W_t, described in section 3.3. It is useful to reveal periods of risk exposure associated with VaR model. Recall that whenever W_t is below the benchmark level τ^*, the VaR model increases the risk exposure by underestimating the related conditional quantile of returns, since (ideally) W_t should be as close as possible to τ^*. To illustrate the methodology, the estimated W_t series as well as the periods of risk exposure for the RiskMetrics-VaR(99%) model are depicted in Figures 5 and 6, where the gray bars indicate periods in which $W_t < \tau^*$.

Figure 5 - W_t (RiskMetrics-VaR 99%)



Notes: a) The black series is the computed W_t;
$W_t \equiv \{\widetilde{\tau} \in [0,1] \mid V_t = \widehat{Q}_{R_t}(\widetilde{\tau} \mid \mathcal{F}_{t-1})\}$

Figure 6 - R_t and V_t (RiskMetrics-VaR 99%)



Note: Gray bars indicate $W_t < \tau^*$;

In other words, gray bars suggest periods in which the VaR measure underestimates the risk exposure. Since the RiskMetrics-VaR(99%) model is rejected by the VQR test, the risk exposure periods could be very useful for risk managers interested in improving the accuracy of the underlying model. For instance, a visual inspection on figure 6 indicates that the RiskMetrics model usually underestimates (gray bars) the degree of risk for high volatility periods. Therefore, we are able to unmask the bad performance of the RiskMetrics model in our empirical exercise based on a local behavior analysis, which brings some additional (and important) information to the backtest investigation by exposing some "reasons of rejection". Note that this local behavior investigation could only be conducted through our proposed quantile regression methodology, which we believe to be a novelty in the backtest literature.

Another relevant issue regarding VaR evaluation is the comparison among several competing models. Although it is not the main objective of this paper, we outline (for the sake of completion of our empirical exercise) a simple nonparametric decision rule for model selection and apply it to our empirical exercise (see Giacomini and Komunjer (2005) for a detailed discussion of model comparison). We are, thus, concerned with relative evaluation, which involves comparing the performance of competing models and choosing the one that performs the best according to our suggested criterion of section 2.3. The main results are next summarized:

Table 4 - Loss Function $L(V_t)$ for $\tau^* = 99\%$

CAViaR	GARCH	RM	HS1m	HS3m
0.00320	0.00497	0.00560	0.06843	0.02099

Notes: a) Recall that $C_t \equiv |W_t - \tau^*|$; $I_t = \begin{cases} 1 & ; \text{ if } W_t > \tau^* \\ 0 & ; \text{ if } W_t \leq \tau^* \end{cases}$;

and $L(V_t) \equiv \frac{1}{T}\sum_{t=1}^{T} C_t * (\gamma_1 I_t + \gamma_2(1 - I_t))$;

b) We adopted $\gamma_1 = 1.0$ and $\gamma_2 = 1.5$.

Based on this procedure, one should choose the model in which W_t best tracks the desired τ^* level, according to the asymmetric weights γ_1 and γ_2. In our exercise, the CAViaR model exhibits the best performance (i.e., lowest value of $L(V_t)$), which is a natural result, given that it is exactly designed to produce Value-at-Risk measures, whereas the other discussed VaRs are only obtained from conditional volatility models with specific distributional assumptions.

8 Conclusions

Backtesting could prove very helpful in assessing Value-at-Risk models and is nowadays a key component for both regulators and risk managers. Since the first procedures suggested by Kupiec (1995) and Christoffersen (1998), a lot of research has been done in the search for adequate methodologies to assess and help improve the performance of VaRs, which (preferable) do not require the knowledge of the underlying model.

As noted by the Basle Committee (1996), the magnitude as well as the number of exceptions of a VaR model is a matter of concern. The so-called "conditional coverage" tests indirectly investigate the VaR accuracy, based on a "filtering" of a serially correlated and heteroskedastic time series (R_t) into a serially independent sequence of indicator functions (hit sequence H_t). Thus, the standard procedure in the literature is to verify whether the hit sequence is iid. However, an important piece of information might be lost in that process: not only is the sequence of past hits that matters, but also the magnitude of H_t is of vital importance, since the conditional distribution of returns is dynamically updated. This issue is also discussed by Campbell (2005), which states that the reported quantile provides a quantitative and continuous measure of the magnitude of realized profits and losses, while the hit indicator only signals whether a particular threshold was exceeded. In this sense, the author suggests that quantile tests can provide additional power to detect an inaccurate risk model.

That is exactly the objective of this paper: to provide a VaR-backtest fully based on a quantile regression framework. Our proposed methodology enables us to: (i) formally conduct a Wald-type hypothesis test to evaluate the performance of VaRs; and (ii) identify periods of an increased risk exposure. We illustrate the usefulness of our setup through an empirical exercise with daily S&P500 returns, in which we construct five competing VaR models and evaluate them through our proposed backtest (and through other standard backtests). In addition, we also suggest a simple nonparametric procedure to rank the competing models.

Since a Value-at-Risk model is implicitly defined as a conditional quantile function, the quantile approach provides a natural environment to study and investigate VaRs. One of the advantages of our approach is the increased power of the suggested quantile-regression backtest in comparison to some established backtests in the literature, as suggested by a Monte Carlo simulation. Perhaps most importantly, our backtest is applicable under a wide variety of structures, since it does not depend on the underlying VaR model, covering either cases where the VaR comes from a conditional volatility model, or it is directly constructed (e.g., CAViaR or ARCH-quantile methods) without relying on a conditional volatility model. We also introduce a main innovation: based on the quantile estimation, one can also identify periods in which the VaR model might increase the risk exposure, which is a key issue to improve the risk model, and probably a novelty in the literature. A final advantage is that our approach can easily be computed through standard quantile regression softwares.

Although the proposed methodology has several appealing properties, it should be viewed as complementary rather than competing with the existing approaches, due to the limitations of the quantile regression technique discussed along this paper. Furthermore, several important topics remain for future research, such as: (i) time aggregation: how to compute and properly evaluate a 10-day regulatory VaR? Risk models constructed through QAR (Quantile Autoregressive) technique can be quite promising due to the possibility of recursively generation of multiperiod density forecast (see Koenker and Xiao (2006a,b)); (ii) Our randomness approach of VaR also deserves an extended treatment and leaves room for weaker conditions; (iii) multivariate VaR: although the extension of the analysis for the multivariate quantile regression is not straightforward, several proposals have already been suggested in the literature (see Chaudhuri (1996) and Laine (2001)); (iv) inclusion of other variables to increase the power of VQR test in other directions; (v) improvement of the BIS formula for market required capital; (vi) nonlinear quantile regressions; among many others.

According to the Basel Committee (2006), new approaches to backtesting are still being developed and discussed within the broader risk management community. At present, different banks perform different types of backtesting comparisons, and the standards of interpretation also differ somewhat across banks. Active efforts to improve and refine the methods currently in use are underway, with the goal of distinguishing more sharply between accurate and inaccurate risk models. We aim to contribute to the current debate by providing a quantile technique that can be useful as a valuable diagnostic tool, as well as a mean to search for possible model improvements.

Acknowledgements

We would like to thank the Joint Editor Arthur Lewbel, an anonymous Associate Editor and two anonymous referees for their helpful comments and suggestions. We are also grateful to the seminar participants of the Joint PhD afternoon - LSE Economics and Statistics departments, where this paper was first presented (May 2007), specially Qiwei Yao and Myung Seo, and seminar participants at the 2nd AlBan Conference (France), Forecasting in Rio (Brazil), Central Bank of Brazil - DEPEP/RJ Seminar, EEA-ESEM 2008 Congress (Italy), and the Empirical Finance for Central Banks course at CCBS - Bank of England, specially Ibrahim Stevens, Francoise Ben Zur, David Delgado and Sarat Dhal. Wagner is also grateful to the members of the PhD thesis committee: André Minella, João Victor Issler, Maria Cristina Terra and Tomás Málaga for their insightful comments. Parts of this article were written while Wagner visited the Department of Economics at The London School of Economics and Political Science (LSE), whose hospitality is gratefully acknowledged. Research support of Programme Alban - The European Union Programme of High Level Scholarships for Latin America (grant n. E06D100111BR) and CAPES scholarship BEX1142/06-2. The opinions in this paper are those of the authors and do not necessarily reflect the point of view of the Central Bank of Brazil. Any remaining errors are ours.

Appendix A. Proofs of Propositions

Proof of Proposition 1. By Assumption (1), $\alpha_i(U_t)$ are increasing functions of the iid standard uniform random variable U_t and, thus, $Q_{\alpha_i(U_t)} = \alpha_i(Q_{U_t}) = \alpha_i(\tau)$, since for any monotone increasing function g and a standard uniform random variable, U, we have $Q_{g(U)}(\tau) = g(Q_U(\tau)) = g(\tau)$, where $Q_U(\tau) = \tau$ is the quantile function of U_t. By comonotonicity, we have that $Q_{\sum_{i=1}^{p} \alpha_i(U_t)} = \sum_{i=1}^{p} Q_{\alpha_i(U_t)}$. This way, by also considering assumption (1), we guarantee that the conditional quantile function is monotone increasing in τ, which is a crucial property of Value-at-Risk models. In other words, we have that $Q_{R_t}(\tau_1 \mid \mathcal{F}_{t-1}) < Q_{R_t}(\tau_2 \mid \mathcal{F}_{t-1})$ for all $\tau_1 < \tau_2 \in (0,1)$. Assumptions (2)-(4) are regularity conditions necessary to define the asymptotic covariance matrix, and a continuous conditional quantile function, needed for the CLT (7) of Koenker (2005, Theorem 4.1). A sketch of the proof of this CLT, via a Bahadur representation, is also presented in Hendricks and Koenker (1992, Appendix). Given that we established the conditions for the CLT (7), our proof is concluded by using standard results on quadratic forms: For a given random variable $z \sim N(\mu, \Sigma)$ it follows that $(z-\mu)'\Sigma^{-1}(z-\mu) \sim \chi^2_r$ where $r = rank(\Sigma)$. See Johnson and Kotz (1970, p. 150) and White (1984, Theorem 4.31) for further details. ■

Proof of Proposition 2. (i) $P1 \Leftrightarrow S1$ Lets assume that the nominal quantile level of the VaR model is τ^*, i.e., $\Pr[R_t \leq V_t | \mathcal{F}_{t-1}] = \tau^*$. If assumptions (1)-(4) hold, then, it follows that $Q_{R_t}(\tau \mid \mathcal{F}_{t-1}) = \inf\{R_t : F(R_t \mid \mathcal{F}_{t-1}) \geq \tau\}$ and, thus, $\Pr(R_t \leq Q_{R_t}(\tau \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}) = \tau$. In particular, for $\tau = \tau^*$, we have that $\Pr(R_t \leq Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}) = \tau^*$. Therefore, it follows that $\tau^* = \Pr(R_t \leq V_t \mid \mathcal{F}_{t-1}) = \Pr(R_t \leq Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}) \Leftrightarrow V_t = Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$.

(ii) Since we assumed that the null hypothesis of Berkowitz et al. (2006) $E(Hit_t \mid \mathcal{F}_{t-1}) = 0$ is linear, then, we can prove that tests based on binary variables (as the one proposed by Engle and Manganelli (2004)) do not provide a sufficient condition to assess the performance of a quantile model. We make such a comparison by using the optimality (or subgradient) condition of the quantile regression problem. To do so, we firstly follow Koenker (2005, pp. 32-37) to show how quantile regression works. Recall that in quantile regression setup the objective function is given by $R(\beta) = \sum_{t=1}^{n} \rho_\tau(y_t - x_t'\alpha(\tau))$, where $\rho_\tau(u) = \begin{cases} \tau u, u \geq 0 \\ (\tau-1)u, u < 0 \end{cases}$. Notice that ρ_τ is piecewise linear and continuous function, and it is differentiable except at the points at which one or more residuals are zero. At such points, $R(\beta)$ has directional derivatives in all directions. The directional derivative of R in direction w ($\nabla R(\beta, w)$) is given by $\nabla R(\beta, w) = -\sum_{t=1}^{n} \psi(y_t - x_t'\beta, -x_t'w)\, x_t'w$, where $\psi(u,v) = \begin{cases} \tau - I(u<0),\ if\ u \neq 0 \\ \tau - I(v<0),\ if\ u = 0 \end{cases}$. If at point β_0, $\nabla R(\beta_0, w) \geqslant 0$ for all $w \in \mathbb{R}^p$ with $\|w\| = 1$, then β_0 minimizes $R(\beta)$.

(iia) $S2 \nRightarrow S1$ Since in $S2$ we assumed a linear model, the conditional expectation of Berkowitz et al. (2006) can be represented by the orthogonality condition $E(Hit_t X_t') = 0$, in which X_t is measurable with respect to $\mathcal{F}_{t-1}$. Now, consider the quantile problem which the VQR is based on. Therein, if we set $y_t = R_t$, $x_t = [\mathbf{1}\ V_t]'$ and $\beta_0 = [0\ 1]'$, then the directional derivative becomes $\nabla R(\beta_0, w) = -\sum_{t=1}^{n} \psi(y_t - x_t'\beta_0, -x_t'w)\, x_t'w$, where $\psi(y_t - x_t'\beta_0, -x_t'w)\, x_t'w =$

$\begin{cases} Hit_t x_t' w, \text{ if } u_t \neq 0 \\ Hit_t^* x_t' w, \text{ if } u_t = 0 \end{cases}$, and $Hit_t^* = I([1\ V_t]\, w < 0) - \tau$. Notice that the function Hit is the same one defined by Engle and Manganelli (2004) and that $Hit_t \neq Hit_t^*$. Now, we can see that the orthogonolaty condition $n^{-1} \sum_{t=1}^{n} Hit_t X_t' = 0$, does not imply $\triangledown R(\beta_0, w) \geqslant 0$ for all $w \in \mathbb{R}^2$ with $\|w\| = 1$. In this case, the orthogonality condition is **not** sufficient to assess a performance of a quantile model V_t. In practice, this result implies that there can exist a misspecified model that is not rejected by backtests based on binary variable Hit_t

(iib) $S3 \nRightarrow S1$ If we set $X_t = x_t = [\mathbf{1}\ V_t]'$ in the previous proof, then the orthogonality condition becomes $n^{-1} \sum_{t=1}^{n} Hit_t\, [\mathbf{1}\ V_t] = 0$, which corresponds to the null hypothesis of the DQ_{oos} test with instrument $X_t = [\mathbf{1}\ V_t]'$. Again, such a orthogonality condition does not imply $\triangledown R(\beta_0, w) = \sum_{t=1}^{n} \begin{cases} Hit_t\, [1\ V_t]\, w, \text{ if } u_t \neq 0 \\ Hit_t^*\, [1\ V_t]\, w, \text{ if } u_t = 0 \end{cases} \geqslant 0$ for all $w \in \mathbb{R}^2$ with $\|w\| = 1$. In this case, the orthogonality condition of the DQ_{oos} test is **not** sufficient to assess a performance of a quantile model V_t. In practice, this result implies that there can exist a misspecified model that is not rejected by the DQ_{oos} test. We can obviously generalize this result for a list of instruments larger than $X_t = [\mathbf{1}\ V_t]'$.

(iic) $S4 \nRightarrow S1$ Based on the previous argument, if we simply set $X_t = [Hit_{t-1}]'$, we obtain a similar result, in which $S4$ holds, whereas, $S1$ does not hold.

(iid) $S5 \nRightarrow S1$ From $S5$, we have that $E(Hit_t) = 0$, which is merely a special case of (iib) in which $X_t = [\mathbf{1}]'$.

(iii) Now, we use Corollary 2.1 (Koenker 2005, p.37) which states that a consequence of the optimality condition $\triangledown R(\beta_0, w) \geqslant 0$ is that the proportion of negative residuals is approximately τ and the proportion of positive residuals is $1 - \tau$. This can be used to show that under the null hypothesis of the VQR test (i.e., $V_t = Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1})$), we have that $E(Hit_t \mid \mathcal{F}_{t-1}) = 0$, which means that the existing Backtests (such as Kupiec (1995), Christoffersen (1998) and Engle and Manganelli (2004)) are **implied** by the VQR test.

(iiia) $S1 \Rightarrow S2$ From the definition of H_t, it follows that $E(H_t \mid \mathcal{F}_{t-1}) = 1 * \Pr(R_t > V_t \mid \mathcal{F}_{t-1}) + 0 * \Pr(R_t \leq V_t \mid \mathcal{F}_{t-1}) = \Pr(R_t > V_t \mid \mathcal{F}_{t-1}) = \Pr(R_t > Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1})$, where the last equality is due to $S1$. This way, $E(H_t \mid \mathcal{F}_{t-1}) = 1 - \Pr(R_t \leq Q_{R_t}(\tau^* \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}) = 1 - \tau^*$, based on the previously mentioned Corollary 2.1 and the definition of the conditional quantile function. Therefore, $E(H_t - (1 - \tau^*) \mid \mathcal{F}_{t-1}) = E(Hit_t \mid \mathcal{F}_{t-1}) = 0$.

(iiib) $S1 \Rightarrow S3$ Firstly, notice that from assumption DQ9 of Engle & Manganelli (2004), it follows that $E[Hit_t X_t'] = 0$, in which $X_t \in \mathcal{F}_{t-1}$. For instance, $X_t = [\mathbf{1}, V_t, H_{t-1}, \ldots, H_{t-p}]$, in which V_t is also measurable (by construction) in respect to $\mathcal{F}_{t-1}$. This way, it naturally follows that $S2 \Rightarrow S3$ and, since $S1 \Rightarrow S2$ from (iiia), we have that $S1 \Rightarrow S3$.

(iiic) $S1 \Rightarrow S4$ From (iiia) it follows that $S1 \Rightarrow S2$. Following Berkowitz et al. (2006), the martingale difference hypothesis ($S2$) naturally implies that the demeaned violation sequence is uncorrelated at all leads and lags which implies in the null hypothesis of the Markov test of Christoffersen (1998). In other words, $S2 \Rightarrow S4$ and, therefore, $S1 \Rightarrow S4$.

(iiid) $S1 \Rightarrow S5$ From item (iiia), it follows that $S1 \Rightarrow S2$. Applying the law of iterated expectations on $S2$, it follows that $E(Hit_t) = 0$. ■

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent Measures of Risk. Mathematical Finance 9, 203-228.

[2] Basle Committee on Banking Supervision, 1996. Amendment to the Capital Accord to Incorporate Market Risks. BIS - Bank of International Settlements.

[3] Berkowitz, J., O'Brien, J., 2002. How Accurate are the Value-at-Risk Models at Commercial Banks? Journal of Finance 57, 1093-1111.

[4] Berkowitz, J., Christoffersen, P., Pelletier, D., 2006. Evaluating Value-at-Risk Models with Desk-Level Data, mimeo.

[5] Campbell, S.D., 2005. A Review of Backtesting and Backtesting Procedures. Finance and Economics Discussion Series, Working Paper 21.

[6] Charles, A., Darné, O., 2005. Outliers and GARCH models in financial data. Economic Letters 86, 347-352.

[7] Chaudhuri, P., 1996. On a geometric notion of quantiles for multivariate data. Journal of the American Statistical Association 91, 862-872.

[8] Chen, J.E., 2001. Investigations on Quantile Regression: Theories and Applications for Time Series Models. National Chung-Cheng University, mimeo.

[9] Chen, M.Y., Chen, J.E., 2002. Application of Quantile Regression to Estimation of Value at Risk. National Chung-Cheng University, mimeo.

[10] Chernozhukov, V., Umantsev, L., 2001. Conditional Value-at-Risk: Aspects of Modelling and Estimation. Empirical Economics 26 (1), 271-292.

[11] Christoffersen, P.F., 1998. Evaluating interval forecasts, International Economic Review 39, 841-862.

[12] ______, 2003. Elements of Financial Risk Management, San Diego: Academic Press.

[13] Christoffersen, P.F., Goncalves, S., 2005. Estimation Risk in Financial Risk Management. Journal of Risk 7, 1–28.

[14] Christoffersen, P.F., Hahn, J., Inoue, A., 2001. Testing and Comparing Value-at-Risk Measures. Journal of Empirical Finance 8, 325-342.

[15] Christoffersen, P.F., Pelletier, D., 2004. Backtesting Value-at-Risk: A Duration-Based Approach. Journal of Financial Econometrics 2 (1), 84-108.

[16] Crnkovic, C., Drachman, J., 1997. Quality Control in VaR: Understanding and Applying Value-at-Risk. London: Risk Publications.

[17] Daníelsson, J., de Vries, C., 1997. Tail index and quantile estimation with very high frequency data. Journal of Empirical Finance 4, 241-257.

[18] Daníelsson, J., Jorgensen, B.N., Samorodnitsky, G., Sarma, M., de Vries, C.G., 2005. Suadditivity Re-examined: the case for Value-at-Risk. Manuscript. London School of Economics. Downloaded from www.RiskResearch.org.

[19] Dowd, K., 2005. Measuring Market Risk. John Wiley and Sons Ltd., 2nd edition.

[20] Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer-Verlag.

[21] Embrechts, P., Resnick, Samorodnitsky, G., 1999. Extreme value theory as a risk management tool. North American Actuarial Journal 3, 30-41.

[22] Engle, R.F., 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. Econometrica 50, 987-1008.

[23] Engle, R.F., Manganelli, S., 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. Journal of Business and Economic Statistics 22 (4), 367-381.

[24] Engle, R.F., Patton, A.J., 2001. What good is a volatility model? Quantitative Finance, Institute of Physics Publishing 1, 237–245.

[25] Franses, P.H., Ghijsels, H., 1999. Additive outliers, GARCH and forecasting volatility. International Journal of Forecasting 15, 1-9.

[26] Giacomini, R., Komunjer, I., 2005. Evaluation and Combination of Conditional Quantile Forecasts. Journal of Business and Economic Statistics 23 (4), 416-431.

[27] Hafner, C.M., Linton, O., 2006. Comment on "Quantile Autoregression". Journal of the American Statistical Association 101 (475), 998-1001.

[28] Hartz, C., Mittnik, S., Paolella, M., 2006. Accurate Value–at–Risk Forecasting Based on the Normal–GARCH Model. Computational Statistics and Data Analysis 51 (4), 2295–2312.

[29] Hendricks, W., Koenker, R., 1992. Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity. Journal of the American Statistical Association 87 (417), 58–68.

[30] Huber, P.J., 1967. The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. University of California Press, Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability 4, 221-233.

[31] Huisman, R., Koedijk, K.G., Kool, C.J.M., Palm, F. 2001. Tail-Index Estimates in Small Samples. Journal of Business & Economic Statistics 19(2), 208-216.

[32] Ibragimov, R., Walden, J., 2007. Value at Risk under dependence and heavy-tailedness: models with common shocks. Manuscript, Harvard University.

[33] Johnson, N.L., Kotz, S., 1970. Distributions in statistics: Continuous univariate distributions. Wiley Interscience.

[34] Jorion, P., 2007. Value-at-risk: The new benchmark for managing financial risk. McGraw Hill, 3rd edition.

[35] J.P. Morgan, 1996. RiskMetrics, Technical Document. New York, 4th edition.

[36] Koenker, R., 2005. Quantile Regression. Cambridge University Press.

[37] Koenker, R., Bassett, G., 1978. Regression Quantiles. Econometrica 46, 33-50.

[38] ______, 1982a. Robust Tests for Heteroscedasticity Based on Regression Quantiles. Econometrica 50 (1), 43–62.

[39] ______, 1982b. Tests of Linear Hypotheses and L1 Estimation. Econometrica 50 (6), 1577–1584.

[40] Koenker, R., Machado, J.A.F., 1999. Goodness of Fit and Related Inference Processes for Quantile Regression. Journal of the American Statistical Association 94 (448), 1296–1310.

[41] Koenker, R., Portnoy, S., 1999. Quantile Regression. Unpublished Manuscript, University of Illinois.

[42] Koenker, R., Xiao, Z., 2002. Inference on the Quantile Regression Process. Econometrica 70 (4), 1583–1612.

[43] ______, 2006a. Quantile Autoregression. Journal of the American Statistical Association 101 (475), 980-990.

[44] ______, 2006b. Rejoinder of "Quantile Autoregression". Journal of the American Statistical Association 101 (475), 1002-1006.

[45] Koenker, R., Zhao, Q., 1996. Conditional Quantile Estimation and Inference for ARCH Models. Econometric Theory 12, 793-813.

[46] Kuester, K., Mittik, S., Paolella, M. S., 2005. Value-at-risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics 4 (1), 53–89.

[47] Kupiec, P., 1995. Techniques for Verifying the Accuracy of Risk Measurement Models. Journal of Derivatives 3, 73-84.

[48] Laine, B., 2001. Depth contours as multivariate quantiles: a directional approach. Master's thesis. Université Libre de Bruxelles.

[49] Laurent S., Peters, J.-P. 2006. G@RCH 4.2, Estimating and Forecasting ARCH Models. Timberlake Consultants Press. London.

[50] Lima, L.R., Neri, B.A.P., 2006. Comparing Value-at-Risk Methodologies. Ensaios Economicos EPGE, 629. Mimeo. Getulio Vargas Foundation.

[51] Linton, O., Whang, Y.J., 2007. The quantilogram: With an application to evaluating directional predictability. Journal of Econometrics 141(1), 250-282

[52] Lopez, J.A., 1999. Methods for Evaluating Value-at-Risk Estimates. Federal Reserve Bank of San Francisco, Economic Review 2, 3-17.

[53] Machado, J.A.F., Mata, J., 2001. Earning functions in Portugal 1982-1994: Evidence from quantile regressions. Empirical Economics 26, 115-134.

[54] McNeil, A.J., Frey, R. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7(3), 271-300.

[55] Mincer, J., Zarnowitz, V., 1969. The evaluation of economic forecasts and expections. In: Mincer, J. (ed.), Economic Forecasts and Expections. National Bureau of Economic Research, New York.

[56] Nankervis, J., Sajjad, R., Coakley, J., 2006. Value-at-Risk for long and short positions: A comparison of regime-switching GARCH and ARCH family models. University of Essex, mimeo.

[57] Pascual, L., Romo, J., Ruiz, E., 2006. Bootstrap prediction for returns and volatilities in garch models. Computational Statistics & Data Analysis 50, 2293–2312.

[58] Pritsker, M., 2001. The Hidden Dangers of Historical Simulation. Washington: Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 27.

[59] Robinson, P.M., 2006. Comment on "Quantile Autoregression". Journal of the American Statistical Association 101 (475), 1001-1002.

[60] Schulze, N., 2004. Applied Quantile Regression: Microeconometric, Financial, and Environmental Analyses. Dissertation zur Erlangung des Doktorgrades, Universität Tübingen.

[61] White, H., 1984. Asymptotic Theory for Econometricians. Academic Press, San Diego.