UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 09-46
Economic Series (25)
May 2009

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Evaluating Value-at-Risk models via Quantile Regression

Wagner Piazza Gaglianone* Luiz Renato Lima† Oliver Linton‡ Daniel Smith§

14th May 2009

Abstract

This paper is concerned with evaluating value at risk estimates. It is well known that using only binary variables, such as whether or not there was an exception, 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 variables. It is shown that the new backtest provides a sufficient condition to assess the finite sample 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).
§ Faculty of Business Administration, Simon Fraser University, 8888 University Drive, Burnaby BC V6P 2T8, Canada (e-mail: drsmith@sfu.ca).

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 the 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 continuous 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).

Daníelsson et al. (2005) and Ibragimov and Walden (2007) show that for continuous random variables either 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 (e.g., Cauchy distribution). In 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.

A crucial issue 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 (i.e., an indicator for whether the loss exceeded the VaR), from which statistical properties are derived and further tested. More recently, Berkowitz et al. (2009) developed a unified framework for VaR evaluation and showed that the existing backtests can be interpreted as a Lagrange Multiplier (LM) type of tests.

The existing backtests are based on orthogonality conditions between a binary variable and some instruments. If the VaR model is correctly specified, then this orthogonality condition holds and one can therefore use it to derive an LM test. Monte carlo simulations have shown that such tests have low power in finite samples against a variety of misspecified models. The problem arises because binary variables are constructed to represent rare events. In finite samples, it may be the case that there are few extreme events, leading to a lack of the information needed to reject a misspecified model. In this case, a typical solution would be to either increase the sample size or construct a new test that uses more information than the existing ones to reject a misspecified model.

The contribution of this paper is twofold. First we propose a random coefficient model that can be used to construct a Wald test for the null hypothesis that a given VaR model is correctly specified. To the best of our knowledge, no such framework exists in the current literature. It is well known that although LM and Wald tests are asymptotically equivalent under the null hypothesis and local alternatives, they can yield quite different results in finite samples. We show in this paper that the new test uses more information to reject a misspecified model which makes it deliver more power in finite samples than any other existing test.

Another limitation of the existing tests is that they do not give any guidance as to how the VaR models are wrong. The second contribution of this paper is to develop a mechanism by which we can evaluate the local and global performance of a given VaR model and, therefore, find out why and when it is misspecified. By doing this, we can unmask the reasons of rejection of a misspecified model. This information could be used to reveal if a given model is under or over estimating risk, if it reacts quickly to increases in volatility, or even to suggest possible model combinations that would result in more accurate VaR estimates. Indeed, it has been proven that model combination (see Issler and Lima, 2009) can increase the accuracy of forecasts. Since VaR is simply a conditional quantile, it is also possible that model combination can improve the forecast of a given conditional quantile or even of an entire density.

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:

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

where $\mathcal{F}_{t-1}$ is the information set available at time $t-1$. From the above definition, it is clear that V_t is the $\tau^* th$ conditional quantile of R_t. In other words, V_t is the one-step ahead forecast of the $\tau^* th$ quantile of R_t based on the information available up to period $t-1$.

From Equation (1) it is clear that finding a VaR is equivalent to finding the conditional quantile of R_t. Following the idea of Christoffersen et al. (2001), one can think of generating a VaR measure as the outcome of a quantile regression, treating volatility as a regressor. 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. 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:

$$\begin{aligned} R_t &= \alpha_0(U_t) + \alpha_1(U_t) V_t & (2) \\ &= x_t' \beta(U_t), & (3) \end{aligned}$$

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 $x_t' = [1, V_t]$.

Proposition 1 *Given the random coefficient model (2) and the comonotonicity assumption of $\alpha_i(U_t)$, $i = 0,1$, the τth conditional quantile of R_t can be written as*

$$Q_{R_t}\left(\tau \mid \mathcal{F}_{t-1}\right) = \alpha_0(\tau) + \alpha_1(\tau) V_t \ \text{; for all } \tau \in (0,1). \tag{4}$$

Proof. *See appendix.* ■

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 (4), a natural way to test for the overall performance of a VaR model is to test the null hypothesis

$$H_o : \begin{cases} \alpha_0(\tau^*) = 0 \\ \alpha_1(\tau^*) = 1 \end{cases} \tag{5}$$

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 should be interpreted as a Mincer and Zarnowitz (1969) type-regression framework for a conditional quantile model.

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), and it is normal with covariance matrix that takes the form of a Huber (1967) sandwich:

$$\sqrt{T}(\widehat{\theta}(\tau^*) - \theta(\tau^*)) \xrightarrow{d} N(0, \tau^*(1-\tau^*)H_{\tau^*}^{-1} J H_{\tau^*}^{-1}) = N(0, \Lambda_{\tau^*}), \tag{6}$$

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

$$\zeta_{VQR} = T[\widehat{\theta}(\tau^*)'(\tau^*(1-\tau^*)H_{\tau^*}^{-1} J H_{\tau^*}^{-1})^{-1}\widehat{\theta}(\tau^*)]. \tag{7}$$

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 for all $\tau \in (0,1)$:

$$J = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' \quad ; \quad H_\tau = p\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' [f_t(Q_{R_t}(\tau \mid x_t))].$$

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 2 *(VQR test) Consider the quantile regression (4). Under the null hypothesis (5), 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 it 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 Existing Backtests

Recall that a correctly specified VaR model at level τ^* is nothing other 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 review some of the existing backtests, which are based on an orthogonality condition between a binary variable and some instruments. This framework is the basis of GMM estimation and offers a natural way to construct a Lagrange Multiplier (LM) test. Indeed, the current literature on backtesting is mostly represented by LM type of tests. In finite samples, the LM test and the Wald test proposed in this paper can perform quite different. In particular, we will show that the test proposed in this paper consider more information than some of the existing test and, therefore, it can deliver more power in finite sample.

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

$$H_t = \begin{cases} 1 \;\; ; \;\; \text{if } R_t < V_t \\ 0 \;\; ; \;\; \text{if } R_t \geq V_t \end{cases} . \tag{8}$$

By definition, the probability of violating the VaR should always be

$$\Pr(H_t = 1 | \mathcal{F}_{t-1}) = \tau^*. \tag{9}$$

Based on these definitions, we now present some backtests usually mentioned in the literature to identify misspecified VaR models:

(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 $N = \sum_{t=1}^{T} H_t$. The objective of the test is to know whether $\widehat{p} \equiv N/T$ is statistically equal to τ^*:

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

The probability of observing N violations over a sample size of T is driven by a Binomial distribution and the null hypothesis $H_o : p = \tau^*$ can be verified through a LR test (also known as the unconditional coverage test). This test rejects the null hypothesis of an accurate VaR if the actual fraction of VaR violations in a sample is statistically different than τ^*. 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.

(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

$$\Pi = \begin{array}{c} \text{Previous day} \\ \left[\begin{array}{cc} 1-\pi_0 & 1-\pi_1 \\ \pi_0 & \pi_1 \end{array} \right] \end{array} \begin{array}{l} \text{current day (violation)} \\ \text{no violation} \end{array} \tag{11}$$

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:

$$LR_{ind.} = 2\ln\left(\frac{(1-\pi_0)^{T_{00}}\pi_0^{T_{01}}(1-\pi_1)^{T_{10}}\pi_1^{T_{11}}}{(1-\pi)^{(T_{00}+T_{10})}\pi^{(T_{01}+T_{11})}}\right). \tag{12}$$

The joint test, also known as the "conditional coverage test", includes the unconditional coverage and independence properties. An interesting feature of this test is that a rejection of the conditional 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.

Berkowitz et al (2009) extended and unified the existing tests by noting that the de-meaned violations $Hit_t = H_t - \tau^*$ form a martingale difference sequence (m.d.s.). By the definition of the violation, equations (8) and (9) imply that

$$E[Hit_t|\mathcal{F}_{t-1}] = 0,$$

that is, the de-meaned violations form an m.d.s. with respect to $\mathcal{F}_{t-1}$. This implies that Hit_t is uncorrelated at all leads and lags. In other words, for any vector X_t in $\mathcal{F}_{t-1}$ we must have

$$E[Hit_t \otimes X_t] = 0, \tag{13}$$

which constitutes the basis of GMM estimation. The framework based on such orthogonality conditions offers a natural way to construct a Lagrange Multiplier (LM) test. Indeed, if we allow X_t to include lags of Hit_t, V_t and its lags, then we obtain the well-known DQ test proposed by Engle and Manganelli (2004), which we describe below.

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

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

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 test statistic 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).

Several other related procedures can be immediately derived from the orthogonality condition (13). For example, Chirstoffersen and Pelletier (2004) and Haas (2005) propose duration-based

tests to the problem of assessing VaR accuracy. However, as shown by the Monte-carlo experiment in Berkowitz et. al. (2009), the DQ test in which $X_t = V_t$ appears to be the best backtest for 1% VaR models, and other backtests generally have much lower power against misspecified VaR models. In the next section, we provide a link between the proposed VQR test and the DQ test, and we give some reasons that suggest the VQR test should be more powerful than the DQ test in finite samples.

4.1 DQ versus VQR test

In the previous section we showed that the DQ test proposed by Engle and Manganelli (2004) can be interpreted as a LM test. The orthogonality condition (13) constitutes the basis of GMM estimation and therefore can be used to estimate the quantile regression model (4) under the null hypothesis (5). It is well know that OLS and GMM are asymptotically equivalent. By analogy, quantile regression estimation and GMM estimation are also known to be asymptotically equivalent (see for instance, footnote 5 of Buchinsky, 1998, and the proofs of Powell, 1984). However, if the orthogonality condition (13) provides a poor finite-sample approximation of the objective function, then GMM and quantile regression estimation will be quite different and the LM tests and Wald test proposed in this paper will yield very different results. In order to compare the DQ test and our VQR test, we consider the simplest case in which $X_t = [\mathbf{1}\ V_t]'$. As showed in Koenker (2005, pp. 32-37), the quantile regression estimator minimizes the objective function $R(\beta) = \sum_{t=1}^{n} \rho_\tau \left(y_t - x_t'\alpha\left(\tau\right)\right)$, where $\rho_\tau\left(u\right) = \tau u 1(u \geq 0) + \left(\tau - 1\right) u 1(u < 0)$. 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 $\nabla R(\beta, w) = -\sum_{t=1}^{n} \psi\left(y_t - x_t'\beta, -x_t'w\right) x_t'w$, 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 $R(\beta)$. 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 $\nabla R(\beta_0, w) = -\sum_{t=1}^{n} \psi\left(y_t - x_t'\beta_0, -x_t'w\right) x_t'w$, where

$$\psi\left(y_t - x_t'\beta_0, -x_t'w\right) x_t'w = \begin{cases} Hit_t\left[\mathbf{1}\ V_t\right] w, \text{ if } u_t \neq 0 \\ Hit_t^* x_t'w, \text{ if } u_t = 0, \end{cases}$$

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 $Hit_t \neq Hit_t^*$. According to Koenker (2005), there will be at least

p zero residuals where p is the dimension of β. This suggests that the orthogonality condition $n^{-1}\sum_{t=1}^{n} Hit_t\left[\mathbf{1}\ V_t\right] = 0$, does not imply $\nabla R(\beta_0, w) \geqslant 0$ 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 finite samples. In practice, the VQR test is using more information than the DQ test to reject a misspecified model and, therefore, it can exhibit superior power in finite samples than the DQ test. Finally, it is possible to generalize this result for a list of instruments larger than $X_t = [\mathbf{1}\ V_t]'$ as long as we redefine our quantile regression (4) to include other regressors besides an intercept and V_t

Although the first order condition $n^{-1}\sum_{t=1}^{n} Hit_t\left[\mathbf{1}\ V_t\right] = 0$ does not strictly imply in the optimality condition $\nabla R(\beta_0, w) \geqslant 0$ in finite samples, this does not prevent it from holding asymptotically if one makes an assumption that the number of zero residuals are $o_p(1)$ (see the discussion in the handbook chapter of Newey and McFadden (1994). Hence, the DQ test proposed by Engle and Manganelli and our VQR test are asymptotically equivalent under the null hypothesis and local alternatives**.** However, in finite samples the orthogonality and the optimality conditions can be quite different and the two tests can therefore yield very different results

Finally, notice that if $\nabla R(\beta_0, w) \geqslant 0$ for all $w \in \mathbb{R}^2$ with $\|w\| = 1$, then $n^{-1}\sum_{t=1}^{n} Hit_t\left[\mathbf{1}\ V_t\right] = 0$. Therefore, the DQ test is implied by the VQR when the VaR model is correctly specified. 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.

In sum, the above results suggest that there are some reasons to prefer the VQR test over the DQ test since the former can have more power in finite sample and be equivalent to the latter under the null hypothesis. 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.

5 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 are observed in finite samples. We consider, besides the VQR test, the unconditional coverage test of Kupiec (1995), the conditional coverage test of Christoffersen (1998) and the out-of-sample DQ test of Engle and Manganelli (2004), in which we considered the instruments $X_t = [1\ V_t]'$. We look at the one-day ahead forecast and simulate 5,000 sample paths of length $T + T_e$ observations, using the first $T_e = 250$ observations to compute the initial values of (18) with $T = \{250, 500, 1000, 2500\}$

(i.e., approximately 1, 2, 4, and 10 years of daily data).

Asymptotic critical values are used in this monte carlo experiment, but bootstrap methods could also be considered to compute finite-sample critical values. For example, for the VQR test we could have employed the bootstrap method proposed by Parzen et al (1994). For the Kupiec, Christorffersen and DQ tests, finite-sample critical values could be computed by using the method proposed by Dufour (2006). Bootstrap based critical values are more difficult to be computed by a financial institution on a daily basis. On the other hand, asymptotic critical values may not be accurate enough in small samples. As the results in this section show, asymptotic critical values are reasonably accurate for sample size of 4 years of daily observations, which is not a restriction for many financial institutions.

Nonetheless, the lack of accuracy in asymptotic critical values can give rise to size distortions especially for samples as small as $T = 250$. In order to avoid that these size distortions favor some tests in terms of power, we will compute the so called size-adjusted power. In the size-adjusted power, the 5% critical value is going to correspond to the 5% quantile of the test statistic distribution under the null hypothesis. In doing this, we eliminate any power performance of tests caused by the use of asymptotic critical values. Of course, if the empirical size is close to the nominal one, which is 5%, then size-adjusted and asymptotic power will be close to each other as well.

We start evaluating the empirical size of 5% tests. To assess the size of the tests we will assume that the correct data-generating process is a zero mean, unit unconditional variance normal innovation-based GARCH model:

$$R_t = \sigma_t \varepsilon_t, \ t = 1, ..., T, \tag{15}$$

$$\sigma_t^2 = (1 - \alpha - \beta) + \alpha \cdot y_{t-1}^2 + \beta \cdot \sigma_{t-1}^2, \tag{16}$$

where $\alpha = 0.05$ and $\beta = 0.90$, and $\varepsilon_t \sim iidN(0, 1)$. We therefore compute the true VaR as

$$V_t = \sigma_t \Phi_{\tau^*}^{-1} \tag{17}$$

where $\Phi_{\tau^*}^{-1}$ denotes the τ^* quantile of a standard normal random variable. Since we are computing 5% tests, we should expect that under the null hypothesis each test would reject the correctly specified model $V_t = \sigma_t \Phi_{\tau^*}^{-1}$ 5% of times.

Table 1 reports the size of tests for $\tau^* = 1\%$ and $\tau^* = 5\%$, one can see that the asymptotic critical values do not do a good job when $\tau^* = 1\%$ and sample size as small as $T = 250$. In this case there are few observations around the 1% quantile of R_t implying an inaccurate estimation of $Q_{R_t}(0.01 \mid \mathcal{F}_{t-1})$. The VQR test responds to it by rejecting the null model more frequently than 5% of the time. However, we can see that if we allow for larger samples, the empirical size of the

VQR test converges to its nominal size of 5%. It is important to mention that 1000 observations correspond to only four years of daily observations and therefore it is a common sample size used in practice by financial institutions. For instance, Berkowitz et al (2009) used daily data ranging from 2001 to 2005 (about four years) to test VaR models used by different business lines of a financial institution.

The Kupiec and Christorffersen tests seem to be undersized when $\tau^* = 1\%$ and the sample is as small as $T = 250$ but their empirical size converges to the nominal size as the sample increases. The same thing happens to the DQ test, which is slightly oversized for small samples but has correct size as the sample size increases. When $\tau^* = 0.05$ observations become more dense and consequently the quantiles are estimated more accurately. Since the 5% conditional quantile is now estimated more precisely, the improvement in the VQR test is relatively larger than in other tests. Indeed, when $\tau^* = 5\%$ and sample is as small as $T = 250$ the empirical size of the VQR and DQ tests are almost the same. As the sample increases, the empirical size of all tests converge to the nominal size of 5% In sum, if the sample has a reasonable size then our theoretical findings are confirmed in the sense that all tests have correct size under the null hypothesis.

Table 1: Empirical Size of 5% tests

Panel A: $\tau^* = 0.01$				
Sample Size	250	500	1000	2500
ζ_{Kupiec}	0.036	0.042	0.046	0.049
$\zeta_{\text{Christ.}}$	0.038	0.045	0.057	0.058
ζ_{DQ}	0.088	0.068	0.063	0.058
ζ_{VQR}	0.125	0.106	0.076	0.059
Panel B: $\tau^* = 0.05$				
Sample Size	250	500	1000	2500
ζ_{Kupiec}	0.036	0.043	0.045	0.050
$\zeta_{\text{Christ.}}$	0.040	0.048	0.052	0.054
ζ_{DQ}	0.071	0.067	0.059	0.053
ζ_{VQR}	0.070	0.063	0.058	0.051

The main difference between the tests relates to their power to reject a misspecified model. To investigate power, we will assume that the correct data-generating process is again given by (15). We must also choose a particular implementation for the VaR calculation that is misspecified in that sense it differs from (17). Following the industry practice (see Pérignon and Smith, 2006), we

assume that the Bank uses an 1-year Historical Simulation method to compute its VaR. Specifically,

$$V_t = percentile\left(\{R_s\}_{s=t-T_e}^{t}, 100\tau^*\right). \tag{18}$$

Historical simulation is by far the most most popular VaR model used by commercial banks. Indeed, Pérignon and Smith (2006) document that almost three-quarters of banks that disclose their VaR method report using historical simulation. Historical simulation is the empirical quantile and therefore does not respond well to increases in volatility (see Pritsker, 2001). Combining its popularity and this weakness it is the natural choice for the misspecified VaR Model (we note that Berkowitz, Christorffersen, and Pelletier, 2006, use a similar experiment).

Table 2 reports the size-adjusted power. When $\tau^* = 1\%$ and the sample size is as small as 250, all tests present low power against the misspecified historical simulation model. Even in this situation, the VQR test is more powerful than any other test. When the sample size increases, the VQR test becomes even more powerful against the historical simulation model rejecting such a misspecified model 17%, 49% and 80% of the times when $T = 500, 1000$ and 2500, respectively. This performance is by far the best one among all backtests considered in this paper. The Kupiec test rejects the misspecified model only 32,2% of the times when $T = 2500$, which is close to the rate of the Christorffersen test (39,6%), but below the DQ test (64,4%). If a model is misspecified, then a backtest is supposed to use all available information to reject it. The Kupiec test has low power because it ignores the information about the time dependence of the hit process. The conditional coverage tests (Christorffersen and DQ) sacrifice information that comes from zero residuals and therefore fail to reject a misspecified model when the orthogonality condition holds, but the optimality condition does not. The VQR test simply compares the empirical conditional quantiles that is estimated from the data with the conditional quantile estimated by using the historical simulation model. If these two estimates are very different from each other, then the null hypothesis is going to be rejected and the power of the test is delivered.

There are theoretically few observations below V_t when $\tau^* = 1\%$, which explains the low power exhibited by all tests. When $\tau^* = 5\%$ the number of observations below V_t increases, giving to the tests more information that can be used to reject a misspecified model. Hence, we expect that the power of each test will increase when one considers $\tau^* = 5\%$ rather than $\tau^* = 1\%$. This additional information is used differently by each test. Here again the VQR benefits from using more information than any other test to reject a misspecified model. Indeed, even for $T = 250$ the VQR rejects the null hypothesis 21,5% of the times, above Kupiec (6,8%), Christorffersen (14,2%) and DQ (17,9%) test. For $T = 2500$ the power of the VQR approaches 90% against 42,3%, 50,9% and 76,9% for Kupiec, Chistorffersen and DQ respectively.

Table 2: Size-adjusted Power of 5% tests

Panel A: $\tau^* = 1\%$				
Sample Size	250	500	1000	2500
ζ_{Kupiec}	0.059	0.113	0.189	0.322
$\zeta_{\text{Christ.}}$	0.087	0.137	0.216	0.396
ζ_{DQ}	0.084	0.171	0.402	0.644
ζ_{VQR}	0.091	0.174	0.487	0.800
Panel B: $\tau^* = 5\%$				
Sample Size	250	500	1000	2500
ζ_{Kupiec}	0.068	0.146	0.267	0.423
$\zeta_{\text{Christ.}}$	0.142	0.227	0.319	0.509
ζ_{DQ}	0.179	0.316	0.454	0.769
ζ_{VQR}	0.215	0.366	0.644	0.883

Another advantage of the random coefficient framework proposed in this paper is that it can be used to derive a statistical device that indicates why and when a given VaR model is misspecified. In the next section we introduce this device and show how to use it to identify periods of risk exposure.

6 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: $W_t \equiv \{Q_{U_t}(\widetilde{\tau}) = \widetilde{\tau} \in [0,1] \mid V_t = Q_{R_t}(\widetilde{\tau} \mid \mathcal{F}_{t-1})\}$, representing the empirical quantile of the standard iid uniform random variable, U_t, such that the equality $V_t = Q_{R_t}(\widetilde{\tau} \mid \mathcal{F}_{t-1})$ holds at period t.

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 $Q_{R_t}(\tau \mid \mathcal{F}_{t-1})$ 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 $Q_{R_t}(\tau \mid \mathcal{F}_{t-1})$ across all quantiles τ inside the adopted grid.

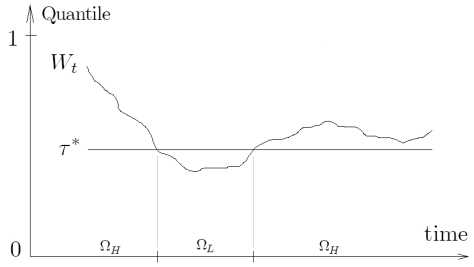
Now consider the set of all observations $\Omega = 1, \ldots, T$, in which T is the sample size, and define the following partitions of Ω:

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



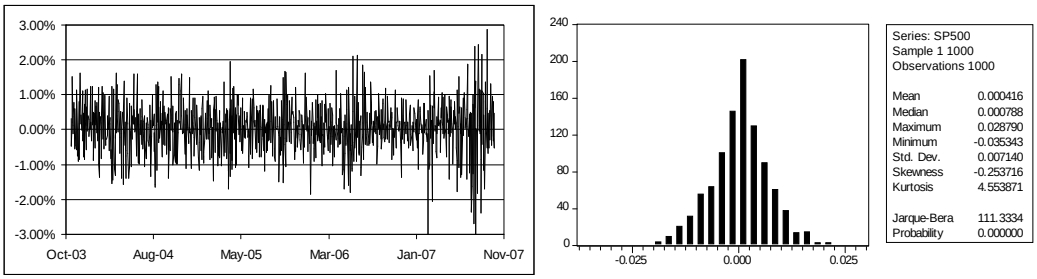
It should be mentioned that a VaR model that exhibits a good performance in the VQR test (i.e., in which H_o is not rejected) is expected to exhibit W_t as close as possible to τ^*, fluctuating around τ^*, in which periods of W_t below τ^* are balanced by periods above this threshold. On the other hand, 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.

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 two VaR models, based on the VQR test and other competing backtests commonly presented in the 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 2 - 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 two Value-at-Risk models adopted in our evaluation procedure are the popular 12-months historical simulation (HS12M) and the GARCH (1,1) model. According to Perignon and Smith (2006) about 75% of financial institutions in the US, Canada and Europe that disclose their VaR model report using historical simulation methods. The choice of a GARCH (1,1) model with Gaussian innovations is motivated by the work of Berkowitz and O'Brien (2002) who documented that the performance of the GARCH(1,1) is highly accurate even as compared to more sophisticated structural models.

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).

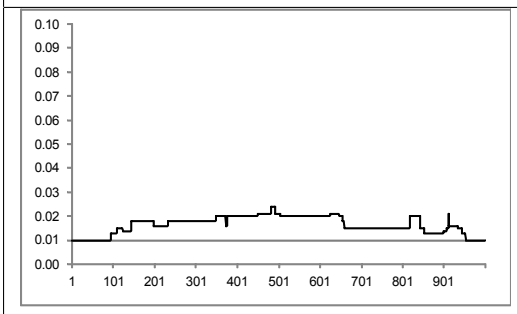
Practice generally shows that different models lead to widely different VaR time series for the same considered return series, leading us to the crucial issue of model comparison and hypothesis testing. The HS12M method has serious drawbacks and is expected to generate poor VaR measures, since it ignores 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.

7.2 Results

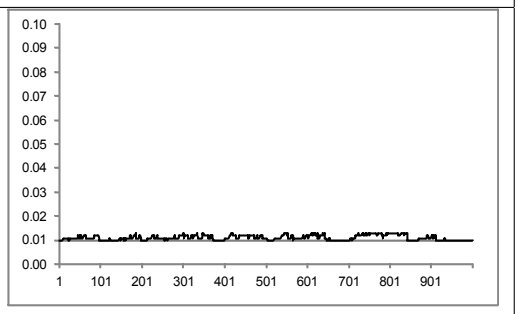
We start by showing the local analysis results for each VaR model. When $\tau^* = 1$ figure 3 shows that the VaR computed by the HS12M method is above the 1% line most of the times, indicating that such a model over estimate 1% VaR very frequently. The same does not happen to the GARCH(1,1) model, which seems to yield VaR estimates that are very close to the 1% line. If we now turn our attention to the 5% value-at-risk, we can see that both models have a poor performance. Indeed, The VaR estimated from the HS12M model seems to be quite erratic, with some periods in which 1% VaR is under estimated and other periods in which it is over estimated. The GARCH(1,1) model seems to under estimate 1% VaR most of the times, although it does not seem to yield a VaR estimate that is far below 0.03. Therefore, if we backtest these two models. we should expect to reject the 1% VaR estimated by HS12M, and the 5% VaR estimated by HS12M and GARCH(1,1). The results from our backtesting is exhibited in Table 3. 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.

Figure 3 - Local Analysis of VaR Models.

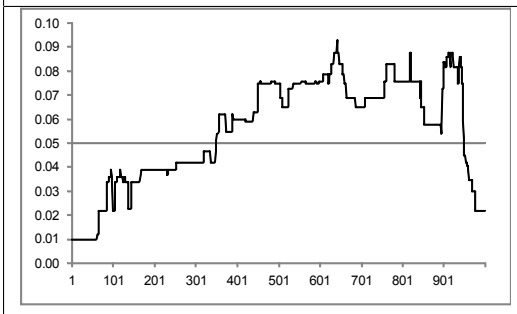
HS12m - 1% VaR



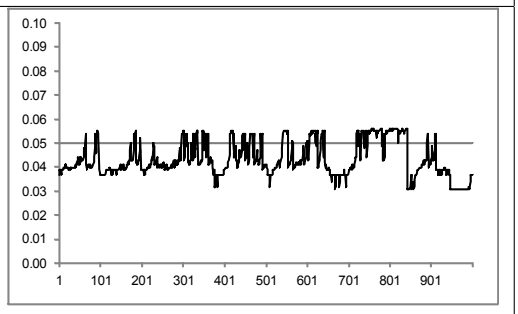
GARCH - 1% VaR



HS12m - 5% VaR



GARCH - 5% VaR



Despite its poor local performance, the HS12M model is not rejected by Kupiec and Christoffersen tests at 5% level of significance. In part, this can explained by the fact that the Kupiec test is mainly concerned on the percentage of hits generated by the VaR model, which in the case of the HS12M model reported in table 3, does not seem to be far away from the hypothetical levels The null hypothesis of the Christoffersen test is very restrictive, being rejected when there is strong evidence against first order dependence of the hit function or when the percentage of hits is far away from τ^*. It is well known that there many other forms of time dependence that is not accounted by the null hypothesis of the Christoffersen test. The more powerful test of Engle and Manganelli with the ex-ante VaR as the only instrument performs quite well and rejects the misspecified HS12M model for $\tau^* = 1\%$ and 5%. As expected, the VQR test performs quite well, easily rejecting the misspecified HS12M.

In our local analysis, the GARCH (1,1) model seems to work well when we use it to estimate a 1% VaR but fails to estimate a 5% VaR accurately. The backtest analysis in Table 3 indicates that neither tests rejects the null hypothesis that the GARCH(1,1) is a correctly specified model for a 1% VaR. However, as documented by Kuester et al. (2005) it is possible that for different τ^*'s (target probabilities) the model can do well at a given target probability, but otherwise poorly at another target probability. Here, the GARCH(1,1) model seems to predict the 1% VaR quite

well but not the 5% VaR. However, when we backtest the GARCH(1,1) for $\tau^* = 5\%$ by using the existing backtests, we fail to reject the null hypothesis. This results suggest that the GARCH(1,1) model correctly predicts the 5% VaR despite the clear evidence against it showed in Figure 3. The results of Table 3 indicate that GARCH(1,1) model for the 5% VaR is only rejected by the VQR test, which is compatible with the previous evidence in figure 3. Our methodology is, therefore, able to reject more misspecified VaR models in comparison to other backtests.

Table 3: Backtesting Value-at-Risk Models

Model	% of Hits	ζ_{Kupiec}	$\zeta_{\text{Christ.}}$	ζ_{DQ}	ζ_{VQR}
$\tau^* = 1\%$					
HS12M	1.6	0.080	0.110	0.000	0.000
GARCH(1,1)	1.1	0.749	0.841	0.185	0.010
$\tau^* = 5\%$					
HS12M	5.5	0.466	0.084	0.000	0.000
GARCH(1,1)	4.5	0.470	0.615	0.841	0.042
Notes: P-values are shown in the ζ's columns					

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 Hit_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, 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).

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, two anonymous referees, Ignacio Lobato, Richard Smith, J.C. Escanciano and Keith Knight for their helpful comments and suggestions. Lima would like to thank CNPq for financial support as well as the University of Illinois. Linton would like to thank the ESRC and Leverhulme foundations for financial support as well as the Universidad Carlos III de Madrid-Banco Santander Chair of Excellence.

Appendix A. Proofs of Propositions

Proof of Proposition 1. Given the random coefficient model (2), we can compute the conditional quantile of R_t as $Q_{R_t}(\tau \mid \mathcal{F}_{t-1}) = Q_{[\alpha_0(U_t)+\alpha_1(U_t)VaR_t]}(\tau \mid \mathcal{F}_{t-1})$. Comonotonicity implies that $Q_{\sum_{i=1}^{p}\alpha_i(U_t)} = \sum_{i=1}^{p} Q_{\alpha_i(U_t)}$. Therefore we can write $Q_{R_t}(\tau \mid \mathcal{F}_{t-1}) = Q_{\alpha_0(U_t)}(\tau \mid \mathcal{F}_{t-1}) + Q_{\alpha_1(U_t)VaR_t}(\tau \mid \mathcal{F}_{t-1})$. Since $\alpha_i(U_t)$ are increasing functions of the iid standard uniform random variable U_t, then we know that $Q_{\alpha_i(U_t)} = \alpha_i(Q_{U_t})$ and therefore $Q_{R_t}(\tau \mid \mathcal{F}_{t-1}) = \alpha_0(Q_{U_t}(\tau)) + \alpha_1(Q_{U_t}(\tau))VaR_t$. Finally, recall that $U_t \sim iid\ U(0,1)$ and therefore $Q_{U_t}(\tau) = \tau$, $\tau \in (0,1)$. which implies $Q_{R_t}(\tau \mid \mathcal{F}_{t-1}) = \alpha_0(\tau) + \alpha_1(\tau)VaR_t,\ \ \tau \in (0,1)$. ■

Proof of Proposition 2. By Assumption (1), we have 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 (6) 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 (6), 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. ■

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. (2009), "Evaluating Value-at-Risk Models with Desk-Level Data", forthcoming in *Management Science*.

[5] Buchinsky, M. (1998), "Recent Advances in Quantily Regression Models: A Practical Guideline for Empirical Research," *The Journal of Human Resources*, 33, 88-126.

[6] Campbell, S.D. (2005), "A Review of Backtesting and Backtesting Procedures," *Finance and Economics Discussion Series*, Working Paper 21.

[7] Charles, A., Darné, O. (2005), "Outliers and GARCH models in financial data," *Economic Letters* 86, 347-352.

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

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

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

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

[12] Christoffersen, P.F. (1998), "Evaluating interval forecasts," *International Economic Review* 39, 841-862.

[13] _____ (2003), "Elements of Financial Risk Management," *San Diego: Academic Press*.

[14] Christoffersen, P.F., Goncalves, S. (2005), "Estimation Risk in Financial Risk Management," *Journal of Risk* 7, 1–28.

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

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

[17] Crnkovic, C., Drachman, J. (1997), "Quality Control in VaR: Understanding and Applying Value-at-Risk," London: *Risk Publications.*

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

[19] 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.

[20] Dowd, K. (2005), "Measuring Market Risk," *John Wiley and Sons Ltd.*, 2nd edition.

[21] Embrechts, P., Kluppelberg, C., Mikosch, T. (1997), "Modelling Extremal Events," *Springer-Verlag.*

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

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

[24] 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.

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

[26] Franses, P.H., Ghijsels, H. (1999), "Additive outliers, GARCH and forecasting volatility," *International Journal of Forecasting* 15, 1-9.

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

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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

[34] Issler, J., Lima., L. R. (2009), "A panel data approach to economic forecasting. the bias-corrected average forecast," Forthcoming in *Journal of Econometrics*.

[35] Johnson, N.L., Kotz, S. (1970), "Distributions in statistics: Continuous univariate distributions," Wiley Interscience.

[36] Jorion, P. (2007), "Value-at-risk: The new benchmark for managing financial risk," *McGraw Hill, 3rd edition*.

[37] J.P. Morgan. (1996), "RiskMetrics", Technical Document. New York, 4th edition.

[38] Koenker, R. (2005), "Quantile Regression," Cambridge University Press.

[39] Koenker, R., Bassett, G. (1978), "Regression Quantiles," *Econometrica* 46, 33-50.

[40] _____ (1982a), "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica* 50 (1), 43–62.

[41] _____ (1982b), "Tests of Linear Hypotheses and L1 Estimation," *Econometrica* 50 (6), 1577–1584.

[42] 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.

[43] Koenker, R., Portnoy, S. (1999), "Quantile Regression," Unpublished Manuscript, University of Illinois.

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

[45] _____ (2006a), "Quantile Autoregression," *Journal of the American Statistical Association* 101 (475), 980-990.

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

[47] Koenker, R., Zhao, Q. (1996), "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory* 12, 793-813.

[48] 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.

[49] Kupiec, P. (1995), "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives* 3, 73-84.

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

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

[52] Lima, L.R., Neri, B.A.P. (2007), "Comparing Value-at-Risk Methodologies," *Brazilian Review of Econometrics*, 27(1), 1-25.

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

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

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

[56] 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.

[57] 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.

[58] 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.

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

[60] Péreginon, C., Smith, D. (2006), "The Level and Quality of Value-at-Risk Disclosure by Commercial Banks," *Manuscript, Simon Fraser University.*

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

[62] Powell, J. L. (1984). "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-325.

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

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

[65] White, H. (1984), "Asymptotic Theory for Econometricians," Academic Press, San Diego.