CENTER FOR APPLIED ECONOMICS AND POLICY RESEARCH

CAEPR Working Paper
#2008-021

Specification Tests of Parametric Dynamic Conditional Quantiles

Juan Carlos Escanciano
Indiana University Bloomington

Carlos Velasco
Universidad Carlos III de Madrid

August 15, 2008

SPECIFICATION TESTS OF PARAMETRIC DYNAMIC CONDITIONAL QUANTILES*

J. Carlos Escanciano†
Indiana University

Carlos Velasco‡
Universidad Carlos III de Madrid

August 4, 2008

Abstract

This article proposes omnibus specification tests of parametric dynamic quantile regression models. Contrary to the existing procedures, we allow for a flexible and general specification framework where a possibly continuum of quantiles are simultaneously specified. This is the case for many econometric applications for both time series and cross section data which require a global diagnostic tool. We study the asymptotic distribution of the test statistics under fairly weak conditions on the serial dependence in the underlying data generating process. It turns out that the asymptotic null distribution depends on the data generating process and the hypothesized model. We propose a subsampling procedure for approximating the asymptotic critical values of the tests. An appealing property of the proposed tests is that they do not require estimation of the non-parametric (conditional) sparsity function. A Monte Carlo study compares the proposed tests and shows that the asymptotic results provide good approximations for small sample sizes. Finally, an application to some European stock indexes provides evidence that our methodology is a powerful and flexible alternative to standard backtesting procedures in evaluating market risk by using information from a range of quantiles in the lower tail of returns.

Keywords and Phrases: Omnibus tests; Conditional quantiles; Nonlinear time series; Empirical processes; Quantile processes; Subsampling; Value-at-Risk; Tail Risk.

---

*We would like to thank Roger Koenker and participants at the Econometrics in Rio Conference for helpful comments. All errors are our own. Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908.

†Parts of this paper were written while J. Carlos Escanciano was visiting the Cowles Foundation at Yale University and the Economics Department at Cornell University, whose hospitality is gratefully acknowledged. This author thanks Professor Peter C.B. Phillips and Professor Yongmiao Hong the possibility of these visits. Research funded by the Spanish Ministerio de Educación y Ciencia, reference number SEJ2005-07657/ECON.

‡Corresponding author: Departamento de Economía. Universidad Carlos III de Madrid. Calle Madrid 126, 28903 Getafe (Madrid), Spain. e-mail: carlos.velasco@uc3m.es

1. INTRODUCTION

Quantile regression is a powerful alternative to least squares regression in a wide range of econometric applications that vary from labor economics or demand analysis to finance; see the special issue of Empirical Economics (2001, vol.26) and the references therein. The conditional quantile has the advantage over its natural competitor, the conditional mean, of being more robust to outliers and imposing less restrictions on the data generating process (DGP). Rather than relying on a single measure of conditional location, the quantile regression approach allows the researcher to explore a continuous range of conditional quantile functions, thereby providing a more complete and flexible analysis of the conditional dependence structure of the variables under consideration. A researcher interested in the whole conditional distribution will consider the specification of the conditional quantile at all quantile levels, requiring some diagnostic on the global suitability of the model. Thus, conditional Goodness-of-fit tests are of paramount importance in econometrics and finance, see e.g. Andrews (1997) and Corradi and Swanson (2006). On the other hand, a risk manager will not be interested in the whole Profit&Loss account's distribution but mainly in its left tail, and hence she or he will consider a set of small values of quantile levels, usually 1% or 5% as recommended by the Basel Accord (1996a). The methods developed here have important applications to measuring market risk; see Section 5. Obviously, one can envision many situations in economics where the interest is in the lower and upper parts of the distribution; see e.g. studies of unemployment duration (e.g. Koenker and Xiao (2002) and references therein), and wage inequalities (e.g. Machado and Mata, 2005). It is well-known that inference procedures within parametric quantile models depend crucially on the validity of the specified parametric functional forms for the range of quantiles under consideration. For instance, the counterfactual decomposition described in Machado and Mata (2005), that has been recently used in many studies to analyze the gender gap in log wages across the distribution (see e.g. Albrecht, van Vuuren and Vroman, 2007), and the Martingale transform methods in Koenker and Xiao (2002) depend crucially on the linear quantile specification. Therefore, it is important to develop powerful tests for the correct specification of parametric conditional quantiles over a possibly continuous range of quantiles of interest and under fairly general conditions on the underlying DGP. This is the main purpose of the present paper.

More precisely, suppose we observe a real-valued dependent variable Y_t, and the explanatory vector $I_{t-1} = (W'_{t-1}, Z'_t)' \in \mathbb{R}^d$, $d = s + m$, where $Z_t \in \mathbb{R}^m$, $m \in \mathbb{N}$, is an observable random vector (r.v) and $W_{t-1} = (Y_{t-1}, ..., Y_{t-s})' \in \mathbb{R}^s$, where A' denotes the matrix transpose of A. We assume throughout the article that the time series process $\{(Y_t, Z'_t)' : t = 0, \pm 1, \pm 2, ...\}$, defined on the probability space $(\Omega, \mathcal{A}, P)$, is strictly stationary and ergodic. Assuming that the conditional distribution of Y_t given I_{t-1} is continuous, we define the α-th conditional quantile of Y_t given I_{t-1}

as the measurable function q_α satisfying the conditional restriction

$$P(Y_t \leq q_\alpha(I_{t-1}) \mid I_{t-1}) = \alpha, \text{ almost surely (a.s.).} \tag{1}$$

In *parametric* quantile regression modeling one assumes the existence of a family of functions $\mathcal{M} = \{m(\cdot, \theta(\alpha)) : \theta(\cdot) : \mathcal{T} \to \Theta \subset \mathbb{R}^p\}$, where $\mathcal{T}$ is a compact set which comprises the range of quantiles of interest, $\mathcal{T} \subset [0,1]$, and one proceeds to make inference on θ or to test if $q_\cdot \in \mathcal{M}$, i.e., if there exists some $\theta_0 : \mathcal{T} \to \Theta$ such that $m(\cdot, \theta_0(\alpha)) = q_\alpha(\cdot)$ a.s. *for all* $\alpha \in \mathcal{T}$[1].

Leading examples of specifications $\mathcal{M}$ are the Linear Quantile Regression (LQR) model

$$m(I_{t-1}, \theta_0(\alpha)) \equiv m(Z_t, \theta_0(\alpha)) = Z_t'\theta_0(\alpha), \qquad \alpha \in \mathcal{T},$$

with the *location-scale* regression model as the prominent example, in which $\theta_0(\alpha) = (\beta_0, \gamma_0 F_0^{-1}(\alpha)) \in \Theta \subset \mathbb{R}^p$, and where $F_0^{-1}(\alpha)$ denotes a univariate quantile function, see, e.g., Koenker and Xiao (2002), or the Linear Quantile Autoregression model of order s (LQAR(s)),

$$m(I_{t-1}, \theta_0(\alpha)) \equiv m(W_{t-1}, \theta_0(\alpha)) = \theta_{01}(\alpha) + W_{t-1}'\theta_{02}(\alpha), \qquad \theta_0(\alpha) = (\theta_{01}(\alpha), \theta_{02}'(\alpha))',$$

which results, for instance, from the random coefficient model

$$Y_t = \theta_{01}(U_t) + W_{t-1}'\theta_{02}(U_t), \tag{2}$$

where $\theta_{01}(\cdot)$ and $\theta_{02}(\cdot)$ are such that the right hand side of (2) is monotone increasing in U_t, and $\{U_t\}$ are independent and identically distributed (*iid*) $U[0,1]$ random variables; see Koenker and Xiao (2006) for inferences on the LQAR(s) model.

Much effort has been devoted to inferences on $\theta_0(\alpha)$ for the aforementioned models based on the associated quantile processes $Q_n(\alpha) := \sqrt{n}\,(\theta_n(\alpha) - \theta_0(\alpha))$, for $\theta_n(\alpha)$ a $\sqrt{n}$-consistent estimator of $\theta_0(\alpha)$. It is well-known, however, that inferences based on $Q_n(\alpha)$ will heavily depend on the correct specification of the parametric quantile regression model. Although there exist some works on quantile regression model checks, to the best of our knowledge no consistent test for $q_\cdot \in \mathcal{M}$ has been proposed. The existing literature has been mostly limited to *iid* observations, linear models, and to a fixed quantile level $\alpha \equiv \alpha_0 \in (0,1)$. Zheng (1998) has proposed a quantile regression specification test based on kernel smoothing estimators of the conditional moment $E[1(Y_t \leq m(I_{t-1}, \theta_0(\alpha_0))) - \alpha_0 \mid I_{t-1}]$; see also Horowitz and Spokoiny (2002) for the median function (i.e., $\alpha_0 = 0.5$). Recently, Whang (2005), using empirical likelihood methods, proposed a specification test for quantile regression and censored quantile regression for *iid* data. Tests based

[1] We can actually take $\mathcal{T} = [0,1]$ in our theory provided the centered estimator $\sqrt{n}(\theta_n - \theta_0)$ is asymptotically tight on the whole interval $[0,1]$. To the best of our knowledge, such result is, however, not available in the literature for most popular estimators. Thus, we do not pursue such generality in this paper and we restrict our analysis to $\mathcal{T} \subset [0,1]$, in accordance with the econometrics literature.

on smoothers usually have known asymptotic null distributions after an appropriate choice of the bandwidth sequence, but they are not consistent against Pitman's local alternatives.

Using an integrated approach, Bierens and Ginther (2001) proposed a diagnostic test for a linear quantile regression. These authors consider *iid* observations and do not take into account the uncertainty due to parameter estimation. Their test is consistent against $n^{-1/2}$ local alternatives, with n the sample size, but it relies on an upper bound on the asymptotic critical value, which might be too conservative. To solve this deficiency, Whang (2004) considers a subsampling approach to approximate the asymptotic critical values. Koul and Stute (1999) introduced asymptotic pivotal tests for parametric conditional quantiles of first-order nonlinear autoregressive processes. To obtain the pivotal property of the test they use a martingale transform (cf. Khmaladze, 1981). Alternatively, He and Zhu (2003) develop a bootstrap-based test for linear and nonlinear quantile regressions. Our paper also contributes to this literature of specification tests for a unique quantile, since our methods trivially apply to the unique quantile case in a more general framework than these aforementioned works. By extending the scope of conditional quantile specifications to a, possibly, continuum of quantiles we provide a very flexible specification procedure.

In the present article we propose omnibus tests for $q_{\cdot} \in \mathcal{M}$ that are valid for general linear and nonlinear quantile models under time series. Our tests are based on the fact that $q_{\cdot} \in \mathcal{M}$ is characterized by the *infinite* number of conditional moment restrictions

$$E[1(Y_t \leq m(I_{t-1}, \theta_0(\alpha))) - \alpha \mid I_{t-1}] = 0 \text{ a.s. for some } \theta_0(\cdot) : \mathcal{T} \rightarrow \Theta \subset \mathbb{R}^p, \forall \alpha \in \mathcal{T}. \quad (3)$$

The proposed tests are functionals of a quantile-marked empirical process that characterizes condition (3). The asymptotic theory is largely complicated by the fact that (3) involves an infinite number of conditional moment restrictions, indexed by $\alpha \in \mathcal{T}$. We solve this technical difficulty using delicate weak convergence results for empirical processes under martingale conditions. It turns out that the asymptotic null distributions of test statistics depend on the specification under the null and the DGP. Therefore, we propose to implement the test with the assistance of the subsampling.

The rest of the article is organized as follows. In Section 2 we introduce the quantile-marked process, which is the basis upon which the new test statistics for testing (3) are developed. We study the asymptotic distribution of the proposed tests under the null, fixed and local alternatives. In Section 3 a subsampling procedure for approximating the asymptotic null distribution of tests is considered. In Section 4 we present a simulation exercise assessing the finite-sample performance of tests. Finally, in Section 5 an application to some European stock indexes provides evidence that our methodology can serve as powerful and flexible alternative to standard backtesting procedures in evaluating market risk. Proofs are deferred to an appendix. Throughout the article A^c and $|A|$ denote the complex conjugate and Euclidean norm of A, respectively. In the sequel C is a generic constant that may change from one expression to another. All limits are taken as $n \rightarrow \infty$.

2. TEST STATISTICS AND ASYMPTOTIC THEORY

We aim to test the null hypothesis

$$H_0 : E[\Psi_\alpha(Y_t - m(I_{t-1}, \theta_0)) \mid I_{t-1}] = 0 \text{ a.s. for some } \theta_0 \in \mathcal{B} \text{ and for all } \alpha \in \mathcal{T},$$

against the nonparametric alternatives

$$H_A : P(E[\Psi_\alpha(Y_t - m(I_{t-1}, \theta(\alpha))) \mid I_{t-1}] \neq 0) > 0, \text{ for some } \alpha \in \mathcal{T} \text{ and for all } \theta(\alpha) \in \Theta \subset \mathbb{R}^p,$$

where $\Psi_\alpha(\varepsilon) = 1(\varepsilon \leq 0) - \alpha$, and $\mathcal{B}$ is a family of uniformly bounded functions from $\mathcal{T}$ to $\Theta \subset \mathbb{R}^p$. To simplify notation denote $\Psi_{\alpha,t}(\theta) \equiv \Psi_\alpha(Y_t - m(I_{t-1}, \theta))$ and $m_{t-1}(\theta) \equiv m(I_{t-1}, \theta)$. Note that under H_0 (and a mild continuity condition), $m_{t-1}(\theta_0)$ is identified as the α-th quantile of the conditional distribution of Y_t given I_{t-1}, for all $\alpha \in \mathcal{T}$. Testing for H_0 is a challenging testing problem since it involves an infinite number of non-smooth conditional moments parametrized by $\alpha \in \mathcal{T}$.

Our first aim is to characterize H_0 by the infinite number of *unconditional* moment restrictions

$$E[\Psi_{\alpha,t}(\theta_0) \exp(ix' I_{t-1})] = 0, \ \forall x \in \mathbb{R}^d, \text{ for some } \theta_0 \in \mathcal{B} \text{ and for all } \alpha \in \mathcal{T}, \tag{4}$$

where $i = \sqrt{-1}$ is the imaginary unit; see Bierens (1982). Instead of the exponential function we may also use, e.g., any of the parametric families considered in Escanciano (2006).

Given a sample $\{(Y_t, I'_{t-1})' : 1 \leq t \leq n\}$ and a parameter value $\theta \in \mathcal{B}$, we consider the quantile-marked empirical process indexed by $x \in \mathbb{R}^d$, $\alpha \in \mathcal{T}$ and $\theta \in \mathcal{B}$,

$$S_n(x, \alpha, \theta) := n^{-1/2} \sum_{t=1}^{n} \Psi_{\alpha,t}(\theta) \exp(ix' I_{t-1}).$$

Associated to S_n are the quantile-marked *error* and *residual* processes, respectively, defined by

$$R_n(x, \alpha) \equiv S_n(x, \alpha, \theta_0) \qquad \text{and} \qquad R_n^1(x, \alpha) \equiv S_n(x, \alpha, \theta_n),$$

for a $\sqrt{n}-$consistent estimator $\theta_n(\alpha)$ of $\theta_0(\alpha)$, say. The null hypothesis is likely to hold when the process $R_n^1(x, \alpha)$ is close to zero for almost all $(x', \alpha)' \in \mathbb{R}^d \times \mathcal{T}$.

The most popular estimator of θ_0 is the Quantile Regression Estimator (QRE), initially proposed by Koenker and Basset (1978) for linear models, and subsequently generalized to other frameworks by numerous authors, see references below. The QRE is defined as any solution $\theta_{KB,n}(\alpha)$ minimizing

$$\beta \longmapsto \sum_{t=1}^{n} \rho_\alpha(Y_t - m(I_{t-1}, \beta))$$

with respect to $\beta \in \Theta \subset \mathbb{R}^p$, where $\rho_\alpha(\varepsilon) = -\Psi_\alpha(\varepsilon)\,\varepsilon$. Koenker and Park (1996) discussed the existence of $\theta_{KB,n}(\alpha)$ and an interior point algorithm for its computation.

Basset and Koenker (1978) proved the consistency and asymptotic normality of $\theta_{KB,n}(\alpha)$ in the linear regression model, including the least absolute deviation estimator, see also Pollard (1991). The asymptotic theory for $Q_n(\cdot) = \sqrt{n}(\theta_{KB,n}(\cdot) - \theta_0(\cdot))$ as a process indexed by the parameter $\alpha \in \mathcal{T}$, has been considered, among others, in Gutenbrunner and Jurečkova (1992) and Gutenbrunner, Jurečkova, Koenker and Portnoy (1993) for linear models, in Koul and Saleh (1994) and Jurečkova and Hallin (1999) for linear autoregressions, and in Mukherjee (1999) for nonlinear autoregressions. For early contributions see Portnoy (1984). In the present article we do not restrict ourselves to $\theta_{KB,n}$ and we consider any estimator θ_n satisfying some mild conditions, see A4 below. For instance, our results apply to the Quasi-Maximum Likelihood Estimator in Komunjer (2005).

The process R_n^1 is a mapping from $(\Omega, \mathcal{A}, P)$ with values in $\ell^\infty(\Pi)$, where $\ell^\infty(\Pi)$ is the space of all complex-valued functions that are uniformly bounded on Π, with $\Pi := \Upsilon \times \mathcal{T}$, and Υ a generic compact subset of $\mathbb{R}^d$ containing the origin. The space $\ell^\infty(\Pi)$ is furnished with the supremum metric, say d_∞, and let $\mathcal{B}_{d_\infty}$ be the corresponding Borel σ-algebra. Let $\Longrightarrow$ denote weak convergence on $(\ell^\infty(\Pi), \mathcal{B}_{d_\infty})$ in the sense of J. Hoffmann-Jørgensen, see, e.g., Dudley (1999, p. 94), or Definition 1.3.3 in van der Vaart and Wellner (1996). Since Υ is generic, $\Longrightarrow$ is indeed weak convergence on compacta.

After (4), test statistics are based on a distance from the standardized sample analogue of $E[\Psi_{\alpha,t}(\theta_0)\exp(ix'I_{t-1})]$ to zero, i.e., on a norm of R_n^1, say $\Gamma(R_n^1)$. A popular norm is the Cramér-von Mises (CvM) functional

$$CvM_n := \int_{\Pi} \left|R_n^1(x,\alpha)\right|^2 d\Phi(x)dW(\alpha), \tag{5}$$

where Φ and W are some integrating measures on Υ and $\mathcal{T}$, respectively. Other continuous (with respect to d_∞) functionals Γ from $\ell^\infty(\Pi)$ to $\mathbb{R}$ are of course possible. For instance, we can consider tests combining sup- and L_2-norms, as in the Kolmogorov-type (K) functional

$$KS_n := \sup_{\alpha\in\mathcal{T}} \int_{\Upsilon} \left|R_n^1(x,\alpha)\right|^2 d\Phi(x). \tag{6}$$

Then, the omnibus tests we proposed in this article reject the null hypothesis H_0 for "large" values of $\Gamma(R_n^1)$. Practical issues about the computation of the test statistics CvM_n and KS_n are discussed in Section 4.

2.1 Asymptotic null distribution.

In this subsection we establish the limit distribution of the quantile-marked empirical process R_n^1 under the null hypothesis H_0. The null limit distributions of the tests are the limit distributions of some continuous functionals of R_n^1. To derive asymptotic results we consider the following notation and assumptions. Throughout the paper the family $\mathcal{B}$, in which the parameter θ_0 takes values, is endowed with the sup norm, i.e., $\|\theta\|_{\mathcal{B}} = \sup_{\alpha\in\mathcal{T}} |\theta(\alpha)|$. Let, for each $t \in \mathbb{Z}$, $\mathcal{F}_t = \sigma(I_t', I_{t-1}', ...)$, be the σ-field generated by the information set obtained up to time t. Define for each $t \in \mathbb{Z}$, the quantile "innovation" $\varepsilon_{t,\alpha} := Y_t - q_\alpha(I_{t-1})$ and the parametric quantile "error" $e_t(\theta) \equiv e_t(\theta(\alpha)) := Y_t - m(I_{t-1}, \theta(\alpha))$. Define also the family of conditional distributions

$$F_x(y) := P(Y_t \leq y \mid I_{t-1} = x). \tag{7}$$

Let f_x be the density function of the cumulative distribution function (cdf) F_x. Let $N_{[\cdot]}(\delta, \mathcal{H}, \|\cdot\|)$ be the δ-bracketing number of a class of functions $\mathcal{H}$ with respect to a norm $\|\cdot\|$, i.e., the smallest number r such that there exist $f_1, ..., f_r$ and $\Delta_1, ..., \Delta_r$ such that $\max_{1\leq i\leq r} \|\Delta_i\| < \delta$ and for all $f \in \mathcal{H}$, there exists an $1 \leq i \leq r$ such that $\|f - f_i\| < \Delta_i$, see Definition 2.1.6 in van der Vaart and Wellner (1996).

Assumption A1:

A1(a): $\{(Y_t, Z_t')' : t = 0, \pm 1, \pm 2, ...\}$ is a strictly stationary and erdogic process. Under H_0, $\{\Psi_{\alpha,t}(\theta_0), \mathcal{F}_t\}$ is a martingale difference sequence for all $\alpha \in \mathcal{T}$.

A1(b): The parametric family $m(\cdot, \theta_0(\alpha))$ is nondecreasing in α a.s.

A1(c): $E[|I_0|^2] < C$.

A1(d): The family of distributions functions $\{F_x, x \in \mathbb{R}^d\}$ has Lebesgue densities $\{f_x, x \in \mathbb{R}^d\}$ that are uniformly bounded,

$$\sup_{x\in\mathbb{R}^d, y\in\mathbb{R}} |f_x(y)| \leq C,$$

and equicontinuous: for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\sup_{x\in\mathbb{R}^d, |y-z|\leq\delta} |f_x(y) - f_x(z)| \leq \epsilon.$$

Assumption A2: For each $\theta_1 \in \mathcal{B}$,

A2(a): There exists a vector of functions $g_{t-1} : \Theta \to \mathbb{R}^q$ such that $g_{t-1}(\theta_1(\alpha))$ is $\mathcal{F}_{t-1}$-measurable for each $t \in \mathbb{Z}$, and satisfies, for all $k < \infty$,

$$\sup_{1\leq t\leq n, \|\theta_1-\theta_2\|_{\mathcal{B}}\leq kn^{-1/2}} n^{1/2} \|m_{t-1}(\theta_2) - m_{t-1}(\theta_1) - (\theta_2 - \theta_1)' g_{t-1}(\theta_1)\|_{\mathcal{B}} = o_P(1).$$

A2(b): For a sufficiently small $\delta > 0$,

$$E\left[\sup_{\|\theta_1-\theta_2\|_{\mathcal{B}}\leq\delta} |1(Y_t \leq m_{t-1}(\theta_1(\alpha))) - 1(Y_t \leq m_{t-1}(\theta_2(\alpha)))|\right] \leq C\delta, \ \forall\alpha \in \mathcal{T} \text{ and}$$

$$E\left[\sup_{|\alpha_1-\alpha_2|\leq\delta} |m_{t-1}(\theta_1(\alpha_1)) - m_{t-1}(\theta_1(\alpha_2))|\right] \leq C\delta.$$

A2(c): Uniformly in $\alpha \in \mathcal{T}$, $E\left|g_{t-1}(\theta_1(\alpha))\right|^2 < \infty$, and uniformly in $(x', \alpha)' \in \Pi$,

$$\left|\frac{1}{n}\sum_{t=1}^{n} g_{t-1}(\theta_0(\alpha)) \exp(ix' I_{t-1}) f_{I_{t-1}}(m_{t-1}(\theta_0)) - E\left[g_{t-1}(\theta_0(\alpha)) \exp(ix' I_{t-1}) f_{I_{t-1}}(m_{t-1}(\theta_0))\right]\right| = o_P(1).$$

Assumption A3: The parametric space Θ is compact in $\mathbb{R}^p$. The true parameter $\theta_0(\alpha)$ belongs to the interior of Θ for each $\alpha \in \mathcal{T}$, and $\theta_0 \in \mathcal{B}$. The class $\mathcal{B}$ satisfies

$$\int_0^\infty \left(\log(N_{[\cdot]}(\delta^2, \mathcal{B}, \|\cdot\|_{\mathcal{B}}))\right)^{1/2} d\delta < \infty.$$

Assumption A4: The estimator $\theta_n \in \mathcal{B}$, for all n sufficiently large, and satisfies the following asymptotic expansion under H_0 uniformly in $\alpha \in \mathcal{T}$,

$$Q_n(\alpha) = \sqrt{n}(\theta_n(\alpha) - \theta_0(\alpha)) = \frac{1}{\sqrt{n}} \sum_{t=1}^n l_\alpha(Y_t, I_{t-1}, \theta_0(\alpha)) + o_P(1),$$

where $l_\alpha(\cdot)$ is such that $E[l_\alpha(Y_1, I_0, \theta_0(\alpha))] = 0$, $L_\alpha(\theta_0(\alpha)) = E[l_\alpha(Y_1, I_0, \theta_0(\alpha))l'_\alpha(Y_1, I_0, \theta_0(\alpha))]$ exists and is positive definite, and $E[l_\alpha(Y_t, I_{t-1}, \theta_0(\alpha))\Psi_\alpha(Y_s - m(I_{s-1}, \theta_0(\alpha)))] = 0$ if $t \neq s$. Furthermore, as a process in $\ell^\infty(\mathcal{T})$, $Q_n(\alpha)$ converges weakly to a Gaussian process $Q(\cdot)$ with zero mean and covariance function

$$K_Q(\alpha_1, \alpha_2) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n E[l_{\alpha_1}(Y_t, I_{t-1}, \theta_0(\alpha_1))l_{\alpha_2}(Y_s, I_{s-1}, \theta_0(\alpha_2))].$$

Assumption A1(a) is standard in the model checks literature under time series, see, e.g., Bierens and Ploberger (1997). A1(b) is natural in the present context. A1(c) is needed to prove the equicontinuity of the limit process of R_n and can be avoided using $\exp(ix'\phi(I_{t-1}))$, with $\phi(\cdot)$ a one-to-one bounded mapping (see e.g. Bierens and Ginther, 2001), instead of $\exp(ix'I_{t-1})$. A1(d) is necessary for the asymptotic tightness of the process R_n^1 and is required in Koul and Stute (1999). Assumptions A2(a)-A2(c) are classical in inference about nonlinear models, see Koul's (2002) monograph. A2 is satisfied for all models considered in the literature under mild moment assumptions, e.g. LQR and LQAR models. Conditions for the satisfaction of A3 can be found in van der Vaart and Wellner (1996), see e.g. their Theorem 2.7.5 for monotone classes of functions which applies to LQAR models. The condition $\theta_n \in \mathcal{B}$, for all n sufficiently large, can be weakened to $P(\theta_n \in \mathcal{B}) \to 1$ as $n \to \infty$, at the cost of complicating the proofs. A4 has been established in the literature under a variety of conditions and different models and DGP's, see, for instance, Theorem 1 in Gutenbrunner and Jurečkova (1992) or Theorem 3.2 in Mukherjee (1999). For nonlinear models with *iid* innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$ distributed as F_ε, Mukherjee (1999) proved A4 for $\theta_{KB,n}(\alpha)$. Under some mild additional assumptions, including that $\Sigma_{\theta_0(\alpha)} := E\left[g\left(I_1, \theta_0(\alpha)\right) g\left(I_1, \theta_0(\alpha)\right)'\right]$ exists and is positive definite, Mukherjee (1999) showed that A4 holds for the QRE under H_0 with

$$l_\alpha(Y_t, I_{t-1}, \theta_0(\alpha)) = -\frac{\Sigma^{-1}_{\theta_0(\alpha)} g(I_{t-1}, \theta_0(\alpha))\Psi_\alpha(\varepsilon_t)}{q(\alpha)},$$

where $q(\alpha) = f_\varepsilon(F_\varepsilon^{-1}(\alpha))$ is the reciprocal of the sparsity function and f_ε is the density of F_ε. The quantile limit process $Q(\cdot)$ in that case is $\Sigma_{\theta_0(\cdot)}^{-1} W(\cdot)/q(\cdot)$, where $W(\cdot)$ denotes a vector of p independent Brownian bridges on $\mathcal{T}$.

We establish now the limit distribution of R_n. Under A1(a) and H_0, because $R_n(v)$ is a zero-mean square-integrable martingale for each $v = (x', \alpha)' \in \Pi$, using a suitable Central Limit Theorem (CLT) for stationary ergodic martingale difference sequences, cf. Billingsley (1961), we have that the finite-dimensional distributions of R_n converge to those of a multivariate normal distribution with a zero mean vector and variance-covariance matrix given by the covariance function

$$K_\infty(v_1, v_2) = (\alpha_1 \wedge \alpha_2 - \alpha_1\alpha_2) E[\exp(i(x_1 - x_2)' I_0)], \tag{8}$$

where from now on $v_1 = (x_1', \alpha_1)'$ and $v_2 = (x_2', \alpha_2)'$ represent generic elements of Π, and $\wedge$ denotes the minimum, i.e., $a \wedge b = \min\{a, b\}$. The next result is an extension of the convergence of the finite-dimensional distributions of R_n to weak convergence in the space $\ell^\infty(\Pi)$. We stress that no mixing conditions are required for the weak convergence to hold.

Theorem 1: *Under the null hypothesis H_0 and Assumptions A1(a-c)*

$$R_n \Longrightarrow R_\infty,$$

where R_∞ is a Gaussian process with zero mean and covariance function (8).

In practice, θ_0 is unknown and has to be estimated from a sample $\{(Y_t, I_{t-1}')' : 1 \leq t \leq n\}$ by an estimator θ_n. When we replace θ_0 in R_n by θ_n, resulting in R_n^1, we need to investigate how the estimation error will affect the asymptotic properties of R_n^1. The next result shows this effect on the asymptotic null distribution of R_n^1. Define the function

$$G(x, \theta_0(\alpha)) := E[g_{t-1}(\theta_0(\alpha)) f_{I_{t-1}}(m_{t-1}(\theta_0)) \exp(ix' I_{t-1})], \qquad x \in \Upsilon,\ \alpha \in \mathcal{T}.$$

Theorem 2: *Under the null hypothesis H_0 and Assumptions A1-A4*

$$\sup_{x \in \Upsilon, \alpha \in \mathcal{T}} \left| R_n^1(x, \alpha) - R_n(x, \alpha) + G'(x, \theta_0(\alpha)) n^{-1/2} \sum_{t=1}^{n} l_\alpha(Y_t, I_{t-1}, \theta_0(\alpha)) \right| = o_P(1).$$

As a consequence, we obtain the following corollary.

Corollary 1: *Under the assumptions of Theorem 2*

$$R_n^1 \Longrightarrow R_\infty^1,$$

where $R_\infty^1(\cdot) = R_\infty(\cdot) - G'(\cdot, \theta_0(\cdot)) Q(\cdot)$ (in distribution).

Now, using the last corollary and the Continuous Mapping Theorem (CMT) we obtain the asymptotic null distribution of continuous functionals such as CvM_n and KS_n.

Corollary 2: *Under the assumptions of Theorem 2, for any continuous functional* $\Gamma(\cdot)$ *from* $\ell^\infty(\Pi)$ *to* $\mathbb{R}$,

$$\Gamma(R_n^1) \xrightarrow{d} \Gamma(R_\infty^1).$$

2.2 Consistency and Pitman's local alternatives.

In this section we study the consistency properties of tests based on functionals $\Gamma(R_n^1)$. First, we show that these tests are consistent against all fixed alternatives provided a mild regularity condition is satisfied.

Assumption A5: Under H_A, (i) there exists a $\theta_1 \in \mathcal{B}$ such that $\|\theta_n - \theta_1\|_{\mathcal{B}} = o_P(1)$; (ii) $E[\Psi_\cdot(e_t(\theta_1(\cdot)))\exp(i \cdot I_{t-1})]$ is different from zero in a subset with positive Lebesgue measure on Π.

See Kim and White (2003) for conditions on $\theta_{KB,n}$ to satisfy Assumption A5(i), see also Section 3 in Angrist, Chernozhukov and Fernández-Val (2006). A sufficient condition for A5(ii) is that I_{t-1} is bounded. Notice that this condition always holds if we replace I_{t-1} by $\phi(I_{t-1})$, with ϕ a one-to-one bounded mapping, as in Bierens and Ginther (2001). Henceforth, almost sure convergence of nonmesurable maps is understood, as usual, as outer almost sure convergence, see van der Vaart and Wellner (1996) for definitions.

Theorem 3: *Under the alternative hypothesis* H_A *and Assumptions A1, A2, A3 and A5,*

$$n^{-1/2}R_n^1(\cdot) \xrightarrow{a.s} E[\Psi_\cdot(e_t(\theta_1(\cdot)))\exp(i \cdot I_{t-1})].$$

A consequence of Theorem 3 and the CMT is that (under the assumptions of Theorem 3),

$$\int_\Pi \left|n^{-1/2}R_n^1(x,\alpha)\right|^2 d\Phi(x)dW(\alpha) \xrightarrow{P} \int_\Pi \left|E[\Psi_\alpha(e_t(\theta_1(\alpha)))\exp(ix'I_{t-1})]\right|^2 d\Phi(x)dW(\alpha) > 0,$$

provided that Φ and W are absolute continuous with respect to the Lebesgue measure on Π. In such a situation, the test statistic CvM_n will diverge to $+\infty$ under any fixed alternative, and the test will be consistent against all directions in the alternative hypothesis.

Now we analyze the asymptotic distribution of R_n^1 under a sequence of local alternatives converging to null at a parametric rate $n^{-1/2}$. We consider the DGP generating the local alternatives

$$H_{A,n} : E[\Psi_\alpha(Y_t - m_{t-1}(\theta_0)) \mid \mathcal{F}_{t-1}] = \frac{a_\alpha(I_{t-1})}{n^{1/2}} \text{ a.s. for some } \theta_0 \in \mathcal{B} \text{ and for all } \alpha \in \mathcal{T}, \quad (9)$$

where the function $a_\alpha(\cdot) : \mathbb{R}^d \longrightarrow \mathbb{R}$ satisfies the following assumption.

Assumption A6: $a_\alpha(\cdot)$ is such that $E\sup_{\alpha\in\mathcal{T}} |a_\alpha(I_{t-1})| < \infty$. There exists a $\mathcal{F}_{t-1}$-measurable r.v. C_{t-1} with $E[C_{t-1}^2] < \infty$, such that for all $t \in \mathbb{Z}$ and for all $\alpha_1, \alpha_2 \in \mathcal{T}$,

$$|a_{\alpha_1}(I_{t-1}) - a_{\alpha_2}(I_{t-1})| \leq C_{t-1} |\alpha_1 - \alpha_2| \,, \text{ a.s.}$$

To derive the next result we need the following assumption on the behaviour of the estimator under the local alternatives.

Assumption A4': The estimator $\theta_n(\alpha)$ satisfies the following asymptotic expansion under $H_{A,n}$, uniformly in α,

$$\sqrt{n}(\theta_n(\alpha) - \theta_0(\alpha)) = \xi_a(\alpha) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n} l_\alpha(Y_t, I_{t-1}, \theta_0(\alpha)) + o_P(1),$$

where the function $l_\alpha(\cdot)$ is as in A4 and $\xi_a(\alpha) \in \mathbb{R}^p$ for each $\alpha \in \mathcal{T}$.

Assumption A4' holds for most estimators considered in the literature. For instance, in the nonlinear time series context of Mukherjee (1999), the corresponding term $\xi_a(\alpha)$ to $\theta_{KB,n}(\alpha)$ is

$$\xi_a(\alpha) = -q^{-1}(\alpha)\Sigma^{-1}_{\theta_0(\alpha)} E[f_{I_{t-1}}(m_{t-1}(\theta_0))g_{t-1}(\theta_0)a_\alpha(I_{t-1})].$$

The shift in charge of local power against alternatives in $H_{A,n}$ is given by

$$D_a(x, \theta_0(\alpha), \alpha) := E[a_\alpha(I_0)\exp(ix'I_0)] - \xi_a'(\alpha)G(x, \theta_0(\alpha)).$$

THEOREM 4: *Under the local alternatives (9), Assumptions A1-A3, A6 and A4'*

$$R_n^1 \Longrightarrow R_\infty^1 + D_a,$$

where R_∞^1 is the process defined in Theorem 2.

It is not difficult to show that

$$D_a \equiv 0 \text{ a.e.} \Longleftrightarrow a_\alpha(I_{t-1}) = \xi_a'(\alpha)g(I_{t-1}, \theta_0(\alpha)) \text{ for all } \alpha \in \mathcal{T} \text{ a.s.}$$

Therefore, for directions $a_\alpha(\cdot)$ not collinear to the score $g(I_{t-1}, \theta_0(\alpha))$, the shift function D_a is non-trivial and test statistics based on $\Gamma(R_n^1)$ for a symmetric functional Γ are asymptotically strictly unbiased against the local alternatives (9); see Escanciano (2008).

3. SUBSAMPLING APPROXIMATION

We have seen before that the asymptotic null distribution of continuous functionals of R_n^1 depends in a complex way of the DGP and the specification under the null. Therefore, critical values for the test statistics can not be tabulated for general cases. In this section we overcome this problem with the assistance of the subsampling methodology. Resampling methods have been used extensively in the literature of quantile regression models, see, e.g., Hahn (1995), Horowitz (1998), Bilias, Chen and Ying (2000), Sakov and Bickel (2000) or He and Hu (2002). These articles consider *iid* sequences. When time series are involved the bootstrap approximation becomes more challenging. Subsampling is a powerful resampling scheme that allows an asymptotically valid inference under very general conditions on the DGP, see the monograph by Politis, Romano and Wolf (1999). Chernozhukov (2002) and Whang (2004) considered subsampling approximation for LQR model checks. In this section we apply the subsampling methodology to approximate the critical values of continuous functionals of R_n^1. With an abuse of notation we write the test statistic as a function of the data $\{X_t = (Y_t, Z_{t+1}')' : t = 0, \pm 1, \pm 2, ...\}$, $\Gamma(R_n^1) = \Gamma(R_n^1(X_1, ..., X_n))$. Let $G_n^\Gamma(w)$ be the test statistic's cdf,

$$G_n^\Gamma(w) = P(\Gamma(R_n^1) \leq w).$$

We describe the subsampling approximation for the time series case; see the aforementioned references for *iid* sequences. Let $\Gamma(R_{b,i}^1) = \Gamma(R_b^1(X_i, ..., X_{i+b-1}))$ be the test statistic computed with the subsample $(X_i, ..., X_{i+b-1})$ of size b. We note that each subsample of size b (taken without replacement from the original data) is indeed a sample of size b from the true DGP. Hence, it is clear that one can approximate the sampling distribution $G_n^\Gamma(w)$ using the distribution of the values of $\Gamma(R_{b,i}^1)$ computed over the $n - b + 1$ different subsamples of size b (or the $\binom{n}{b}$ different subsamples of size b in the cross-section case). That is, we approximate $G_n^\Gamma(w)$ by

$$G_{n,b}^\Gamma(w) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1(\Gamma(R_{b,i}^1) \leq w), \qquad w \in [0, \infty). \tag{10}$$

Let $c_{n,1-\tau,b}^\Gamma$ be the $(1-\tau)$-th sample quantile of $G_{n,b}^\Gamma(w)$, i.e.,

$$c_{n,1-\tau,b}^\Gamma = \inf\{w : G_{n,b}^\Gamma(w) \geq 1 - \tau\}.$$

Thus, our subsampling tests reject the null hypothesis if $\Gamma(R_n^1) > c_{n,1-\tau,b}^\Gamma$. Let $c_{1-\tau}^\Gamma$ be the $(1-\tau)$-th quantile of $G_\infty^\Gamma(w) = P(\Gamma(R_\infty^1) \leq w)$. To justify theoretically this resampling approximation we need an additional assumption on the serial dependence of the DGP. Define the α-mixing coefficients as

$$\alpha(m) = \sup_{n \in \mathbb{Z}} \sup_{B \in \mathcal{F}_n, A \in \mathcal{P}_{n+m}} |P(A \cap B) - P(A)P(B)|, \; m \geq 1$$

where the σ-fields $\mathcal{F}_n$ and $\mathcal{P}_n$ are $\mathcal{F}_n := \sigma(X_t, t \leq n)$ and $\mathcal{P}_n := \sigma(X_t, t \geq n)$, respectively, with $X_t = (Y_t, Z_{t+1}')'$.

Assumption A7: $\{X_t = (Y_t, Z_{t+1}')' : t = 0, \pm 1, \pm 2, ...\}$ is a strictly stationary strong mixing process with α-mixing coefficients satisfying

$$\sum_{m=1}^{n} \alpha(m) = o(n).$$

The mixing assumption in A6 is sufficient but not necessary for the validity of the subsampling, see Politis, Romano and Wolf (1999). This subsampling procedure allows us to approximate the asymptotic critical values of the tests based on $\Gamma(R_{n,w}^1)$. The next result justifies theoretically the subsampling approximation. Its proof follows closely that of Theorem 2 in Whang (2004).

Theorem 5: *Assume Assumptions A1-A7 and that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$. Then,*

(i) Under the null hypothesis H_0,

$$c_{n,1-\tau,b}^{\Gamma} \xrightarrow{P} c_{1-\tau}^{\Gamma}.$$

and

$$P(\Gamma(R_n^1) > c_{n,1-\tau,b}^{\Gamma}) \longrightarrow \tau.$$

(ii) Under any fixed alternative hypothesis,

$$P(\Gamma(R_n^1) > c_{n,1-\tau,b}^{\Gamma}) \longrightarrow 1.$$

(iii) Under the local alternatives (9),

$$P(\Gamma(R_n^1) > c_{n,1-\tau,b}^{\Gamma}) \longrightarrow P(\Gamma(R_\infty^1 + D_a) > c_{1-\tau}^{\Gamma}).$$

Theorem 5 implies that the proposed subsampling tests have a correct asymptotic level, are consistent and are able to detect alternatives tending to the null at the parametric rate $n^{-1/2}$. An appealing property of our subsampling tests is that they do not need estimation of the nonparametric (conditional) sparsity function $f_{I_{t-1}}(m_{t-1}(\theta_0))$, which results in a substantial simplification of the tests. In practice, the empirical size and power of the tests depend on the choice of the parameter b. For this choice the reader is referred to Politis, Romano and Wolf (1999) or Sakov and Bickel (2000). In the present article, we follow the suggestion of Sakov and Bickel (2000) and we chose $b = \lfloor kn^{2/5} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part, which yields the optimal minimax accuracy under certain conditions. Section 5 below shows that this resampling procedure provides good approximations in finite samples for a variety of values for k.

It is sometimes argued that some kind of recentering might improve the power performance of subsampling-based tests. We explored two possibilities in our simulations below. First, we may consider to replace $R_{b,i}^1$ in (10) by $R_{b,i}^1 - b^{1/2}n^{-1/2}R_n^1$; see Chernozhukov (2002) for an example of this centering. Alternatively, we may recenter the test statistic $\Gamma(R_{b,i}^1) - \Gamma(b^{1/2}n^{-1/2}R_n^1)$. We found in our simulations below, that with the DGPs considered the power improvement is not significative, although there was a small positive improvement in all cases. This improvement is not without cost;

the empirical size performance became more sensitive to the choice of b and computationally, the test statistic is much more difficult to compute. Of course, these results don't need to hold for other DGPs. Therefore, based on our experience, in applications we recommend to compute the uncentered version for computational reasons; see next section for the computation of the test statistics.

4. FINITE SAMPLE PERFORMANCE

We investigate in this section, by means of a Monte Carlo experiment, the finite sample performance of the proposed tests. The aim is to provide evidence of the good finite-sample performance of the new test statistics.

We describe our simulation setup. The choice of $\Phi(\cdot)$ in (5) is up-to the practitioner and gives flexibility to direct the power against some preferred alternatives. Following Escanciano and Velasco (2006) and references therein, we choose $\Phi(\cdot)$ equal to the $d-$variate standard normal random vector[2]. We consider as W a uniform discrete distribution over a grid of $\mathcal{T}$ in $m = 21$ equidistributed points from $\epsilon = 0.1$ to $1 - \epsilon = 0.9$. Denote by $\mathcal{T}_m = \{\alpha_j\}_{j=1}^m$ the points in the grid, with $\epsilon = \alpha_1 < \cdots < \alpha_m = 1 - \epsilon$. Let $W_{\exp}$ be the $n \times n$ matrix with elements $w_{\exp,t,s} = \exp(-\frac{1}{2}\left|I_{t-1} - I_{s-1}\right|^2)$ and let Ψ be the $n \times m$ matrix with elements $\psi_{ij} = \Psi_{\alpha_j}(Y_i - m(I_{i-1}, \theta_n))$. Hence, the CvM test statistic is computed as

$$CvM_n = m^{-1}\sum_{j=1}^{m} \psi'_{\cdot j} W_{\exp} \psi_{\cdot j}, \tag{11}$$

where $\psi_{\cdot j}$ denotes the j column of Ψ. Therefore, the computation of CvM_n is straightforward. Similarly, we can compute

$$KS_n = \max_{1 \le j \le m} \psi'_{\cdot j} W_{\exp} \psi_{\cdot j} \tag{12}$$

Our theory would allow for $m \to \infty$ as $n \to \infty$ and the $\{\alpha_j\}_{j=1}^m$ generated independently from a distribution on $\mathcal{T}$. For simplicity in the computations we have considered m fixed and $\{\alpha_j\}_{j=1}^m$ deterministic throughout this section.

For the simulations, we examined two data generating processes that have been previously considered in Zheng (1998) and Whang (2004):

$$DGP1 : Y_t = X_{1t} + X_{2t} + c_1 \sigma_t^{3/2} + u_{1t}, \qquad t = 1, \ldots, n,$$

where $\sigma_t = X_{1t}^2 + X_{2t}^2 + X_{1t}X_{2t}$ and X_{1t}, X_{2t} and $u_{1t} \sim iid\ N(0,1)$, mutually independent. The null hypothesis corresponds to the location model with $c_1 = 0$, so the null quantile model is a LQR model

$$m(I_{t-1}, \theta(\alpha)) = Z_t'\theta_0(\alpha), \qquad \alpha \in \mathcal{T},$$

[2] Strictly speaking our present theory does not allow to integrate in whole $\mathbb{R}^d$ in the CvM test, but our theory can be easily adapted, see e.g. Escanciano's (2006) Hilbert space approach, to allow for the present definition of the CvM test. In any case, there is no practical difference.

with $Z_t = (1, X_{1t}, X_{2t})'$ and $\theta_0(\alpha) = (\phi^{-1}(\alpha), 1, 1)'$, with $\phi^{-1}(\alpha)$ the quantile function of the standard normal r.v.

The second design is a time series model:

$$DGP2 : Y_t = 0.6Y_{t-1} + X_t + c_2 X_t^2 + u_{2t}, \qquad t = 1, \ldots, n,$$

where $X_t = 0.5X_{t-1} + \varepsilon_t$ with both u_{2t} and ε_t are sampled independently from $N(0,1)$ and $Y_0 = X_0 = 0$. Here, the null model corresponds to $c_2 = 0$. Under H_0, a LQR model holds with $I_{t-1} = (1, Y_{t-1}, X_t)'$, and $\theta_0(\alpha) = (\phi^{-1}(\alpha), 0.6, 1)'$.

We consider two sample sizes $n = 100$ and $n = 300$ and a quantile interval $[0.1, 0.9]$. As the number of subsamples, we follow the suggestion of Sakov and Bickel (2000) and we chose $b = \lfloor kn^{2/5} \rfloor$, with several choices of k. For DGP1 we consider k from 7 to 9. These values correspond to $b = 42, 48$ and 54 for $n = 100$ and $b = 63, 72$ and 81 for $n = 300$, respectively. For DGP2, k is chosen to be from 3 to 5 ($b =$ 18, 24 and 30 for $n = 100$, and $b =$ 27, 36 and 45 for $n = 300$). We set the number of Monte Carlo repetitions to 1,000. The parameter $\theta_0(\alpha)$ is estimated by the QRE of Koenker and Bassett (1978). In all experiments, the nominal probability of rejecting a correct null hypothesis is 0.05. The results with other nominal values are similar. For simplicity in the computations we consider the same subsampling approximation in the cross-section and time series examples.

Table I and Table II provide the rejection probabilities of the tests for DGP1 and DGP2 for both statistics, respectively. When $c_1 = 0$, the results show that the size performance of the subsampling-based tests is good for all the subsample sizes considered and both statistics. We observe that to achieve appropriate empirical sizes the choice of b for the DGP1 should be larger than for the DGP2. When $c_1 \neq 0$, the results show the power performance of the tests. The rejection probabilities increase as n increases, as expected, showing that the tests are consistent against these fixed alternatives. The CvM test statistic CvM_n has higher power than the Kolmogorov-type test KS_n. The power does not depend substantially on the choice of b. For DGP2 we obtain similar conclusions to those under DGP1. This limited simulation study suggests that even with relative small sample sizes the subsampling tests exhibit fairly good size accuracy and power performance.

Please insert Table I and Table II about here.

Unreported simulations using the indicator weight function $1(I_{t-1} \leq x)$, instead of $\exp(ix'I_{t-1})$, confirm that exponential-based tests have higher power than indicator-based tests for these alternatives. In fact, this was our motivation for the use of the exponential weight in the CvM test. These unreported simulations can be obtained from the authors upon request.

5. APPLICATION TO MARKET RISK MANAGEMENT

The quantification of market risk for derivative pricing, portfolio optimization and pricing risk purposes has generated a large amount of theoretical and practical work. One of the implications of the creation of the Basel Committee on Banking Supervision was the implementation of Value-at-Risk (VaR) as the standard tool for measuring market risk. In financial terms, VaR is the maximum loss on a trading portfolio for a period of time given a confidence level. In statistical terms, VaR is the (conditional) quantile of the conditional distribution of returns on the portfolio given agent's information set. Nowadays, VaR has become a standard risk measure due its universality, conceptual simplicity and easy computation and evaluation.

The evaluation of VaR measures has become of paramount importance in risk management. In fact, for banks with sufficiently highly developed risk management systems the implementation of VaR techniques was a priori the only restriction set by the Basel Accord (1996a) for computing capital reserves. Thus, in order to monitor and assess the accuracy and quality of the different VaR forecasts techniques the Basel Accord (1996a) and the Amendment of Basel Accord (1996b) developed a diagnostic testing procedure that was denominated *backtesting*. To explain formally what backtesting is, let us consider the following implication of (1),

$$E[\Psi_{\alpha,t}(\theta_0) \mid \widetilde{I}_{t-1,\alpha}(\theta_0)] = 0, \text{ a.s. for some } \theta_0(\alpha) \in \Theta \text{ and some } \alpha \in (0,1), \tag{13}$$

where $\widetilde{I}_{t-1,\alpha}(\theta_0) := (\Psi_{\alpha,t-1}(\theta_0), \Psi_{\alpha,t-2}(\theta_0), ...)'$. The popularity of condition (13) is mostly due to the discrete character and ease of interpretation of the variables $\{H_{t,\alpha}(\theta_0)\}$, with $H_{t,\alpha}(\theta_0) = 1(Y_t \leq m(I_{t-1}, \theta_0(\alpha)))$, which are the so-called *hits* or *exceedances*. In particular, the discreteness of the exceedances implies that condition (13) is equivalent to

$$\{H_{t,\alpha}(\theta_0)\} \text{ are } iid \text{ Ber}(\alpha) \text{ random variables } (r.v.) \text{ for some } \theta_0 \in \Theta, \tag{14}$$

where Ber(α) stands for a Bernoulli *r.v.* with parameter α. In the VaR literature, the satisfaction of condition (14) has been taken as the criteria for the out-of-sample evaluation of VaR forecasts, leading to the so-called unconditional backtesting (i.e. tests for $E[H_{t,\alpha}(\theta_0)] = \alpha$) and conditional tests or tests of independence (i.e. tests for $\{H_{t,\alpha}(\theta_0)\}$ being *iid*).

The unconditional backtest is carried out with the so-called Kupiec-test statistic (cf. Kupiec, 1995), see also Christoffersen (1998) and Escanciano and Olmo (2008), based on the absolute value of the standardized sample mean, i.e.

$$K_{n,\alpha} := \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{H_{t,\alpha}(\theta_n) - \alpha\} \right|. \tag{15}$$

For the conditional hypothesis, Christoffersen (1998) introduces a likelihood ratio (LR) test which is equivalent to a test based on the autocovariance

$$C_{n,\alpha} = \left| \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \{H_{t,\alpha}(\theta_n) - \alpha\}\{H_{t-1,\alpha}(\theta_n) - \alpha\} \right| .$$

Berkowitz, Christoffersen and Pelletier (2006) review some of the existing methods for testing the conditional and unconditional hypotheses.

In this paper we propose an alternative methodology to the mentioned classical backtesting methods that overcomes some of their important deficiencies. First, it is important to stress that tests based on R_n^1 are expected to be more powerful than standard backtesting techniques. This is so because we incorporate more (possibly nonlinear) information in the test statistic. In particular, the unconditional backtest statistic coincides with $R_n^1(0, \alpha)$, whereas we exploit a continuum number of $x's$, thereby leading to a more powerful test. This is confirmed in the applications below. Second, by using only one quantile level, VaR only tells us the most we can lose if a tail events does not occur; if a tail event does occur, we can expect to lose more than the VaR, but the VaR itself gives us no indication of how much that might be. Therefore, two positions can have the same VaR at a given quantile level α and yet have very different risk exposures. This is the so-called *tail risk* problem in VaR. Our methodology solves this deficiency by taking a larger, possibly infinite, number of quantiles in the tail, thereby giving a more complete picture of the underlying risk exposure and leading to a better understanding of the fitting properties of the associated risk model.

In this section, we compare the new methodology with the aforementioned standard backtesting techniques. For simplicity in the arguments, we only consider in-sample comparisons. The extension to out-of-sample exercises poses no extra difficulties, and hence it is omitted. The data sets we consider are daily closed European stock indexes returns from the Frankfurt DAX Index (DAX), the London FTSE-100 Index (FTSE) and Paris CAC-40 Index (CAC) from 1 January 2003 to 9 June 2008, with a total of n =1417 observations. We consider the returns of the indexes obtained as the log differences of the data.

We entertain a pure Gaussian GARCH(1,1) model with AR(1) conditional mean for the log-returns Y_t, leading to the quantile model

$$\begin{aligned} m(I_{t-1}, \theta_0(\alpha)) &= \mu + \beta_0 Y_{t-1} + \sigma_t \Phi_\varepsilon^{-1}(\alpha), \\ \text{with } \sigma_t^2 &= \eta_{00} + \eta_{10}(Y_t - \mu - \beta_0 Y_{t-1})^2 + \eta_{20}\sigma_{t-1}^2, \end{aligned}$$

where $\Phi_\varepsilon^{-1}(\alpha)$ is the α-quantile of the standard Gaussian error distribution and the parameters $(\mu, \beta_0, \eta_{00}, \eta_{10}, \eta_{20})'$ are estimated by Quasi-Maximum Likelihood (QML). This specification is standard in the econometrics literature. We also entertained other specifications, like pure Gaussian GARCH(1,1) and Student-t GARCH(1,1) models with degrees of freedom estimated by MLE, and we obtained similar conclusions. For the sake of exposition we omit these alternative specifications.

The Basel Accord (1996a) and the Amendment of Basel Accord (1996b) recommends to carry out backtesting procedures with quantile levels $\alpha = 0.01$ or $\alpha = 0.05$. Here, we take as $\mathcal{T}$ a grid of $m = 10$ equidistributed points $\{\alpha_j\}_{j=1}^{m}$ from $\alpha_1 = 0.005$ to $\alpha_m = 0.05$, in intervals of length 0.005, covering the region recommended by Basel Accord (1996a). We apply our CvM test in (5) and the K test in (6) with $I_{t-1} = (Y_{t-1}, ..., Y_{t-d})$ for $d = 1$ and 2, and denote the corresponding test statistics by $CvM_{n,d}$ and $KS_{n,d}$. We compute these tests following (11) and (12). For a better comparison with our tests, we also consider aggregated standard backtests given by

$$K_n = m^{-1} \sum_{j=1}^{m} K_{n,\alpha_j}, \qquad C_n = m^{-1} \sum_{j=1}^{m} C_{n,\alpha_j}.$$

In Table III we report the subsampling $p-$values for several choices of k in $b = \lfloor kn^{2/5} \rfloor$.

Please insert Table III about here.

We can draw several conclusions from the results of Table III. First, our results indicate that the AR(1)-GARCH(1,1) model with Gaussian innovations is not able to adequately fit the tails of these stock returns. Our tests strongly reject this model for the CAC and FTSE stocks, and it is dubious for the DAX index, with rejections at 10% when $d = 2$ with $CvM_{n,2}$ and at 5% when $d = 1$ with both, $CvM_{n,1}$ and $KS_{n,1}$. Second, the cumulative conditional backtest has rather low power and indeed, it is not able to detect any of these alternatives. This result is consistent with other finite sample studies using this test, see Escanciano and Olmo (2008). Third, it is apparent from the results for CAC that in order to detect this alternative it is important to consider a larger information set containing the second lag. Traditional backtests only use limited information, no conditional information for K_n and the information provided for the previous hit for C_n, which results in a lack of power, as can be seen from the results with the CAC index return.

We complement the previous analysis with the marginal tests for each α_j, $j = 1, ..., m$, in Figures 1 to 3 for the subsampling size $b = \lfloor kn^{2/5} \rfloor$ with $k = 4$, i.e., $b = 73$. We take $d = 1$ for the DAX and FTSE indexes, whereas for CAC we consider $d = 2$ for a better understanding of the cause of rejection. We observe that conditional marginal backtests are more sensitive to α than our marginal tests and unconditional backtests. The rejection for the DAX index is mostly due to the low quantile levels (from 0.005 to 0.025) which is the most relevant part in case of an extreme event. For a risk manager applying classical backtesting techniques at the usual level $\alpha = 0.05$, the risk model provided by the AR(1)-GARCH(1,1) would seem appropriate. Using our more powerful test, he or she would conclude that this is not the case. This model fails to fit quantiles in the range $\alpha \in [0.005, 0.025]$ and $\alpha = 0.05$.

Please insert Figure 1 and Figure 2 about here.

Figure 2 reveals that the rejection for the CAC index of our CvM test is due to the misspecification of the conditional quantiles at large levels $[0.02, 0.05]$. Again, traditional backtests are not able to reject this alternative at $\alpha = 0.05$. The reason being the inefficient use (or not use at all) of conditional information from values of the index at higher lags than one. Figure 3 shows for the FTSE index the low power of conditional backtests for moderate values of α, even for alternatives that can be easily detected with alternative tests.

Please insert Figure 3 about here.

For a better understanding of the cause of rejection, we report in Table IV the number of violations $Viol_{\alpha} = \sum_{t=1}^{n} H_{t,\alpha}(\theta_n)$ for each α_j, $j = 1, ..., m$, as well as the number of expected violations $EViol_{\alpha} = n \cdot \alpha$. We observe that in all cases with the DAX and FTSE indexes the number of violations is higher than the its expected value, indicating fatter tails than the Gaussian AR-GARCH model, especially within the quantile region $[0.005, 0.025]$. For the CAC index we observe a similar pattern but with a smaller number of violations, which is consistent with our previous results with $d = 1$. Unreported simulations with a Student-t distribution showed that an AR(1)-GARCH(1,1) model with Student-t innovations is still not able to fit the tails of these data sets, although the number of violations reduced considerably in all cases. We omitt these additional simulations for the sake of space.

Please insert Table IV about here.

This application to stock returns shows that our methods have important implications for evaluating market risk measures such as VaR. We stress that our methodology can be seen as a general framework to analyze market risk. For instance, there is now an important growing literature in finance, proposing the Conditional Expected Shortfall (CES) as an alternative to the VaR for measuring market risk in financial data. The CES is defined as

$$\rho_{W_{t-1},\alpha}(Y_t) = \alpha^{-1} \int_0^{\alpha} q_{\nu}(I_{t-1}) d\nu, \qquad \alpha \in (0, 1). \tag{16}$$

Therefore, in modeling the CES the interest is only in the range of quantiles $[0, \alpha]$ and not on the whole conditional distribution; see Escanciano and Mayoral (2008) for discussion of parametric CES models. The methods proposed in this section can be also seen as model specification tools of CES models.

We finish this section with some final conclusions. Econometric modeling often requires the specification of conditional quantile models for a range of quantiles of the conditional distribution. For the evaluation of models for quantile regression we propose and justify a general and flexible method which compares favorably with single quantile techniques and ad-hoc tests. We have shown in this paper that our tests have higher power than the standard unconditional and conditional backtesting procedures commonly used by banks and regulators to assess dynamic parametric VaR estimates. In particular, we find that the standard conditional backtesting procedure has rather low power in detecting misspecifications of an AR(1)-GARCH(1,1) VaR model for three major European stock indexes. Our methods provide flexible and powerful tools that can be used by practitioners to asses the plausibility of standard market risk models.

APPENDIX. PROOFS

First, we shall state a weak convergence theorem which is an extension of Theorem A1 in Delgado and Escanciano (2007) and that is of independent interest. Let for each $n \geq 1$, $I'_{n,0},...,I'_{n,n-1}$, be an array of random vectors in $\mathbb{R}^p$, $p \in \mathbb{N}$, and $Y_{n,1},...,Y_{n,n}$, be an array of real random variables (r.v.'s). Denote by $(\Omega_n, \mathcal{A}_n, P_n)$, $n \geq 1$, the probability space in which all the r.v.'s $\{Y_{n,t}, I'_{n,t-1}\}_{t=1}^n$ are defined. Let $\mathcal{F}_{n,t}$, $0 \leq t \leq n$, be a double array of sub σ-fields of $\mathcal{A}_n$ such that $\mathcal{F}_{n,t-1} \subset \mathcal{F}_{n,t}$, $t = 1,...,n$ and such that for each $n \geq 1$ and each $\gamma \in \mathcal{H}$,

$$E[w(Y_{n,t}, I_{n,t-1}, \gamma) \mid \mathcal{F}_{n,t-1}] = 0 \text{ a.s.} \qquad 1 \leq t \leq n, \ \forall n \geq 1. \tag{17}$$

Moreover, we shall assume that $\{w(Y_{n,t}, I_{n,t-1}, \gamma), \mathcal{F}_{n,t}, 0 \leq t \leq n\}$ is a square-integrable martingale difference sequence for each $\gamma \in \mathcal{H}$, that is, (17) holds, $Ew^2(Y_{n,t}, I_{n,t-1}, \gamma) < \infty$ and $w(Y_{n,t}, I_{n,t-1}, \gamma)$ is $\mathcal{F}_{n,t}$-measurable for each $\gamma \in \mathcal{H}$ and $\forall t, 1 \leq t \leq n, \forall n \in \mathbb{N}$. The following result gives sufficient conditions for the weak convergence of the empirical process

$$\alpha_{n,w}(\gamma) = n^{-1/2} \sum_{t=1}^{n} w(Y_{n,t}, I_{n,t-1}, \gamma) \qquad \gamma \in \mathcal{H}.$$

Under mild conditions the empirical process $\alpha_{n,w}$ can be viewed as a mapping from Ω_n to $\ell^\infty(\mathcal{H})$, the space of all complex-valued functions that are uniformly bounded on $\mathcal{H}$, with $\mathcal{H}$ a generic metric space. The weak convergence theorem that we present here is funded on results by Levental (1989), Bae and Levental (1995) and Nishiyama (2000). In Theorem A1 in Delgado and Escanciano (2007) $\mathcal{H}$ was finite-dimensional, but here we allow for an infinite-dimensional $\mathcal{H}$. The proof of theorem does not change by this possibility, however.

An important role in the weak convergence theorem is played by the conditional quadratic variation (CV) of the empirical process $\alpha_{n,w}$ on a finite partition $\mathcal{B} = \{H_k; 1 \leq k \leq N\}$ of $\mathcal{H}$, which is defined as

$$CV_{n,w}(\mathcal{B}) = \max_{1 \leq k \leq N} n^{-1} \sum_{t=1}^{n} E\left[\sup_{\gamma_1, \gamma_2 \in H_k} |w(Y_{n,t}, I_{n,t-1}, \gamma_1) - w(Y_{n,t}, I_{n,t-1}, \gamma_2)|^2 \mid \mathcal{F}_{n,t-1} \right]. \tag{18}$$

Then, for the weak convergence theorem we need the following assumptions.

W1: For each $n \geq 1$, $\{(Y_{n,t}, I_{n,t-1})' : 1 \leq t \leq n\}$ is a strictly stationary and ergodic process. The sequence $\{w(Y_{n,t}, I_{n,t-1}, \gamma), \mathcal{F}_{n,t}, 1 \leq t \leq n\}$ is a square-integrable martingale difference sequence for each $\gamma \in \mathcal{H}$. Also, there exists a function $C_w(\gamma_1, \gamma_2)$ on $\mathcal{H} \times \mathcal{H}$ to $\mathbb{R}$ such that uniformly in $(\gamma_1, \gamma_2) \in \mathcal{H} \times \mathcal{H}$

$$n^{-1} \sum_{t=1}^{n} w(Y_{n,t}, I_{n,t-1}, \gamma_1) w^c(Y_{n,t}, I_{n,t-1}, \gamma_2) = C_w(\gamma_1, \gamma_2) + o_{P_n}(1).$$

W2: The family $w(Y_{n,t}, I_{n,t-1}, \gamma)$ is such that $\alpha_{n,w}$ is a mapping from Ω_n to $\ell^\infty(\mathcal{H})$ and for every $\delta > 0$ there exists a finite partition $\mathcal{B}_\delta = \{H_k; 1 \leq k \leq N_\delta\}$ of $\mathcal{H}$, with N_δ being the number of elements of such partition, such that

$$\int_0^\infty \sqrt{\log(N_\delta)} d\delta < \infty \tag{19}$$

and

$$\sup_{\delta \in (0,1) \cap \mathbb{Q}} \frac{CV_{n,w}(\mathcal{B}_\delta)}{\delta^2} = O_{P_n}(1). \tag{20}$$

Let $\alpha_{\infty,w}(\cdot)$ be a Gaussian process with zero mean and covariance function given by $C_w(\gamma_1, \gamma_2)$. We are now in position to state the following

Theorem A1: *If Assumptions W1 and W2 hold, then it follows that*

$$\alpha_{n,w} \Longrightarrow \alpha_{\infty,w} \text{ in } \ell^\infty(\mathcal{H}).$$

Proof of Theorem A1: Theorem A1 in Delgado and Escanciano (2007).

Corollary A1: *Assuming that W1 holds for* $w(Y_{n,t}, I_{n,t-1}, v) = \Psi_\alpha(Y_{n,t} - m(I_{n,t-1}, \theta_0(\alpha))) \exp(ix' I_{n,t-1})$, $v = (x', \alpha)' \in \Pi$, *A1(b) and that*

$$n^{-1} \sum_{t=1}^{n} |I_{n,t-1}|^2 = O_{P_n}(1),$$

then the weak convergence of Theorem A1 holds.

Proof of Corollary A1: We shall apply Theorem A1. Let us define the metric

$$d(v_1, v_2) := \sqrt{|\alpha_1 - \alpha_2| + |x_1 - x_2|^2}, \qquad v_1, v_2 \in \Pi.$$

Then, we define an δ-bracket as an interval $[v_1, v_2]$ such that $v_1 \leq v_2$ and $d(v_1, v_2) \leq \delta$. The bracketing number $N(\delta, \Pi, d)$ is the minimum number of δ-brackets needed to cover Π. Then, it is easy to show that

$$\int_0^\infty \sqrt{\log(N(\delta, \Pi, d))} d\delta < \infty$$

holds. It remains to show that (20) holds. Consider a partition $\mathcal{B}_\delta = \{H_k; 1 \leq k \leq N(\delta, \Pi, d) \equiv N_\delta\}$ of Π in δ-brackets $H_k = [\underline{v}_k, \overline{v}_k]$, with $\underline{v}_k = (\underline{x}_k', \underline{\alpha}_k)'$ and $\overline{v}_k = (\overline{x}_k', \overline{\alpha}_k)'$, $\underline{x}_k \leq \overline{x}_k$ and $\underline{\alpha}_k \leq \overline{\alpha}_k$. Define $\varepsilon_{n,t}(\alpha) = Y_{n,t} - m(I_{n,t-1}, \theta_0(\alpha))$. Then, by simple algebra and the monotonicity of $1(\varepsilon_{n,t}(\alpha) \leq 0)$ due to A1(b), $CV_{n,w}(\mathcal{B}_\delta)$ in (18) is bounded by

$$
\begin{aligned}
& 2 \max_{1 \leq k \leq N_\epsilon} n^{-1} \sum_{t=1}^{n} E\left[\sup_{v_1, v_2 \in H_k} \left|1(\varepsilon_{n,t}(\alpha_1) \leq 0) - \alpha_1 - 1(\varepsilon_{n,t}(\alpha_2) \leq 0) + \alpha_2\right|^2 \mid \mathcal{F}_{n,t-1} \right] \\
& +2 \max_{1 \leq k \leq N_\epsilon} n^{-1} \sum_{t=1}^{n} \left[\sup_{v_1, v_2 \in H_k} \left|\exp(ix_1' I_{n,t-1}) - \exp(ix_2' I_{n,t-1})\right|^2 \right] \\
\leq\ & C \max_{1 \leq k \leq N_\epsilon} \left\{ |\underline{\alpha}_k - \overline{\alpha}_k| + |\underline{x}_k - \overline{x}_k|^2 n^{-1} \sum_{t=1}^{n} |I_{n,t-1}|^2 \right\}.
\end{aligned}
$$

Hence, (20) holds for the partition $\mathcal{B}_\delta$. Therefore, W2 of Theorem A1 holds and the corollary is proved. □

Proof of Theorem 1. Follows from Corollary A1. □

Theorem A2. *Assume Assumptions A1(c-d), A2, A3, and that there exists a $\theta_1 \in \mathcal{B}$ such that $\|\theta_n - \theta_1\|_{\mathcal{B}} = o_P(1)$. Then, uniformly in $(x', \alpha)' \in \Pi$,*

$$
\begin{aligned}
R_n^1(x, \alpha) \quad = \quad & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{\Psi_\alpha(e_t(\theta_1)) - E[\Psi_\alpha(e_t(\theta_1)) \mid \mathcal{F}_{t-1}]\} \exp(ix' I_{t-1}) \qquad (21) \\
& + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{E[\Psi_\alpha(e_t(\theta)) \mid \mathcal{F}_{t-1}]_{\theta=\theta_n} - E[\Psi_\alpha(e_t(\theta_1)) \mid \mathcal{F}_{t-1}]\} \exp(ix' I_{t-1}) \\
& + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E[\Psi_\alpha(e_t(\theta_1)) \mid \mathcal{F}_{t-1}] \exp(ix' I_{t-1}) - E\left[E[\Psi_\alpha(e_t(\theta_1)) \mid \mathcal{F}_{t-1}] \exp(ix' I_{t-1})\right] \\
& + \sqrt{n} E\left[E[\Psi_\alpha(e_t(\theta_1)) \mid \mathcal{F}_{t-1}] \exp(ix' I_{t-1})\right] + o_P(1).
\end{aligned}
$$

Proof of Theorem A2: Write $w_{t-1}(v, \theta) := \{\Psi_\alpha(e_t(\theta)) - E[\Psi_\alpha(e_t(\theta)) \mid \mathcal{F}_{t-1}]\} \exp(ix' I_{t-1})$. First we shall show that the process

$$
S_n(v, \theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} w_{t-1}(v, \theta)
$$

is asymptotically tight with respect to $(v, \theta) \in \mathcal{W} = \Pi \times \mathcal{B}$.

Let us define the class $\mathcal{K} = \{w_\cdot(v, \theta) : (v, \theta) \in \mathcal{W}\}$. Denote $X_{t-1,\infty} = (I_{t-1}, I_{t-2}, ...)'$. Let $\mathcal{B}_\delta = \{B_k; 1 \leq k \leq N_\delta \equiv N_{[]}(\delta, \mathcal{K}, \|\cdot\|_2\}$, with $B_k = [\underline{w}_k(Y_t, X_{t-1,\infty}), \overline{w}_k(Y_t, X_{t-1,\infty})]$, be a partition of $\mathcal{K}$ in δ-brackets with respect to $\|\cdot\|_2$, where $\|\cdot\|_2$ denotes the L_2 norm of random variables, i.e., $\|X\|_2 = \left(E[X^2]\right)^{1/2}$.

Conditions A1(c-d) and A2 imply that for a sufficiently small $\delta > 0$,

$$\begin{aligned}
& \left\| \sup_{\substack{(v_2,\theta_2)\in\mathcal{W}:d(v_1,v_2)\le\delta \\ \|\theta_1-\theta_2\|_{\mathcal{B}}\le\delta}} |w_{t-1}(v_1,\theta_1) - w_{t-1}(v_2,\theta_2)| \right\|_2 \\
\le\ & C \left\| \sup_{\substack{(v_2,\theta_2)\in\mathcal{W}:d(v_1,v_2)\le\delta \\ \|\theta_1-\theta_2\|_{\mathcal{B}}\le\delta}} |\Psi_{\alpha_1}(e_t(\theta_1)) - \Psi_{\alpha_2}(e_t(\theta_2))| \right\|_2 + C\delta \\
\le\ & C \left\| \sup_{|\alpha_1-\alpha_2|\le\delta} |1(Y_t \le m_{t-1}(\theta_1(\alpha_1))) - 1(Y_t \le m_{t-1}(\theta_1(\alpha_2)))| \right\|_2 \\
& + C\left(E\left[\sup_{\|\theta_1-\theta_2\|_{\mathcal{B}}\le\delta} |1(Y_t \le m_{t-1}(\theta_1(\alpha))) - 1(Y_t \le m_{t-1}(\theta_2(\alpha)))| \right] \right)^{1/2} + C\delta \\
\le\ & C\delta^{1/2}.
\end{aligned}$$

Theorem 3 in Chen et al. (2003) and A3 yield that (19) holds for such partition. Therefore, by similar arguments as in Corollary A1, (20) follows, and condition W2 of Theorem A1 holds. The asymptotically tightness of $S_n(v,\theta)$ is then proved. As a result,

$$\sup_{v\in\Pi} |S_n(v,\theta_n) - S_n(v,\theta_1)| = o_P(1),$$

which can be rewritten as

$$\begin{aligned}
R_n^1(\cdot) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \{\Psi_\alpha(e_t(\theta_1)) - E[\Psi_\alpha(e_t(\theta_1)) \mid \mathcal{F}_{t-1}]\} \exp(ix' I_{t-1}) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n E[\Psi_\alpha(e_t(\theta)) \mid \mathcal{F}_{t-1}]_{\theta=\theta_n} + o_P(1),
\end{aligned}$$

from which (21) follows. □

Proof of Theorem 2: Under the null $\theta_1 = \theta_0$ and $E[\Psi_\alpha(e_t(\theta_0)) \mid \mathcal{F}_{t-1}] = 0$ a.s. From the expansion in (21), it follows that, uniformly in $v \in \Pi$,

$$\begin{aligned}
R_n^1(\cdot) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \Psi_\alpha(e_t(\theta_0)) \exp(ix' I_{t-1}) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n \{E[\Psi_\alpha(e_t(\theta)) \mid \mathcal{F}_{t-1}]_{\theta=\theta_n} - E[\Psi_\alpha(e_t(\theta_0)) \mid \mathcal{F}_{t-1}]\} \exp(ix' I_{t-1}) + o_P(1) \\
&= R_n(\cdot) + \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ F_{I_{t-1}}(m(I_{t-1},\theta_n)) - F_{I_{t-1}}(m_{t-1}(\theta_0)) \right\} \exp(ix' I_{t-1}) + o_P(1).
\end{aligned}$$

Now, from A1(d) and Koul and Stute (1999, pp. 228-229), uniformly in $v \in \Pi$,

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ F_{I_{t-1}}(m(I_{t-1},\theta_n)) - F_{I_{t-1}}(m_{t-1}(\theta_0)) \right\} \exp(ix' I_{t-1}) \\
=\ & \sqrt{n}(\theta_n - \theta_0) \frac{1}{n} \sum_{t=1}^n g(I_{t-1},\theta_0) f_{I_{t-1}}(m_{t-1}(\theta_0)) \exp(ix' I_{t-1}) + o_P(1).
\end{aligned}$$

This together with Theorem 1, A2(c) and A4 proves the theorem. □

Proof of Theorem 3: Let $\mathcal{W} = \Pi \times \mathcal{B}$. Let $w = (x', \alpha, \theta'(\cdot))'$ be a general element of $\mathcal{W}$. The space $\mathcal{W}$ is endowed with the metric

$$\rho(w_1, w_2) = |x_1 - x_2| + |\alpha_1 - \alpha_2| + \sup_{\alpha \in \mathcal{T}} |\theta_1(\alpha) - \theta_2(\alpha)|,$$

where $w_1 = (x_1', \alpha_1, \theta_1'(\cdot))'$ and $w_2 = (x_2', \alpha_2, \theta_2'(\cdot))'$ belong to $\mathcal{W}$. Let $B(w, \delta)$ be the open ball of radius δ around w, i.e., $B(w, \delta) = \{w_1 \in \mathcal{W} : \rho(w_1, w) < \delta\}$. Note that A1-A3 yield that for each $w = (x', \alpha, \theta'(\cdot))' \in \mathcal{W}$ it holds that

$$\lim_{\delta \to 0} E\left[\sup_{w_1 \in B(w,\delta)} \left|\Psi_{\alpha_1}(e_t(\theta_1(\alpha_1))) \exp(ix_1' I_{t-1}) - \Psi_\alpha(e_t(\theta(\alpha))) \exp(ix' I_{t-1})\right|^2\right] = 0.$$

Therefore, $E[\Psi_\alpha(e_t(\theta_1(\alpha))) \exp(ix' I_{t-1})]$ is a continuous function of $v = (x', \alpha)'$. Therefore, a uniform version of the Ergodic Theorem

$$\sup_{\theta \in \mathcal{B}} \sup_{v \in \Pi} \left| \frac{1}{n} \sum_{t=1}^{n} [\Psi_\alpha(e_t(\theta(\alpha))) \exp(ix' I_{t-1}) - E[\Psi_\alpha(e_t(\theta(\alpha))) \exp(ix' I_{t-1})] \right| = o_P(1).$$

Hence, from the last display and A5

$$\sup_{v \in \Pi} \left| \frac{1}{n} \sum_{t=1}^{n} [\Psi_\alpha(e_t(\theta_n(\alpha))) \exp(ix' I_{t-1}) - E[\Psi_\alpha(e_t(\theta_1(\alpha))) \exp(ix' I_{t-1})] \right| = o_P(1).$$

and the function $E[\Psi_\cdot(e_t(\theta_1(\cdot)))1(I_{t-1} \leq \cdot)]$ is different from zero in a subset with positive Lebesgue measure on Π. □

Proof of Theorem 4: The proof follows from Theorem A2 and Assumptions A5 and A6 jointly with A4' in a routine fashion, and then, it is omitted. □

Proof of Theorem 5. The proof follows the same steps as Theorems 2, 3 and 4 of Whang (2004) and then, it is omitted. □

REFERENCES

Albrecht, J., van Vuuren, A. and Vroman, S., (2007): "Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands", unpublished manuscript.

Angrist, J. Chernozhukov, V. and Fernandez-Val, I. (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure", forthcoming in *Econometrica.*

Bae, J. and Levental, S. (1995): "Uniform CLT for Markov chains and its invariance principle: a martingale approach", *Journal of Theoretical Probability*, 8, 549-570.

Basel Committee on Banking Supervision (1996a): "Amendment to the Capital Accord to Incorporate Market Risks, Bank for International Settlements. "

Basel Committee on Banking Supervision (1996b): "Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements, Bank for International Settlements. "

Bassett, G. and Chen, H.L. (2001): "Portfolio style: return-based attribution using regression quantiles", *Empirical Economics*, 26, 293-305.

Bassett, Jr., G., and Koenker, R. (1978): "Asymptotic theory of least absolute error regression," *Journal of the American Statistical Association*, 73, 618–622.

Behnen, K. and Neuhaus, G. (1975): "A central limit theorem under contiguous alternatives", *The Annals of Statistics,* 3, 1349-1353.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2006): "Evaluating Value-at-Risk models with desk-level data. " *Working Paper Series, 010* Department of Economics, North Carolina State University.

Bierens, H.J. (1982): "Consistent model specification tests", *Journal of Econometrics*, 20, 105-134.

Bierens, H., and Ginther, D.K. (2001): "Integrated conditional moment testing of quantile regression models", *Empirical Economics*, 26, 307-324.

Bierens, H.J. and Ploberger, W. (1997): "Asymptotic theory of integrated conditional moment test", *Econometrica*, 65, 1129-1151.

Bilias, Y., Chen, S. and Ying, Z. (2000): "Simple resampling methods for the censored regression quantiles", *Journal of Econometrics*, 99, 373–386.

Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12, 788-792.

Chen, X., Linton, O. and van Keilegom, I. (2003): "Estimation of semiparametric models when the criterion function is not smooth", *Econometrica*, 71, 1591-1608.

CHERNOZHUKOV, V. (2002): "Inference on the quantile regression process, an alternative," Working article 02-12 (February), MIT (www.ssrn.com).

CHRISTOFFERSEN, P. (1998): "Evaluating Interval Forecasts. "*International Economic Review,* 39, 841-862.

CORRADI, V. AND N.R. SWANSON (2006): "Predictive Density and Conditional Confidence Intervals Accuracy Tests."*Journal of Econometrics*, 135, 187-228.

DELGADO, M.A. AND ESCANCIANO, J.C. (2007): "Nonparametric tests for conditional symmetry in dynamic models."*Journal of Econometrics*, 127, 652-682.

DUDLEY, R.M. (1999): *Uniform Central Limit Theorems.* Cambridge. University Press, Massachusetts.

ESCANCIANO, J. C. (2006): "Goodness-of-fit tests for linear and non-linear time series models´´, *Journal of the American Statistical Association* 101, 531-541.

ESCANCIANO, J. C. (2008): "On the Lack of Power of Omnibus Specification Tests", Forthcoming in Econometric Theory. (2008).

ESCANCIANO, J. C. AND MAYORAL, S. (2008): "Semiparametric Estimation of Dynamic Conditional Expected Shortfall Models", *The International Journal of Monetary Economics and Finance*, Vol. 1, No. 2, 2008.

ESCANCIANO, J. C. AND OLMO, J. (2008): "Backtesting VaR with Estimation Risk", forthcoming in *Journal of Business and Economics Statistics.*

ESCANCIANO, J.C. AND VELASCO, C. (2006): "Generalized spectral tests for the martingale difference hypothesis", *Journal of Econometrics*, 134, 151-185.

GUTENBRUNNER, C. AND JUREČKOVA, J. (1992): "Regression rank scores and regression quantiles", *The Annals of Statistics*, 20, 305–329.

GUTENBRUNNER, C., JUREČKOVA, J., KOENKER, R. AND PORTNOY, S. (1993): "Test of linear hypotheses based on regression rank scores", *Journal of Nonparametric Statistics*, 2, 307-331.

HAHN, J. (1995): "Bootstrapping Quantile Regression Estimators", *Econometric Theory*, 11, 105-121.

HE, X., AND HU, F. (2002): "Markov chain marginal bootstrap", *Journal of the American Statistical Association*, 97, 783–795

HE, X. AND ZHU, L.X (2003): "A Lack-of-Fit test for Quantile Regression", *Journal of the American Statistical Association*, 98, 1013-1022.

HOROWITZ, J.L. (1998): "Bootstrap methods for median regression models", *Econometrica*, 66, 1327-1351.

Horowitz, J.L. and Spokoiny, V.G. (2002): "An adaptive, rate-optimal test of linearity for median regression models", *Journal of the American Statistical Association*, 97, 822-835.

Jennrich, R.I. (1969): "Asymptotic properties of nonlinear least squares estimators", *Annals of Mathematical Statistics*, 40, 633-643.

Jurečkova, J. and Hallin, M. (1999): "Optimal tests for autoregressive models based on autoregression rank scores", *The Annals of Statistics*, 27, 1385–1414.

Jurečkova, J. and Prochazka (1994): "Regression quantiles and trimmed least squares estimation in nonlinear regression models," *Journal of Nonparametric Statistics*, 3, 201–222.

Khmaladze, E.V. (1981): "Martingale approach to the goodness of fit tests", *Theory of Probability and Its Applications*, 26, 246-265.

Kim,T.H. and White,H. (2003): "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regressions", in T. Fomby and R.C. Hill, eds., Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later. New York: Elsevier, 107-132.

Koenker, R. and Bassett, G. (1978): "Regression quantiles", *Econometrica*, 46, 33-50.

Koenker, R., and Machado, J.A. (1999): "Goodness of fit and related inference processes for quantile regression," *Journal of the American Statistical Association*, 94, 1296-1310.

Koenker, R. and Park, B.J. (1996): "An interior point algorithm for nonlinear quantile regression", *Journal of Econometrics*, 71, 265-285.

Koenker, R. and Xiao, Z. (2002): "Inference on the quantile regression process", *Econometrica*, 70, 1583–1612.

Koenker, R. and Xiao, Z. (2006): "Quantile autoregression." *Journal of the American Statistical Association*, 101, 980-990.

Komunjer, I. (2005): "Quasi-Maximum Likelihood Estimation for Conditional Quantiles", *Journal of Econometrics*, 128, 137-164.

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* 2nd ed, Lecture Notes in Statistics, Vol. 166, Springer.

Koul, H. L, and Ling, S. (2006): "Fitting an error distribution is some heteroscedastic time series models", forthcoming in *The Annals of Statistics.*

Koul, H.L., and Saleh, A.K.M.E (1994) "Autoregression quantile and related rank scores processes", *The Annals of Statistics*, 23, 670-689.

Koul, H.L., and Stute W. (1999): "Nonparametric model checks for time series", *The Annals of Statistics*, 27, 204-236.

Kupiec, P. (1995): "Techniques for Verifying the Accuracy of Risk Measurement Models, "*Journal of*

Derivatives, 3,73-84.

Levental, S. (1989): "A uniform CLT for uniformly bounded families of martingale differences", *Journal of Theoretical Probability*, 2, 271-287.

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

Machado, J.A.F. and Mata, J. (2005): "Counterfactual decomposition of changes inwage distributions using quantile regression", *Journal of Applied Econometrics*, 20, 445-65.

Mukherjee, K. (1999): "Asymptotics of quantiles and rank scores in nonlinear time series", *Journal of Time Series Analysis*, 20(2), 173–192.

Nishiyama, Y. (2000): "Weak convergence of some classes of martingales with jumps", *The Annals of Probability*, 28, 685-712.

Pollard, D. (1991): "Asymptotics for least absolute deviation regression estimators", *Econometric Theory*, 7, 186-199.

Politis, D. N., Romano, J.P. and Wolf, M. (1999): *Subsampling.* Springer-Verlag, New York.

Portnoy, S. (1984): "Tightness of the sequence of empiric cdf processes defined from regression fractiles", in Robust and Nonlinear Time Series Analysis, eds. J. Franke, W. Hardle, and D. Martin, Springer-Verlag: New York.

Sakov, A. and Bickel, P. (2000): "An Edgeworth expansion for the m out of n bootstrap median", *Statistical and Probability Letters*, 49, 217-223.

Sharpe, W.F. (1988): "Determining a Fund's Effective Asset Mix", *Investment Management Review*, December, 59-69.

Sharpe, W.F. (1992): "Asset Allocation: Management Style and Performance Measurement", *Journal of Portfolio Management*, 18, 7-19.

Stinchcombe, M., and White, H. (1998): "Consistent specification testing with nuisance parameters present only under the alternative", *Econometric Theory*, 14, 295-325.

van der Vaart, A.W., and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer. New York.

Whang, Y-J. (2004): "Consistent specification tests for quantile regression models", forthcoming in Frontiers of Analysis and Applied Research: Essays in Honor of Peter C.B. Phillips, Cambridge University Press.

Whang, Y-J. (2005): "Smoothed Empirical Likelihood Methods for Quantile Regression Models", forthcoming in *Econometric Theory.*

Zheng, J.X. (1998): "A consistent nonparametric test of parametric regression models under conditional quantile restrictions", *Econometric Theory*, 14, 123-138.

Table I: Empirical size and power. 5% of significance level.

CvM_n		**DGP1**			**DGP2**		
c_1	n	$k=7$	$k=8$	$k=9$	$k=3$	$k=4$	$k=5$
0.0	**100**	4.6	5.1	6.5	5.2	5.9	5.7
	300	4.0	4.3	5.2	4.9	4.9	4.6
0.1	**100**	48.3	48.3	48.9	9.9	10.1	10.6
	300	98.8	98.3	97.8	21.5	21.9	21.2
0.2	**100**	87.2	86.1	83.9	23.3	22.4	23.2
	300	100.0	100.0	100.0	71.0	71.1	68.9
0.3	**100**	97.4	96.4	95.7	41.8	39.8	40.2
	300	100.0	100.0	100.0	96.3	96.4	64.6

Table II: Empirical size and power. 5% of significance level.

KS_n		**DGP1**			**DGP2**		
c_1	n	$k=7$	$k=8$	$k=9$	$k=3$	$k=4$	$k=5$
0.0	**100**	3.7	3.9	5.7	5.9	5.9	6.4
	300	3.6	3.9	4.4	6.3	6.7	6.9
0.1	**100**	34.0	33.9	33.8	8.1	7.9	8.5
	300	96.2	95.2	74.1	17.1	17.3	16.2
0.2	**100**	74.1	71.3	67.5	17.2	17.2	16.7
	300	100.0	100.0	100.0	57.0	58.6	56.8
0.3	**100**	91.2	89.4	86.2	30.1	29.0	30.9
	300	100.0	100.0	100.0	91.8	90.5	89.4

Table III: Aggregated tests: Subsampling $p-$values.

$n=1471$		$CvM_{n,1}$	$CvM_{n,2}$	$KS_{n,1}$	$KS_{n,2}$	K_n	C_n
DAX	$k=3$	0.027	0.062	0.021	0.127	0.040	0.186
	$k=4$	0.049	0.071	0.046	0.145	0.064	0.261
	$k=5$	0.048	0.079	0.058	0.195	0.064	0.284
CAC	$k=3$	0.241	0.000	0.194	0.000	0.148	0.326
	$k=4$	0.298	0.000	0.233	0.000	0.250	0.407
	$k=5$	0.370	0.000	0.283	0.000	0.278	0.476
FTSE	$k=3$	0.000	0.000	0.000	0.001	0.000	0.292
	$k=4$	0.000	0.000	0.000	0.002	0.000	0.417
	$k=5$	0.000	0.000	0.000	0.000	0.000	0.535

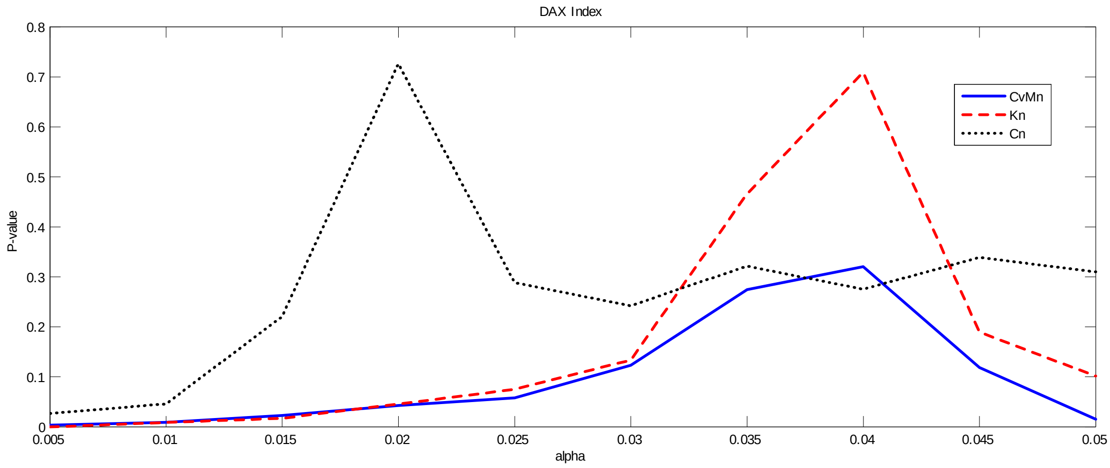

Figure 1. Subsampling p-values for $CvM_{n,1,\alpha}$ test (solid line), unconditional backtest $K_{n,\alpha}$ (dashed line), and the conditional backtest $C_{n,\alpha}$ (dotted line) as a function of alfa. Subsample size $b = 73$.

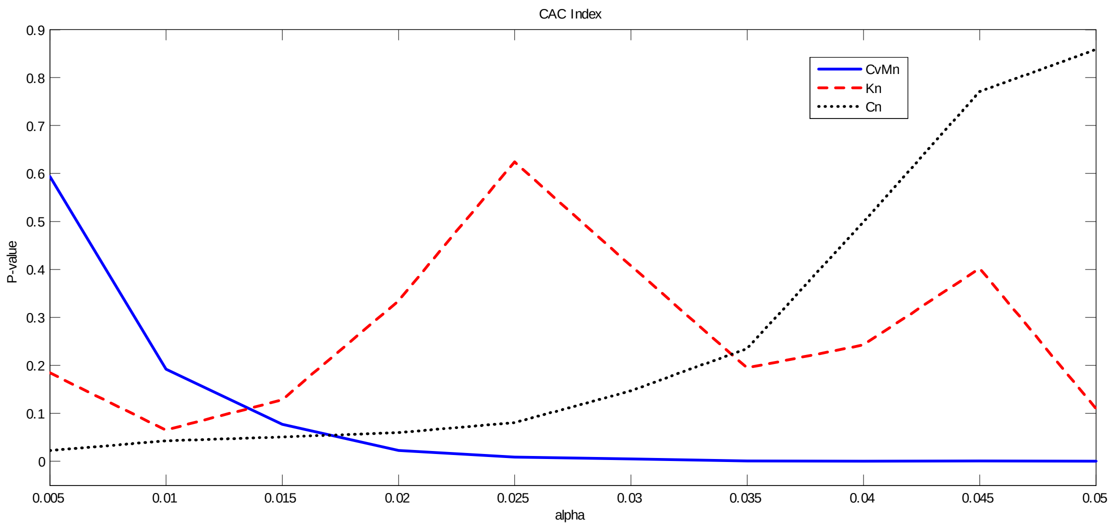

Figure 2. Subsampling p-values for $CvM_{n,2,\alpha}$ test (solid line), unconditional backtest $K_{n,\alpha}$ (dashed line), and the conditional backtest $C_{n,\alpha}$ (dotted line) as a function of alfa. Subsample size $b = 73$.

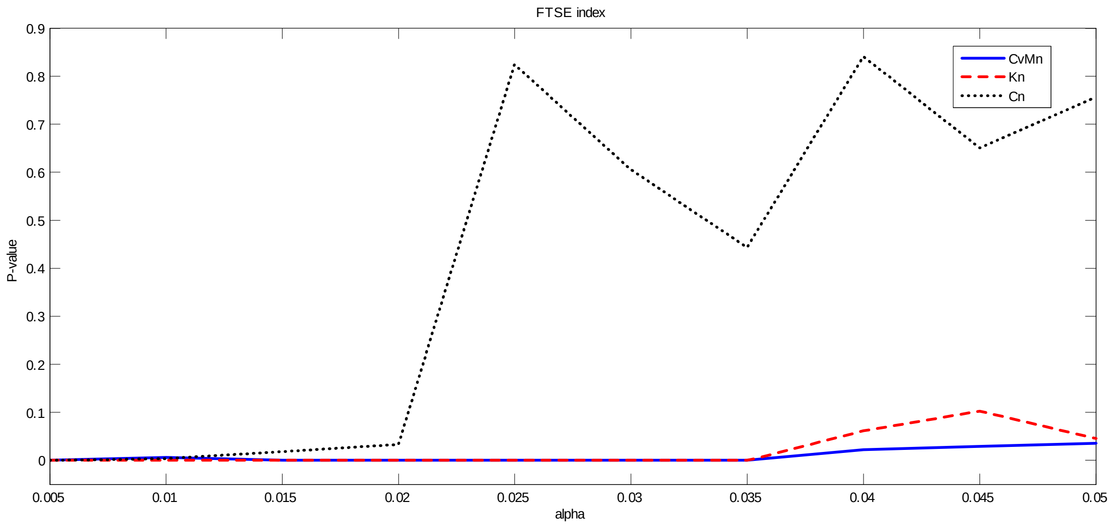

Figure 3. Subsampling p-values for $CvM_{n,1,\alpha}$ test (solid line), unconditional backtest $K_{n,\alpha}$ (dashed line), and the conditional backtest $C_{n,\alpha}$ (dotted line) as a function of alfa. Subsample size $b = 73$.

TABLE IV: Number of violations ($Viol_\alpha$) and expected violations ($EViol_\alpha$)

		DAX	CAC	FTSE
α_j	$EViol_\alpha$	$Viol_\alpha$	$Viol_\alpha$	$Viol_\alpha$
0.005	7.0	20	12	20
0.010	14.1	25	21	29
0.015	21.2	35	29	37
0.020	28.3	43	32	48
0.025	35.4	48	33	55
0.030	42.4	52	38	62
0.035	49.5	54	41	65
0.040	56.6	59	48	69
0.045	63.7	72	57	75
0.050	70.8	84	58	82