Simple Bootstrap Tests for Conditional Moment Restrictions

JUAN CARLOS ESCANCIANO
Indiana University

First version: March 1, 2008. This version: January 13, 2009.

Abstract: This paper proposes consistent tests for conditional moment restrictions of stationary time series. The proposed tests are based on continuous functionals of a projected integrated conditional moment process which acknowledges that deviations in the direction of the score function cannot be differentiated from deviations within the parametric model. As a result, the new tests are expected to have better power properties than existing tests and allow for a simple multiplier-type-bootstrap approximation. Thereby extending the scope of the wild-bootstrap and related methods to general conditional moment restrictions, including quantile regressions for which wild-bootstrap methods were not available. The new tests have the following remarkable properties: (i) they do not need to choose a tuning parameter, other than the number of bootstrap replications; (ii) there is no need for re-estimating the unknown parameters in each bootstrap replication; (iii) the new tests are valid under higher order conditional moments of unknown form; and (iv) they allow for non-separable and non-smooth moment functions and any root-n consistent estimator. A Monte Carlo experiment shows that the new method presents more accurate size and higher power than subsampling procedures. In an empirical application we study the dynamics in mean and variance of the Hong Kong stock market index and we evaluate models for the Value-at-Risk of the S&P 500. These applications highlight the merits of our approach and show that the new methods have higher power than popular backtesting methods.

Key words and phrases: Conditional moment restrictions; Omnibus tests; Asset Pricing; Value-at-Risk; Quantile Regression; Wild bootstrap; Non-separable models; Backtesting.

JEL Classification Number: C12; C14; C32.

1Department of Economics, Indiana University, email: jescanci@indiana.edu. Research funded by the Spanish Plan Nacional de I+D+I, reference number SEJ2007-62908. I thank seminar participants at Penn State University.
1. Introduction

Economic theories in dynamic contexts often suggest Conditional Moment Restrictions (CMR) in the underlying economic variables. Rational expectations and asset pricing equations are prominent examples of these behavioral assumptions. Alternatively, exogeneity or other statistical assumptions can also lead to these CMR. A general framework for analyzing such CMR is given by

\[ E[h(Y_t, \theta_0) | X_t] = 0 \text{ almost surely (a.s.), for a unique } \theta_0 \in \Theta \subset \mathbb{R}^p, \]

for a measurable, possibly non-smooth, moment function \( h : \mathbb{R}^{d_y} \times \Theta \to \mathbb{R}^{d_h} \) that is assumed to be known up-to the finite dimensional parameter \( \theta_0 \in \Theta \), with \( \Theta \) a compact subset of \( \mathbb{R}^p, p \in \mathbb{N} \). The vector-valued stochastic process \( \{Z_t \equiv (Y'_t, X'_t)\}_{t \in \mathbb{Z}} \) is a strictly stationary and ergodic time series defined on the probability space \( (\Omega, \mathcal{F}, P) \). Henceforth, \( |A| \) and \( A' \) denote the Euclidean norm \( |A| = (tr(A'A))^{1/2} \) and the transpose of a matrix \( A \), respectively. The conditioning variable \( X_t \) takes values in \( \mathbb{R}^{d_x}, d_x < \infty \), and it can contain lagged values of \( Y_t \) and other exogenous variables. Throughout the paper we shall assume that equation (1) identifies uniquely the parameter \( \theta_0 \). The following examples illustrate the general applicability of our methods. Further examples abound in the literature; see examples in Chen and Fan (1999), Li (1999), Delgado et al. (2006), Anatolyev (2007) and references therein, to mention just a few. In these examples \( X_t \) denotes the agent’s information set available at period \( t - 1 \), and \( X_t \) is a finite-dimensional subset of \( X_t \).

Example 1: (Asset Pricing Models, Hansen and Singleton, 1982) The consumption-based capital asset market model studied in Hansen and Singleton (1982) leads to the CMR

\[ E[h(Y_t, \theta_0) | X_t] = 0, \quad \theta_0 = (\beta, \gamma), \]

with \( h(Y_t, \theta_0) \) an \( s \times 1 \) vector with the \( i-th \) component

\[ h_i(Y_t, \theta_0) = \left[ \beta(1 + r_{i,t}) \left( \frac{c_t}{c_{t-1}} \right)^{-\gamma} - 1 \right], \quad 1 \leq i \leq s, \]

where \( c_t \) denotes spending on consumption in period \( t \), \( r_{i,t} \) is the net return of the asset \( i \), \( \beta \in (0, 1) \) is a discount factor, \( \gamma \) is the coefficient of risk aversion and \( Y_t = (r_{1,t}, ..., r_{s,t}, c_t/c_{t-1})' \). Hansen and Singleton (1982) proposed estimating \( \theta_0 \) by GMM. Different specifications of the agent’s preferences lead to different CMR; see e.g. Abel (1990) and Constantinides (1990) for extensions of this basic framework.
**Example 2:** (Joint specification of conditional mean and variance models) Correct specification of volatility models plays an important role in financial time series. Although it is not standard practice, a joint specification of the conditional mean and variance is appropriate in this context. This correct specification corresponds to the CMR with moment functions

\[
h(Y_t, \theta_0) = \begin{bmatrix} Y_{1t} - f(X_{t}, \theta_0) \\ (Y_{1t} - f(X_{t}, \theta_0))^2 - \sigma^2(X_{t}, \theta_0) \end{bmatrix},
\]

where \( Y_t = (Y_{1,t}, X_{t}')' \), \( Y_{1t} \) is the dependent random variable, \( f(X_{t}, \theta_0) \) is the conditional mean and \( \sigma^2(X_{t}, \theta_0) \) is the conditional variance.

**Example 3:** (Portfolio conditional mean-variance efficiency) There is an extensive literature on mean-variance analysis in finance and econometrics; see Sentana (2008) for a survey. Let \( r_{p,t} \) be the return on the portfolio \( p \) in excess of the riskless rate and \( r_t = (r_{1,t}, ..., r_{q,t}) \) be a \( q \times 1 \) vector of excess returns of the other assets. The portfolio \( p \) is conditionally mean-variance efficient if and only if the following CMR holds

\[
E[E[r_{p,t}^2 | X_{t}] - E[r_{p,t} | X_{t}]r_{p,t} | X_{t}] = 0.
\]

In a parametric approach one considers parametric models for \( E[r_{p,t}^2 | X_{t}] \) and \( E[r_{p,t} | X_{t}] \), say \( \mu_2(X_{t}, \theta_0) \) and \( \mu_1(X_{t}, \theta_0) \), respectively, and proceeds with a joint test using the \( q + 2 \) moments

\[
h(Y_t, \theta_0) = \begin{bmatrix} r_{p,t} - \mu_1(X_{t}, \theta_0) \\ r_{p,t}^2 - \mu_2(X_{t}, \theta_0) \\ \{\mu_2(X_{t}, \theta_0) - \mu_1(X_{t}, \theta_0)r_{p,t}\} r_{t} \end{bmatrix}.
\]

Nonparametric versions of (4) have been tested in Wang (2002, 2003) and Chen and Fan (1999). When the dimension of \( X_t \) is high, the resulting tests based on smoothers will be affected by the lack of precision of the nonparametric estimators. Alternative parametric approaches have been considered using GMM techniques, see e.g. Cochrane (1996), generally focussing just on the restriction (4). Joint tests can be constructed using our methods, leading to consistent and simple tests of conditional mean-variance efficiency.

**Example 4:** (Evaluating Value-at-Risk models) Value-at-Risk (VaR) has become the most popular measure of market risk used by financial institutions to determine the amount of capital on hold. 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 a quantile of the conditional distribution of returns on the portfolio given agent’s information set. Denote the real-valued time series of portfolio returns by \( Y_{1,t} \). Assuming that the conditional distribution
of $Y_{1,t}$ given $X_t$ is continuous, we define the $\alpha$-th conditional VaR $m_{\alpha}(X_t, \theta_0)$ of $Y_{1,t}$ given $X_t$ by the CMR

$$P(Y_t \leq m_{\alpha}(X_t, \theta_0) \mid X_t) = \alpha, \text{ a.s.}$$

where $\theta_0 \in \Theta \subset \mathbb{R}^p$ is an unknown parameter, and $\alpha \in (0, 1)$ is the VaR level, usually $\alpha = 0.01$. Therefore, the correct specification of the VaR model implies the CMR

$$E[h(Y_t, \theta_0) \mid X_t] = 0,$$

(5)

where $h(Y_t, \theta) = 1(Y_{1,t} \leq m_{\alpha}(X_t, \theta)) - \alpha$, $1(A)$ the indicator function of the event $A$, i.e., $1(A) = 1$ if $A$ occurs and zero otherwise, and $Y_t = (Y_{1,t}, X'_t)'$. The parameter $\theta_0$ can be estimated by the Quantile Regression Estimator (QRE), see Basset and Koenker (1978).

**Example 5**: (Evaluating Expected Shortfall models) Despite its universality, several authors have pointed out the deficiencies of VaR, including its lack of subadditivity; see Artzner et al. (1997, 1999). These authors proposed the Expected Shortfall (ES) as an alternative to VaR. We can specify a parametric conditional ES as

$$S_{\alpha}(X_t, \theta) = E[Y_{1,t} \mid Y_{1,t} \leq m_{\alpha}(X_t, \theta), X_t].$$

The parametric model will be correctly specified if and only if

$$E[\{Y_{1,t}1(Y_{1,t} \leq m_{\alpha}(X_t, \theta_0)) - \alpha S_{\alpha}(X_t, \theta_0)\} \mid X_t] = 0 \text{ a.s.}$$

(6)

To the best of our knowledge, no consistent test has been proposed for testing (6).

Tests for (1) have been traditionally tested using the the so-called Moment (M) and related specification tests (cf. Newey (1985) and Tauchen (1985)); see e.g. the J-test of Hansen (1982) for (2) and (4), Wooldridge (1990) test for (3) or the Backtesting techniques in Engle and Manganelli (2004) for (5). It is well-known, however, that M-tests and J-tests are inconsistent for testing CMR. Bierens (1982) proposed the Integrated Conditional Moment test (ICM) as a solution to this problem in a regression context; see also Stute (1997) for a related approach. A tenet for ICM tests and other omnibus tests is that the asymptotic null distributions depend in a complicated way on the data generating process (DGP), the model and the unknown parameter $\theta_0$, making it difficult the computation of asymptotic critical values. This problem has hampered the application of these consistent tests in applied econometric work.

Several solutions have been proposed to overcome the practical problem of computing the critical values in some specific cases, with special focus on regression models. Existing meth-
ods are, however, computationally expensive to implement and usually require the choice of tuning parameters (e.g. subsampling size, block size, etc.) Therefore, often the researcher is presented with the following dilemma: either to choose an inconsistent test (e.g. M-test) with simple asymptotic approximation or to choose a consistent test, whose asymptotic distribution is difficult to approximate and whose approximation depends substantially on some tuning parameter, without much guidance for its choice.

This paper proposes new tests for the general CMR described in (1) against nonparametric alternatives of unknown form, and which are straightforward to implement. The new tests are motivated from power considerations. They are based on continuous functionals of a projected integrated conditional moment process which acknowledges that deviations in the direction of the score function cannot be differentiated from deviations within the parametric model. Existing tests do not take into account this information, and hence, are expected to have less power than our tests. As a by-product, the new tests allow for a simple multiplier-type-bootstrap approximation, thereby extending the scope of the wild-bootstrap to general CMR, including cases with non-smooth moments such as quantile regression, and overcoming the practical limitations of the existing procedures. Our construction leads to powerful tests for (1), where no tuning parameter is necessary, other than the number of bootstrap replications. The new tests allow for any $\sqrt{n}$-consistent estimator, and do not require the estimation of the asymptotic distribution of the estimator, nor estimating the parameters in each bootstrap replication. Therefore, the proposed tests are easy to implement in examples where estimators (or their asymptotic variances) are difficult to obtain, as in e.g. nonlinear or quantile models. Finally, our tests allow for non-separable and non-smooth moments, and hence, they have general applicability.

The rest of the paper is organized as follows. In Section 2 we introduce some notation and the test statistics. Section 3 formally establishes the asymptotic null distribution of tests and compares the power of the new tests with classical tests. Section 4 introduces the multiplier-type-bootstrap approximation, proves its consistency and establishes connections with existing methods. Section 5 provides some Monte Carlo experiments and Section 6 considers an application to the Hong Kong stock index and the S&P 500, comparing the new tests with subsampling-based tests. Extensions to nonparametric (infinite-dimensional) scores are discussed in Section 7. Finally, in Section 8, we conclude with future research. Mathematical proofs of our results are gathered in an Appendix.

\footnote{An exception is Wooldridge (1990). However, it should be noted that Wooldridge’s motivation was different to ours. His motivation was to construct tests robust to departures from distributional assumptions that were not being tested.}
2. Testing Problem and Test Statistics

Following the integrated approach for testing CMR (see e.g. Bierens (1982), Stute (1997), Escanciano (2007a)), we characterize the CMR by an infinite number of unconditional moment restrictions. More specifically, the law of iterated expectations yields

\[ H(x, \theta_0) := E[h(Y_t, \theta_0)1(X_t \leq x)] = \int_{(-\infty,x]} m(\cdot, \theta_0) dP_X \]  \hspace{1cm} (7)

where

\[ m(x, \theta) := E[h(Y_t, \theta) | X_t = x], \hspace{1cm} x \in \mathbb{R}^{d_x}, \theta \in \Theta, \]  \hspace{1cm} (8)

and \( P_X \) is the stationary probability of \( X_t \). Therefore, from (7) and Billingsley (1995, Theorem 16.10iii), we have\(^3\)

\[ m(x, \theta_0) \equiv 0 \text{ } P_X\text{-a.s} \iff H(x, \theta_0) \equiv 0 \hspace{1cm} \text{for all } x \in \mathbb{R}^{d_x}. \]

The function \( H(x, \theta_0) \) is called the integrated CMR, after (7). We aim to test the null hypothesis

\[ H_0 : E[h(Y_t, \theta_0)1(X_t \leq x)] = 0, \hspace{1cm} \text{for all } x \in \mathbb{R}^{d_x} \text{ and some } \theta_0 \in \Theta \subset \mathbb{R}^p, \]

against the class of nonparametric alternatives

\[ H_A : P_X( x : E[h(Y_t, \theta)1(X_t \leq x)] \neq 0) > 0, \hspace{1cm} \text{for all } \theta \in \Theta \subset \mathbb{R}^p. \]

In view of the characterization of the null hypothesis, it is natural to reject \( H_0 \) for large values of the marked empirical process,

\[ R_n(x, \theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h(Y_t, \theta) 1(X_t \leq x). \]  \hspace{1cm} (9)

There is a long history on the use of processes like \( R_n(x, \theta) \) for inference in econometrics and statistics. We do not aim to review this extensive literature here. We rather refer the reader to Bierens (1982), Stute (1997) and Koul and Stute (1999) for important contributions in the context of regression models; see also Escanciano (2007a) for a unified treatment. In a time series context, general CMR with smooth moment functions were studied in Chen and Fan

\(^3\)Our methods can be similarly applied to other weighting functions different from the indicator family \( 1(X_t \leq x) \); see Bierens and Ploberger (1997), Stinchcombe and White (1998) and Escanciano (2007a) for examples.
(1999). For a thorough comparison with some existing methods see Section 4. Alternatively, consistent tests for CMR can be also based on smoothed versions of $R_n$; see e.g. Härdle and Mammen (1993). This literature is too extensive to be summarized here; see Hart (1997) for some review of the smoothing approach when $d = 1$, and Li (1999), Horowitz and Spokoiny (2001) and Guerre and Lavergne (2005) for some recent references. The use of smoothers leads to an ill-posed testing problem where some kind of regularization (i.e. bandwidth) is necessary.

Existing tests for CMR do not acknowledge that $\theta_0$ is a nuisance parameter in the construction of the test. The situation here is similar in spirit to the one addressed in Neyman (1959). Since $\theta_0 \in \Theta$ is unknown, deviations in the direction of the score function cannot be differentiated from deviations within the parametric model, i.e. from local deviations of $\theta_0$. A simple way to incorporate this information in the test statistic is to construct a test such that does not waste power in the direction of the score. More precisely, instead of $R_n(x, \theta_n)$ we consider a feasible version of the projected marked empirical process

$$R^1_n(x, \theta_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(X_t \leq x)I_h - G'(x, \theta_n)\Gamma^{-1}g(X_t, \theta_n)\}h(Y_t, \theta_n),$$

where $\theta_n$ is a $\sqrt{n}$-consistent estimator of $\theta_0$, $g(X_t, \theta) := (\partial/\partial \theta')m(X_t, \theta)$ is the $p \times d_h$ matrix of scores, $I_h$ is the identity matrix of order $d_h$, $G(x, \theta) := E[g(X_t, \theta)1(X_t \leq x)]$ and $\Gamma := E[g(X_t, \theta_0)g'(X_t, \theta_0)]$. We stress that only $m(X_t, \theta)$ is required to be differentiable in $\theta$, not $h(Y_t, \theta)$. Unlike $R_n$, $R^1_n$ satisfies the following convenient property (under some regularity conditions, cf. Theorem 1)

$$\sup_{x \in \mathbb{R}^{d_x}} |R^1_n(x, \theta_n) - R^1_n(x, \theta_0)| = o_P(1). \tag{10}$$

Property (10) is the key for the simple bootstrap approximation developed later. The process $R^1_n$ is not the only one satisfying this property; any test statistic of the form $n^{-1/2} \sum_{t=1}^{n} a(X_t)h(Y_t, \theta_n)$, with $a(X_t)$ orthogonal to the score $g(X_t, \theta_0)$, i.e. $E[a(X_t)g(X_t, \theta_0)] = 0$, satisfies (10). A prominent example satisfying the orthogonality property is the Khmaladze’s (1981) transformation. Our choice of $R^1_n$ is motivated from power arguments and the simplicity of its construction, see Section 3 for further elaboration on these points. In particular, we shall prove that $R^1_n$, or a simple variation of $R^1_n$, has higher power than tests based on the classical marked process $R_n$ or tests based on Khmaladze’s (1981) transformation.
Our test statistics will be continuous functionals of the feasible projected process

$$\hat{R}_n^1(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \{1(X_t \leq x)I_h - G'_n(x, \theta_n)\Gamma_n^{-1} g(X_t, \theta_n)\} h(Y_t, \theta_n),$$

where $G_n(x, \theta) = n^{-1} \sum_{t=1}^n g(X_t, \theta) 1(X_t \leq x)$ and $\Gamma_n = n^{-1} \sum_{t=1}^n g(X_t, \theta_n) g'(X_t, \theta_n)$. An implicit assumption throughout this section is that the score $g(X_t, \theta)$ is known up to the finite-dimensional parameter $\theta_0$; see Section 6 for an extension of our results to the case of infinite-dimensional unknown score $g(X_t, \theta)$.

A popular example of functional is the Cramer-von Mises-type (CvM) test statistic

$$CvM_n := \int_{\mathbb{R}^d} \left| \hat{R}_n^1(x) \right|^2 dF_{n,X}(x),$$

where $F_{n,X}$ is the empirical distribution function of $\{X_t\}_{t=1}^n$. For computation of (12) in regression and quantile regression models see Section 5. We choose $CvM_n$ over other alternative functionals for the simplicity of its computation.

In spite of (10), we shall prove that, like with $R_n$, the limit distribution of functionals of $\hat{R}_n^1$, such as $CvM_n$, will still depend on the DGP and the model, although, unlike with $R_n$, not on the estimator $\theta_n$. The latter advantage suffices to allow for a simple bootstrap approximation in Section 4.

3. Asymptotic Theory

3.1 Asymptotic Null Distribution

This section introduces sufficient conditions to establish the asymptotic null distribution of $\hat{R}_n^1$. Let $\mathcal{F}_t := \sigma(Y_{t-1}, X_t, ...)$ be the $\sigma$-field generated by the information set obtained up to time $t$. Let $F(\cdot)$ be the joint cumulative distribution function (cdf) of $(Y_t, X_t)$, and let $F_Y(\cdot)$ and $F_X(\cdot)$ be their marginal cumulative distributions, respectively. By defining $\hat{R}_n^1(\infty) = \hat{R}_n^1(0) = 0$, the sample paths of $\hat{R}_n^1$ belong to the space $\ell^\infty(\mathbb{R}^d_x)$, the space of all uniformly bounded real functions on $\mathbb{R}^d_x := [-\infty, \infty]^d_x$, which is equipped with the sup-norm. Tests statistics are continuous (in the sup-norm) functionals of $\hat{R}_n^1$, say $\varphi(\hat{R}_n^1)$, for some suitable continuous functional $\varphi : \ell^\infty(\mathbb{R}^d_x) \mapsto \mathbb{R}^+ \equiv [0, \infty)$. Once a Functional Central Limit Theorem for $\hat{R}_n^1$ is provided, the limiting distribution of $\varphi(\hat{R}_n^1)$ under $H_0$ is obtained by applying the Continuous Mapping Theorem.

In this article we consider convergence in distribution of empirical processes in the metric space $\ell^\infty(\mathbb{R}^d_x)$ with the sup-norm in the sense of J. Hoffmann-Jørgensen (see, e.g., Dudley
1999 p. 94). The symbol $\Rightarrow$ denotes weak convergence in $\ell^\infty(\mathbb{R}^d)$ and the corresponding Borel sigma field. Measurability issues are resolved by standard arguments based on outer probabilities and they are not further discussed in this paper; see van der Vaart and Wellner (1996). See also Andrews (1994) for earlier applications of empirical processes theory in econometrics.

Our first result proves the asymptotic uniform equivalence between $\hat{R}_n^1$ and $R_n^1(\cdot, \theta_0)$. To this end, we need the following assumptions. Let $\Theta_0$ be an arbitrary small neighborhood around $\theta_0$. For $x \in \mathbb{R}^d$, we define the random function

$$G_t(x) := E\left[ E[\sup_{\theta \in \Theta_0} |h(Y_t, \theta)|^2 | X_t] 1(X_t \leq x) \mid \mathcal{F}_{t-1} \right]$$

Recall $m(x, \theta) := E[h(Y_t, \theta) \mid X_t = x]$ and $\Gamma = E[g(X_t, \theta_0) g'(X_t, \theta_0)]$. The following regularity condition is necessary for the subsequent asymptotic analysis.

**Assumption A1:**

A1(a): $E[\sup_{\theta \in \Theta_0} |h(Y_t, \theta)|^2] < \infty$. The cdf $F_X$ is absolutely continuous. Under $H_0$, $E[h(Y_t, \theta) \mid \mathcal{F}_t] = E[h(Y_t, \theta) \mid X_t]$ a.s. in an arbitrary neighborhood of $\theta_0$.

A1(b): $\{Z_t\}_{t \in \mathbb{Z}}$ is a strictly stationary and ergodic process.

A1(c): The function $m(x, \theta)$ is once continuously differentiable in $\Theta$ with derivative $g(x, \theta)$ satisfying $E\left[ \sup_{\theta \in \Theta_0} |g(X_t, \theta)|^2 \right] < \infty$. Furthermore, $E[g(X_t, \theta) g'(X_t, \theta)]$ is positive definite in $\Theta_0$ and $E[|g(X_t, \theta_0) h(Y_t, \theta_0)|^2] < \infty$.

A1(d) $|G_t(x_1) - G_t(x_2)| \leq C_t |x_1 - x_2|^{s_1}$, for each $(x_1, x_2) \in \mathbb{R}^d \times \mathbb{R}^d$, some $s_1 > 0$. Moreover, for some $s > 0$, all sufficiently small $\delta > 0$ and with $\overline{g}(x, \theta) = I_h$ or $\overline{g}(x, \theta) = g(x, \theta))$,

$$E\left[ \sup_{|\theta_1 - \theta_2| \leq \delta, \theta_1, \theta_2 \in \Theta_0} |\overline{g}(X_t, \theta_1) h(Y_t, \theta_1) - \overline{g}(X_t, \theta_2) h(Y_t, \theta_2)|^2 \mid \mathcal{F}_t \right] \leq C_t \delta^s \text{ a.s.}$$

for a generic stationary sequence $C_t$ with $E|C_t| < \infty$.

**Assumption A2:** The parametric space $\Theta$ is compact in $\mathbb{R}^p$. The parameter $\theta_0$ belongs to the interior of $\Theta$ and $\sqrt{n}(\theta_n - \theta_0) = O_P(1)$.

Assumptions A1-A2 are considerably weaker than related conditions in the literature of CMR. For instance, Dominguez and Lobato (2004) require the existence of at least fourth moments and smoothness of $h(Y_t, \theta)$ in $\theta$. We only require second moments and conditional Lipschitz conditional moments in A1(d). The Markov assumption in A1 is needed to apply martingale theory and for the score $g(X_t, \theta)$ to depend just on $X_t$ and not the whole history.
\(\mathcal{F}_t\). The local Markov assumption can be relaxed at the cost of longer proofs. Note that we do not need mixing assumptions in A1. In A2 we require \(\sqrt{n}(\theta_n - \theta_0) = O_P(1)\). Without further regularity conditions to those in Theorem 1 below, the minimum-distance estimator (cf. Wolfowitz 1957)

\[
\theta_{n,CGMM} := \arg\min_{\theta \in \Theta} \int |R_n(\cdot, \theta)|^2 dF_{n,X}(x),
\]

satisfies A2. Indeed, the estimator \(\theta_{n,CGMM}\) is a generalization to that proposed in Dominguez and Lobato (2004), allowing for non-smooth moment functions. A direct consequence of our results is that, under A1-A2,

\[
\sqrt{n}(\theta_{n,CGMM} - \theta_0) = -\Sigma_{GG}^{-1} \Sigma_{GR_\infty} + o_P(1),
\]

with \(\Sigma_{GG'} = \int G(x, \theta_0)G'(x, \theta_0)dF_X(x), \Sigma_{GR_\infty} = \int G(x, \theta_0)R_\infty(x, \theta_0)dF_X(x)\), and where \(R_\infty(\cdot, \theta_0)\) is Gaussian process with mean zero and covariance function

\[
K(x, y) := E[h(Y_t, \theta_0) h'(Y_t, \theta_0) 1(X_t \leq x \wedge y)],
\]

with \(x \wedge y\) the coordinate-wise minimum of \(x\) and \(y\). This estimator can be used in our Example 1 on asset pricing to avoid possible identification problems with GMM, that can lead to inconsistency of the GMM estimator and the resulting inferences based on \(\hat{R}^1_n\) (cf. Dominguez and Lobato, 2004).

We are now in position to state the first important result of the paper.

**Theorem 1:** Under Assumptions A1-A2,

\[
\sup_{x \in \mathbb{R}^d_x} \left| \hat{R}^1_n(x) - R^1_n(x, \theta_0) \right| = o_P(1),
\]

We remark that the uniform representation in Theorem 1 holds under both, the null and alternative hypotheses. A consequence of Theorem 1 is the weak convergence of \(\hat{R}^1_n\). To simplify the notation, define \(1^+(X_t \leq x) := \{1(X_t \leq x)I_h - G'(\cdot, \theta_0)\Gamma^{-1}g(X_t, \theta_0)\}\).

**Corollary 1:** Under Assumptions A1-A2, if (1) holds, then

\[
\hat{R}^1_n \Rightarrow R^1_\infty,
\]
where \( R_{\infty}^1 \) is a Gaussian process with mean zero and covariance function:

\[
K^1(x, y) := E[\mathbb{1}^\infty(X_t \leq x)h(Y_t, \theta_0)h'(Y_t, \theta_0)(\mathbb{1}^\infty(X_t \leq y))] .
\]

As a result of Corollary 1, the Continuous Mapping Theorem and Glivenko-Cantelli’s Theorem, the asymptotic distribution of \( \text{CvM}_n \) is

\[
\text{CvM}_n \Rightarrow \text{CvM}_\infty := \int_{R^d_x} |R_{\infty}^1(x)|^2 dF_X(x),
\]

We proceed to study the power properties of tests based on \( \hat{R}_n^1 \) in the next section.

### 3.2 Asymptotic Power Properties

This section discusses both, global and local, power properties of \( \hat{R}_n^1 \). Consider the local (Pitman) alternatives

\[
H_{A_n}(a) : E[h(Y_t, \theta_0) \mid X_t] = \frac{a(X_t)}{\sqrt{n}} \text{ a.s.}
\]

where \( a(X_t) \) is a non-zero \( P_X \)-integrable function such that

\[
E[g(X_t, \theta_0)a(X_t)] = 0. \tag{13}
\]

Note that condition (13) does not entail any loss of generality, see Escanciano (2008b).

A consequence of Theorem 1 is that, under the local alternatives \( H_{A_n}(a) \),

\[
\hat{R}_n^1 \Rightarrow R_{\infty}^1 + D_a, \tag{14}
\]

where \( R_{\infty}^1 \) is defined in Corollary 1 and \( D_a(x) := E[\mathbb{1}(X_t \leq x)a(X_t)] \).

Several conclusions emerge from the convergence in (14). First, our tests are able to detect any Pitman local alternative satisfying (13). Second, suppose instead of \( \hat{R}_n^1 \) one entertains a classical test based on the marked process \( R_n(\cdot, \theta_n) \). Under the standard assumption that \( \theta_n \) satisfies an asymptotic Bahadur expansion under \( H_{A_n}(a) \) such as

\[
\sqrt{n}(\theta_n - \theta_0) = \xi_a + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Lambda^{-1}k(X_t, \theta_0)h(Y_t, \theta_0) + o_P(1), \tag{15}
\]

where \( \xi_a = E[k(X_t, \theta_0)a(X_t)] \), \( \Lambda = E[k(X_t, \theta_0)g'(X_t, \theta_0)] \) and \( k(X_t, \theta_0) \) is a measurable \((p \times d_h)\)-matrix valued function, then it can be similarly proved that under the local alternatives
$H_{A,n}(a)$, the following weak convergence holds

$$R_n(\cdot, \theta_n) \xrightarrow{} \tilde{R}_\infty + D_a,$$

where $\tilde{R}_\infty$ is a zero mean Gaussian process with covariance function

$$\tilde{K}(x, y) := E[\bar{I}(X_t \leq x) h(Y_t, \theta_0) h'(Y_t, \theta_0)(\bar{I}(X_t \leq y))]',$$

where $\bar{I}(X_t \leq x) = 1(X_t \leq x)I_h - G'(x, \theta_0)\Lambda^{-1}k(X_t, \theta_0)$.

Notice that the same shift function $D_a$ appears in (14) and (16). Assuming for the time being that the model is conditionally homocedastic with unit conditional variance, i.e. $\Omega^2(X_t) := E[h(Y_t, \theta_0)h'(Y_t, \theta_0) \mid X_t] \equiv I_h$, it follows from standard properties of orthogonal projections that for all $x \in \mathbb{R}^d$, $K^1(x, x) \leq \tilde{K}(x, x)$. In other words, the process $R^1_\infty$ is of “smaller” magnitude than $\tilde{R}_\infty$. One can prove that, as a result, tests based on $\tilde{R}_\infty$ will have better power properties than tests based on $R_n(\cdot, \theta_n)$. To prove this point formally, given a “check” function $\psi(\cdot)$, define the test, say $T_{n1}$, based on rejecting for large absolute values of

$$\int \psi(x)\hat{R}^1_n(dx, \theta_n).$$

Similarly, denote by $T_{n2}$ the test rejecting for large absolute values of

$$\int \psi(x)R^1_n(dx, \theta_n).$$

Then, for any square integrable functions $\psi$ and under mild regularity conditions, including (13), it can be shown that the Pitman Asymptotic Relative Efficiency of $T_{n1}$ and $T_{n2}$ is

$$ARE_\psi := \frac{E[|P\hat{\psi}(X_t)|^2]}{E[|L\hat{\psi}(X_t)|^2]},$$

where we introduce the linear operators $P\psi(X_t) := \psi(X_t) - E[g'(X_t, \theta_0)\psi(X_t)]\Gamma^{-1}g(X_t, \theta_0)$ and $L\psi := \psi(X_t) - E[g'(X_t, \theta_0)\psi(X_t)]\Lambda^{-1}k(X_t, \theta_0)$, respectively. Since $P$ is an orthogonal projection and $L$ also projects in the space orthogonal to the score, it follows by the least squares property that $ARE_\psi \leq 1$, with $ARE_\psi < 1$ if $k(X_t, \theta_0)$ is not collinear to the score $g(X_t, \theta_0)\psi$. Therefore, we conclude that for a general class of estimators in homocedastic CMR, our construction leads to test with better power properties than classical tests based

$^4$The case $k(X_t, \theta_0) = g(X_t, \theta_0)$ corresponds to the use of the efficient estimator. For nonlinear models our construction is expected to lead to better finite-sample performance even when $ARE_\psi = 1$, since the orthogonality to the score holds in finite samples.
on $R_1^n$.

We remark that we do not need conditional homoscedasticity for Theorem 1 to hold, but only to guarantee that projected tests have higher power than non-projected tests. Under conditional heteroskedasticity projected test may or may not be more powerful than non-projected tests. Our argument, however, does not depend on the conditional variance and it can be readily extended to general conditional heteroskedasticity of unknown form if we modify our projected marked process as

$$\tilde{R}_1^n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(X_t \leq x)I_h - \tilde{G}_n(x, \theta_n)\tilde{\Gamma}_n^{-1}g(X_t, \theta_n)\} \Omega_n^{-2}(X_t)h(Y_t, \theta_n),$$

(17)

where $\tilde{G}_n(x, \theta) = n^{-1} \sum_{t=1}^{n} g(X_t, \theta)\Omega_n^{-2}(X_t)1(X_t \leq x)$, $\tilde{\Gamma}_n = n^{-1} \sum_{t=1}^{n} g(X_t, \theta_n) \Omega_n^{-2}(X_t)g'(X_t, \theta_n)$ and $\Omega_n^2(X_t)$ is a nonparametric estimator $\Omega^2(X_t)$, e.g. a Nadaraya-Watson kernel estimator. The asymptotic theory for $\tilde{R}_1^n$ can be established using the results in Section 7. In particular, from our results it can be shown that nonparametric estimation of $\Omega^{-2}(X_t)$ has no effect in the (first order) asymptotic theory of $\tilde{R}_1^n$, see Theorem 3. In this paper we emphasize $\tilde{R}_n$ over $\tilde{R}_1^n$ for simplicity of computation of the former. Readers concerned with efficiency issues may prefer to use $\tilde{R}_n$ rather than $\tilde{R}_1^n$. A comparison of both procedures is beyond the scope of this paper.

Regarding the global power properties, our Theorem 1 yields that under the alternative hypothesis

$$n^{-1/2} \tilde{R}_n(\cdot) \Rightarrow E[1^\perp(X_t \leq \cdot)h(Y_t, \theta_0)].$$

Therefore, our test is consistent against all alternatives not collinear to the score, that is, all $m$ such that $E[1^\perp(X_t \leq x)m(X_t, \theta_0)] \neq 0$ in a set with positive $P_X$-measure. Note that this is not an important limitation. After all, all tests based on integrated CMR have trivial local power against those directions, see Escanciano (2008b). As a result, the global power of all tests in the direction of the score will be also low (cf. Strasser 1990).

### 4. Bootstrap-Based Tests

In the previous section we have shown that the asymptotic null distribution of $\tilde{R}_n$ depends in a complicated way on the DGP, as well as the specification under the null hypothesis. Therefore, critical values for the tests statistics cannot be tabulated for general cases. There are some available solutions in the literature that are only applicable to or they have been only justified for regression models$^5$. A popular choice within regression models is the wild

$^5$Among these are, the conditional p-value method of Hansen (1996), the Khmaladze’s transformation (cf.
bootstrap used in Stute et al. (1998), Whang (2000) and Escanciano (2007a), among others. For instance, Escanciano (2007a) proposed a wild-bootstrap for general ICM tests in dynamic nonlinear regression models. The wild-bootstrap requires the estimation of parameters \( \theta_0 \) in each bootstrap replication, and more importantly, a separable moment function \( h(Y_t, \theta) \), in the sense that the dependent variable, say \( Y_{1,t} \), has to be recovered from the generalized error \( h(Y_t, \theta) \) in an additive manner\(^6\). In particular, the classical wild-bootstrap procedure cannot be directly applied to quantile regressions.

There are other, more general, resampling methods that can be applied to CMR. For instance, Corradi and Swanson (2002) considered the block bootstrap, whereas Chen and Fan (1999) used the stationary bootstrap of Politis and Romano (1994) and an extension of the conditional p-value method of Hansen (1996). For quantile regression problems, the subsampling approach has been the most popular choice; see Chernozhukov (2002), Whang (2004) and Escanciano and Velasco (2007), among others. The subsampling method requires, however, the choice of a tuning parameter, namely, the size of the subsample, whose choice can change dramatically the outcome of the resulting inferences. Two researches with the same data set and model can reach different conclusions using different tuning parameters. This undesirable property is shared by other available methods such as the block-bootstrap.

Our approach is more related to, but is different from, the bootstrap method in Delgado, Domínguez and Lavergne (2006), who proposed a resampling method based on the asymptotic linear expansion of the classical marked empirical process. Unlike their test, our tests do not need neither to estimate the influence function of the estimator nor require the estimator to have a Bahadur linear expansion. Moreover, since our tests are based on *orthogonal* projections they are expected to be more powerful than tests based on non-orthogonal projections, see Section 3.

In this section we propose a bootstrap method to solve the problem of approximating the asymptotic null distribution of \( \hat{R}_1^n \). Our definition of \( \hat{R}_1^n \) allows for a simple application of the multiplier-type bootstrap principle (see Chapter 2.9 in van der Vaart and Wellner (1996)), without estimating the parameters in each bootstrap replication. The multiplier approximation has been successfully applied in the literature of regression models, see, e.g., Hansen (1996). As we shall show later in this section, our approach can be seen as an extension of the wild bootstrap to general and possibly non-separable CMR. The new approach does not need any tuning parameter and seems to be superior to other competing methods in terms of accuracy of the approximation; see the Monte Carlo section.

---

\(^6\) The standard example is a regression model, where \( h(Y_t, \theta) = Y_{1,t} - m(X_t, \theta) \) and therefore, \( Y_{1,t} = m(X_t, \theta) + h(Y_t, \theta) \).

We shall approximate the asymptotic distribution of \( \varphi(\hat{R}_n^1) \) by that of \( \varphi(\hat{R}_{n1}^1) \), where \( \hat{R}_{n1}^1 \) is a simple multiplier-bootstrap approximation of \( \hat{R}_n^1 \) given by

\[
\hat{R}_{n1}^1(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(X_t \leq x)I_h - C_n'(x, \theta_n)\Gamma_n^{-1}g(X_t, \theta_n)\}V_t h(Y_t, \theta_n),
\]

and where \( \{V_t : 1 \leq t \leq n\} \) is a sequence of independent random variables (r.v’s) with zero mean, unit variance, bounded support and also independent of the sequence \( \{(Y_t', X_t')' : 1 \leq t \leq n\} \). Examples of \( \{V_t\} \) sequences are i.i.d. Bernoulli variates with

\[
P(V_t = 0.5(1 - \sqrt{5})) = b \quad P(V_t = 0.5(1 + \sqrt{5})) = 1 - b,
\]

where \( b = (1+\sqrt{5})/2\sqrt{5} \), or \( P(V_t = 1) = 0.5 \) and \( P(V_t = -1) = 0.5 \), see Wu (1986). To justify theoretically this bootstrap approximation we do not need any further additional assumptions. Notice that other related wild-bootstrap procedures require additional assumptions, for instance, on the bootstrap version of the estimator.

To prove the consistency of the proposed bootstrap we use the concept of convergence in distribution in probability one. A less restrictive concept is convergence in distribution in probability, see Giné and Zinn (1990). The unknown limiting null distribution of \( \varphi(\hat{R}_n^1) \), i.e. the distribution of \( \varphi(R_{n1}^1) \), is approximated by the bootstrap distribution of \( \varphi(\hat{R}_{n1}^1) \). That is, the bootstrap distribution

\[
F^*(x) = \Pr \left( \varphi(\hat{R}_{n1}^1) \leq x \mid \{Y_t, X_t\}_{t=1}^{n} \right)
\]

estimates the asymptotic null asymptotic distribution function

\[
F_{\infty}(x) = \Pr \left( \varphi(R_{n1}^1) \leq x \right).
\]

Thus, \( H_0 \) will be rejected at the 100\( \alpha \)% of significance when \( \varphi(\hat{R}_{n1}^1) \geq c_{n,\alpha}^*, \) where \( F^* \left( c_{n,\alpha}^* \right) = 1 - \alpha \). Also, we can use the bootstrap \( p \) – values, \( p_n^* \) say, rejecting \( H_0 \) when \( p_n^* < \alpha \), where \( p_n^* = \Pr \left( \varphi(\hat{R}_{n1}^1) \geq \varphi(\hat{R}_n^1) \mid \{Y_t, X_t\}_{t=1}^{n} \right) \). The bootstrap assisted test is valid if \( F^* \) is a consistent estimator of \( F_{\infty} \) at each continuity point of \( F_{\infty} \). When consistency is a.s., it is expressed as \( \hat{R}_{n1}^1 \to_d \varphi(R_{n1}^1) \) a.s. See Giné and Zinn (1991) or van der Vaart and Wellner (1996) for discussion.

**Theorem 2:** Assume A1-A2. Then, for any continuous functional \( \varphi \)

\[
\varphi(\hat{R}_{n1}^1) \to_d \varphi(R_{n1}^1) \text{ a.s.}
\]

15
It is straightforward to prove that Theorem 2 implies that our bootstrap test is consistent against all alternatives in $H_A$ not collinear to the score, provided $\varphi$ is such that $\varphi(f) = 0 \iff f = 0$ a.s.-$P_X$. Moreover, it can be proved that our bootstrap test preserves the asymptotic local power properties of $\varphi\left(\hat{R}^1_n\right)$, including its asymptotic admissibility. Details are omitted to save space.

We compare now the new bootstrap procedure with existing bootstrap methods such as the bootstrap test proposed in Delgado, Domínguez and Lavergne (2006) and the wild bootstrap proposed in Wu (1986). Suppose the parameter $\theta_0$ can be estimated by a $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ satisfying (15). As with $\hat{R}^1_n$, it can be shown that under (1) and other regularity conditions, the distribution of $R_n(x, \theta_n)$ can be approximated by that of

$$n^{-1/2} \sum_{t=1}^n \{1(X_t \leq x)I_h - G'_n(x, \theta_n)\Lambda_n^{-1} k(X_t, \theta_n)\} V_t h(Y_t, \hat{\theta}_n), \tag{19}$$

where $\Lambda_n = n^{-1} \sum_{t=1}^n k(X_t, \theta_n) g'(X_t, \theta_n)$. This is a different, valid, bootstrap procedure to the one proposed in this paper. For iid observations and smooth moment functions, this bootstrap has been used in Delgado, Domínguez and Lavergne (2006) to approximate the asymptotic distribution of $R_n(x, \theta_n)$. This bootstrap is also related to the conditional p-value method of Hansen (1996). For generalized linear regression models, this method was used in Su and Wei (1991).

We establish some interesting connections of the bootstrap in (19), the classical wild-bootstrap and our methods. When $k(X_t, \theta_0) = g(X_t, \theta_0)$, our bootstrap approximation is asymptotically equivalent to that in Delgado et al. (2006). Interestingly enough, for a linear regression model using OLS estimators, our bootstrap boils down to the bootstrap approximation in Delgado et al. (2006), which in turn coincides with the classical wild bootstrap. The proof of this statement is straightforward, and hence omitted. For nonlinear models, even when $k(X_t, \theta_0) = g(X_t, \theta_0)$, the three bootstrap methods are different in finite samples, since in general $R_n(\cdot, \theta_n) \neq \hat{R}^1_n$. As discussed above, our construction being based on orthogonal projections it is expected to lead to tests with better power properties.

5. Monte Carlo Simulations

In this section we study, via some Monte Carlo experiments, the finite-sample performance of the proposed bootstrap, in comparison with the subsampling. We consider two sets of simulations. The first experiment corresponds to a mean regression model. The model under the null hypothesis is

$$Y_{1,t} = \beta_0 Y_{1,t-1} + \gamma_0 W_t + \varepsilon_t,$$
where $\theta_0 = (\beta_0, \gamma_0)$ and $E[Y_{1,t} - \beta_0 Y_{1,t-1} - \gamma_0 W_t \mid X_t] = 0$. Henceforth, $\{W_t\}$ and $\{\nu_t\}$ will denote a sequence of mutually independent i.i.d. N(0,1) r.v’s. We examine the adequacy of this model under the following DGPs:

DGP 1. An AR(1) model with Gaussian (N(0,1)) errors: $Y_{1,t} = 0.6Y_{1,t-1} + W_t + \nu_t$.

DGP 2. An AR(1) model with Stochastic Volatility (SV): $Y_{1,t} = 0.6Y_{1,t-1} + W_t + \varepsilon_t$, where $\varepsilon_t = \nu_t \exp(\sigma_t^2)$, $\sigma_t^2 = 0.936\sigma_{t-1}^2 + 0.32U_t$, and $U_t$ are i.i.d. N(0,1) r.v’s, independent of $\nu$’s.

DGP 3. An AR(1) model with Student-t errors with 3 degrees of freedom (t(3)): $Y_{1,t} = 0.6Y_{1,t-1} + Y_{1,t-1} + \varepsilon_t$, where $\varepsilon_t \sim$ i.i.d. t(3).

DGP 4. An AR(1) model with GARCH(1,1) errors (GARCH): $Y_{1,t} = 0.6Y_{1,t-1} + W_t + \varepsilon_t$, where $\varepsilon_t = \nu_t \sigma_t$ and $\sigma_t^2 = 0.001 + 0.05\varepsilon_{t-1}^2 + 0.90\sigma_{t-1}^2$.

DGP 5. Threshold Autoregressive Model (TAR),

\[
    Y_{1,t} = \begin{cases} 
        0.6Y_{1,t-1} + \nu_t & ; \ Y_{1,t-1} < 1, \\
        -0.5Y_{1,t-1} + \nu_t & ; \ Y_{1,t-1} \geq 1.
    \end{cases}
\]

DGP 6. Sign autoregressive model (SIGN), $Y_{1,t} =$sign($Y_{1,t-1}$)+0.43$\nu_t$, where sign($x$) = 1 ($x > 0$)−1 ($x < 0$).

DGP 7. Tem Map model (TM),

\[
    Y_{1,t} = \begin{cases} 
        Y_{1,t-1}/\alpha & ; \ 0 < Y_{1,t-1} < \alpha, \\
        (1 - Y_{1,t-1})/(1 - \alpha) & ; \ \alpha \leq Y_{1,t-1} \leq 1,
    \end{cases}
\]

where $\alpha = .49999$ and $Y_{1,0}$ is generated from $U [0, 1]$.

DGP 8. Exponential AR(1) model (EXPAR), $Y_{1,t} = .5Y_{1,t-1} + 10Y_{1,t-1} \exp(-Y_{1,t-1}^2) + \nu_t$.

Models 1–5 are used to examine the empirical size of the tests, while models 6–10 allow us to examine the empirical power. We study the performance of our test for testing the correct specification of the regression model. Therefore, the corresponding moment function is $h(Y_t, \theta_0) = Y_t - \beta_0 Y_{t-1} - \gamma_0 W_t$, $Y_t = (Y_{1,t}, Y_{1,t-1}, W_t)'$, and $X_t = (Y_{1,t-1}, W_t)'$. The parameter $\theta_0$ is estimated by the ordinary least squares estimator $\theta_{n,1} = \left(\sum_{t=1}^{n} X_t X_t'\right)^{-1} \left(\sum_{t=1}^{n} X_t Y_{1,t}\right)$. Note that for this model

\[
    G_n(x, \theta) = n^{-1} \sum_{t=1}^{n} X_t 1(X_t \leq x) \text{ and } \Gamma_n = n^{-1} \sum_{t=1}^{n} X_t X_t'.
\]
The computation of $CvM_n$ is very simple. After some algebra, we can write

$$CvM_n = n^{-2}h'(\theta_n)PP'h(\theta_n)$$

where $h(\theta_n) = (h(Y_1, \theta_n), \ldots, h(Y_n, \theta_n))'$, $P = HW$, $H = I_n - X(X'X)^{-1}X'$, $I_n$ is the $n \times n$ identity matrix, $X$ is the $n \times 2$ matrix with rows $X_i'$, and $W$ is the $n \times n$ matrix with elements $w_{ts} = 1(X_t \leq X_s)$. Similarly, if we denote by $CvM^*_n$ the bootstrap version, we can compute

$$CvM^*_n = n^{-2}h^*(\theta_n)PP'h^*(\theta_n),$$

where $h^*(\theta_n) = (V_1h(Y_1, \theta_n), \ldots, V_nh(Y_n, \theta_n))'$, and $\{V_i\}_{i=1}^n$ are generated from (18). These formulas show the simple computation of our bootstrap procedure, only the vector of residuals changes in each bootstrap replication, from $h(\theta_n)$ to $h^*(\theta_n)$.

We compare our approximation with the subsampling approximation. Subsampling has been shown to be a powerful resampling scheme that allows for asymptotically valid inference under very general conditions on the DGP, see the monograph by Politis, Romano and Wolf (1999). We describe the subsampling approximation in our present framework. Let $G_n^{CvM}(w)$ be the cdf of the $CvM_n$ test statistic, i.e.

$$G_n^{CvM}(w) = P(CvM_n \leq w).$$

Let $CvM_b(\hat{R}_{b,i}) = CvM_b(\hat{R}_b^i(Z_i, \ldots, Z_{i+b-1}))$ be the test statistic computed with the subsample $(Z_i, \ldots, Z_{i+b-1})$ of size $b$, with $Z_i = (Y_i', X_i')'$. 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 $G_n^{CvM}(w)$ using the distribution of the values of $CvM_b(\hat{R}_{b,i})$ computed over the $n - b + 1$ different subsamples of size $b$. That is, we approximate $G_n^{CvM}(w)$ by the empirical distribution

$$G_{n,b}^{CvM}(w) = \frac{1}{n-b+1} \sum_{i=1}^{n-b+1} 1(CvM_b(\hat{R}_{b,i}) \leq w), \quad w \in [0, \infty).$$

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

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

Thus, the subsampling test rejects the null hypothesis at $\tau \%$ if $CvM_n > c_{n,1-\tau,b}^{\Gamma}$. The validity of the subsampling approximation was proved by Politis, Romano and Wolf (1999) for strong mixing sequences provided $b \to \infty$ as $n \to \infty$ and $b/n \to 0$. For the simulations, the size of subsamples is chosen to be $b = \lfloor kn^{2/5} \rfloor$, with $[\cdot]$ denoting the integer part, as suggested
by Sakov and Bickel (2000), with \( k = 3, 4, \) and 5. We have also considered subsampling simulations based on the centered statistics \( \hat{R}_{b,i}^1 \) and \( n^{-1/2} \hat{R}_n^1 \), as in Chernozhukov (2002), but the difference with respect to the uncentered case is not significant, so these simulations are not reported.

In each of 1000 replications, the CvM test statistic is computed for each DGP and sample sizes \( n = 50, 100, \) and 300. The number of bootstrap replications is \( B = 999 \). The rejection probabilities (RP) for the new proposed bootstrap approximation are presented as \( \text{CvM}_n \), with \( n \) the sample size, whereas \( \text{CvM}_{n,k} \) denotes the ones from the subsampling-based tests based on \( b = \lfloor kn^{2/5} \rfloor \) subsample size.

Table 1 reports the RP (in percentage points) for models 1-4.

Table 1 about here

The new bootstrap exhibits good size accuracy, uniformly in all nominal levels, models and for all sample sizes. It is clearly superior to the subsampling approximation. The subsampling has a special difficulty maintaining an accurate empirical size at 1%, overrejeting in most cases. It presents a slight underrejection for the N(0,1) model at sample sizes \( n = 100 \) and \( n = 300 \), and a large distortion (again underrejection) for the SV and t(3) models for the same sample sizes. The performance of the subsampling is substantially sensitive to the choice of \( b \).

We report in Table 2 the RP for models 5 to 8. As expected, all the tests are able to detect these alternatives and the rejection probabilities increase as \( n \) increases. The new bootstrap test is superior to the subsampling approximation in three of the alternatives considered, namely, the TAR, TM and EXPAR models, and it is comparable, or even better, to the subsampling for the SIGN model. It is remarkable that the superiority of the new method holds even uniformly in the subsample size \( b \) for most of the alternatives considered.

Table 2 about here

The second Monte Carlo experiment deals with a dynamic conditional median model. We entertain the model:

\[
Y_{1,t} = 0.6Y_{1,t-1} + W_t + cW_t^2 + u_{2t}, \quad t = 1, \ldots, n, \tag{21}
\]

where \( W_t = 0.5W_{t-1} + \varepsilon_t, \) and where both \( u_{2t} \) and \( \varepsilon_t \) are sampled independently from \( N(0,1) \) and \( Y_{1,0} = W_0 = 0 \). Here, the null model corresponds to \( c = 0 \). Under \( H_0 \), a linear quantile model holds with \( Y_t = (Y_{1,t-1}, X_t)' , \) \( X_t = (Y_{1,t-1}, W_t)' , \) and hence the moment function for the median regression is \( h(Y_t, \theta_0) = 1(Y_{1,t} \leq \beta_0 Y_{1,t-1} + \gamma_0 W_t) - 0.5, \) with \( \theta_0 = (\beta_0, \gamma_0)' . \)
In this context $\theta_0$ is estimated by the QRE, $\theta_{KB,n}$ say, which is the minimizer of

$$
\theta \mapsto \sum_{t=1}^{n} \rho_{0.5}(Y_{1,t} - m(X_t, \theta))
$$

with respect to $\theta \in \Theta \subset \mathbb{R}^2$, where $\rho_{\alpha}(\varepsilon) = -(1(\varepsilon \leq 0) - \alpha)\varepsilon$ and $m(X_t, \theta) = \theta_0' X_t$. It has been shown in the literature (see e.g. Koenker and Bassett (1978)) that under some mild regularity conditions

$$
\sqrt{n}(\theta_{KB,n} - \theta_0) = -\Gamma^{-1} f_{ut}^{-1}(0) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t h(Y_t, \theta_n) + o_P(1),
$$

where $\Gamma = E[X_t X_t']$, and $f_{ut}(0)$ is the innovation’s density. Therefore, our assumption $\sqrt{n}(\theta_{KB,n} - \theta_0) = O_P(1)$ holds under mild moment conditions. Note also that $g(X_t, \theta_0) = f_{ut}(0) X_t$. Therefore, there is no need for estimating the innovation density in our test and the computation of our test in the conditional median case can be carried out in the same way as in (20), just replacing the OLS estimator by $\theta_{KB,n}$ and the regression residuals by the median residuals $h(Y_t, \theta_{KB,n})$.

We consider two sample sizes $n = 100$ and $n = 300$. Again, we choose $b = \lfloor kn^{2/5} \rfloor$, but now with $k = 6, 7$ and 8. The new values for $k$ are chosen such that the empirical size of the subsampling tests are close to the nominal size\(^7\). To save space, we only report in the simulations the results at 5% nominal level. The conclusions are similar for other nominal levels. We plot in Figure 1 the rejection probabilities for our bootstrap test $CvM_n$ and the subsampling versions $CvM_{n,k}$ for $n = 50$, at 5% for model (21) and several values of $c$. Figure 2 replicates this plot for $n = 300$.

\textit{Figures 1 and 2 about here}

For $c = 0$, the size performance of the bootstrap and subsampling-based tests is good for both sample sizes considered. When $c \neq 0$, all tests show consistency against these fixed alternatives. The larger the $c$, the higher is the power for all tests. The new bootstrap presents a much higher power than subsampling-based tests for all values of $b$ considered. The difference in power is particularly high for $n = 50$ and large values of $c$. The power for the subsampling tests does not depend substantially on the choice of $b$ for this simulation experiment.

These simulations have shown the excellent performance of the new bootstrap-based test, in comparison with the subsampling approach. The results for the median regression are

\(^7\)Of course, this cannot be done for real data, where we do not know the true DGP.
particularly illuminating, since there is substantial improvement in power with our methods. Note that for the quantile median regression subsampling is the most popular choice in the literature of specification tests. Therefore, we believe that the new test can provide a more efficient inference method for regression quantiles.

6. Dynamics in Stock Market Returns

6.1 Mean and Variance Dynamics in the Hong Kong Stock Market.

We now apply our tests to study the dynamics in mean and variance of the Hong Seng Index (HSI) in the Hong Kong stock market. Following Koul and Ling (2006) we split the sample in four periods of two years of daily observations from June 1, 1988 to May 31, 1996. Koul and Ling (2006) proposed a goodness of fit tests for the error distribution in a class of heteroskedastic time series models. They fitted an AR(1)-GARCH(1,1) model for these data sets and tested for Gaussianity. Their test failed to reject the Gaussian distribution for the third and fourth periods, arguing that the Tian An Men Square event may be the cause for the rejection in the first two periods, the effect gradually disappearing for the third and fourth periods. It is important to stress that their test relies on the correct specification of the conditional mean and variance functions. In that sense, our test can be considered to be a useful pre-test in their framework.

We check the AR(1) mean and GARCH(1,1) variance specifications in Table 3. That is, we entertain the model

\[ Y_t = a_1 Y_{t-1} + \varepsilon_t, \quad \varepsilon_t = \sigma_t u_t, \]

\[ \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \]

and apply our CvM test, comparing with subsampling tests, to check the validity of these specifications. We check for correct mean specification using the moment function \( h_m(Y_t, Y_{t-1}, \theta_0) = Y_t - a_1 Y_{t-1}, \) with \( \theta_0 = a_1 \) estimated by OLS. For the variance specification, we consider \( h_v(Y_t, Y_{t-1}, \theta_0) = [(Y_t - a_1 Y_{t-1})^2/\sigma_t^2] - 1, \) with \( \theta_0 = (a_1, \alpha_0, \alpha_1, \beta_1)' \) estimated by Quasi-Maximum Likelihood Estimation (QMLE). The necessary scores to compute our test can be found in Koul and Ling (2006). In the application we have considered \( X_t = (Y_{t-1}, \ldots, Y_{t-d}), \) with \( d = 1, 3 \) and 5. To save space only the results with \( d = 3 \) are reported, the outcomes with other values of \( d \) being similar. The number of bootstrap replications is \( B = 999, \) and for the subsampling test we chose \( b = \lfloor kn^{2/5} \rfloor, \) with \( k \) from 2 to 6. Only the results for \( k = 3 \) and 5 are reported to save space. For completeness we also report the results for the joint test, using the moment \( h = (h_m, h_v)'). \) Table 3 contains the bootstrap p-values for the mean and variance specifications.
Several conclusions can be drawn from Table 3. First, the mean is correctly specified, as suggested by our test and the subsampling tests. Only the second period is rejected at 10% with our test, but not with the subsampling test. Second and more importantly, the variance seems to be misspecified at 5% for all periods but for the third period, as can be seen from the bootstrap p-values of the new CvM test. The joint test confirms the results of the marginal variance test. These rejections stand in contrast with the outcomes of the subsampling tests across several values of $k$, thereby highlighting the high power of our test relative to subsampling methods. These rejections question the suitability of the inferences in Koul and Ling (2006) for all but the third period of the HSI, since as previously discussed, their methods relied on the correct specification of the GARCH model.


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 for banking risk monitoring. As a result of this agreement, financial institutions have to report their VaR and compute their capital reserves according to the outcome of a statistical test for evaluating the VaR model denominated backtesting. The computation of VaR measures and their evaluation have become then of paramount importance in risk management.

In this section we use our methods to provide an alternative to standard backtesting techniques. Our methods are expected to be more powerful than traditional backtesting methods since they account for a larger information set. This higher power is confirmed in the application below. Moreover, in the presence of estimation risk, our methods are much simpler to implement. We illustrate these points with an application to model the VaR at different quantile levels of the well-known and extensively studied S&P500 daily stock index. We consider several periods of this stock index. The first period spans from 2 January 1990 until 31 December 1993, the second period from 3 January 1994 until 31 December 1997 and the third from 2 January 1998 until 28 August 2002. The number of observations in each period is 1013, 1011 and 1170, respectively.

Within the VaR paradigm, parametric location-scale models have been the most popular in attempting to describe the dynamics of VaR measures. In particular, Berkowitz and O’Brien (2002) compared the VaR forecasts obtained from the location-scale family (ARMA-GARCH model) with the internal structural models used by commercial banks. Their main conclusion was that GARCH models generally provide lower VaRs and are better at pre-
dicting changes in volatility, thereby allowing comparable risk coverage with less regulatory capital.

In this application we fit a GARCH(1,1)-VaR model to the log differences of the S&P500 \( (Y_t) \) as

\[
m_o(X_t, \theta) = c + \sigma_t F_{\alpha}^{-1}(\alpha),
\]

(23)

where \( \sigma_t \) is specified by a GARCH(1,1) and where \( F_{\alpha}^{-1}(\alpha) \) is an unconditional quantile function. We assume that this specification follows from a location-scale GARCH(1,1) model with a constant location. We aim to check the conditional quantile specification, and hence the moment function is \( h(Y_t, X_t, \theta) = 1(Y_t \leq m_o(X_t, \theta)) - \alpha \), and \( \theta_0 \) is estimated by QMLE. We entertain two specifications for \( F_{\alpha}^{-1}(\alpha) \), a Gaussian and Student-t distribution, respectively. In the second case we estimated the degrees of freedom parameter by maximum likelihood.

In order to monitor and assess the accuracy and quality of the different VaR techniques the Basel Accord (1996a) developed a statistical testing device that was denominated backtesting. Backtesting methods are statistical tests for the hypothesis that \( \{h(Y_t, X_t, \theta_0)\}_{t=1}^n \) is a sequence of iid Bernoulli variables with zero mean. Among the existing backtesting methods the most popular ones are the so-called unconditional backtest and test of independence, which reject the VaR model for large values of

\[
K_n = \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h(Y_t, X_t, \theta_0) \right|
\]

and

\[
C_n = \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} h(Y_t, X_t, \theta_0) h(Y_{t-1}, X_{t-1}, \theta_0) \right|
\]

respectively. To approximate the critical values of \( K_n \) and \( C_n \) we use a subsampling approximation, and denote the corresponding test statistics by \( K_{n,k} \) and \( C_{n,k} \).

In Table 4 we report the bootstrap p-values for different choices of the VaR level \( \alpha \) for the Gaussian and Student-t distributions. The simulation parameters such as bootstrap replications, subsample sizes and other parameter values are as in the mean-variance specifications.

Table 4 about here

We have several conclusions from the quantile specifications. First, our test suggests that a GARCH model with either Gaussian or Student-t distribution does not fit the data well. Only for the third period the GARCH model seems to be appropriate, specially with tails modelled with the Student-t distribution when \( \alpha = 0.01 \). This application shows that our test
has more power than competing methods such as subsampling or the popular backtesting methods; see for instance the inability of these methods in rejecting the GARCH model when $\alpha = 0.05$. This application suggests that the conclusions in Berkowitz and O’Brien (2002) have to be taken with caution in view of the lack of power of the traditional backtesting techniques. Overall, our simulations and applications show the superiority of our methods in terms of size accuracy and power performance over competing methods, while being simpler to implement.

7. Extensions: The Case of Nonparametric Scores

This section discusses an important extension of our previous theory. There are situations where the score $g(X_t, \theta_0)$ is not known up to the parameter $\theta_0$ and depends in a nonparametric way on the true DGP. We consider in this section the case where $g(X_t, \theta_0)$ can be estimated nonparametrically. A prominent example is quantile regression in a non-location-scale model. It can be shown that for the $\alpha-th$ conditional quantile error $h(Y_t, \theta_0) = 1(Y_{1,t} \leq m_\alpha(X_t, \theta_0)) - \alpha$, the score process in this general case is given by

$$g(X_t, \theta) = f_{Y/X}(m_\alpha(X_t, \theta_0)) \hat{m}_\alpha(X_t, \theta_0), \quad (24)$$

where $f_{Y/X}(x)$ is the density of $Y_{1,t}$ given $X_t$ evaluated at $x$, $m_\alpha(X_t, \theta_0)$ is the specified quantile model and $\hat{m}_\alpha(X_t, \theta_0) = \partial m_\alpha(X_t, \theta_0) / \partial \theta$. Without further assumptions, the quantity $f_{Y/X}(m_\alpha(X_t, \theta_0))$ is unknown, but it can be estimated nonparametrically by e.g. a kernel estimator. This section extends our previous theory to cover this and other examples. This extension is highly technical and requires sophisticated empirical process methods that are of independent interest.

We introduce some notation. For a given $\delta > 0$ and a given family of measurable functions $\mathcal{G}$, define the class of score functions

$$\mathcal{B}_\delta = \{ s(\cdot, \theta) : s(\cdot, \theta) \in \mathcal{G}, \theta \in \Theta : |\theta - \theta_0| < \delta \},$$

and the empirical process

$$\alpha_n(s, \theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^n s(X_t, \theta) h(Y_t, \theta),$$

indexed by $l = (s, \theta) \in \mathcal{H}_\delta = \mathcal{B}_\delta \times \Theta_\delta$, where $\Theta_\delta = \{ \theta \in \Theta : |\theta - \theta_0| < \delta \}$. Let $B_\delta(X_t)$ be the envelope function for the class $\mathcal{B}_\delta$, that is, $\sup_{s \in \mathcal{B}_\delta} |s(X_t, \theta)| < B_\delta(X_t)$. Then, we
define the *conditional quadratic variation* of the empirical process $\alpha_n$ on a finite partition $\Pi = \{H_k; 1 \leq k \leq N\}$ of $\mathcal{H}_\delta$, which is defined as

$$CV_{\alpha, n}(\Pi) := \max_{1 \leq k \leq N} n^{-1} \sum_{t=1}^{n} E \left[ \sup_{l_1, l_2 \in H_k} |l_1(X_t, \theta_1) h(Y_t, \theta_1) - l_2(X_t, \theta_2) h(Y_t, \theta_2)|^2 \mid \mathcal{F}_t \right],$$

(25)

where henceforth $l_i = (s_i, \theta_i) \in \mathcal{H}_\delta$, $i = 1, 2$. In addition to A1 and A2, for this extension we require the following assumption. We assume $\mathcal{B}_\delta$ is endowed with a pseudo-metric $\|\cdot\|_{\mathcal{B}_\delta}$.

Associated to a possibly nonparametric estimator $\hat{g}(X_t, \theta_n)$ define the estimators

$$\hat{G}_{\hat{g}}(x, \theta_n) = n^{-1} \sum_{t=1}^{n} \hat{g}(X_t, \theta_n) 1(X_t \leq x)$$

and

$$\hat{\Gamma}_{\hat{g}} = n^{-1} \sum_{t=1}^{n} \hat{g}(X_t, \theta_n) \hat{g}'(X_t, \theta_n).$$

**Assumption A3**: For each $\delta > 0$ sufficiently small,

A3(a): $E \left[ |B_\delta(X_t)|^2 \right] < \infty$. Furthermore, $E[|B_\delta(X_t)h(Y_t, \theta_0)|^2] < \infty$.

A3(b): $(\mathcal{B}_\delta, \|\cdot\|_{\mathcal{B}_\delta})$ is a totally bounded metric space.

A3(c): For every $\varepsilon > 0$ there exists a finite partition $\Pi_\varepsilon = \{H_k; 1 \leq k \leq N_\varepsilon\}$ of $\mathcal{H}_\delta$, with $N_\varepsilon$ being the number of elements of such partition, such that

$$\int_{0}^{\infty} \sqrt{\log(N_\varepsilon)} d\varepsilon < \infty$$

(26)

and

$$\sup_{\varepsilon \in (0, 1) \cap \mathbb{Q}} \frac{CV_{\alpha, n}(\Pi_\varepsilon)}{\varepsilon^2} = O_P(1).$$

(27)

A3(d): The estimator $\hat{g}(X_t, \theta_n)$ satisfies $\Pr(\hat{g}(X_t, \theta_n) \in \mathcal{B}_\delta) \rightarrow 1$ as $n \rightarrow \infty$, and it is consistent for $g(X_t, \theta_0)$, in the sense that $\|\hat{g} - g\|_{\mathcal{B}_\delta} = o_P(1)$ for all $\delta > 0$. Moreover, $\hat{G}_{\hat{g}}(x, \theta_n)$ is consistent for $G(x, \theta_0)$, uniformly in $x$, and $\hat{\Gamma}_{\hat{g}}$ is consistent for $\Gamma$.

The next theorem extends Theorem 1 to the nonparametric score case. It shows that estimation of the infinite-dimensional score function has no effect on the first order asymptotic theory of the test statistic.
**Theorem 3:** Under Assumptions A1-A3,

\[
\sup_{x \in \mathbb{R}^d} \left| \hat{R}_n^1(x) - R_n^1(x, \theta_0) \right| = o_P(1),
\]

If we assume that \( g(X_t, \theta_0) \) is a smooth function, usual examples of \( \mathcal{B}_\delta \) are spaces of smooth functions such as Sobolev, Hölder, or Besov classes; see for instance van der Vaart and Wellner (1996, pg. 154). Under appropriate smoothness conditions we can take the partitions in A3(c) with respect to covering or bracketing numbers; see Appendix for definitions of these numbers. For smooth classes the entropy condition in (26) holds under mild conditions; see van der Vaart and Wellner (1996), Theorem 2.7.1.

To finish this section, let us consider the conditional quantile example. The score in (24) can be estimated by

\[
\hat{g}(X_t, \theta_n) = \hat{f}_{Y/X}(m_\alpha(X_t, \theta_n)) \hat{m}_\alpha(X_t, \theta_n),
\]

where \( \theta_n \) is a \( \sqrt{n} \)-consistent estimator of \( \theta_0 \) and \( \hat{f}_{Y/X} \) is the popular Rosenblatt estimator (cf. Rosenblatt, 1969)

\[
\hat{f}_{Y/X}(y) = \frac{\sum_{t=1}^n K(x - X_t/h_n)k(y - Y_t/h_n)}{h_n \sum_{t=1}^n K(x - X_t/h_n)}, \tag{28}
\]

where \( K \) and \( k \) are kernel functions and \( h_n \) is a sequence of bandwidths converging to zero at a suitable rate. Under appropriate smoothness and boundness conditions on the joint density of \( Z_t = (Y_t, X'_t)' \) and the kernels \( K \) and \( k \) it is not difficult to show that Assumption A3 holds for this example. The uniform consistency for \( \hat{f}_{Y/X} \) can be obtained from the results in Hansen (2008). We omit details for the sake of space. Specific assumptions for the quantile example can be obtained from the author upon request.

8. Conclusions and future extensions

In this paper we have proposed new bootstrap-based tests for general, possibly non-smooth, conditional moment restrictions. We have shown that an orthogonal projection in the classical marked empirical process allows for a simple multiplier-type bootstrap approximation, and it leads to a simple and powerful testing procedure. It has the practical convenience of being free of tuning parameters such as the block-size or the subsample size, it does not need to re-estimate parameters in each bootstrap replication, and it is more powerful than subsampling approximations, as shown in the simulations. Therefore, the new
method should be useful for practitioners interested in testing CMR, such as Asset Pricing models, regression models or VaR models, among others. One of the most important appealing properties of the new Cramer-von Mises test is its simplicity, being a quadratic form of the generalized residuals.

Our approach is geometric in nature, and therefore, it can be easily extended to more general frameworks, e.g. semiparametric CMR containing infinite-dimensional nuisance parameters, provided the orthogonal projection into the tangent space of nuisance parameters can be constructed. Since the tangent space of a semiparametric model is substantially larger than that of a parametric model, our methods are specially well-suited for semiparametric models as they incorporate a great amount of information in the projection. Extensions to particular semiparametric models are deferred for future research.

The main objective of this paper has been to propose a simple testing procedure for testing general CMR. Some extensions of the basic framework can improve the proposed test, at the cost of complicating the proofs and the implementation of the test, and obscuring the main message of the paper. Most notably, in Section 3 we discussed an extension of our basic framework to a test with better power properties than classical marked process when there is conditional heteroskedasticity of unknown form, see (17). This extension can be carried out using the methods discussed in Section 7. We have preferred to keep the exposition and the ideas as simple as possible.

Mathematical Appendix

To prove Theorem 1 we need a useful lemma and further definitions. Let \((\mathcal{G}, \|\cdot\|_{\mathcal{G}})\) be a subset of a metric space of vector-valued functions \(g\). The covering number \(N(\varepsilon, \mathcal{G}, \|\cdot\|_{\mathcal{G}})\) is the minimal number of \(N\) for which there exist \(\varepsilon\)-balls \(\{f : \|f - g_j\|_{\mathcal{G}} \leq \varepsilon, \|g_j\|_{\mathcal{G}} < \infty, j = 1, \ldots, N\}\) to cover \(\mathcal{G}\). The covering number with bracketing \(N[](\varepsilon, \mathcal{G}, \|\cdot\|_{\mathcal{G}})\) is the minimal number of \(N\) for which there exist \(\varepsilon\)-brackets \(\{[l_j, u_j] : \|l_j - u_j\|_{\mathcal{G}} \leq \varepsilon, \|l_j\|_{\mathcal{G}} < \infty, \|u_j\|_{\mathcal{G}} < \infty, j = 1, \ldots, N\}\) to cover \(\mathcal{G}\). For the definition of asymptotic tightness see van der Vaart and Wellner (1996).

Define the process

\[
\alpha_n(x, c) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ h(Y_t, \theta_0 + cn^{-1/2}) - E[h(Y_t, \theta_0 + cn^{-1/2}) \mid \mathcal{F}_t] \right\} 1(X_t \leq x).
\]

indexed by \((x, c) \in \mathbb{R}^d_x \times \mathcal{C}_K\), where \(\mathcal{C}_K = \{c \in \mathbb{R}^p : |c| \leq K\}\), and \(K > 0\) is an arbitrary but fixed constant.
Lemma B1: Under Assumptions A1-A2, the process $\alpha_n(x, c)$ is asymptotically tight with respect to $(x, c) \in \mathbb{R}^d \times \mathcal{C}_K$.

Proof of Lemma B1: Let us define the class of functions

$$\mathcal{K} = \{ h(Y_t, \theta_0 + cn^{-1/2}) - E[h(Y_t, \theta_0 + cn^{-1/2}) | \mathcal{F}_t] \} 1(X_t \leq x) : (x, c) \in \mathbb{R}^d \times \mathcal{C}_K \}.$$ 

To simplify notation write

$$w_{t-1}(u) \equiv \{ h(Y_t, \theta_0 + cn^{-1/2}) - E[h(Y_t, \theta_0 + cn^{-1/2}) | \mathcal{F}_t] \} 1(X_t \leq x),$$

where $u = (x, c)$. Define the pseudo-metric, with $s$ as in A1(d),

$$d(u_1, u_2) := \left| F_X(x_1) - F_X(x_2) \right|^{1/2} + |c_1 - c_2|^{s/2},$$

where hereafter $u_1 = (x_1, c_1)$ and $u_2 = (x_2, c_2)$ belong to $\Pi_K := \mathbb{R}^d \times \mathcal{C}_K$. From the compactness of $\mathcal{C}_K$ and the continuity of $F_X$ we have

$$\int_0^\infty \sqrt{\log(N[\delta, \Pi_K, d])} d\delta < \infty.$$ 

Let $\mathcal{B}_\delta = \{ B_k; 1 \leq k \leq N_\delta \equiv N[\delta, \mathcal{C}_K, d] \}$ be a partition of $\Pi_K$ in $\delta$-brackets $B_k = [l_k, u_k]$, i.e., \{ $B_k\}_{k=1}^{N_\delta}$ covers $\Pi_K$, $l_k = (x'_k, c'_k)$, $l_k = (y_k, c_k)$, $x_k \leq y_k$, $c_k \leq c_k$, and $d(x_k, y_k) \leq \delta$. The quadratic variation of $\alpha_n(x, c)$ (cf. 25) is bounded by

\[
\max_{1 \leq k \leq N_\delta} \frac{1}{n} \sum_{t=1}^{n} E \left[ \sup_{u_1, u_2 \in B_k} \left| w_{t-1}(u_1) - w_{t-1}(u_2) \right|^2 | \mathcal{F}_t \right] 
\leq C \max_{1 \leq k \leq N_\delta} \frac{1}{n} \sum_{t=1}^{n} E \left[ \sup_{\theta \in \Theta_0} \left| h(Y_t, \theta) \right|^2 | \mathcal{F}_t \right] \{ 1(X_t \leq y_k) - 1(X_t \leq x_k) \}
\]

\[
+ Cn^{-1} \sum_{t=1}^{n} E \left[ \sup_{|\theta_1 - \theta_2| \leq \delta^{2/s}, \theta_1, \theta_2 \in \Theta_0} \left| h(Y_t, \theta_1) - h(Y_t, \theta_2) \right|^2 | \mathcal{F}_t \right]
\]

\[
\leq C\delta^2 O_p(1),
\]

where the first inequality follows from the triangle inequality and the boundness and monotonicity of the indicator function. The last inequality follows from A1(d) and the proof of Theorem 1 in Escanciano (2007b). Hence, the conditions of Theorem A1 in Delgado and Escanciano (2007) hold for the partition $\mathcal{B}_\delta$, and the asymptotically tightness of $\alpha_n$ is then proved. \(\Box\)
Proof of Theorem 1: Without loss of generality we assume $d_h = 1$. We write

$$
\hat{R}_n^1(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n h(Y_t, \theta_n) 1(X_t \leq x) - G'_n(x, \theta_n) \Gamma_n^{-1} \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n g(X_t, \theta_n) h(Y_t, \theta_n) \right)
$$

$$
= R_n(x, \theta_n) - G'_n(x, \theta_n) \Gamma_n^{-1} I_{g,n},
$$

where $I_{g,n}$ is implicitly defined.

Notice that the uniform law of large numbers (ULLN) of Jennrich (1969, Theorem 2) coupled with the Ergodic Theorem, see e.g. Dehling and Philipp (2002), implies that

$$
\sup_{x \in \mathbb{R}^d} |G_n(x, \theta_n) - G_n(x, \theta_0)| = o_P(1) \quad \text{and} \quad |\Gamma_n^{-1} - \Gamma^{-1}| = o_P(1).
$$

(30)

Note that A2 yields that for each $x \in \mathbb{R}^d$ and each $c \in \mathbb{R}^p$

$$
E \left[ |\alpha_n(x, c) - \alpha_n(x, 0)|^2 \right] = o(1),
$$

where $\alpha_n(x, c)$ is defined in (29). Thus, by Lemma B1, for any $K > 0$

$$
\sup_{x \in \mathbb{R}^d, |c| \leq K} |\alpha_n(x, c) - \alpha_n(x, 0)| = o_P(1).
$$

The latter display and (30) imply the following uniform representation for any $K > 0$,

$$
\sup_{|c| \leq K, x \in \mathbb{R}^d} \left| R_n(x, \theta_0 + cn^{-1/2}) - R_n(x, \theta_0) - G'(x, \theta_0)c \right| = o_P(1).
$$

Taking $c = \sqrt{n}(\theta_n - \theta_0)$ and by A2 we conclude

$$
\sup_{x \in \mathbb{R}^d} \left| R_n(x, \theta_n) - R_n(x, \theta_0) - G'(x, \theta_0) \sqrt{n}(\theta_n - \theta_0) \right| = o_P(1).
$$

(31)

Similarly by an application of Lemma B1 and A1(d)

$$
\left| I_{g,n} - \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n g(X_t, \theta_0) h(Y_t, \theta_0) \right) - \Gamma \sqrt{n}(\theta_n - \theta_0) \right| = o_P(1).
$$

(32)
Therefore, (31) and (32) jointly imply that, uniformly in \( x \),

\[
\hat{R}_n^1(x) = R_n(x, \theta_0) + G'(x, \theta_0) \sqrt{n}(\theta_n - \theta_0) + \sum_{t=1}^{n} g(X_t, \theta_0) h(Y_t, \theta_0) \nabla h(Y_t, \theta_0) + o_P(1)
\]

which completes the proof of Theorem 1. \( \square \)

**Proof of Corollary 1:** Part (i) follows from the expansion Theorem 1 and the joint weak convergence of the process

\[
\frac{1}{\sqrt{n}} \left( \sum_{t=1}^{n} h(Y_t, \theta_0) \mathbb{1}(X_t \leq x) \right)
\]

which follows from our Lemma B1 and the Central Limit Theorem (CLT) for stationary and ergodic martingale difference sequences, cf. Billingsley (1961). Part (ii) follows again from Theorem 1, and the uniform version of the Ergodic Theorem. \( \square \)

Remark that we say that the bootstrap statistic \( \eta_n^* \) converges in probability a.s. to \( \eta_n \) if for all \( \delta > 0 \), \( \Pr( |\eta_n^* - \eta_n| \geq \delta | \{Y_t, X_t\}_{t=1}^{n} ) \rightarrow 0 \) a.s., which is expressed as \( \eta_n^* = \eta_n + o_P(1) \) a.s. Also, bootstrap expectations are denoted by \( E^*(\eta_n^*) = E(\eta_n^* | \{Y_t, X_t\}_{t=1}^{n}) \).

**Proof of Theorem 2:** Following the arguments of the proof of Theorem 1, it can easily proved that

\[
\sup_{x \in \mathbb{R}^d} \left| \hat{R}_n^1(x) - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1^+(X_t \leq x) V_i h(Y_t, \theta_0) \right| = o_P(1) \ a.s.
\]

where recall \( 1^+(X_t \leq x) := \{1(X_t \leq x) I_h - G'(\cdot, \theta_0) \nabla g(X_t, \theta_0)\} \). Define

\[
R_n^1(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1^+(X_t \leq x) V_i h(Y_t, \theta_0).
\]

At this point the proof follows standard arguments, see e.g. Delgado et al. (2006). We need to show that (i) the process \( R_n^* \) (conditionally on the sample) has the same asymptotic finite dimensional distributions that the process \( \hat{R}_n^1 \), and that (ii) \( \hat{R}_n^1 \) is asymptotically tight, both with probability one. Part (i) follows from the Lindeberg-Levy CLT, whereas part (ii) follows from our Lemma B1, see also Theorem 2.9.2 in van der Vaart and Wellner (1996). \( \square \)
Proof of Theorem 3: The proof follows the same steps as Theorem 1. We only need to prove (32). By A3(c) and Theorem A1 in Delgado and Escanciano (2007) the empirical process

$$\alpha_n(g, \theta) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} g(X_t, \theta) \{h(Y_t, \theta) - E[h(Y_t, \theta) \mid \mathcal{F}_t]\},$$

converges weakly to a Gaussian process with zero mean, indexed by \(l = (g, \theta) \in \mathcal{H}_\delta = B_\delta \times \Theta_\delta\), where \(\Theta_\delta = \{\theta \in \Theta : |\theta - \theta_0| < \delta\}\). By Theorem 1.5.7 in van der Vaart and Wellner (1996) and by A3(b), the process \(\alpha_n(g, \theta)\) is \(\rho\)-equicontinuous, for the semimetric

\[
\rho((g_1, \theta_1), (g_2, \theta_2)) = \|g_1 - g_2\|_{B_\delta} + |\theta_1 - \theta_2|.
\]

Define the sets \(A_n = \{\hat{g}(X_t, \theta_n) \in B_\delta\}\) and \(B_n = \{\|\hat{g} - g_2\|_{B_\delta} < \delta\}\). Hence, for any \(\varepsilon > 0\) and \(\eta > 0\), we can choose a sufficiently small \(\delta > 0\) such that

\[
\Pr (|\alpha_n(\hat{g}, \theta_n) - \alpha_n(g, \theta_n)| > \varepsilon) = \Pr (|\alpha_n(\hat{g}, \theta_n) - \alpha_n(g, \theta_n)| > \varepsilon \mid A_n \cap B_n) \Pr (A_n \cap B_n)
\]

\[
\leq \Pr \left( \sup_{\|s - \hat{g}\|_{B_\delta} < \delta} |\alpha_n(s, \theta_n) - \alpha_n(g, \theta_n)| > \varepsilon \mid A_n \cap B_n \right) \Pr (A_n \cap B_n)
\]

\[
< \eta.
\]

The rest of the proof follows the same steps as in Theorem 1. \(\square\)
References


Table 1: Empirical Size of Bootstrap and Subsampling Tests.
Rejection Probabilities (in percentage points)

<table>
<thead>
<tr>
<th></th>
<th>N(0,1)</th>
<th>SV</th>
<th>t(3)</th>
<th>GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha%)</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>(CvM_{50})</td>
<td>10.5</td>
<td>5.1</td>
<td>0.8</td>
<td>10.7</td>
</tr>
<tr>
<td>(CvM_{50,3})</td>
<td>10.0</td>
<td>5.4</td>
<td>3.0</td>
<td>9.0</td>
</tr>
<tr>
<td>(CvM_{50,4})</td>
<td>11.0</td>
<td>6.2</td>
<td>3.0</td>
<td>10.3</td>
</tr>
<tr>
<td>(CvM_{50,5})</td>
<td>11.6</td>
<td>8.4</td>
<td>5.4</td>
<td>11.7</td>
</tr>
<tr>
<td>(CvM_{100})</td>
<td>10.9</td>
<td>5.5</td>
<td>1.0</td>
<td>10.3</td>
</tr>
<tr>
<td>(CvM_{100,3})</td>
<td>7.8</td>
<td>3.3</td>
<td>1.4</td>
<td>7.6</td>
</tr>
<tr>
<td>(CvM_{100,4})</td>
<td>8.6</td>
<td>4.9</td>
<td>1.3</td>
<td>9.4</td>
</tr>
<tr>
<td>(CvM_{100,5})</td>
<td>10.4</td>
<td>5.9</td>
<td>2.2</td>
<td>10.5</td>
</tr>
<tr>
<td>(CvM_{300})</td>
<td>10.6</td>
<td>4.5</td>
<td>1.1</td>
<td>9.7</td>
</tr>
<tr>
<td>(CvM_{300,3})</td>
<td>7.4</td>
<td>3.2</td>
<td>0.7</td>
<td>5.3</td>
</tr>
<tr>
<td>(CvM_{300,4})</td>
<td>8.3</td>
<td>4.2</td>
<td>1.1</td>
<td>4.3</td>
</tr>
<tr>
<td>(CvM_{300,5})</td>
<td>8.8</td>
<td>5.4</td>
<td>2.0</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Monte Carlo simulations 1000, bootstrap replications 999, and subsample sizes \(b = \lfloor kn^{2/5} \rfloor\).

Table 2: Empirical Power of Bootstrap and Subsampling Tests.
Rejection Probabilities (in percentage points)

<table>
<thead>
<tr>
<th></th>
<th>TAR</th>
<th>SIGN</th>
<th>TM</th>
<th>EXPAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha%)</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>(CvM_{50})</td>
<td>32.7</td>
<td>18.5</td>
<td>3.3</td>
<td>29.7</td>
</tr>
<tr>
<td>(CvM_{50,3})</td>
<td>20.5</td>
<td>10.2</td>
<td>5.8</td>
<td>32.3</td>
</tr>
<tr>
<td>(CvM_{50,4})</td>
<td>20.6</td>
<td>12.7</td>
<td>7.1</td>
<td>29.8</td>
</tr>
<tr>
<td>(CvM_{50,5})</td>
<td>23.4</td>
<td>16.0</td>
<td>9.1</td>
<td>28.0</td>
</tr>
<tr>
<td>(CvM_{100})</td>
<td>65.3</td>
<td>46.2</td>
<td>13.7</td>
<td>56.3</td>
</tr>
<tr>
<td>(CvM_{100,3})</td>
<td>37.3</td>
<td>18.7</td>
<td>5.7</td>
<td>55.2</td>
</tr>
<tr>
<td>(CvM_{100,4})</td>
<td>36.4</td>
<td>21.2</td>
<td>8.4</td>
<td>54.1</td>
</tr>
<tr>
<td>(CvM_{100,5})</td>
<td>37.4</td>
<td>25.1</td>
<td>13.2</td>
<td>53.0</td>
</tr>
<tr>
<td>(CvM_{300})</td>
<td>100.0</td>
<td>99.9</td>
<td>95.1</td>
<td>94.2</td>
</tr>
<tr>
<td>(CvM_{300,3})</td>
<td>99.6</td>
<td>94.0</td>
<td>57.4</td>
<td>94.3</td>
</tr>
<tr>
<td>(CvM_{300,4})</td>
<td>99.5</td>
<td>92.2</td>
<td>66.1</td>
<td>92.7</td>
</tr>
<tr>
<td>(CvM_{300,5})</td>
<td>99.0</td>
<td>91.2</td>
<td>69.0</td>
<td>89.9</td>
</tr>
</tbody>
</table>

Monte Carlo simulations 1000, bootstrap replications 999, and subsample sizes \(b = \lfloor kn^{2/5} \rfloor\).
Figure 1. Empirical power function for the median quantile model as a function of $c$. The new bootstrap test $CvM_n$ in solid line. Subsampling tests $CvM_{n,b}$ in dashed, dotted and dash-dotted lines for $b = 37$, 44 and 50, respectively. Sample size $n = 50$, Monte Carlo simulations 1000 and bootstrap replications 999. Nominal level 5%. 

Figure 2. Empirical power function for the median quantile model as a function of $c$. The new bootstrap test $CvM_n$ in solid line. Subsampling tests $CvM_{n,b}$ in dashed, dotted and dash-dotted lines for $b = 37$, 44 and 50, respectively. Sample size $n = 300$, Monte Carlo simulations 1000 and bootstrap replications 999. Nominal level 5%. 

37
**Table 3: Empirical P-Values, Hong Seng Index Fitted**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
<td>493</td>
<td>495</td>
<td>498</td>
<td>497</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CvM_n$</td>
<td>0.492</td>
<td>0.112</td>
<td>0.756</td>
<td>0.370</td>
</tr>
<tr>
<td>$CvM_{n,3}$</td>
<td>0.394</td>
<td>0.277</td>
<td>0.485</td>
<td>0.126</td>
</tr>
<tr>
<td>$CvM_{n,5}$</td>
<td>0.340</td>
<td>0.288</td>
<td>0.503</td>
<td>0.142</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CvM_n$</td>
<td>0.028</td>
<td>0.034</td>
<td>0.716</td>
<td>0.048</td>
</tr>
<tr>
<td>$CvM_{n,3}$</td>
<td>0.114</td>
<td>0.078</td>
<td>0.325</td>
<td>0.182</td>
</tr>
<tr>
<td>$CvM_{n,5}$</td>
<td>0.122</td>
<td>0.055</td>
<td>0.382</td>
<td>0.158</td>
</tr>
<tr>
<td>Joint</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CvM_n$</td>
<td>0.034</td>
<td>0.030</td>
<td>0.686</td>
<td>0.028</td>
</tr>
<tr>
<td>$CvM_{n,3}$</td>
<td>0.114</td>
<td>0.078</td>
<td>0.327</td>
<td>0.182</td>
</tr>
<tr>
<td>$CvM_{n,5}$</td>
<td>0.122</td>
<td>0.055</td>
<td>0.384</td>
<td>0.160</td>
</tr>
</tbody>
</table>

Bootstrap replications 999, and subsample sizes $b = \lfloor kn^{2/5} \rfloor$, $k = 3, 5$. $d = 3$.

Mean: AR(1) model, estimated by OLS. Variance: GARCH(1,1) model estimated by QMLE.
### Table 4: Empirical P-Values. S&P500 Fitted.

<table>
<thead>
<tr>
<th>α</th>
<th>Gaussian</th>
<th>Student-t</th>
<th>Gaussian</th>
<th>Student-t</th>
<th>Gaussian</th>
<th>Student-t</th>
<th>Gaussian</th>
<th>Student-t</th>
<th>Gaussian</th>
<th>Student-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>$CvM_n$</td>
<td>0.012</td>
<td>0.000</td>
<td>0.048</td>
<td>0.056</td>
<td>0.000</td>
<td>0.124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$CvM_{n,3}$</td>
<td>0.045</td>
<td>0.040</td>
<td>0.186</td>
<td>0.102</td>
<td>0.084</td>
<td>0.311</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$CvM_{n,5}$</td>
<td>0.049</td>
<td>0.035</td>
<td>0.195</td>
<td>0.036</td>
<td>0.045</td>
<td>0.338</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_n$</td>
<td>0.028</td>
<td>0.000</td>
<td>0.061</td>
<td>0.024</td>
<td>0.003</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$K_{n,3}$</td>
<td>0.004</td>
<td>0.000</td>
<td>0.091</td>
<td>0.157</td>
<td>0.000</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$C_n$</td>
<td>0.007</td>
<td>0.000</td>
<td>0.012</td>
<td>0.005</td>
<td>0.002</td>
<td>0.367</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$C_{n,3}$</td>
<td>0.043</td>
<td>0.045</td>
<td>0.005</td>
<td>0.026</td>
<td>0.012</td>
<td>0.188</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 0.05  | $CvM_n$  | 0.034     | 0.000    | 0.006     | 0.032    | 0.004     | 0.002    |           |          |           |
|       | $CvM_{n,3}$ | 0.179     | 0.100    | 0.075     | 0.188    | 0.113     | 0.074    |           |          |           |
|       | $CvM_{n,5}$ | 0.118     | 0.103    | 0.025     | 0.121    | 0.102     | 0.028    |           |          |           |
|       | $K_n$    | 0.420     | 0.649    | 0.074     | 0.403    | 1.000     | 0.082    |           |          |           |
|       | $K_{n,3}$ | 0.355     | 0.907    | 0.159     | 0.391    | 0.656     | 0.099    |           |          |           |
|       | $C_n$    | 0.150     | 0.667    | 0.475     | 0.159    | 0.694     | 0.487    |           |          |           |
|       | $C_{n,3}$ | 0.124     | 0.868    | 0.549     | 0.150    | 0.897     | 0.566    |           |          |           |

| 0.1   | $CvM_n$  | 0.000     | 0.008    | 0.002     | 0.000    | 0.058     | 0.000    |           |          |           |
|       | $CvM_{n,3}$ | 0.004     | 0.081    | 0.025     | 0.029    | 0.205     | 0.029    |           |          |           |
|       | $CvM_{n,5}$ | 0.008     | 0.109    | 0.061     | 0.027    | 0.231     | 0.062    |           |          |           |
|       | $K_n$    | 0.009     | 0.011    | 0.280     | 0.569    | 0.073     | 0.043    |           |          |           |
|       | $K_{n,3}$ | 0.061     | 0.000    | 0.109     | 0.469    | 0.141     | 0.093    |           |          |           |
|       | $C_n$    | 0.004     | 0.037    | 0.572     | 0.014    | 0.055     | 0.330    |           |          |           |
|       | $C_{n,3}$ | 0.001     | 0.093    | 0.523     | 0.041    | 0.089     | 0.395    |           |          |           |

| 0.2   | $CvM_n$  | 0.000     | 0.000    | 0.086     | 0.000    | 0.000     | 0.092    |           |          |           |
|       | $CvM_{n,3}$ | 0.009     | 0.000    | 0.286     | 0.050    | 0.082     | 0.306    |           |          |           |
|       | $CvM_{n,5}$ | 0.001     | 0.000    | 0.322     | 0.032    | 0.095     | 0.336    |           |          |           |
|       | $K_n$    | 0.037     | 0.000    | 0.789     | 0.140    | 0.001     | 0.812    |           |          |           |
|       | $K_{n,3}$ | 0.052     | 0.000    | 0.675     | 0.208    | 0.020     | 0.868    |           |          |           |
|       | $C_n$    | 0.169     | 0.019    | 0.466     | 0.339    | 0.069     | 0.696    |           |          |           |
|       | $C_{n,3}$ | 0.155     | 0.018    | 0.368     | 0.388    | 0.043     | 0.550    |           |          |           |

Bootstrap replications 999, and subsample sizes $b = \lfloor kn^{2/5} \rfloor$, $k = 3, 5$. $d = 3$.

Conditional Quantile: GARCH(1,1) Location model.