

A Specification Test of Dynamic Conditional Distribution and Quantile Models*

(Job Market Paper)

Victor Troster[†]

November 14, 2014

Abstract

An important part of the empirical economic and finance research is conducted under the assumption of the correct specification of dynamic conditional distribution models. This paper proposes a practical and consistent specification test of conditional distribution models for dependent data in a very general setting. Our approach covers conditional distribution models possibly indexed by function-valued parameters, which allows for a wide range of important empirical applications, such as the linear quantile auto-regressive, the CAViaR, and the distributional regression models. Our test statistic is based on a comparison between the estimated parametric and the empirical distribution functions. The new specification test (i) is valid for general linear and nonlinear dynamic models under parameter estimation error, (ii) is robust to dynamic misspecification, (iii) is consistent against fixed alternatives, and (iv) has nontrivial power against \sqrt{T} -local alternatives, with T the sample size. As the test statistic is non-pivotal, we propose and theoretically justify a block bootstrap approach to obtain valid inference. Monte Carlo simulations illustrate that the proposed test has good finite sample properties for different data generating processes and sample sizes. Finally, an empirical application to models of Value-at-Risk (VaR) highlights the benefits of our approach.

JEL Classification: C14, C22, C52

Keywords: Quantile regression; Empirical processes; Block bootstrap; Dynamic misspecification; Functional delta method.

*I thank Carlos Velasco, Abderrahim Taamouti, Jesús Gonzalo, Juan Carlos Escanciano, Miguel Delgado, Jesús Carro, Emre Ergemen, Pedro Sant'Anna, Anett Erdmann, Peter Eccles, and Nora Wegner for helpful comments and discussions.

[†]Correspondence: Department of Economics, Universidad Carlos III de Madrid, Getafe, Madrid 28903. Tel.: (+34) 916 249367. Email: vtroster@eco.uc3m.es. Web: <https://sites.google.com/site/victortroster>

1 Introduction

Many important economic and finance hypotheses are investigated through testing the specification of restrictions on the conditional distribution of a time series, such as conditional goodness-of-fit (Box and Pierce (1970)), conditional quantiles (Koenker and Machado (1999)), and distributional Granger non-causality (Taamouti, Bouezmarni, and El Ghouch, 2014). After the landmark work of Hausman (1978), numerous authors have developed specification tests under i.i.d. observations. White (1982) proposed a comparison of different variance matrix estimators to detect misspecification of econometric models. Newey (1985) constructed tests of conditional moment restrictions that generalized the approach of Hausman (1978) and White (1982). Although these tests can also be applied in a time series context, none of them is consistent against all possible sources of misspecification. Despite Andrews (1997) developed a consistent test statistic for testing conditional distribution specifications, his approach can be applied only for i.i.d. data.

This paper proposes a practical and consistent specification test of conditional distribution models for dependent data in a very general setting. Bai (2003) developed a Kolmogorov-Smirnov type test of conditional distribution specifications for time series based on the comparison of an estimated conditional distribution function with the distribution function of a uniform on $[0, 1]$. To overcome the parameter error estimation effect, Bai (2003) proposed a martingale transformation that delivers a nuisance free limiting distribution for the test statistic. However, Bai (2003)'s test is inconsistent as it cannot detect lag order misspecification of a linear autoregressive model with elliptically distributed innovations (see Corradi and Swanson (2006) and Delgado and Stute (2008)). Corradi and Swanson (2006) modified the approach of Bai (2003) allowing for dynamic misspecification of the past information set under the null hypothesis. They proposed a consistent test of correct specification for a given information set. There exists dynamic misspecification when the conditional distribution of the variable of interest Y_t given a past information set X_t is not equivalent to the conditional distribution of Y_t given all the "relevant" past information set \mathcal{F}_{t-1} of the conditioning variable, with $X_t \subset \mathcal{F}_{t-1}$, i.e. $Y_t|X_t$ is not equal in distribution as $Y_t|\mathcal{F}_{t-1}$. This is relevant when a dynamic specification

test is developed, as one generally has the problem of defining the relevant past information \mathcal{F}_{t-1} (e.g. how many lags to include), which may involve pre-testing and imply a sequential test bias. Besides, critical values derived under correct specification given \mathcal{F}_{t-1} are not in general valid in the case of correct specification given a subset of \mathcal{F}_{t-1} .

In this paper, we construct a specification test for time series models that takes into account dynamic misspecification and parameter error estimation effect. We build on the work of Chernozhukov, Fernández-Val, and Melly (2013) and Rothe and Wied (2013) that develop specification tests of conditional distribution models indexed by possibly function-valued parameters for i.i.d. data. We generalize their approach to testing the specification of dynamic conditional distribution models indexed by function-valued parameters in contexts with dependent data.

Allowing the parameters to be function-valued is important for many empirical applications. For example, our approach covers the linear quantile autoregressive (QAR) of Koenker and Xiao (2006), which implies a linear structure for the inverse of the dynamic conditional distribution $F^{-1}(\tau|\theta_0, \mathbf{Y}_{t-p}) = \mathbf{Y}'_{t-p}\theta_0(\tau)$, for the quantile $\tau \in (0, 1)$, with $\mathbf{Y}_{t-p} = \{Y_{t-1}, \dots, Y_{t-p}\} \in \mathcal{F}_{t-1}$, and a functional parameter $\theta_0(\tau)$ strictly monotone in τ . Our procedure also considers testing the specification of nonlinear quantile autoregressive models, such as the CAViaR model of Engle and Manganelli (2004), that directly measures the market risk of financial institutions by estimating a particular quantile of future portfolio values - the Value-at-Risk (VaR). Our proposed test statistic checks the validity of the distributional regression model introduced by Foresi and Peracchi (1995), where the conditional distribution is modeled through a family of binary response models for the event that the variable of interest Y_t exceeds some threshold $y \in \mathbb{R}$. To the best of our knowledge, it has not been developed yet a consistent specification test of conditional distribution models indexed by function-valued parameters under dependent data.

An additional benefit of our approach is that it permits us to test conditional quantile models over a continuum of quantiles under time series. Koenker and Machado (1999) considered tests for the specification of regression quantile location-scale models for independent observations. Koenker and Xiao (2002) applied the “Khmaladze” transformation

to test the specification of linear quantile models under i.i.d data. However, none of these tests are justified for dependent data, and they do not check for the validity of the quantile regression model itself. Whang (2006) proposed a specification test of conditional quantile models for a given quantile τ for time series data, while Escanciano and Velasco (2010) generalized this approach by providing consistent tests of dynamic quantile regression models over a continuum of quantiles under dependent data. Our new test provides a further advantage: it also checks the validity of models for the whole conditional distribution and distributional regression specifications, while the framework Escanciano and Velasco (2010) considers only conditional quantile regression models. Andrews (2012), Bierens and Wang (2014), and Kheifets (2014) have also developed consistent specification tests for conditional distribution models for dependent data, but these methods cannot be applied to evaluate models indexed by function-valued parameters. In sum, we believe that our approach is a useful alternative to existing specification methods for dynamic conditional models under dependent data because it allows for models indexed by possibly function-valued parameters, covering the setups of Corradi and Swanson (2006), Escanciano and Velasco (2010), and Rothe and Wied (2013) in a unified way.

Our test statistic is a Cramér-von-Mises (CVM) functional of the discrepancy between the empirical distribution function and a restricted estimate imposing the structure implied by the dynamic conditional distribution model, and we reject the null hypothesis of correct specification if this discrepancy is “large”. Since its asymptotic distribution under general time series assumptions is non-pivotal, we propose and justify a block bootstrap resampling scheme to estimate the critical values. This is likely to be computationally intensive, but it delivers a test statistic that (i) is robust to dynamic misspecification, (ii) does not require the estimation of smoothing parameters or nuisance functions used in a Khmaladze transformation as in Bai (2003) or in Koenker and Xiao (2002), and (iii) is consistent against all fixed alternatives. Besides, our test statistic has nontrivial power against \sqrt{T} -local alternatives, with T the sample size.

As further contributions, we investigate the finite sample performance of our method on simulated data and we illustrate the empirical applicability of our setting by verifying

the specification of conditional distribution models for Value-at-Risk (VaR), which is the most used measure of market risk in the financial industry. Using data on two major stock return indexes, we show that our test statistic rejects some widely used specifications of VaR models.

The plan of the paper is as follows. In Section 2, we propose a test statistic for the null hypothesis of correct specification of dynamic conditional distribution models indexed by function-valued parameters under time series and dynamic misspecification. In Section 3, we derive the asymptotic limit distribution of our test statistic under the null and the alternative hypotheses. We also prove that our test statistic has nontrivial power against \sqrt{T} -local alternatives, with T the sample size. In Section 4, we theoretically justify the validity of the block bootstrap in our framework. Section 5 provides some examples of conditional distribution and quantile models that are covered by our setting. Section 6 presents Monte Carlo simulation results. In Section 7, we present an empirical application of our proposed test. Finally, Section 8 concludes the paper.

2 A General Approach to Testing Dynamic Conditional Distribution and Quantile Models

Suppose we observe a sample $\{(Y_t, X_t) \in \mathbb{R} \times \mathbb{R}^d, t = 1, \dots, T\}$ from a stationary process $\{Y_t, X_t\}_{t=-\infty}^{\infty}$, with joint distribution F_{YX} , where X_t may contain lags of Y_t and/or of other variables. Let $\mathcal{F}_{t-1} := \{X_s\}_{s=-\infty}^t$ be the information set including all relevant past information. Let \mathcal{G} be a parametric family of conditional distribution models on the support of Y given X satisfying

$$\mathcal{G} = \{F(\cdot|\theta, \cdot) \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta)\}, \quad (1)$$

where $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ is a function-valued parameter such that $\theta(\tau) \in \Theta \subset \mathbb{R}^K$, for each $\tau \in \mathcal{T} \subset \mathbb{R}$. The null hypothesis of correct specification could be written as

$$H_0 : F(y|\mathcal{F}_{t-1}) = F(y|\theta_0, \mathcal{F}_{t-1}), \text{ a.s. for all } y \in \mathbb{R} \text{ and for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta), \quad (2)$$

against

$$H_A : \Pr [F(y|\mathcal{F}_{t-1}) \neq F(y|\theta, \mathcal{F}_{t-1})] > 0, \text{ for some } y \in \mathbb{R} \text{ and for all } \theta \in \mathcal{B}(\mathcal{T}, \Theta), \quad (3)$$

focusing on the whole information set \mathcal{F}_{t-1} . Instead, in this paper we are interested in the distribution of Y_t given a finite dimensional vector of conditioning variables $X_t \in \mathbb{R}^d$, for $X_t \subset \mathcal{F}_{t-1}$. If $Y_t|\mathcal{F}_{t-1}$ is not equal in distribution to $Y_t|X_t$, then X_t is dynamically misspecified. However, in empirical applications we do not know a priori what is the “relevant” past information set \mathcal{F}_{t-1} , and finding out how much information to include may involve pre-testing (Corradi and Swanson, 2006). Moreover, the critical values for specification tests obtained under H_0 of (2) given \mathcal{F}_{t-1} are not in general valid in the case of correct specification given X_t , for $X_t \subset \mathcal{F}_{t-1}$. Thus, we allow for dynamic misspecification of X_t and even in the presence of it, we obtain an asymptotically consistent test statistic for the correct specification of Y_t given X_t . Thus we want to test null hypotheses of correct specification of conditional distribution models of the form

$$\mathcal{H}_0 : F(y|x) = F(y|\theta_0, x), \text{ for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \text{ and for all } (y, x) \in \mathcal{W}, \quad (4)$$

versus

$$\mathcal{H}_A : F(y|x) \neq F(y|\theta, x), \text{ for some } (y, x) \in \mathcal{W} \text{ and for all } \theta \in \mathcal{B}(\mathcal{T}, \Theta), \quad (5)$$

where \mathcal{W} is the support of $W_t := (Y_t, X_t)'$.

We assume that the functional parameter $\theta_0(\cdot)$ solves a sequence of moment equalities. Let $\psi : \mathcal{W} \times \Theta \times \mathcal{T} \mapsto \mathbb{R}^K$ be a uniformly integrable function. For every $\tau \in \mathcal{T}$, we assume that the function-valued parameter $\theta_0(\cdot)$ solves

$$\Psi(\theta_0, \tau) := \mathbb{E} [\psi(W_t, \theta_0, \tau)] = 0, \quad (6)$$

where $\Psi(\theta, \tau)$ is a function $\Psi : \Theta \times \mathcal{T} \mapsto \mathbb{R}^K$ that satisfies some regularity conditions described in Section 3. In this paper, we assume that under \mathcal{H}_A in equation (5), there exists a “pseudo”-true functional parameter $\theta_1(\cdot)$ solving the moment conditions (6). Chernozhukov et al. (2013) developed theoretical results for Z -estimators of the moment conditions of (6) for i.i.d. data. We provide conditions for the estimation of function-valued parameters in a context of dependent observations in Section 3.

To test \mathcal{H}_0 defined in equation (4), we first restate our null hypothesis into an equality of unconditional distributions by integrating-up both sides of \mathcal{H}_0 with respect to the marginal distribution of the conditioning variable F_X ; see Theorem 16.10 (iii) in Billingsley (1995) and Andrews (1997). As $F(y|x) = E(\mathbb{1}\{Y_t \leq y\}|X_t = x)$, where $\mathbb{1}\{A\}$ is the indicator function of the event A , the null hypothesis \mathcal{H}_0 of (4) can be equivalently restated as

$$\int F(y|\bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x}) = \int F(y|\theta_0, \bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x}),$$

for some $\theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$ and for all $(y, x) \in \mathcal{W}$,

where $F_{YX}(y, x) := \int F(y|\bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ is the unconditional joint distribution function, and $F(y, x, \theta_0) := \int F(y|\theta_0, \bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ is the unconditional distribution function implied by the parametric conditional distribution model. Let $\hat{Z}_T(y, x)$ and $\hat{F}_T(y, x, \hat{\theta}_T)$ be the joint empirical distribution function and the semi-parametric estimated distribution function of $\{Y_t, X_t\}_{t=1}^T$ respectively,

$$\hat{Z}_T(y, x) = \frac{1}{T} \sum_{t=1}^T \mathbb{1}\{Y_t \leq y\} \mathbb{1}\{X_t \leq x\}, \text{ for } (y, x) \in \mathbb{R}^{1+d}, \quad (7)$$

and

$$\hat{F}_T(y, x, \hat{\theta}_T) = \int F(y|\hat{\theta}_T, \bar{x}) \mathbb{1}\{\bar{x} \leq x\} d\hat{F}_X(\bar{x}), \text{ for } (y, x) \in \mathbb{R}^{1+d}, \quad (8)$$

where $\hat{F}_X(x)$ is the empirical distribution function of $\{X_t\}_{t=1}^T$,

$$\hat{F}_X(x) = \frac{1}{T} \sum_{t=1}^T \mathbb{1}\{X_t \leq x\}, \text{ for } x \in \mathbb{R}^d. \quad (9)$$

Under \mathcal{H}_0 of (4), we assume there is a \sqrt{T} -consistent estimator $\hat{\theta}_T(\tau)$ of $\theta_0(\tau)$, for each $\tau \in \mathcal{T}$, that minimizes the empirical analog $\hat{\Psi}_T(\hat{\theta}_T, \tau)$ of the moment conditions in (6):

$$\left\| \hat{\Psi}_T(\hat{\theta}_T, \tau) \right\|^2 \leq \inf_{\theta \in \Theta} \left\| \hat{\Psi}_T(\theta, \tau) \right\|^2 + \hat{u}(\tau)^2, \quad (10)$$

where $\|\hat{u}\|_{\mathcal{T}} = o_P(T^{-1/2})$, and $\|\cdot\|$ denotes the supremum norm. Our proposed test statistic of \mathcal{H}_0 is the functional norm of the distance between $\hat{Z}_T(y, x)$ and $\hat{F}_T(y, x, \hat{\theta}_T)$, similar to the approach of Andrews (1997) and Rothe and Wied (2013). To this purpose we consider

$$D_T(y, x) = \frac{1}{T} \sum_{t=1}^T \left(\mathbb{1}\{Y_t \leq y\} - F(y|\hat{\theta}_T, X_t) \right) \mathbb{1}\{X_t \leq x\}, \quad (11)$$

and to test the null hypothesis \mathcal{H}_0 we propose a T -scaled Cramér-von Mises functional norm of $D_T(y, x)$:

$$S_T = T \int_{\mathcal{W}} (D_T(y, x))^2 d\hat{Z}_T(y, x). \quad (12)$$

The test statistic S_T should be small if the null hypothesis is correct, while “large” values of S_T imply the rejection of \mathcal{H}_0 in (4). In Section 3, we develop an asymptotic

theory that covers the case of serial dependence, extending the analysis of Rothe and Wied (2013) for S_T to the specification of time series models and the approach of Corradi and Swanson (2006) to specification testing under dynamic misspecification for function-valued parameter models. It is possible to apply other functional norms to $D_T(y, x)$, such as the Kolmogorov-Smirnov functional norm: $\sqrt{T} \sup_{(y,x) \in \mathcal{W}} |D_T(y, x)|$. However, unreported simulations suggested that the S_T test statistic outperforms in terms of size and power other alternative functionals such as the Kolmogorov-Smirnov. Therefore, we focus in this paper on S_T .

3 Asymptotic Theory

In this section, we derive the asymptotic distributions of our test statistic S_T under the null and alternative hypothesis. Let $\{Y_{Tt} : t \leq T, T = 1, 2, \dots\}$ be a triangular array with stationary rows of random variables defined on a complete probability space (Ω, \mathcal{A}, P) . Let \mathcal{M} be a permissible class of functions, and denote by $Pf = \int f(y, x)dP(y, x)$, for $f \in \mathcal{M}$. We consider $\mathcal{A}_T(m)$ as the σ -field generated by Y_{Tt} for $t \leq m$, and $\mathcal{B}_T(m+d)$ to be the σ -field generated by the variables Y_{Tt} for $t \geq m+d$. The sequence $\{Y_{Tt}\}$ is α -mixing if there is a sequence of numbers $\{\alpha(d)\}$ converging to zero for which

$$|\Pr(AB) - \Pr(A)\Pr(B)| \leq \alpha(d), \text{ for all } A \in \mathcal{A}_T(m), \text{ all } B \in \mathcal{B}_T(m+d), \text{ all } m, d, T.$$

Our test statistic S_T in (12) is based on an empirical process indexed by a class of functions $\ell^\infty(\mathcal{H})$, which is the class of real-valued functions that are uniformly bounded on \mathcal{H} , with $\mathcal{H} := \mathcal{W} \times \mathcal{T}$, equipped with the supremum norm $\|\cdot\|_{\ell^\infty(\mathcal{H})}$. To simplify notation, we use $\|\cdot\|$ to denote the supremum norm. The class $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ has a finite and integrable envelope function $\mathbb{F}(x) = \sup_{f \in \mathcal{M}} |f(x)|$ and can be covered by a finite number of elements, not necessarily in \mathcal{M} . Given $\varepsilon > 0$, we define the covering number $N(\varepsilon, \mathcal{M}, \|\cdot\|)$ as the minimal number of $L_2(P)$ -balls of radius ε needed to cover \mathcal{M} , where a $L_2(P)$ -ball of radius ε around a function $g \in L_2(P)$ is the set $\{h \in L_2(P) : \|h - g\| < \varepsilon\}$.

We define the uniform covering numbers as $\sup_P N(\varepsilon\|\mathbb{F}\|, \mathcal{M}, L_2(P))$, with \mathbb{F} the square-integrable envelope of \mathcal{M} . Finally, the \mathcal{M} is assumed in this paper to form a so-called Vapnik-Chervonenkis (VC) class of functions (see Dudley, 1978, Pollard, 1984). The VC class is an extension of the class of indicator functions and has the interesting property that for $1 \leq p < \infty$, there are constants C_1 and C_2 satisfying

$$N(\varepsilon, \mathcal{M}, \|\cdot\|) \leq C_1 \left(\frac{(P(\mathbb{F})^p)^{1/p}}{\varepsilon} \right)^{C_2},$$

for all $\varepsilon > 0$ and all probability measures P (see Lemmas II.25 and II.32 in Pollard, 1984). Throughout the paper we use “ \xrightarrow{d} ” and “ \implies ” to denote convergence in distribution of random variables and weak convergence of stochastic processes, respectively. We write $Z_T \implies Z$ in $\ell^\infty(\mathcal{H})$ to denote weak convergence of a stochastic process Z_T to a random element Z in the function space $\ell^\infty(\mathcal{H})$ (in the Hoffmann-Jørgensen sense, cf. Alexander (1987)) for the metric induced by $\|\cdot\|$. Let $B_\varepsilon(\theta)$ be a closed ball of radius ε centered at θ . All limits are taken as $T \rightarrow \infty$, where T is the sample size. We maintain the following main assumptions to analyse the asymptotic behavior of our test statistic:

Assumption 1. $\{(Y_{Tt}, X_{Tt}) : t \leq T, T = 1, 2, \dots\}$ is an α -mixing triangular array with stationary rows, satisfying $E(|Y_1|^{2+\gamma}) < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\gamma/(4+\gamma)} < \infty$ for some $\gamma \in (0, 2)$.

Assumption 2. The parametric space Θ is compact in \mathbb{R}^K and \mathcal{T} is a compact set of some metric space.

Assumption 3. For each $\tau \in \mathcal{T}$, $\Psi(\theta, \tau) : \Theta \mapsto \mathbb{R}^K$ possess a unique zero at $\theta_0(\tau)$, and for some $\varepsilon > 0$, $\bigcup_{\tau \in \mathcal{T}} B_\varepsilon(\theta_0(\tau))$ is a compact subset of \mathbb{R}^K contained in Θ . Moreover, the class of functions $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ is a permissible and VC class of measurable functions with a square integrable envelope function \mathbb{F} satisfying $P(\mathbb{F})^p < \infty$, for $2 < p < \infty$.

Assumption 4. *The mapping $\Psi(\theta, \tau) : \Theta \times \mathcal{I} \mapsto \mathbb{R}^K$ is continuous, where \mathcal{I} is an open set containing \mathcal{T} . Besides, $\frac{\partial}{\partial \theta} \Psi(\theta, \tau) := \dot{\Psi}_{\theta, \tau}$ exists at $(\theta_0(\tau), \tau)$ and is continuous at $(\theta_0(\tau), \tau)$, for each $\tau \in \mathcal{T}$, with $\inf_{\tau \in \mathcal{T}} \inf_{\|h\|=1} \|\dot{\Psi}_{\theta_0, \tau} h\| > 0$.*

Assumption 5. *For each $\tau \in \mathcal{T}$, the map $\theta \mapsto F(\cdot | \theta, \cdot)$ is Hadamard differentiable at all $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ with derivative $h \mapsto \dot{F}(\cdot | \theta, \cdot)[h]$.*

Assumption 1 is needed to restrict the dependence of $\{Y_{Tt}, X_{Tt}\}$ and holds for many relevant econometric models in practice, including ARMA and GARCH processes under mild additional assumptions; see e.g. Carrasco and Chen (2002). It enables us to establish weak convergence of the empirical process $Z_T(y, x)$ under a variety of situations, see Theorem 7.2 in Rio (2000). Assumptions 2-4 provide conditions to guarantee that a functional central limit theorem holds to the Z -estimator process $\tau \mapsto \sqrt{T}(\hat{\theta}_T(\tau) - \theta_0(\tau))$ for strong mixing processes. Assumption 5 is a smoothness condition required to establish a functional delta-method for the bootstrap of our test statistic (see Theorem 3.9.11 in Van der Vaart and Wellner, 2000).

In comparison with the framework of Rothe and Wied (2013), we need to impose Assumption 1 to establish the asymptotic theory of our test statistic under dependence, while this assumption is not needed in contexts with independent data. In addition, Assumption 4 requires that the class of functions $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ is a permissible and VC class of measurable functions, while Rothe and Wied (2013) work with Donsker class of functions in an i.i.d. setting. Assumptions 1-5 imply the following theorem, which describes the limit distribution of the proposed test statistic S_T under the null and the alternative.

Theorem 1. *Under Assumptions 1-5, the following hold:*

(i) *Under the null hypothesis \mathcal{H}_0 in (4),*

$$S_T \xrightarrow{d} \int (\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x))^2 dF_{YX}(y, x),$$

where $(\mathbb{H}_1, \mathbb{H}_2)$ are tight mean zero Gaussian processes.

(ii) Under the alternative hypothesis \mathcal{H}_A in (5), there exists an $\varepsilon > 0$ such that

$$\lim_{T \rightarrow \infty} \Pr(S_T > \varepsilon) = 1.$$

Theorem 1 shows that the asymptotic null distribution of S_T is a functional of the zero-mean Gaussian processes $(\mathbb{H}_1, \mathbb{H}_2)$. By Theorem 1, we expect that S_T is significantly positive whenever the null hypothesis \mathcal{H}_0 is violated. However, the asymptotic distribution of S_T varies with the conditional distribution model, the parameter $\theta_0(\cdot)$, and with the serial dependence in the data. As a result, S_T is not asymptotically pivotal and we cannot tabulate critical values. Since $\hat{Z}_T(y, x)$ is an integrating measure on \mathcal{W} depending on T and on data, $\hat{Z}_T(y, x) \implies F_{YX}(y, x)$ in $\ell^\infty(\mathcal{W})$, as T goes to infinity (see Lemma A.1 in the Appendix). In Section 4, we justify a block bootstrap approach that provides critical values for S_T and does not require the estimation of nuisance functions.

3.1 Local Power of the Test Statistic

Now we analyze the asymptotic power of S_T against a sequence of Pitman's local alternatives converging to the null hypothesis at rate \sqrt{T} , where T denotes the sample size. Let $J(\cdot|\cdot)$ be an alternative conditional distribution function such that $J(\cdot|\cdot) \notin \mathcal{G}$ of (1). For any $0 < \delta \leq \sqrt{T}$, we consider that under a sequence of local alternatives $\mathcal{H}_{A,T}$ the data are distributed accordingly to the following conditional distribution

$$\mathcal{H}_{A,T} : F_T(y|x) = \left(1 - \frac{\delta}{\sqrt{T}}\right) F(y|\theta_0, x) + \left(\frac{\delta}{\sqrt{T}}\right) J(y|x), \quad (13)$$

for all $(y, x) \in \mathcal{W}$ and for some $\theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$. To ensure nontrivial local power of our test statistic, we make the following assumption:

Assumption 6. *Under the local alternative in (13), the conditional distribution under the local alternative in (13) implies a sequence of distribution functions $F_T^A(y, x) =$*

$\int F_T(y|\bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ that is contiguous to the distribution function $F(y, x, \theta_0) = \int F(y|\theta_0, \bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ based on $F(y|\theta_0, X_t)$.

Assumption 6 is standard in the study of the asymptotic power under a sequence of Pitman's local alternatives. Andrews (1997) shows that when $F(\cdot|\theta_0, \cdot)$ and $J(\cdot|\cdot)$ have density functions $f(\cdot|\theta_0, \cdot)$ and $j(\cdot|\cdot)$ with respect to the same σ -finite measure, then a sufficient condition for Assumption 6 is

$$\sup_{(y,x):f(y|\theta_0,x)>0} \frac{j(y|x)}{f(y|\theta_0, x)} < \infty.$$

Let $\Psi_J(\theta, \tau) := E_J[\psi(W_t, \theta, \tau)]$ and $\Psi_F(\theta, \tau) := E_F[\psi(W_t, \theta, \tau)]$, where $E_J[\cdot]$ and $E_F[\cdot]$ denote expectation w.r.t. $J = J(y|X_t)$ and $F = F(y|\theta_0, X_t)$, respectively in (13). We consider $\theta_0(\cdot)$ and $\theta_1(\cdot)$ as solutions to

$$\Psi_F(\theta_0, \tau) = 0, \tag{14}$$

and

$$\Psi_J(\theta_1, \tau) = 0, \tag{15}$$

for all $\tau \in \mathcal{T}$ respectively. Let $\frac{\partial}{\partial \theta} \Psi_F(\theta_0, \tau)$ satisfy Assumption 4 for the functional parameter θ_0 solving the moment conditions in (14). The following theorem sheds light on the asymptotic power of the test statistic S_T under a sequence of local alternatives satisfying (13).

Theorem 2. *Under the local alternative $\mathcal{H}_{A,T}$ in (13) and Assumptions 1-6*

$$S_T \xrightarrow{d} \int (\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x) + \Delta(y, x))^2 dF_{YX}(y, x),$$

with $\Delta(y, x) = \delta \int (J(y|\bar{x}) - F(y|\theta_0, \bar{x}) + \dot{F}(y|\theta_0, \bar{x})[h])\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$, and h is the function $h(\tau) = [\frac{\partial}{\partial \theta} \Psi_F(\theta_0, \tau)]^{-1} \Psi_J(\theta_0, \tau)$.

Theorem 2 implies that the test statistic S_T has non-trivial local power when $\Delta(y, x) \neq 0$. Note that the choice of θ_0 affects the asymptotic power, since $\Delta(y, x)$ is a function of θ_0 . This follows because we cannot choose θ_0 under the local alternatives, and θ_1 corresponds to the value that makes $J(\cdot|\cdot)$ as “close” as possible to $F(\cdot|\theta_0, \cdot)$ in the sense of the Kullback-Leibler information distance (Andrews, 1997). For a functional parameter θ_1 solving (15), we may choose $F(\cdot|\theta_1, \cdot)$ as the probability limit under J to which the sequence of local alternatives $F_T(\cdot|\cdot)$ shrinks as the sample size grows. Then $[\frac{\partial}{\partial \theta} \Psi_F(\theta_0, \tau)]^{-1} \Psi_J(\theta_0, \tau) = 0$, and we have a simpler drift term

$$\Delta(y, x) = \delta \int (J(y|\bar{x}) - F(y|\theta_0, \bar{x})) \mathbb{1}(\bar{x} \leq x) dF_X(\bar{x}).$$

4 Bootstrap Tests

As the test statistic S_T has an asymptotic distribution under \mathcal{H}_0 that depends on the data-generating process, we propose a block bootstrap approach to obtain critical values. We also derive its asymptotic properties under the null and alternative hypothesis. We could consider a subsampling approach, for which similar asymptotic results can be shown to hold as well, see e.g. Chernozhukov and Fernández-Val (2005). However, we choose a block bootstrap because we expect it to have more power asymptotically and in finite samples. The block bootstrap is a resampling method with replacement extended to time series observations. It consists of splitting the data into consecutive blocks of observations with length ℓ - $(X_t, X_{t+1}, \dots, X_{t+\ell-1})$ - and resampling the blocks with replacement from all blocks and joining them to create a bootstrap sample; for a review of block bootstrap and other resampling methods for dependent data, see Kreiss and Paparoditis (2011). Although the block bootstrap is computationally demanding, the estimated asymptotic critical values are robust to dynamic misspecification.

Block bootstrap approaches differ on whether the blocks are overlapping or non-overlapping and on whether the length of the blocks is deterministic or random. We apply a block bootstrap with an overlapping block length - since it is more efficient than

the non-overlapping one - and with non-random block length, which has a smaller first order variance (Lahiri, 1999). In what follows, P^*, E^*, F^*, \dots denote probability laws, expectations, distribution functions, etc. in the block bootstrap, i.e., conditionally on the observed data. The algorithm for computing a fixed block bootstrap realization of our test statistic S_T has the following steps.

1. Let $\ell \in \mathbb{N}$, $\ell \ll T$, $b = \lceil T/\ell \rceil$ and $k = T - b\ell$. Let I_1, I_2, \dots, I_{b+1} be discrete independent random variables taking values in the set $\{1, 2, \dots, T - \ell + 1\}$, that describe the set of starting indexes of the selected blocks;
2. From the sample $W_t = (Y_t, X_t)$, lay the blocks $(W_{I_s}, W_{I_s+1}, \dots, W_{I_s+\ell-1})$, $s = 1, 2, \dots, b+1$, end to end in the ordered sample and drop the last $\ell - k$ observations to obtain a resampled series $W_t^* = (Y_t^*, X_t^*)$, with $W_1^*, W_2^*, \dots, W_T^*$, that can also be written as

$$W_{I_1}, W_{I_1+1}, \dots, W_{I_1+\ell-1}, W_{I_2}, W_{I_2+1}, \dots, W_{I_2+\ell-1}, \dots, W_{I_b}, W_{I_b+1}, \dots, W_{I_b+\ell-1}.$$

3. Given the block bootstrap data $\{W_t^* = (Y_t^*, X_t^*)\}$, we compute the bootstrap equivalents $\hat{Z}_T^*(y, x)$ and $\hat{F}_T^*(y, x, \hat{\theta}_T^*)$ of $\hat{Z}_T(y, x)$ and $\hat{F}_T(y, x, \hat{\theta}_T)$, respectively. Then we obtain the following bootstrap version of the empirical process $\sqrt{T}(\hat{Z}_T(y, x) - F_{YX}(y, x))$ defined in $\ell^\infty(\mathcal{W})$ as

$$\sqrt{T}(\hat{Z}_T^*(y, x) - \hat{Z}_T(y, x)) := \frac{1}{\sqrt{T}} \sum_{t=1}^T [\mathbb{1}(Y_t^* \leq y) \mathbb{1}(X_t^* \leq x) - \mathbb{1}(Y_t \leq y) \mathbb{1}(X_t \leq x)],$$

and the re-centered bootstrap statistic S_T^* :

$$S_T^* = \sum_{t=1}^T \left[\left(\hat{Z}_T^*(Y_t, X_t) - \hat{F}_T^*(Y_t, X_t, \hat{\theta}_T^*) \right) - \left(\hat{Z}_T(Y_t, X_t) - \hat{F}_T(Y_t, X_t, \hat{\theta}_T) \right) \right]^2.$$

Given a significance level $\alpha \in (0, 1)$, our test rejects \mathcal{H}_0 if $S_T > c_T^*(\alpha)$, where the bootstrap critical value $c_T^*(\alpha)$ is the lowest value that satisfies $\Pr^* [S_T^* \leq c_T^*(\alpha)] \geq 1 -$

α , and this is estimated through Monte Carlo simulations. To justify theoretically the block bootstrap resampling in our setting, we need an additional assumption on the serial dependence on the data. We define the k -th beta mixing coefficient $\beta(k)$ by

$$\beta(k) = \frac{1}{2} \sup_{(i,j) \in I \times J} |\Pr(A_i \cap B_j) - \Pr(A_i) \Pr(B_j)|,$$

where the supremum is taken over all finite measurable partitions $\{A_i\}_{i \in I}$ and $\{B_j\}_{j \in J}$ with $A_i \in \sigma(Y_m : m \leq 1)$ and $B_j \in \sigma(Y_m : m \geq 1 + k)$. We say that a sequence $\{Y_t\}$ is beta mixing if $\lim_{k \rightarrow \infty} \beta_k \rightarrow 0$.

Assumption 7. $\{Y_{Tt}, X_{Tt}, t \leq T, T \geq 1\}$ is a β -mixing triangular array with stationary rows and β -mixing coefficients satisfying

$$\Gamma(\{\beta_k\}_{k \geq T}) \rightarrow 0, \text{ as } T \rightarrow \infty,$$

where $\Gamma : \mathbb{R}^\infty \mapsto \mathbb{R}$ is a monotone mapping such that $a_i \leq b_i$ for $i \geq 0$ implies $\Gamma(\{a_i\}_{i \geq 0}) \leq \Gamma(\{b_i\}_{i \geq 0})$.

Assumption 7 generalizes most of the commonly used mixing conditions in time series processes. Let $P^*(\cdot)$ be the probability law in the block bootstrap, i.e., conditionally on the observed data. We follow the approach of Radulović (1996), which delivers a Block Bootstrap Central Limit Theorem for the class of M-estimators (see Theorem 2 in Radulović (1996)), and justify the block bootstrap approach for our proposed test statistic in the following theorem.

Theorem 3. Under Assumptions 2-7, let W_1^*, \dots, W_T^* be generated according to the block bootstrap with block size $\ell := \ell(T)$, with $\ell(T) \rightarrow \infty$ as $T \rightarrow \infty$, conditional on the data W_1, \dots, W_T . Let $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ be a permissible VC class of measurable functions with a square integrable envelope function \mathbb{F} . If we also assume:

(i) $\limsup_{k \rightarrow \infty} k^q \beta(k) < \infty$, for some $q > p/(p-2)$, for $2 < p < \infty$ such that $P^*(\mathbb{F})^p < \infty$, and

(ii) $\ell(T) = O(T^\rho)$ for some $0 < \rho < (p-2)/[2(p-1)]$,

then:

(i) Under the null hypothesis \mathcal{H}_0 of (4),

$$\Pr(S_T > c_T^*(\alpha)) \rightarrow \alpha.$$

(ii) Under the fixed alternative hypothesis \mathcal{H}_A of (5),

$$\Pr(S_T > c_T^*(\alpha)) \rightarrow 1.$$

(iii) Under the local alternative $\mathcal{H}_{A,T}$ of (13),

$$\lim_{T \rightarrow \infty} \Pr(S_T > c_T^*(\alpha)) \geq \alpha.$$

Theorem 3 is an application of the functional delta method for bootstrap. It shows that our test based on the block bootstrap critical value has asymptotically correct size, is consistent, and is able to detect alternatives tending to the null at the parametric rate \sqrt{T} . Bradley (1985) showed that $P^*(\mathbb{F})^p < \infty$ and $\sum_{k=1}^{\infty} \beta(k)^{p/(2-p)}$ for some $p > 2$ is close to the weakest sufficient conditions for an original (non-bootstrap) central limit theorem for empirical processes for VC-subgraph classes of functions. As the optimal length is $\ell = CT^{1/3}$, for a constant $C > 0$ (see Künsch, 1989, Remark 3.3), the condition on the block length is not too restrictive.

5 Examples

In this section, we consider certain conditional distribution and quantile models that are covered by our setting. We choose those models since they can be used in many

relevant empirical applications.

5.1 Linear Quantile Autoregressive Processes

Many papers in the literature deal with the linear quantile autoregression model, see for example Weiss (1991), Koul and Mukherjee (1994), and Hallin and Jurečková (1999). In the linear quantile autoregression model, the τ -quantile of $Y_t|X_t$ is a linear function of X_t , where X_t can take the lagged values of Y_t as arguments. Koenker and Xiao (2006) investigated quantile autoregressive models in which all of the autoregressive coefficients are τ -dependent and able to change the location, scale, and shape of the conditional densities, provided that the τ -conditional quantile of Y_t is monotone in τ . For example, the quantile autoregression (QAR) of order p of Koenker and Xiao (2006) can be written as

$$\begin{aligned} Q_\tau(Y_t|Y_{t-1}, \dots, Y_{t-p}) &= \theta_0(\tau) + \theta_1(\tau)Y_{t-1} + \dots + \theta_p(\tau)Y_{t-p} \\ &= X_t'\theta(\tau), \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta), \end{aligned} \quad (16)$$

where $F^{-1}(\tau|Y_{t-1}, \dots, Y_{t-p}, \theta(\tau)) = Q_\tau(Y_t|Y_{t-1}, \dots, Y_{t-p})$, and $X_t = (1, Y_{t-1}, \dots, Y_{t-p})'$. If the τ -conditional quantile of Y_t is correctly specified by a QAR model, then there exists a $F(y|\theta, x) \subset \mathcal{G}$ such that the null hypothesis of (4) is not rejected, with \mathcal{G} satisfying

$$\mathcal{G} = \{F(\cdot|\theta, \cdot) \mid F^{-1}(\cdot|\theta, X_t) = X_t'\theta \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta)\}.$$

We consider estimators of the QAR model in (16) as any solution $\hat{\theta}_T(\tau)$ of the problem

$$\arg \min_{\theta \in \Theta} \sum_{t=1}^T \psi(W_t, \theta, \tau),$$

where $\psi(W_t, \theta, \tau) := (\tau - \mathbb{1}\{Y_t - X_t'\theta \leq 0\})$ is the check function. Given the solutions

$\hat{\theta}_T(\tau)$, the τ -quantile of $Y_t|X_t$ can be estimated by $\hat{Q}_\tau(Y_t|X_t) = X_t'\hat{\theta}_T(\tau)$. In our setup, $\hat{\theta}_T$ belong to the class of Z-estimators with $\psi(W_t, \hat{\theta}_T, \tau) = (\tau - \mathbb{1}\{Y_t - X_t'\hat{\theta}_T(\tau) \leq 0\})$. If the conditional distribution of Y_t is monotone in τ , the QAR model in (16) implies a conditional distribution function that can be estimated by $F(y|\hat{\theta}_T(\cdot), x) = \int_{\mathcal{T}} \mathbb{1}\{x'\hat{\theta}_T(\tau) \leq y\}d\tau$. Now we establish the conditions that allows us to apply our test statistic S_T to check the specification of a QAR model.

Proposition 1. *Let:*

(i) *The parameter $\theta_0(\tau)$ is identifiably unique with respect to*

$$\mathbb{E}(\tau - \mathbb{1}\{Y_t - X_t'\theta_0(\tau) \leq 0\}) = 0,$$

and $\theta_0(\tau)$ is interior to Θ for every $\tau \in \mathcal{T}$;

(ii) *$\{(Y_{Tt}, X_{Tt}) : t \leq T, T = 1, 2, \dots\}$ is an α -mixing triangular array with stationary rows satisfying $E(|Y_1|^{2+\gamma}) < \infty$ and $\sum_{j=1}^{\infty} j^2\alpha(j)^{\gamma/(4+\gamma)} < \infty$ for some $\gamma \in (0, 2)$;*

(iii) *The conditional distribution function of Y_t given X_t , $F(\cdot|\cdot)$, and its density function $f(\cdot|\cdot)$ have continuous derivatives up to the 2nd-order denoted respectively by $F^{(s)}(\cdot|\cdot)$ and $f^{(s)}(\cdot|\cdot)$, $s = 1, 2$;*

(iv) *$f(\cdot|\cdot)$ is Lipschitz continuous and bounded away from zero on $X_t'\theta_0(\tau)$ a.s., uniformly over $\tau \in \mathcal{T}$, and $F^{(2)}(\cdot|\cdot)$ and $f^{(2)}(\cdot|\cdot)$ are bounded and uniformly continuous on \mathbb{R} a.s.;*

(v) *The matrix $\mathbb{E}(X_t X_t')$ is finite and has full rank.*

Then Assumptions 1-5 hold for the linear quantile autoregression model.

Proposition 1 provide conditions for identifiability of the moment conditions in (6) and the validity of a functional central limit for a dependent stochastic process $\sqrt{T}(\hat{\theta}_T(\tau) - \theta(\tau))$ (Andrews and Pollard, 1994, Chernozhukov et al., 2013). The Lipschitz condition in (iv) provides a sufficient condition for the class of functions $\{\psi(W_t, \theta, \tau) = (\tau - \mathbb{1}\{Y_t - X_t'\theta(\tau) \leq 0\}) : \theta \in \Theta, \tau \in \mathcal{T}\}$ to be a VC class.

5.2 Nonlinear Quantile Autoregressive Models

We can apply our test to check the correct specification of a nonlinear quantile regression model such as the Conditional Autoregressive Value at Risk (CAViaR) model proposed by Engle and Manganelli (2004). Conditional Value at Risk (VaR) is the standard measure of market risk used by financial institutions and market regulators. Let Y_t be a return on a portfolio series. Given a significance level τ , the VaR of a portfolio is the level of return Y_t^T over the period $[t, T]$ that is exceeded with probability τ : $VaR_t^T(\tau|x) := \inf_L \{L : \Pr(Y_t^T \leq L|x) \geq 1 - \tau\}$. Analogously, we can also write the VaR as $VaR_t^T(\tau|x) = Q_\tau(Y_t|x)$. Since the VaR is a quantile of the conditional distribution of returns, the quantile regression model is a powerful tool to model VaR, using only information pertaining to the quantiles of the distribution.

Rather than imposing a linear quantile regression model, we may assume a nonlinear functional dependence on the quantiles of $Y_t|X_t$:

$$Q_\tau(Y_t|X_t = x) = m(x, \theta(\tau)), \quad (17)$$

where $m : \mathbb{R}^d \times \Theta \times \mathcal{T} \mapsto \mathbb{R}$ is a known function. Under our setup, we have

$$\mathcal{G} = \{F(\cdot|\theta, \cdot) | F^{-1}(\cdot|\theta(\cdot), x) = m(x, \theta(\cdot)) \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta)\}.$$

Similarly to the linear QAR process, we can estimate the parameters $\hat{\theta}_T(\cdot)$ of a nonlinear quantile regression model in (17) by solving

$$\arg \min_{\theta \in \Theta} \sum_{t=1}^T \rho_\tau(Y_t - m(X_t, \theta(\tau))), \quad (18)$$

with $\rho_\tau(u) = u(\tau - \mathbb{1}\{u \leq 0\})$. For sufficient conditions on $m(\cdot, \cdot)$ for the existence of a solution of (18), see Koenker and Park (1996). Given the solutions $\hat{\theta}_T(\cdot)$, the conditional

distribution function can be obtained as $F(y|\hat{\theta}_T(\cdot), x) = \int_{\mathcal{T}} \mathbb{1}\{m(x, \hat{\theta}_T(\tau)) \leq y\} d\tau$, assuming that $F(y|\hat{\theta}_T(\cdot), x)$ is monotone in y . Nonlinear dynamic models allow the inclusion of past values of the quantiles of $Y_t|X_t$. A general CAViaR specification for the quantile regression can be the following

$$Q_\tau(Y_t|\theta(\tau), \Omega_{t-1}^p) = \theta_0(\tau) + \sum_{i=1}^p \theta_i(\tau) Q_\tau(Y_{t-i}|\Omega_{t-i-1}^p) + \sum_{j=1}^q \theta_j(\tau) \ell_{t-j}(\mathbf{x}_{t-j}), \quad (19)$$

where $\Omega_{t-1}^p := \{Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}\}$ is the lagged-value vector of Y_t from $t-p$ up to time $t-1$, the parameter vector θ has a dimension of $r = p + q + 1$, and $\ell(\cdot)$ is a function of a vector of lagged values of observables $\mathbf{x}_{t-j} \in X_{t-j}$, which could be the lagged returns Y_{t-1} for instance. Let $X_t = (1, Y_{t-1}, \dots, Y_{t-p})'$, then the associated estimator θ_T is in the class of Z-estimators with $\psi(W_t, \theta, \tau) = \varepsilon_t(\tau)(\tau - \mathbb{1}\{Y_t - Q_\tau(Y_t|\theta(\tau), X_t) \leq 0\})$, where $\varepsilon_t(\tau) = Y_t - Q_\tau(Y_t|\theta(\tau), X_t)$. The following proposition provides conditions for applying our proposed test to the CAViaR model described in (19).

Proposition 2. *Let Assumptions C0-C7 and AN1-AN3 of Engle and Manganelli (2004) hold. Then Assumptions 1-5 hold for the CAViaR model.*

5.3 Distributional Regression Models

In distributional regression models (DR models), the conditional distribution function of Y_t is model through a family of binary response models for the event that Y_t exceeds some threshold $y \in \mathbb{R}$, as follows:

$$F(y|x) = \Lambda(x'\theta(y)), \text{ for some } \theta(y) \in \mathcal{B}(\mathbb{R}, \Theta) \subset \mathbb{R}^K \text{ and all } y \in \mathbb{R}, \quad (20)$$

where $\Lambda(\cdot)$ is a known strictly increasing link function (e.g., the logistic or standard normal distribution), and $\theta(\cdot)$ is a functional parameter taking values in $\mathcal{B}(\mathbb{R}, \Theta)$. The DR approach was introduced by Foresi and Peracchi (1995), and it has been analysed by

Fortin, Lemieux, and Firpo (2011), Rothe (2012), Rothe and Wied (2013), and Chernozhukov et al. (2013). One of the advantages of the distributional regression approach is that we do not need to assume that the dependent variable is continuously distributed, which could be useful in some empirical applications. One can also run a distributional regression model of Y_t conditional on its lagged values:

$$F(y|Y_{t-1}) = \Lambda(Y'_{t-1}\theta(y)), \text{ for some } \theta(y) \in \mathcal{B}(\mathbb{R}, \Theta) \subset \mathbb{R}^K \text{ and all } y \in \mathbb{R}, \quad (21)$$

For a given cut-off $y \in \mathbb{R}$, the estimator $\hat{\theta}_T(y)$ is given by

$$\hat{\theta}_T(y) := \arg \max_{\theta \in \mathcal{B}(\mathbb{R}, \Theta)} \frac{1}{T} \sum_{t=1}^T \left[\mathbb{1}\{Y_t \leq y\} \ln [\Lambda(Y'_{t-1}\theta(y))] + (1 - \mathbb{1}\{Y_t \leq y\}) \ln [1 - \Lambda(Y'_{t-1}\theta(y))] \right]. \quad (22)$$

Then, the conditional distribution of Y_t given Y_{t-1} is estimated as follows:

$$F_T(y|\hat{\theta}_T(y), Y_{t-1}) = \Lambda(Y'_{t-1}\hat{\theta}_T(y)), \text{ for all } y \in \mathbb{R}. \quad (23)$$

The following proposition provides the conditions for the distributional autoregressive model in (21) to satisfy the Assumptions 1-5, and hence the application of our test statistic S_T .

Proposition 3. *Let:*

(i) $\{(Y_{Tt}, X_{Tt}) : t \leq T, T = 1, 2, \dots\}$ is an α -mixing triangular array with stationary rows satisfying $E(|Y_1|^{2+\gamma}) < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{\gamma/(4+\gamma)} < \infty$ for some $\gamma \in (0, 2)$.

The support of Y , $\text{Supp}(Y)$, is a finite set or a bounded open subset of \mathbb{R} ;

(ii) The parameter $\theta_0(y)$ is identifiably unique with respect to

$$E \left[\mathbb{1}\{Y_t \leq y\} \ln (\Lambda(Y'_{t-1}\theta(y))) + (1 - \mathbb{1}\{Y_t \leq y\}) \ln (1 - \Lambda(Y'_{t-1}\theta(y))) \right] = 0,$$

and $\theta_0(y)$ is interior to Θ for every $y \in \text{Supp}(Y)$;

(iii) The conditional distribution function of Y_t given X_t , $F(\cdot|\cdot)$, has a density function $f(\cdot|\cdot)$ that is continuous, bounded, and bounded away from zero at all $y \in \text{Supp}(Y)$ a.s.;

(iv) $\Lambda(Y'_{t-1}\theta(\cdot))$ is bounded away from zero and one uniformly over $\theta \in \Theta$ a.s.;

(v) The matrix $E(X_t X'_t)$ is finite and has full rank.

Then Assumptions 1-5 hold for the distributional autoregressive model in (21).

Under Assumptions 1-5, we can apply our test statistic S_T to distributional regression models in dependent data settings, such as in (21).

6 Finite-Sample Performance

To examine the finite-sample performance of our proposed test statistic and its bootstrap procedure, we perform simulation experiments with data generating processes (DGPs) under the null and the alternative hypothesis. The data are generated from the processes

below.

Size DGPs :

$$\text{DGP.1 (AR(1)) : } Y_t = 0.2Y_{t-1} + u_t,$$

$$\text{DGP.2 (AR(2)) : } Y_t = 0.2Y_{t-1} + 0.2Y_{t-2} + u_t,$$

Power DGPs :

$$\text{DGP.3 (TAR) : } \begin{cases} Y_t = 1 + 0.6Y_{t-1} + u_t, & \text{if } Y_{t-1} \leq 1, \\ Y_t = 1 - 0.5Y_{t-1} + u_t, & \text{if } Y_{t-1} \geq 1, \end{cases}$$

$$\text{DGP.4 (Bilinear) : } Y_t = 0.8Y_{t-1}u_{t-1} + u_t,$$

$$\text{DGP.5 (Nonlinear MA) : } Y_t = 0.8u_{t-1}^2 + u_t,$$

$$\text{DGP.6 (Logistic Map) : } Y_t = 4Y_{t-1}(1 - Y_{t-1}),$$

$$\text{DGP.7 (GARCH(1,1)) : } Y_t = h_t u_t, \quad h_t^2 = 0.02 + 0.11Y_{t-1}^2 + 0.93h_{t-1}^2,$$

where u_t follows an i.i.d process with distribution $\mathcal{N}(0, 1)$. We want to test the null hypothesis that the quantiles of Y_t follow a AR(1) process:

$$\mathcal{H}_0 : F_{Y_t}^{-1}(\tau | \theta_0(\tau), Y_{t-1}) = \alpha + \beta Y_{t-1} + \Phi_u^{-1}(\tau), \text{ a.s.},$$

where $\Phi_u^{-1}(\tau)$ is the τ -quantile of the standard Normal error distribution. We use DGP.1 and DGP.2, described in Corradi and Swanson (2006), to check the size performance of our test statistic. While a $QAR(1)$ model correctly specifies the conditional distribution in DGP.1, we allow for dynamic misspecification in DGP.2, as $F(y|\theta_0, Y_{t-1}) \neq F(y|\theta_0^*, Y_{t-1}, Y_{t-2})$ with $\theta_0 \neq \theta_0^*$. The DGPs 3-7 allows us to see the empirical power performance and have been considered by Hong and Lee (2003) and Escanciano and Velasco (2010). In these experiments, rejection arises because of misspecification of the

conditional distribution model. DGP.4 and DGP.5 are second-order stationary, though they are not invertible (Granger and Andersen, 1978). DGP.6 follows a process similar to a white noise, but it has autocorrelations in squares similar to ARCH(1) (Granger and Teräsvirta, 2010). DGP.7 examine the power of our test against misspecifications in the conditional variance.

We also design a DGP for testing the specification of a Distributional Regression model in the form of (21). The data are generated as in DGP.5, a Nonlinear MA(1) model, and we are interested in testing the null hypothesis that the Distributional Regression model is correctly specified conditioning Y_t only on Y_{t-1} :

$$\mathcal{H}_0^{DR} : F(y|Y_{t-1}) = \Lambda(Y'_{t-1}\theta(y)), \text{ a.s.}, \quad (24)$$

where $\Lambda(\cdot)$ is specified as a logistic distribution function. For all the experiments, we consider the empirical rejection frequencies for 5% and 10% nominal level tests with different sample sizes ($T = 100$ and 300), and choose a grid $\mathcal{T} = [0.01, 0.99]$. In calculating the test statistics, we use an equally spaced grid of 100 quantiles $\mathcal{T}_n \subset \mathcal{T}$. We perform 1,000 Monte Carlo repetitions in each of the simulations, and apply $B = 399$ block bootstrap replications in each of the simulations to calculate the critical values. Then the maximal simulation standard error for the tests empirical sizes and powers is $\max_{0 \leq p \leq 1} \sqrt{p(1-p)/1000} \approx 0.016$. For each bootstrap replicate, we use three different block lengths $\ell = \{2, 4, 6\}$, which are close to the block length of $CT^{1/3}$, for a constant $C > 0$, suggested by Künsch (1989). In all the replications, we generated and discarded 200 pre-sample data values. Except for the Distributional Regression specification test, we compare our results with the test proposed by Escanciano and Velasco (2010) (EV henceforth), based on

$$EV := \int \int \left| \left(\mathbb{1}(Y_t - m(X_t, \hat{\theta}_T(\tau)) \leq 0) - \tau_j \right) \exp(\mathbf{i}\mathbf{x}'X_t) \right|^2 dW(\mathbf{x})d\Phi(\alpha), \quad (25)$$

where W and Φ are some integrating measures on \mathbb{R} and \mathcal{T} , and $m(X_t, \hat{\theta}_T(\tau))$ is the

estimated parametric QAR(1) model for each τ -quantile, for $\tau \in \mathcal{T}$. The critical values of the test (25) are obtained by subsampling. In each Monte Carlo replication, $T - b - 1$ subsamples of size b were generated. We apply the EV test for two different subsample sizes $b = \lceil kT^{(2/5)} \rceil$, for $k = 3$ and 4, following the suggestion of Sakov and Bickel (2000).

Tables 1 and 2 report the rejection frequencies of the S_T test associated with the DGPs 1-7, for sample sizes $T = 100$ and $T = 300$ respectively. The empirical level of the S_T test is generally close to the nominal level under the null hypothesis, disregarding whether there is dynamic misspecification (DGP.2) or not (DGP.1). On the other hand, the EV test of Escanciano and Velasco (2010) presents size distortions for both sample sizes, increasing in the presence of dynamic misspecification (DGP.2). Those results are robust for different subsample sizes b . Thus, our test has the correct asymptotic size even in the presence of dynamic misspecification.

In terms of power, the S_T test exhibits good power and reliable inference even when using a small sample size $T = 100$. Comparing with the EV test, the S_T test performs well: it is the most powerful test for DGP.3, DGP.4, DGP.6, and DGP.7; it has less power than the EV test only against DGP.5, when the subsample size is $b = 18$, but it still has more power than the EV test for a subsample size of $b = 25$. In addition, the power of both tests converge to 1 for $T = 300$. Our test statistic is also powerful against misspecifications in the distributional regression, as the power for testing \mathcal{H}_0^{DR} in (24) is 1 for a small sample size of $T = 100$ (Table 1). To the best of our knowledge, no specification test for Distributional Regression models has been developed for a time series setting. In sum, our proposed test seems to perform quite well in finite samples.

Table 1. Monte Carlo empirical rejection frequencies of specification tests: $T = 100$

	$S_T(\ell = 2)$		$S_T(\ell = 4)$		$S_T(\ell = 6)$		$EV(b = 18)$		$EV(b = 25)$	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.1	0.030	0.110	0.060	0.110	0.036	0.066	0.075	0.129	0.077	0.120
DGP.2	0.030	0.080	0.040	0.100	0.052	0.102	0.091	0.160	0.084	0.143
DGP.3	0.960	0.990	0.990	0.990	0.920	0.960	0.888	0.931	0.847	0.913
DGP.4	0.962	0.990	1.000	1.000	1.000	1.000	0.997	0.999	0.984	0.993
DGP.5	0.912	0.952	0.864	0.916	0.900	0.954	0.944	0.969	0.913	0.944
DGP.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.7	0.216	0.304	0.260	0.360	0.200	0.380	0.095	0.149	0.101	0.150
\mathcal{H}_0^{DR}	1.000	1.000	1.000	1.000	1.000	1.000	-	-	-	-

Note: S_T denotes our proposed test statistic with $B = 399$ bootstrap replications with block lengths $\ell = \{2, 4, 6\}$. EV denotes the subsampling specification test of Escanciano and Velasco (2010). The null hypothesis \mathcal{H}_0^{DR} test the specification of a Distributional Regression model specified in (24), under DGP.5. We use 1,000 Monte Carlo repetitions based on the DGPs 1-7 described above.

Table 2. Monte Carlo empirical rejection frequencies of specification tests: $T = 300$

	$S_T(\ell = 2)$		$S_T(\ell = 4)$		$S_T(\ell = 6)$		$EV(b = 29)$		$EV(b = 39)$	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.1	0.043	0.087	0.020	0.080	0.031	0.070	0.061	0.107	0.057	0.108
DGP.2	0.053	0.107	0.067	0.107	0.049	0.122	0.092	0.147	0.074	0.134
DGP.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DGP.7	0.970	0.980	0.960	0.980	0.980	1.000	0.186	0.272	0.189	0.260
\mathcal{H}_0^{DR}	1.000	1.000	1.000	1.000	1.000	1.000	-	-	-	-

Note: S_T denotes our proposed test statistic with $B = 399$ bootstrap replications with block lengths $\ell = \{2, 4, 6\}$. EV denotes the subsampling specification test of Escanciano and Velasco (2010). The null hypothesis \mathcal{H}_0^{DR} test the specification of a Distributional Regression model specified in (24), under DGP.5. We use 1,000 Monte Carlo repetitions based on the DGPs 1-7 described above.

7 An Empirical Application

Many empirical papers have proposed methods to precisely check the specification of models for Value-at-Risk (VaR). Since VaR determines the regulatory risk capital of all regulated financial institutions (see Basel Committee on Banking Supervision 1996), the outcome of a VaR model determines the multiplication factors for market risk capital requirements of financial institutions. Thus, an inaccurate VaR model leads to an underestimated multiplicative factor, that delivers an insufficient reserve of capital risk for financial institutions. Therefore, the specification of VaR models is crucial for risk managers, regulators, and financial institutions.

Since the VaR is a quantile of the portfolio returns, conditional on past information, and as the distribution of portfolio returns evolves over time, it is challenging to model time-varying conditional quantiles. An accurate VaR model satisfies $\Pr(Y_t \leq -VaR_t | \mathcal{F}_{t-1}) = \tau$, for a portfolio return series Y_t , a past information set \mathcal{F}_{t-1} , and a quantile $\tau \in (0, 1)$. The dynamic conditional quantile regression approach specifies a conditional VaR model using only the relevant past information that influence the quantiles of interest, and many applications support this methodology (Chernozhukov and Umantsev, 2001, Engle and Manganelli, 2004, Escanciano and Olmo, 2010).

To illustrate the performance of our proposed test statistic, we test different specifications of conditional quantile regression models for estimating the VaR of stock returns. We estimate the VaR of the returns of two major stock indexes, the Frankfurt Dax Index (DAX) and the London FTSE-100 Index (FTSE-100). The DAX and the FTSE-100 daily stock indexes are two representatives of the data for which linear and non-linear quantile regression models have been widely used, see e.g. Escanciano and Velasco (2010), Iqbal and Mukherjee (2012), and Jeon and Taylor (2013). The dataset consists of 2,981 daily observations - from January 2003 to June 2014 - on Y_t , the one-day returns, and X_t , the lagged returns $(Y_{t-1}, \dots, Y_{t-p})$.

Figure 1 displays the daily log-return series of the two series. It shows that both log-return series display calm as well as volatile periods and also single outlying log-return observations. Table 3 presents the summary statistics of the series. Both log-returns series

are highly leptokurtic and present autocorrelation.

Table 3. Summary statistics: DAX and FTSE-100 daily log-returns

	DAX	FTSE-100
Mean	0.02	0.01
Std. Dev.	0.61	0.51
Median	0.03	0.01
Skewness	0.01	-0.12
Kurtosis	9.14	11.71
Minimum	-3.23	-4.02
Maximum	4.69	4.08
Autocorrelation	-0.01	-0.06
LB(10)	21.34	62.35

Note: The Autocorrelation is the first-order autocorrelation coefficient, and $LB(10)$ denotes the Ljung-Box Q-statistic of order 10.

For each series, we estimate a Gaussian AR(1)-GARCH(1,1) of the VaR, $VaR_t(\tau)$, as follows:

$$\text{AR(1)-GARCH(1,1): } F_{Y_t}^{-1}(\tau|\theta_0(\tau), Y_{t-1}, \sigma_t) = \beta_0 + \beta_1 Y_{t-1} + F_{\varepsilon}^{-1}(\tau)\sigma_t,$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 Y_{t-1}^2 + \gamma_2 \sigma_{t-1}^2,$$

where $F_{\varepsilon}^{-1}(\tau)$ is the τ -quantile of the standard Gaussian error distribution. Thus, we test the hypothesis \mathcal{H}_0 : the VaR of the log-return Y_t follows an AR(1)-GARCH(1,1) Gaussian process. We choose this specification as GARCH models have provided appropriate specifications of the VaR of stock returns in the literature (Escanciano and Olmo, 2010). We also entertain other models: GARCH(1,1), AR(2)-GARCH(2,2), E-GARCH(1,1), AR(1)-GARCH(1,1) with Student-t5 distribution, and GARCH(1,1) with Student-t5 distribution. We apply GARCH(1,1) and AR(1)-GARCH(1,1) with a Student-t5 distribution because they are valid models for the distribution of monthly stock returns in Bai (2003) and Kheifets (2014). To present results with a different GARCH specification, we estimate

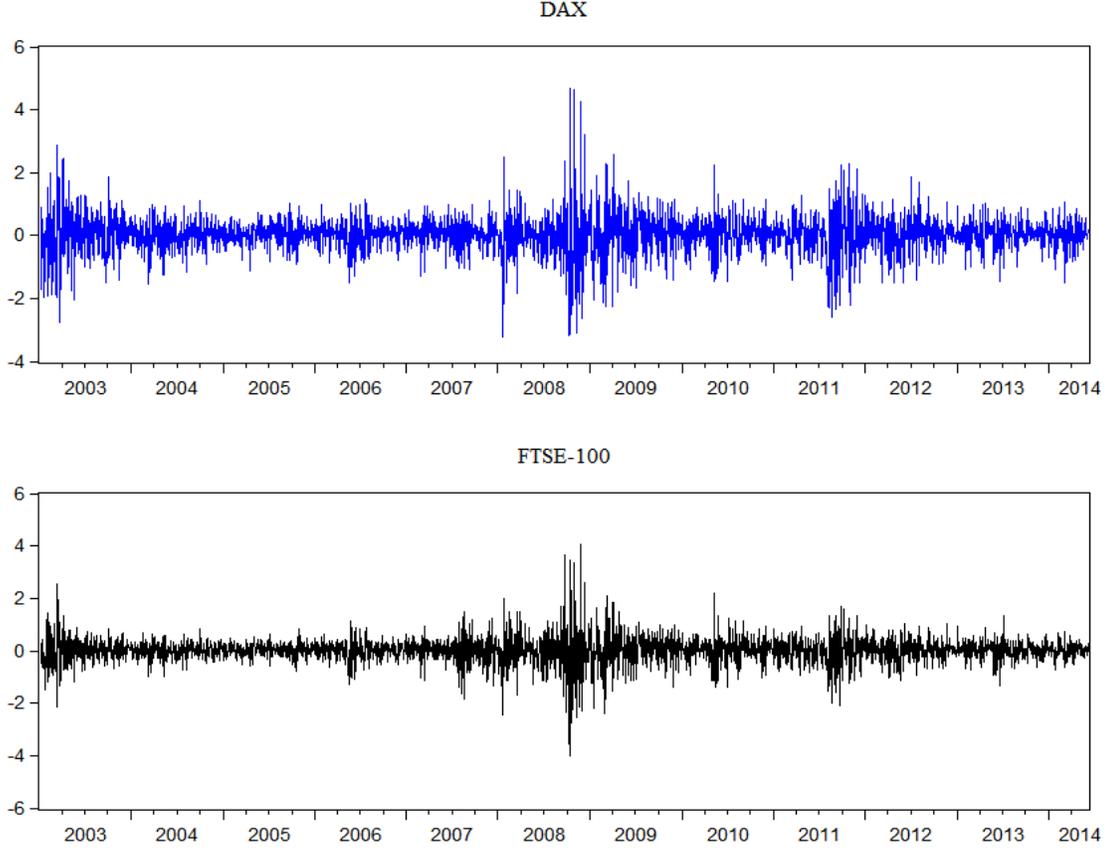


Figure 1. Daily log-returns of DAX and FTSE-100 indexes in the period January 6th, 2003 to July 9th, 2014

an E-GARCH(1,1) model for the VaR as:

$$\begin{aligned} \text{E-GARCH}(1,1): F_{Y_t}^{-1}(\tau|\theta_0(\tau), Y_{t-1}, h_t) &= F_{\varepsilon}^{-1}(\tau)h_t, \\ \ln h_t^2 &= \alpha_0 + \alpha_1 \ln h_{t-1}^2 + \alpha_2 \left(|Y_{t-1}^2| - (2/\pi)^{\frac{1}{2}} \right) - \alpha_3 Y_{t-1}^2. \end{aligned} \quad (26)$$

As we want to compare our methodology with standard specification tests for conditional quantile regression models in the literature, we perform the EV test described in (25), with two different subsample sizes $b = \lceil kT^{2/5} \rceil$ for $k = 3$ and $k = 4$.

Table 4 shows the p -values of the specification tests for all the VaR models. For the DAX index series, our test S_T rejects the specifications of all proposed models to

fitting a VaR for the log-returns at 1% significance level. These results are robust to three different block lengths. On the other hand, the EV test of Escanciano and Velasco (2010) do not reject an AR(1)-GARCH(1,1) specification with Student-t5 distribution at 1% significance level. Regarding the FTSE-100 series, the S_T test does not reject a AR(1)-GARCH(1,1) model at 1% significance level, while the EV test does not reject a AR(1)-GARCH(1,1) model with Student-t5 distribution at the 1% significance level. We note that the AR(1)-GARCH(1, 1) family of models is the only class of models that is not rejected for these returns series, but this result is not robust to different block lengths. Thus, the empirical application shows the ability of our test to detect possibly misspecified conditional distribution models. This is useful for risk managers and financial institutions to apply a valid VaR model and obtain the correct multiplicative factors for their market risk capital requirements.

Table 4. Specification tests p -values of VaR models of DAX and FTSE-100 returns

DAX					
	$S_{T,6}$	$S_{T,8}$	$S_{T,16}$	EV($b=98$)	EV($b=122$)
GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.000	0.000
GARCH(1,1)-t5 - CAViaR	0.001	0.001	0.001	0.001	0.000
AR(1)-GARCH(1,1) - CAViaR	0.002	0.001	0.001	0.000	0.000
AR(1)-GARCH(1,1)-t5 - CAViaR	0.001	0.001	0.002	0.010	0.007
AR(2)-GARCH(2,2) - CAViaR	0.001	0.001	0.001	0.000	0.000
E-GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.001	0.001
FTSE-100					
	$S_{T,6}$	$S_{T,8}$	$S_{T,16}$	EV($b=98$)	EV($b=122$)
GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.000	0.000
GARCH(1,1)-t5 - CAViaR	0.001	0.001	0.001	0.010	0.002
AR(1)-GARCH(1,1) - CAViaR	0.002	0.011	0.003	0.009	0.004
AR(1)-GARCH(1,1)-t5 - CAViaR	0.004	0.005	0.004	0.010	0.007
AR(2)-GARCH(2,2) - CAViaR	0.009	0.004	0.003	0.000	0.000
E-GARCH(1,1) - CAViaR	0.001	0.001	0.001	0.001	0.001

Note: $S_{T,\ell}$ is the S_T test with block length $\ell = \{6, 8, 16\}$. We denote EV as the specification test of Escanciano and Velasco (2010), with sub-sample size b . The E-GARCH(1,1) is estimated as in (26).

8 Conclusion

In this paper, we present a practical and consistent specification test of conditional distribution and quantile models in a very general setting for dependent observations. Our setting covers conditional distribution models possibly indexed by function-valued parameters, which allows for a wide range of important empirical applications in economics and finance, such as the linear quantile auto-regressive, the CAViaR, and the distributional regression models. Based on a comparison between an estimated parametric distribution and the empirical distribution function, our proposed bootstrap test has the correct asymptotic size and is consistent against fixed alternatives. In addition, our test has non-trivial power against \sqrt{T} -local alternatives, with T the sample size.

Finite sample experiments suggest that our proposed test has good size and power properties, and is more powerful than other comparable specification tests in the literature against almost all alternatives. In addition, our approach has the correct asymptotic size under dynamic misspecification. An empirical application illustrates the practical importance of our setting in risk management. The use of misspecified VaR models may lead to the acceptance of a sub-optimal model for VaR, underestimating the multiplicative factors of the reserve of capital risk of financial institutions. Therefore, checking the validity of a VaR model is of crucial importance for monitoring risk of financial institutions. We observe that the AR(1)-GARCH(1, 1) family of models provided valid specifications for the VaR of two major stock returns indexes.

A possible direction for future work is to extend this study to test Granger-causality in distribution. Although the concept of Granger-causality is defined in terms of the conditional distribution, the majority of papers have tested Granger-causality using conditional mean regression models in which the causal relations are linear. As a result, a conditional mean regression model cannot assess a tail causal relation or nonlinear causalities. Our proposed approach allows us to evaluate nonlinear causalities, causal relations in conditional quantiles, and Granger-causality in distribution through an application of distributional regression in a time series context. One could also extend our approach to the class of multivariate models, providing specification tests for vector autoregressions

and multivariate linear and non-linear models, see e.g. Francq and Raïssi (2007) and Escanciano, Lobato, and Zhu (2013).

Appendix

A.1 Lemmas

In this section, we introduce some auxiliary results.

Lemma A.1. *Given Assumption 1, under \mathcal{H}_0 of (4) or \mathcal{H}_A of (5),*

$$v_T(y, x) := \sqrt{T}(\hat{Z}_T(y, x) - F_{YX}(y, x)) \implies \mathbb{H}_1(y, x), \text{ in } \ell^\infty(\mathcal{W}),$$

where \mathbb{H}_1 is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{W})$ with covariance function

$$\text{Cov}(\mathbb{H}_1(y, x), \mathbb{H}_1(y', x')) = \sum_{k=-\infty}^{\infty} \text{Cov}(\mathbb{1}\{Y_0 \leq y\} \mathbb{1}\{X_0 \leq x\}, \mathbb{1}\{Y_k \leq y'\} \mathbb{1}\{X_k \leq x'\}).$$

Proof. Assumption 1 implies strong mixing coefficients $\alpha(j) = O(j^{-k})$, for some $k > 1$. Then the result follows from a direct application of Theorem 7.2 in Rio (2000). \square

Lemma A.2. *Given Assumptions 1-5, under \mathcal{H}_0 of (4) or \mathcal{H}_A of (5), we have*

$$r_T(\theta, \tau) := \sqrt{T}(\hat{\Psi}_T(\theta, \tau) - \Psi(\theta, \tau)) \implies \tilde{\mathbb{H}}_2(\theta, \tau), \text{ in } \ell^\infty(\mathcal{T} \times \Theta),$$

$$\sqrt{T}(\hat{\theta}_T(\cdot) - \theta_0(\cdot)) \implies -\dot{\Psi}_{\theta_0, \cdot}^{-1}[\tilde{\mathbb{H}}_2(\theta_0(\cdot), \cdot)] \text{ in } \ell^\infty(\mathcal{T}),$$

where $\tilde{\mathbb{H}}_2$ is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{T} \times \Theta)$ with covariance function

$$\text{Cov}(\tilde{\mathbb{H}}_2(\theta, \tau), \tilde{\mathbb{H}}_2(\theta', \tau')) = \sum_{k=-\infty}^{\infty} \text{Cov}(\psi(W_0, \theta, \tau), \psi(W_k, \theta', \tau')).$$

Proof. Under Assumptions 1-5, we may apply Lemma E.2 in Chernozhukov et al. (2013), and the map $\theta \mapsto \Psi(\theta, \cdot)$ is Hadamard differentiable at θ_0 with continuously invertible derivative $\dot{\Psi}_{\theta_0, \cdot}$. Then $r_T(\theta, \tau)$ weakly converges to $\tilde{\mathbb{H}}_2(\theta, \tau)$ in $\ell^\infty(\mathcal{T} \times \Theta)$. Assumptions 2-4 provide sufficient conditions for applying Lemma E.1 in Chernozhukov et al. (2013), and the result follows. \square

Lemma A.3. *Given Assumptions 1-5, under \mathcal{H}_0 of (4) or \mathcal{H}_A of (5), we have*

$$v_T^{\theta_0}(y, x) := \sqrt{T}(\hat{F}_T(y, x, \hat{\theta}_T) - F_T(y, x, \theta_0)) \implies \mathbb{H}_2(y, x) \text{ in } \ell^\infty(\mathcal{W}),$$

where \mathbb{H}_2 is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{W})$.

Proof. From Lemma A.2, $\sqrt{T}(\hat{\theta}_T(\cdot) - \theta_0(\cdot)) \implies -\dot{\Psi}_{\theta_0, \cdot}^{-1}[\tilde{\mathbb{H}}_2(\theta_0(\cdot), \cdot)]$ in $\ell^\infty(\mathcal{T})$, where $\tilde{\mathbb{H}}_2$ is a Gaussian process in $\ell^\infty(\mathcal{T} \times \Theta)$. We can rewrite $v_T^{\theta_0}(y, x)$ as

$$\begin{aligned} \sqrt{T}(\hat{F}_T(y, x, \hat{\theta}_T) - F_T(y, x, \theta_0)) &= \int (F(y|\hat{\theta}_T, \bar{x}) - F(y|\bar{x})) \mathbb{1}\{\bar{x} \leq x\} \sqrt{T} dF_X(\bar{x}) \\ &\quad + \int F(y|\bar{x}) \mathbb{1}\{\bar{x} \leq x\} \sqrt{T} d[\hat{F}_X(\bar{x}) - F_X(\bar{x})]. \end{aligned}$$

By the Hadamard differentiability of the map $\theta \mapsto F(\cdot|\theta(\cdot), \cdot)$ in Assumption 5, we can apply the functional delta method as follows:

$$\sqrt{T}(F(y|\hat{\theta}_T, x) - F(y|x)) \implies -\dot{F}^{-1}(y|\theta_0, x) \left[-\dot{\Psi}_{\theta_0, \cdot}^{-1}[\tilde{\mathbb{H}}_2(\theta_0(\cdot), \cdot)] \right] := \mathbb{H}_2^*(y, x) \text{ in } \ell^\infty(\mathcal{W}).$$

By Lemma A.1, $\sqrt{T}(\hat{F}_X(\bar{x}) - F_X(\bar{x}))$ weakly converges to the functional of a tight mean zero Gaussian process. Let the measurable functions $\Gamma : \mathcal{W} \mapsto [0, 1]$ be defined by $(y, x) \mapsto \Gamma(y, x)$ and the bounded maps $\Pi : \mathcal{H} \mapsto \mathbb{R}$ be defined by $f \mapsto \int f d\Pi$. Then it follows from Lemma D.1 in Chernozhukov et al. (2013) that the mapping $(\Gamma, \Pi) \mapsto \int \Gamma(\cdot, x) d\Pi(x)$ - with $\Gamma(\cdot, x) = \mathbb{1}\{\cdot \leq x\} F(\cdot|x)$ and $\Pi = F_X(\cdot)$ - is well defined and Hadamard differentiable. Then the result follows from an application of the functional

delta method, where the Gaussian process \mathbb{H}_2 is given by

$$\mathbb{H}_2(y, x) := \int \mathbb{H}_2^*(y, \bar{x}) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x}) + \int F(y|\bar{x}) \mathbb{1}\{\bar{x} \leq x\} d\mathbb{H}_1(\infty, \bar{x}),$$

and \mathbb{H}_1 is the same process as in Lemma A.1. \square

Lemma A.4. *Under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (13) and Assumptions 1-6,*

$$\sqrt{T}(\hat{Z}_T(y, x) - F_T^A(y, x)) \implies \mathbb{H}_1(y, x), \text{ in } \ell^\infty(\mathcal{W}),$$

$$\sqrt{T}(\hat{\Psi}_T(\theta, \tau) - \Psi_{F_T}(\theta, \tau)) \implies \hat{\mathbb{H}}_2(\theta, \tau), \text{ in } \ell^\infty(\mathcal{T} \times \Theta),$$

where $F_T^A(y, x) = \int F_T(y|\bar{x}) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x})$, $\Psi_{F_T}(\theta, \tau) = \mathbb{E}_{F_T} [\psi(W_t, \theta, \tau)]$, and $(\mathbb{H}_1, \hat{\mathbb{H}}_2)$ are the tight mean zero Gaussian processes derived in Lemmas A.1-A.2.

Proof. First, under Assumption 6, $F_T^A(y, x)$ is contiguous to $F(y, x, \theta_0)$, then Lemma 6.2 in Van der Vaart (2000) and Lemma A.1 imply that $\sqrt{T}(\hat{Z}_T(y, x) - F_T^A(y, x)) \implies \mathbb{H}_1(y, x)$ in $\ell^\infty(\mathcal{W})$. Under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (13) and Assumptions 1-6, $F_T(y|X_t)$ of (13) is a linear combination of two measures that are VC class with a square integrable envelope, then the result follows from an application of Theorem 3.2.10 in Van der Vaart and Wellner (2000). \square

Now we define weak convergence conditional on the data in probability ($\xrightarrow[M]{\mathbb{P}}$ -convergence) in the Hoffmann-Jørgensen sense as defined in Section 11.3 in Kosorok (2007).

Lemma A.5. *Let $W_t = \{Y_{Tt}, X_{Tt}\}$ be a $(1+d)$ -dimensional triangular array with stationary rows satisfying Assumption 7 with marginal distribution P , and let $\mathcal{M} := \{\Psi(\theta, \tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ be a permissible VC class of measurable functions with a square integrable envelope function \mathbb{F} satisfying $P(\mathbb{F})^p < \infty$, for $2 < p < \infty$. Conditional on the data W_1, \dots, W_T , let W_1^*, \dots, W_T^* be generated according to the block bootstrap with block size $\ell := \ell(T)$, with $\ell(T) \rightarrow \infty$ as $T \rightarrow \infty$. Let $v_T^*(y, x) := \sqrt{T}(\hat{Z}_T^*(y, x) - \hat{Z}_T(y, x))$ denote*

the block bootstrap version of the empirical process $v_T(y, x) = \sqrt{T}(\hat{Z}_T(y, x) - F_{YX}(y, x))$.

If we also assume additional conditions:

(i) $\limsup_{k \rightarrow \infty} k^q \beta(k) < \infty$, for some $q > p/(p-2)$, for $2 < p < \infty$ such that $P^*(\mathbb{F})^p < \infty$, and

(ii) $\ell(T) = O(T^\rho)$ for some $0 < \rho < (p-2)/[2(p-1)]$,

then

$$v_T^*(y, x) \xrightarrow[M]{\mathbb{P}} \mathbb{H}_1(y, x), \text{ in } \ell^\infty(\mathcal{W}),$$

where \mathbb{H}_1 is a tight mean zero Gaussian process as defined in Lemma A.1.

Proof. It follows directly from an application of Theorem 1 in Radulović (1996) or Theorem 11.24 in Kosorok (2007), slightly modified to address measurability. \square

Lemma A.6. Under Assumptions 2-7, under \mathcal{H}_0 of (4), or \mathcal{H}_A of (5), or under the local alternative $\mathcal{H}_{A,T}$ of (13),

$$\sqrt{T}(\hat{F}_T^*(y, x, \hat{\theta}_T^*) - \hat{F}_T(y, x, \hat{\theta}_T)) \xrightarrow[M]{\mathbb{P}} \mathbb{H}_2(y, x) \text{ in } \ell^\infty(\mathcal{W}),$$

where \mathbb{H}_2 is the tight mean zero Gaussian process defined in Lemma A.3.

Proof. Since $F(\cdot|\theta, \cdot)$ is Hadamard differentiable, by the chain rule for the Hadamard derivative and Lemma A.5 we can apply a functional delta-method for bootstrap in probability defined in Theorem 3.9.11 of Van der Vaart and Wellner (2000) that yields the result. \square

A.2 Proofs

Proof of Theorem 1. To prove part (i), we consider the empirical processes $v_T(y, x)$ and $v_T^{\theta_0}(y, x)$. Under \mathcal{H}_0 of (4), $F_{YX}(y, x) \equiv F(y, x, \theta_0)$, and we have

$$\begin{aligned}
S_T &= \int (v_T(y, x) - v_T^{\theta_0}(y, x))^2 dF_{YX}(y, x) \\
&\quad + \int (v_T(y, x) - v_T^{\theta_0}(y, x))^2 d(\hat{Z}_T(y, x) - F_{YX}(y, x)).
\end{aligned}$$

By Lemma A.1, we have

$$S_T = \int (v_T(y, x) - v_T^{\theta_0}(y, x))^2 dF_{YX}(y, x) + o_P(1).$$

From Lemmas A.1 and A.3, $(v_T(y, x), v_T^{\theta_0}(y, x)) \implies (\mathbb{H}_1(y, x), \mathbb{H}_2(y, x))$ in $\ell^\infty(\mathcal{W} \times \mathcal{W})$.

Then the result follows by the continuous mapping theorem.

In part (ii), under the alternative hypothesis \mathcal{H}_A of (5), $F_{YX}(y, x) \neq F(y, x, \theta_1)$ for some $(y, x) \in \mathcal{W}$ and for all $\theta_1 \in \mathcal{B}(\mathcal{T}, \Theta)$, and we have

$$\begin{aligned}
S_T &= T \int \left(\hat{Z}_T(y, x) - \hat{F}_T(y, x, \hat{\theta}_T) \pm F_{YX}(y, x) \pm F(y, x, \theta_1) \right)^2 dF_{YX}(y, x) \\
&= \int \left(v_T(y, x) - v_T^{\theta_0}(y, x) + \sqrt{T} (F_{YX}(y, x) - F(y, x, \theta_1)) \right)^2 dF_{YX}(y, x) + o_P(1).
\end{aligned}$$

Therefore, for any fixed constant $\varepsilon > 0$, $\lim_{T \rightarrow \infty} \Pr(S_T > \varepsilon) = 1$ and the result follows. \square

Proof of Theorem 2. Consider the empirical processes

$$v_T^1(y, x) = \sqrt{T} \left(\hat{Z}_T(y, x) - \int F(y|\theta_0, \bar{x}) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x}) \right), \text{ and}$$

$$r_T^1(\theta, \tau) = \sqrt{T} (\hat{\Psi}_T(\theta, \tau) - \mathbf{E}_F[\psi(W_t, \theta, \tau)]),$$

where $\Psi_F(\theta, \tau) := \mathbf{E}_F[\psi(W_t, \theta, \tau)]$ as defined in (14). Then it follows from Lemma A.4 that

$$\begin{aligned} v_T^1(y, x) &\implies \mathbb{H}_1(y, x) + \delta \int (J(y|\bar{x}) - F(y|\theta_0, \bar{x})) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x}), \\ r_T^1(\theta, \tau) &\implies \tilde{\mathbb{H}}_2(\theta, \tau) + \delta [\mathbf{E}_J[\psi(W_t, \theta, \tau)] - \mathbf{E}_F[\psi(W_t, \theta, \tau)]], \end{aligned}$$

where $\Psi_J(\theta, \tau) := \mathbf{E}_J[\psi(W_t, \theta, \tau)]$ as defined in (15). Now taking the empirical process $v_T^{1\theta_0}(y, x) = \sqrt{T} \left(\int F(y|\hat{\theta}_T, \bar{x}) \mathbb{1}\{\bar{x} \leq x\} d\hat{F}_X(\bar{x}) - \int F(y|\theta_0, \bar{x}) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x}) \right)$, we have

$$v_T^{1\theta_0}(y, x) \implies \mathbb{H}_2(y, x) + \delta \int \dot{F}(y|\bar{x})[h] \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x}),$$

with $h(\tau) = [\frac{\partial}{\partial \theta} \Psi_F(\theta_0, \tau)]^{-1} \Psi_J(\theta_0, \tau)$. Similarly to the proof of Theorem 1, under $\mathcal{H}_{A,T}$ of (13), we have

$$S_T = \int (v_T^1(y, x) - v_T^{1\theta_0}(y, x))^2 dF_{YX}(y, x) + o_P(1),$$

then the result follows from the continuous mapping theorem. \square

Proof of Theorem 3. For part (i), from Lemma A.5, $\hat{c}_T^*(\alpha) = c(\alpha) + o_P(1)$, where $c(\alpha)$ satisfies $\Pr(S_T > c(\alpha)) = \alpha + o(1)$. Then as $T \rightarrow \infty$, $\Pr(S_T > \hat{c}_T^*(\alpha)) = \alpha + o(1)$. For part (ii), there exists a fixed constant $C > 0$ such that

$$\begin{aligned} \Pr(S_T \leq \hat{c}_T^*(\alpha)) &= \Pr(S_T \leq \hat{c}_T^*(\alpha), S_T \leq C) + \Pr(S_T \leq \hat{c}_T^*(\alpha), S_T > C) \\ &\leq \Pr(S_T \leq C) + \Pr(\hat{c}_T^*(\alpha) > C) \\ &\leq o(1) + \varepsilon + o(1), \end{aligned}$$

where the first element of the third line follows from Theorem 1 - $\Pr(S_T \leq C) = o(1)$ - and the rest of the third line is due to Lemmas A.5-A.6, that imply for any $\varepsilon > 0$, there exists a fixed constant C such that $\Pr(\hat{c}_T^*(\alpha) > C) < \varepsilon + o(1)$. The result follows from an arbitrary choice of $\varepsilon > 0$. Part (iii) follows from an application of Theorem 4 of Andrews (1997) and Anderson's Lemma in Ibragimov and Has'minskii (1981). By Anderson's Lemma, since $\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x)$ has mean zero $\forall (y, x) \in \mathcal{W}$, under \mathcal{H}_0 we have

$$\begin{aligned} & \Pr \left(\int (\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x) + \Delta(y, x))^2 dF_{YX}(y, x) \geq c(\alpha) \right) \\ & \geq \Pr \left(\int (\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x))^2 dF_{YX}(y, x) \geq c(\alpha) \right) \\ & = \Pr(S_T \geq c(\alpha)) = \alpha. \end{aligned}$$

Under Assumption 6, the conditional distribution under a local alternative $F_T(\cdot|\cdot)$ implies a sequence of distribution functions $Z_T(y, x)$ that is contiguous to the distribution function $F(y, x, \theta_0)$ given by $\int F(y|\theta_0, \bar{x}) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x})$, under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (13). Since contiguity preserves convergence in probability to constants, under the local alternatives we have

$$\begin{aligned} & \Pr \left(\int (\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x) + \Delta(y, x))^2 dF_{YX}(y, x) \geq \hat{c}_T^*(\alpha) \right) \\ & \geq \Pr \left(\int (\mathbb{H}_1(y, x) - \mathbb{H}_2(y, x))^2 dF_{YX}(y, x) \geq c(\alpha) \right) \\ & = \Pr(S_T \geq c(\alpha)) \geq \alpha. \end{aligned}$$

where the third line follows from $\Pr(S_T > c(\alpha)) \geq \alpha + o(1)$ under a sequence of local alternatives. □

Proof of Propositions 1-3. The proofs follow directly from an application of the results contained in Section D of Chernozhukov et al. (2013). □

References

- Alexander, K. S. (1987), “The Central Limit Theorem for Empirical Processes on Vapnik-Cervonenkis Classes,” *The Annals of Probability*, 15, 178–203.
- Andrews, D. W. K. (1997), “A Conditional Kolmogorov Test,” *Econometrica*, 65, 1097–1128.
- Andrews, D. W. K. and Pollard, D. (1994), “An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes,” *International Statistical Review*, 62, 119–132.
- Andrews, I. (2012), “A Mean Likelihood Ratio Specification Test,” *Working Paper*.
- Bai, J. (2003), “Testing Parametric Conditional Distributions of Dynamic Models,” *The Review of Economics and Statistics*, 85, 531–549.
- Bierens, H. J. and Wang, L. (2014), “Weighted Simulated Integrated Conditional Moment Tests for Parametric Conditional Distributions of Stationary Time Series Processes,” *Forthcoming in Econometric Reviews*.
- Billingsley, P. (1995), *Probability and Measure*, John Wiley & Sons.
- Box, G. E. P. and Pierce, D. A. (1970), “Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models,” *Journal of the American statistical Association*, 65, 1509–1526.

- Bradley, R. C. (1985), “On the Central Limit Question under Absolute Regularity,” *The Annals of Probability*, 13, 1314–1325.
- Carrasco, M. and Chen, X. (2002), “Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models,” *Econometric Theory*, 18, 17–39.
- Chernozhukov, V. and Fernández-Val, I. (2005), “Subsampling Inference on Quantile Regression Processes,” *Sankhya*, 67, 253–276.
- Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013), “Inference on Counterfactual Distributions,” *Econometrica*, 81, 2205–2268.
- Chernozhukov, V. and Umantsev, L. (2001), “Conditional Value-at-Risk: Aspects of Modeling and Estimation,” *Empirical Economics*, 26, 271–292.
- Corradi, V. and Swanson, N. R. (2006), “Bootstrap Conditional Distribution Tests in the Presence of Dynamic Misspecification,” *Journal of Econometrics*, 133, 779–806.
- Delgado, M. A. and Stute, W. (2008), “Distribution-Free Specification Tests of Conditional Models,” *Journal of Econometrics*, 143, 37–55.
- Dudley, R. M. (1978), “Central Limit Theorems for Empirical Measures,” *The Annals of Probability*, 6, 899–929.
- Engle, R. and Manganelli, S. (2004), “CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles,” *Journal of Business & Economic Statistics*, 22, 367–381.
- Escanciano, J. and Velasco, C. (2010), “Specification Tests of Parametric Dynamic Conditional Quantiles,” *Journal of Econometrics*, 159, 209–221.
- Escanciano, J. C., Lobato, I. N., and Zhu, L. (2013), “Automatic Specification Testing for Vector Autoregressions and Multivariate Nonlinear Time Series Models,” *Journal of Business & Economic Statistics*, 31, 426–437.

- Escanciano, J. C. and Olmo, J. (2010), “Backtesting Parametric Value-at-Risk with Estimation Risk,” *Journal of Business & Economic Statistics*, 28, 36–51.
- Foresi, S. and Peracchi, F. (1995), “The Conditional Distribution of Excess Returns: An Empirical Analysis,” *Journal of the American Statistical Association*, 90, 451–466.
- Fortin, N., Lemieux, T., and Firpo, S. (2011), “Decomposition Methods in Economics,” in *Handbook of Labor Economics*, eds. Ashenfelter, O. and Card, D., North Holland, Amsterdam, vol. 4 A, 1st ed., pp. 1–102.
- Francq, C. and Raïssi, H. (2007), “Multivariate Portmanteau Test for Autoregressive Models with Uncorrelated but Nonindependent Errors,” *Journal of Time Series Analysis*, 28, 454–470.
- Granger, C. W. J. and Andersen, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoeck und Ruprecht Göttingen.
- Granger, C. W. J. and Teräsvirta, T. (2010), *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
- Hallin, M. and Jurečková, J. (1999), “Optimal Tests for Autoregressive Models Based on Autoregression Rank Scores,” *The Annals of Statistics*, 27, 1385–1414.
- Hausman, J. A. (1978), “Specification Tests in Econometrics,” *Econometrica*, 46, 1251–1271.
- Hong, Y. and Lee, T. H. (2003), “Diagnostic Checking for Adequacy of Nonlinear Time Series Models,” *Econometric Theory*, 19, 1065–1121.
- Ibragimov, I. A. and Has’minskii, R. Z. (1981), *Statistical Estimation: Asymptotic Theory*, New York: Springer.

- Iqbal, F. and Mukherjee, K. (2012), “A Study of Value-at-Risk Based on M-Estimators of the Conditional Heteroscedastic Models,” *Journal of Forecasting*, 31, 377–390.
- Jeon, J. and Taylor, J. W. (2013), “Using CAViaR Models with Implied Volatility for Value-at-Risk Estimation,” *Journal of Forecasting*, 32, 62–74.
- Kheifets, I. (2014), “Specification Tests for Nonlinear Dynamic Models,” *Forthcoming in The Econometrics Journal*, 1–38.
- Koenker, R. and Machado, J. A. F. (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, 265–283.
- Koenker, R. and Xiao, Z. (2002), “Inference on the Quantile Regression Process,” *Econometrica*, 70, 1583–1612.
- (2006), “Quantile Autoregression,” *Journal of the American Statistical Association*, 101, 980–990.
- Kosorok, M. R. (2007), *Introduction to Empirical Processes and Semiparametric Inference*, Springer.
- Koul, H. and Mukherjee, K. (1994), “Regression Quantiles and Related Processes Under Long Range Dependent Errors,” *Journal of multivariate analysis*, 51, 318–337.
- Kreiss, J.-P. and Paparoditis, E. (2011), “Bootstrap Methods for Dependent Data: A Review,” *Journal of the Korean Statistical Society*, 40, 357–378.
- Künsch, H. R. (1989), “The Jackknife and the Bootstrap for General Stationary Observations,” *The Annals of Statistics*, 3, 1217–1241.

- Lahiri, S. N. (1999), “Theoretical Comparisons of Block Bootstrap Methods,” *Annals of Statistics*, 27, 386–404.
- Newey, W. K. (1985), “Maximum Likelihood Specification Testing and Conditional Moment Tests,” *Econometrica*, 53, 1047–1070.
- Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer, New York.
- Radulović, D. (1996), “The Bootstrap for Empirical Processes Based on Stationary Observations,” *Stochastic Processes and Their Applications*, 65, 259–279.
- Rio, E. (2000), *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*, vol. 31, Berlin: Springer-Verlag.
- Rothe, C. (2012), “Partial Distribution Policy Effects,” *Econometrica*, 80, 2269–2301.
- Rothe, C. and Wied, D. (2013), “Misspecification Testing in a Class of Conditional Distributional Models,” *Journal of the American Statistical Association*, 108, 314–324.
- Sakov, A. and Bickel, P. (2000), “An Edgeworth Expansion for the m Out of n Bootstrap Median,” *Statistics & Probability Letters*, 49, 217–223.
- Taamouti, A., Bouezmarni, T., and El Ghouch, A. (2014), “Nonparametric Estimation and Inference for Conditional Density Based Granger Causality Measures,” *Journal of Econometrics*, 180, 251–264.
- Van der Vaart, A. and Wellner, J. (2000), *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer.
- Van der Vaart, A. W. (2000), *Asymptotic Statistics*, vol. 3, Cambridge University Press.
- Weiss, A. (1991), “Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation,” *Econometric Theory*, 7, 46–68.

Whang, Y.-J. (2006), “Smoothed Empirical Likelihood Methods for Quantile Regression Models,” *Econometric Theory*, 22, 173–205.

White, H. (1982), “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50, 1–25.