A Specification Test of Dynamic Conditional Distributions

Victor Troster*
Department of Applied Economics, Universitat de les Illes Balears

February 11, 2016

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. The new specification test (i) is valid for general linear and nonlinear dynamic models under parameter estimation error, (ii) allows for 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. Finally, an empirical application to models of Value-at-Risk (VaR) highlights the benefits of our approach.

Keywords Block bootstrap; Distributional regression; Dynamic misspecification; Empirical processes; Quantile regression.

JEL Classification C14; C22; C52.

*I gratefully acknowledge Carlos Velasco and Abderrahim Taamouti for helpful comments and suggestions. I am also grateful to Miguel Delgado and Juan Carlos Escanciano for their closing read of the paper. I gratefully acknowledge Ministerio de Educación y Ciencia ECO2014-58991-C3-3-R for financial support. I also thank Jesús Gonzalo, Jesús Carro, and participants at presentations at the XXXIX Symposium of the Spanish Economic Association, V_t Workshop in Time Series Econometrics in Zaragoza for their insightful comments and discussions.

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. Our approach covers dynamic conditional distribution models possibly indexed by function-valued parameters. The difference between our approach and that taken elsewhere is motivated within the framework used by Corradi and Swanson (2006) and Rothe and Wied (2013). First, we generalize the approach of Rothe and Wied (2013) to testing the specification of dynamic conditional distribution models indexed by function-valued parameters in contexts with dependent data. This allows for a wide range of models that have been shown to be very useful for risk management and macroeconomic forecasting within a time series framework, such as the linear quantile auto-regressive, the CAViaR, and the distributional regression models. Second, we extend the validity of the block bootstrap for Kolmogorov-type conditional distribution tests proposed by Corradi and Swanson (2006) to the context of dynamic conditional distribution models indexed by function-valued parameters. Rather than analysing models indexed by finite-dimensional parameters as in Corradi and Swanson (2006), we derive a test statistic for conditional distribution models indexed by function-valued parameters that is valid under dynamic misspecification and parameter estimation error. 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.

Dynamic misspecification 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. 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 $Y_t|\mathcal{F}_{t-1}$. 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 $[0, 1]$. However, Bai (2003)'s test is inconsistent as it cannot detect lag order misspecification of a linear autoregressive model with elliptically distributed innovations (see e.g., Corradi and Swanson, 2006, 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. In this paper, we extend the approach of Corradi and Swanson (2006) to construct a specification test for time series models that takes into account dynamic misspecification and parameter error estimation effect, in a context of conditional distribution models indexed by function-valued parameters.

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 $F^{-1}(\tau|\theta_0, \boldsymbol{Y}_{t-p}) = \boldsymbol{Y}'_{t-p}\theta_0(\tau)$, for the quantile $\tau \in (0, 1)$, with $\boldsymbol{Y}_{t-p} = \{Y_{t-1}, \ldots, Y_{t-p}\} \in \mathcal{F}_{t-1}$, 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 $y \in \mathbb{R}$. The distributional regression approach uncovers higher-order multidimensional structure that cannot be found by modeling only the first two moments of the conditional distribution. This has important implications to forecasting excess stock market returns and finding an optimal portfolio (Foresi and Peracchi, 1995). Mean-variance analysis of excess stock market returns works only under special assumptions, like multivariate normality of asset returns or quadratic utility function of investors. In general, a precise definition of risk and an unambiguous ranking of portfolio strategies requires the entire distribution of future returns (Rothschild and Stiglitz, 1970). Besides, focusing on location - for example, on the conditional mean regression - may lead to overlook the impact of certain predictors of excess stock market returns, whose effect is mostly on high-order aspects of the conditional distribution. To the best of our knowledge, we are not aware of a framework to testing for the correct specification of distributional regression models 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. Koul and Stute (1999), Neumann and Paparoditis (2008), Bierens and Wang (2014), and Kheifets (2015), among others, 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) allows for 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.

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 presents Monte Carlo simulation results. In Section 6, we present an empirical application of our proposed test. Finally, Section 7 concludes the paper.

2. A GENERAL APPROACH TO TESTING DYNAMIC CONDITIONAL DISTRIBUTIONS

Suppose we observe a sample $\{(Y_t, X_t) \in \mathbb{R} \times \mathbb{R}^d, t = 1, \ldots, T\}$ 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

$$\mathcal{G} = \left\{ F\left(\bullet|\theta, \bullet\right) \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta) \right\}, \tag{1}$$

where $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ is a function-valued parameter, a class of mappings $\tau \mapsto \theta(\tau)$ such that $\theta(\tau) \in \Theta \subset \mathbb{R}^K$, for each $\tau \in \mathcal{T} \subset \mathbb{R}$. Focusing on the whole information set $\mathcal{F}_{t-1}$, the null hypothesis of correct specification could be written as $F(y|\mathcal{F}_{t-1}) = F(y|\theta_0, \mathcal{F}_{t-1})$, a.s. for all $y \in \mathbb{R}$ and for some $\theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$, against $\Pr[F(y|\mathcal{F}_{t-1}) \neq F(y|\theta, \mathcal{F}_{t-1})] > 0$ for some $y \in \mathbb{R}$ and for all $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$. 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 the under correct specification 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. Therefore, we want to test null hypotheses of correct specification of conditional distribution models of the form

$$\mathcal{H}_0: \; F(y|x) = F\left(y|\theta_0, x\right), \text{for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \text{ and for all } (y, x) \in \mathcal{W}, \tag{2}$$

versus

$$\mathcal{H}_A: \; F(y|x) \neq F\left(y|\theta, x\right), \text{for some } (y, x) \in \mathcal{W} \text{ and for all } \theta \in \mathcal{B}(\mathcal{T}, \Theta), \tag{3}$$

where $\mathcal{W}$ is the support of $W_t := (Y_t, X_t')'$. Under the null hypothesis of (2), the functional parameter $\theta_0(\bullet)$ is identified through 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 $\theta_0(\tau)$ solves

$$\Psi(\theta_0, \tau) := \mathrm{E}\left[\psi(W_t, \theta_0, \tau)\right] = 0, \tag{4}$$

where $\Psi(\theta, \tau)$ is a function $\Psi : \Theta \times \mathcal{T} \mapsto \mathbb{R}^K$ that fulfills some regularity conditions described in Section 3. As in Rothe and Wied (2013), we assume that under the null hypothesis, any $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ satisfying $F(y|x) = F(y|\theta, x)$ for all $(y, x) \in \mathcal{W}$ also satisfies $\theta(\tau) = \theta_0(\tau)$, for all $\tau \in \mathcal{T}$. Thus, $\theta_0(\tau)$ is uniquely identified through the moment conditions (4). In this paper, we assume that under $\mathcal{H}_A$ in equation (3), there exists a "pseudo"-true functional parameter $\theta_1(\tau)$ solving the moment conditions (4), for each $\tau \in \mathcal{T}$. Chernozhukov, Fernández-Val, and Melly (2013) developed theoretical results for Z-estimators of the moment conditions of (4) for i.i.d. data. Rothe and Wied (2013) show that a large class of empirically relevant specifications fits into this framework in a context with 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 (2), 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). We emphasize that the idea of comparing the unrestricted and restricted joint distribution functions, under the null and the alternative, is more than twenty years old in the specification testing literature. Stute (1997) and Andrews (1997) apply this idea in the context of testing specifications of parametric conditional expectation functions and conditional distribution functions, respectively, under i.i.d. data. In a time series context, Corradi and Swanson (2006) and Neumann and Paparoditis (2008) also apply this method to consistently check for the correct specification of dynamic conditional distributions indexed by finite-dimensional parameters. However, our null hypothesis tests the validity of a conditional distributional model indexed by function-valued parameters. 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 (2) can be equivalently restated as

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

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

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,

$$\hat{Z}_T(y, x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{Y_t \leq y\}\mathbb{1}\{X_t \leq x\}, \text{ for } (y, x) \in \mathbb{R}^{1+d}, \tag{5}$$

and

$$\hat{F}_T\big(y, x, \hat{\theta}_T\big) = \int F\big(y|\hat{\theta}_T, \bar{x}\big)\mathbb{1}\{\bar{x} \leq x\}d\hat{F}_X(\bar{x}), \text{ for } (y, x) \in \mathbb{R}^{1+d}, \tag{6}$$

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

$$\hat{F}_X(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{X_t \leq x\}, \text{ for } x \in \mathbb{R}^d. \tag{7}$$

Under $\mathcal{H}_0$ of (2), 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 (4):

$$\left\|\hat{\Psi}_T(\hat{\theta}_T, \tau)\right\|^2 \leq \inf_{\theta\in\Theta} \left\|\hat{\Psi}_T(\theta, \tau)\right\|^2 + \hat{u}(\tau)^2, \tag{8}$$

where $\|\hat{u}\|_{\mathcal{T}} = o_P(T^{-1/2})$, and $\|\bullet\|$ 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 $\hat{F}_T\big(y, x, \hat{\theta}_T\big)$, similar to the approach of Andrews (1997) and Rothe and Wied (2013). To this purpose we consider

$$D_T(y,x) = \frac{1}{T}\sum_{t=1}^{T}\Big(\mathbb{1}\{Y_t \le y\} - F\big(y|\hat{\theta}_T, X_t\big)\Big)\,\mathbb{1}\{X_t \le x\}, \tag{9}$$

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

$$S_T = T\int_{\mathcal{W}}\big(D_T(y,x)\big)^2 d\hat{Z}_T(y,x). \tag{10}$$

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 (2). 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 of (10).

3. ASYMPTOTIC THEORY

In this section, we derive the asymptotic distributions of our test statistic S_T under the null and alternative hypothesis. Let $\{Y_{Tt} : t \le T, T = 1,2,\ldots\}$ be a triangular array with stationary rows of random variables defined on a complete probability space $(\Omega, \mathcal{A}, P)$, where T is the sample size. Let $\mathcal{A}_T(m)$ be the σ-field generated by Y_{Tt} for $t \le m$, and $\mathcal{B}_T(m+d)$ be the σ-field generated by the variables Y_{Tt} for $t \ge m+d$. The sequence $\{Y_{Tt}\}$ is α-mixing if there is a sequence of numbers $\{\alpha(d)\}$ converging to zero for which

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

Let $\mathcal{W}$ be the support of $W_t := (Y_t, X_t')'$ and $\mathcal{T} \subset \mathbb{R}$. Our test statistic S_T in (10) 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 $\mathcal{M} := \{\Psi(\theta,\tau) : \theta \in \Theta, \tau \in \mathcal{T}\}$ is a permissible class of functions that has a finite and integrable envelope function $\mathbb{F}(\theta,\tau) := \sup_{\Psi\in\mathcal{M}} |\Psi(\theta,\tau)|$ and can be covered by a finite number of elements, not necessarily in $\mathcal{M}$ (see the Appendix for more details). Let $Pf = \int f(\theta,\tau)dP(\theta,\tau)$, for $f \in \mathcal{M}$. Finally, the $\mathcal{M}$ class of functions is assumed in this paper to form a so-called Vapnik-Chervonenkis (VC) class of functions (see Dudley, 1978, Pollard, 1984).

Throughout the paper we use "$\xrightarrow{d}$" and "$\Longrightarrow$" to denote convergence in distribution of random variables and weak convergence of stochastic processes, respectively. We write $Z_T \Longrightarrow Z$ 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, according to Alexander, 1987) for the metric induced by $||\bullet||$. Let $B_\varepsilon(\theta)$ be a closed ball of radius ε centered at θ. All limits are taken as $T \to \infty$, where T is the sample size. We maintain the following main assumptions to analyse the asymptotic behavior of our test statistic:

Assumption 1. *$\{(Y_{Tt}, X_{Tt}) : t \leq T, T = 1, 2, \ldots\}$ is an α-mixing triangular array with stationary rows, satisfying $E(|Y_{1,1}|^{2+\gamma}) < \infty$ 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. *Let $\mathcal{I}$ be an open set containing $\mathcal{T}$. The mapping $\Psi(\theta,\tau) : \Theta \times \mathcal{I} \mapsto \mathbb{R}^K$ is continuous and $\theta \mapsto \Psi(\theta,\tau)$ is the gradient of a convex function in θ for each $\tau \in \mathcal{T}$. Besides, $\partial\Psi(\theta,\tau)/\partial\theta := \dot{\Psi}_{\theta,\tau}$ exists at $(\theta_0(\tau),\tau)$ and is continuous at $(\theta_0(\tau),\tau)$, for each $\tau \in \mathcal{T}$, with $\inf_{\tau\in\mathcal{T}} \inf_{||h||=1} ||\dot{\Psi}_{\theta_0,\tau}h|| > c_0 > 0$.*

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

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). 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 (2),

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

where $(\mathbb{H}_1, \mathbb{H}_2)$ follow a tight mean zero Gaussian process.

(ii) Under the alternative hypothesis $\mathcal{H}_A$ in (3), there exists an $\varepsilon > 0$ such that

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

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

$$\mathcal{H}_{A,T} : F_T\left(y|x\right) = \left(1 - \frac{\delta}{\sqrt{T}}\right) F\left(y|\theta_0, x\right) + \left(\frac{\delta}{\sqrt{T}}\right) J(y|x), \tag{11}$$

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 (11), the conditional distribution implies a sequence of distribution functions* $F_T^A\big(y, x\big) = \int F_T\big(y|\bar{x}\big)\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ *that is contiguous to the distribution function* $F\big(y, x, \theta_0\big) = \int F(y|\theta_0, \bar{x})\mathbb{1}\{\bar{x} \leq x\}dF_X(\bar{x})$ *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 $F\left(\cdot|\theta_0, \cdot\right)$ and $J(\cdot|\cdot)$ have density functions $f\left(\cdot|\theta_0, \cdot\right)$ 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 $\mathit{\Psi}_J(\theta, \tau) := \mathrm{E}_J[\psi(W_t, \theta, \tau)]$ and $\mathit{\Psi}_F(\theta, \tau) := \mathrm{E}_F[\psi(W_t, \theta, \tau)]$, where $\mathrm{E}_J[\cdot]$ and $\mathrm{E}_F[\cdot]$ denote expectation w.r.t. $J = J(y|X_t)$ and $F = F(y|\theta_0, X_t)$, respectively in (11). We consider $\theta_0(\cdot)$ and $\theta_1(\cdot)$ as solutions to

$$\mathit{\Psi}_F(\theta_0, \tau) = 0, \text{ and} \tag{12}$$

$$\mathit{\Psi}_J(\theta_1, \tau) = 0, \tag{13}$$

for all $\tau \in \mathcal{T}$ respectively. Let $\partial \mathit{\Psi}_F(\theta_0, \tau)/\partial\theta$ satisfy Assumption 4 for the functional parameter θ_0 solving the moment conditions in (12). The following theorem sheds light on the asymptotic power of the test statistic S_T under a sequence of local alternatives satisfying (11).

Theorem 2. *Under the local alternative $\mathcal{H}_{A,T}$ in (11) and Assumptions 1-6*

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

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 $h(\tau) = [\partial\Psi_F(\theta_0,\tau)/\partial\theta]^{-1}\Psi_J(\theta_0,\tau)$.

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 (13), 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 $[\partial\Psi_F(\theta_0,\tau)/\partial\theta]^{-1}\Psi_J(\theta_0,\tau) = 0$, 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. If there were no dynamic misspecification under $\mathcal{H}_0$ of (2), we could apply a parametric bootstrap resampling method on $\hat{F}_T\big(y,x,\theta_0(\cdot)\big)$ to get asymptotic critical values under the null. However, in the presence of dynamic misspecification, $\hat{F}_T\big(y,x,\theta_0(\cdot)\big)$ is not independent and the covariance structure of the bootstrap statistic is not asymptotically valid. Thus, to solve this problem, we extend the block bootstrap approach proposed by Corradi and Swanson (2006) to test the specification of conditional distribution models indexed by function-valued parameters. We compare the empirical distribution of the resampled series, evaluated at the bootstrap estimator, with the empirical distribution of the actual series, evaluated at the estimator based on the actual data. This resampling method takes into account the parameter estimation error effect and allows for dynamic misspecification.

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 - $(X_t, X_{t+1}, \ldots, X_{t+l-\ell})$ - 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 consistent against fixed alternatives and allow for dynamic misspecification.

Block bootstrap approaches differ on whether the blocks are non-overlapping or overlapping and 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, $P^*, E^*, F^*, \ldots$ 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 ℓ be the length of the block, $\ell \in \mathbb{N}$, $\ell << T$, where T is the sample size. At each replication, we draw $\mathbf{b}$ blocks of length ℓ from the sample $W_t = (Y_t, X_t)$, with $\mathbf{b} = [T/\ell]$. For example, for some i with probability $1/(T - \ell - 1)$, the i-th block is $W_{i+1}, W_{i+2}, \ldots, W_{i+\ell}$. Thus, the set of starting indexes of the selected blocks is described by $I_1, I_2, \ldots, I_{\mathbf{b}}$ discrete i.i.d. uniform random variables taking values in the set $\{1, 2, \ldots T - \ell\}$.

2. Conditional on the sample, we join together the uniform i.i.d. random $\mathbf{b}$ blocks to form

a resampled series $W_1^*, W_2^*, \dots, W_\ell^*, W_{\ell+1}^*, \dots, W_T^*$, that can also be written as

$$\underbrace{W_{I_1}, W_{I_1+1}, \dots, W_{I_1+\ell-1}}_{1^{st}\text{ block}}, \underbrace{W_{I_2}, W_{I_2+1}, \dots, W_{I_2+\ell-1}}_{2^{nd}\text{ block}}, \dots, \underbrace{W_{I_\mathbf{b}}, W_{I_\mathbf{b}+1}, \dots, W_{I_\mathbf{b}+\ell-1}}_{\mathbf{b}^{th}\text{ block}}.$$

3. We denote $\hat{\theta}_T^*$ as the estimator obtained using the block bootstrap resampled series $\{W_t^* = (Y_t^*, X_t^*)\}$. Let $\hat{Z}_T^*(y,x)$ and $\hat{F}_T^*(y,x,\hat{\theta}_T^*)$ be the bootstrap equivalents of $\hat{Z}_T(y,x)$ and $\hat{F}_T(y,x,\hat{\theta}_T)$, respectively. Then we obtain the following 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 $\Pr^*[S_T^* \leq c_T^*(\alpha)] \geq 1-\alpha$, and this is estimated through Monte Carlo simulations. In contrast to the block bootstrap statistic of Corradi and Swanson (2006), we deal with the convergence of empirical process indexed by function-valued parameters. Thus, 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 $\beta(k)$ as

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

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 $\lim_{k\to\infty} \beta_k \to 0$. Then we impose the following assumption.

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*

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

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 $P^*(\bullet)$ 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 $W_1^*, \ldots, W_T^*$ be generated according to the block bootstrap with block size $\ell := \ell(T)$, with $\ell(T) \to \infty$ as $T \to \infty$, conditional on the data $W_1, \ldots, W_T$. 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) $\limsup_{k\to\infty} k^q \beta(k) < \infty$, *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 (2),

$$\Pr\left(S_T > c_T^*(\alpha)\right) \to \alpha.$$

(ii) Under the fixed alternative hypothesis $\mathcal{H}_A$ of (3),

$$\Pr\left(S_T > c_T^*(\alpha)\right) \to 1.$$

(iii) Under the local alternative $\mathcal{H}_{A,T}$ of (11),

$$\lim_{T\to\infty} \Pr\left(S_T > c_T^*(\alpha)\right) \geq \alpha,$$

where equality holds when $\Delta(y, x) \equiv 0$ a.e., with $\Delta(y, x)$ the non-trivial shift function defined in Theorem 2.

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, in terms of bias squared and variance of the block bootstrap approximation, 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. 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 :

DGP.1 (AR(1)) : $Y_t = 0.3Y_{t-1} + u_t,$

DGP.2 (AR(2)) : $Y_t = 0.3Y_{t-1} - 0.3Y_{t-2} + u_t,$

Power DGPs :

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

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

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

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

DGP.7 (Sign Autoregressive) : $Y_t = \mathbb{1}\{Y_{t-1} > 0\} - \mathbb{1}\{Y_{t-1} < 0\} + \sigma u_t,\ \sigma = 0.43,$

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 allow 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 is the SIGN model analyzed in Granger and Teräsvirta (1999), which is a first-order nonlinear autoregressive process that has the same autocorrelation function as an AR(1) process.

Our proposed test statistic checks the validity of the distributional regression model introduced by Foresi and Peracchi (1995), where 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:

$$F(y|x) = \Lambda(x'\theta(y)), \text{for some } \theta(y) \in \mathcal{B}(\mathbb{R}, \Theta) \subset \mathbb{R}^K \text{ and all } y \in \mathbb{R}, \tag{14}$$

where $\Lambda(\bullet)$ is a known strictly increasing link function (e.g., the logistic or standard normal distribution), and $\theta(\bullet)$ is a functional parameter taking values in $\mathcal{B}(\mathbb{R}, \Theta)$. One can also run a distributional regression model of Y_t conditional on its lagged values:

$$F(y|Y_{t-1}) = \Lambda\left(Y_{t-1}'\theta(y)\right), \text{for some } \theta(y) \in \mathcal{B}(\mathbb{R}, \Theta) \subset \mathbb{R}^K \text{ and all } y \in \mathbb{R}. \tag{15}$$

We test the specification of a distributional regression model in the form of (15). The data are generated as in DGPs 1-7, 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}:

$$\mathcal{H}_0^{DR} : F\left(y|Y_{t-1}\right) = \Lambda\left(Y_{t-1}'\theta(y)\right), \text{ a.s.,} \tag{16}$$

where $\Lambda(\boldsymbol{\cdot})$ is specified as a logistic distribution function. For all the experiments, we consider the empirical rejection frequencies for 5% 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 $\ell = \{3, 4, 6\}$, 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

$$EV := \int\int \left|\left(\mathbb{1}\left(Y_t - m(X_t, \hat{\theta}_T(\tau)) \leq 0\right) - \tau_j\right)\exp(\boldsymbol{i}\boldsymbol{x}'X_t)\right|^2 dW(\boldsymbol{x})d\Phi(\alpha), \tag{17}$$

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 (17) 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 $b = [kT^{(2/5)}]$, for $k = 3$ and 4, following the suggestion of Sakov and Bickel (2000).

Table 1 reports 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.

Table 1. Empirical rejection frequencies for 5% specification tests

	S_T			EV	
	$\ell = 3$	$\ell = 4$	$\ell = 6$	$b = 18$	$b = 25$
$T = 100$					
DGP.1	0.030	0.060	0.036	0.075	0.077
DGP.2	0.030	0.040	0.052	0.091	0.084
DGP.3	0.960	0.990	0.920	0.888	0.847
DGP.4	0.962	1.000	1.000	0.997	0.984
DGP.5	0.912	0.864	0.900	0.944	0.913
DGP.6	1.000	1.000	1.000	1.000	1.000
DGP.7	0.630	0.598	0.634	0.608	0.606
$T = 300$					
DGP.1	0.043	0.053	0.031	0.061	0.057
DGP.2	0.053	0.067	0.049	0.092	0.074
DGP.3	1.000	1.000	1.000	1.000	1.000
DGP.4	1.000	1.000	1.000	1.000	1.000
DGP.5	1.000	1.000	1.000	1.000	1.000
DGP.6	1.000	1.000	1.000	1.000	1.000
DGP.7	0.949	0.942	0.949	0.942	0.942

NOTE: S_T denotes our proposed test statistic with $B = 399$ bootstrap replications, for block lengths $\ell = \{3, 4, 6\}$. EV denotes the subsampling specification test of Escanciano and Velasco (2010) with subsample sizes $b = \{18, 25\}$. We use 1,000 Monte Carlo repetitions based on the DGPs 1-7 described above.

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 DPG.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 $T = 300$ in most of the DGPs.

Table 2 presents the empirical rejection frequencies for the distributional regression specification of the S_T test, related to the DGPs 1-7, for sample sizes $T = 100$ and $T = 300$. For a distributional regression model, the empirical size of the S_T test is somewhat above the nominal level when the sample size is low, but it is close to the nominal level when $T = 300$. The results for DGP.2 show that our test allows for dynamic misspecification. Our test statistic is also powerful against misspecifications in the distributional regression, as the power for testing $\mathcal{H}_0^{DR}$ in (16) is high for DGPs 3-7 with sample sizes $T = 100$ and $T = 300$ (Table 2). In sum, our proposed test seems to perform quite well in finite samples.

Table 2. Empirical rejection frequencies for 5% S_T test - distributional regression specification

	$T = 100$			$T = 300$		
	$\ell = 3$	$\ell = 4$	$\ell = 6$	$\ell = 3$	$\ell = 4$	$\ell = 6$
DGP.1	0.100	0.070	0.095	0.050	0.057	0.052
DGP.2	0.078	0.089	0.082	0.047	0.067	0.053
DGP.3	0.990	1.000	1.000	1.000	1.000	1.000
DGP.4	0.984	0.990	1.000	1.000	1.000	1.000
DGP.5	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
DGP.7	0.455	0.475	0.471	0.952	0.920	0.940

NOTE: S_T denotes our proposed test statistic with $B = 399$ bootstrap replications with block lengths $\ell = \{3, 4, 6\}$. The null hypothesis $\mathcal{H}_0^{DR}$ test the specification of a distributional regression model specified in (16). We use 1,000 Monte Carlo repetitions based on the DGPs 1-7 described above.

6. 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 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 $(Y_{t-1}, \dots, Y_{t-p})$. Table 3 presents the summary statistics of the series. Both log-returns series are highly leptokurtic and present autocorrelation.

We test the hypothesis $\mathcal{H}_0$: the VaR of the log-return Y_t follows an AR-GARCH process. We choose AR-GARCH specifications as GARCH models have provided appropriately specified the VaR of stock returns in the financial literature (Escanciano and Olmo, 2010). We entertain the following 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 (2015). 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 (17), with two different subsample sizes $b = [kT^{2/5}]$ for $k = 3$ and $k = 4$.

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.

Table 4 shows the p-values of the specification tests for all the VaR models for the full sample from January 2003 to June 2014. 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.

For robustness, we perform the same tests to these models using only one year of data, from June 26^{th}, 2013 to June 9^{th}, 2014. Table 5 displays the results for this period. While the EV test of Escanciano and Velasco (2010) rejects all models, our test S_T does not reject most of the models at the 1% significance level for the DAX daily returns series.

Moreover, the AR(1)-GARCH(1,1) model with Student-t5 distribution has obtained the highest p-value and is the only model that is not rejected at the 10% significance level. For the FTSE-100 returns, our test does not reject all models at the 1% significance level, while the the EV test of Escanciano and Velasco (2010) does not reject only the ARCH(1,1) with Student-t5 distribution at the 1% significance level. Thus, the empirical application shows the ability of our test to detect possibly misspecified conditional distribution models when we have a small sample size. 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: January 6^{th}, 2003-June 9^{th}, 2014

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

Table 5. Specification tests p-values of VaR models of DAX and FTSE-100 returns: June 26^{th}, 2013-June 9^{th}, 2014

	$S_{T,6}$	$S_{T,8}$	$S_{T,16}$	EV(b=98)	EV(b=122)
DAX					
GARCH(1,1)	0.028	0.031	0.035	0.000	0.000
GARCH(1,1) - t5	0.040	0.030	0.033	0.000	0.000
AR(1)-GARCH(1,1)	0.018	0.029	0.001	0.000	0.000
AR(1)-GARCH(1,1) - t5	0.175	0.167	0.159	0.000	0.000
AR(2)-GARCH(2,2)	0.059	0.050	0.044	0.000	0.000
E-GARCH(1,1)	0.034	0.033	0.044	0.001	0.001
FTSE-100					
GARCH(1,1)	0.634	0.608	0.614	0.000	0.000
GARCH(1,1) - t5	0.622	0.582	0.602	0.010	0.002
AR(1)-GARCH(1,1)	0.451	0.443	0.465	0.009	0.004
AR(1)-GARCH(1,1) - t5	0.289	0.302	0.319	0.010	0.007
AR(2)-GARCH(2,2)	0.416	0.426	0.440	0.000	0.000
E-GARCH(1,1)	0.543	0.567	0.536	0.001	0.001

NOTE: $S_{T,\ell}$ is the S_T test with block length $\ell = \{3, 4, 6\}$. We denote EV as the specification test of Escanciano and Velasco (2010), with sub-sample size b.

7. 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.

A possible direction for future work is to extend this study to test Granger-causality in distribution, as in Taamouti et al. (2014). 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, which cannot assess tail causal relations or nonlinear causalities. Our proposed approach allows us to evaluate nonlinear causalities, causal relations in conditional quantiles, and Granger-causality in distribution. One could also extend our approach to the class of multivariate models, providing specification tests for vector autoregressive and multivariate linear and non-linear models, see e.g. Francq and Raïssi (2007) and Escanciano, Lobato, and Zhu (2013).

APPENDIX

A.1 Tools

In this section, we introduce some auxiliary results. Let $\mathcal{M}$ be a permissible class of functions such that it can be indexed by some set $\mathcal{T}$, i.e., $\mathcal{M} = \{\Psi(\cdot, \tau) : \tau \in \mathcal{T}\}$, in such a way that the following holds: (i) $\mathcal{T}$ is a Suslin metric space (a Hausdorff topological space that is the continuous image of a Polish space) with Borel σ-field $\mathcal{B}(\mathcal{T})$, and (ii) $\Psi(\cdot, \cdot)$ is $\times\mathcal{B}(\mathcal{T})$-measurable function from $\mathbb{R}^K \times \mathcal{T}$ to $\mathbb{R}$ (see Kosorok, 2007, Section 11.6). Let $Pf = \int f(\theta, \tau) dP(\theta, \tau)$, for $f \in \mathcal{M}$. 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 $\mathcal{M}$ class of functions 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

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

for all $\varepsilon > 0$ and all probability measures P (see Lemmas II.25 and II.32 in Pollard, 1984). First, we derive a Central Limit Theorem for strong mixing processes for the empirical distribution, $\hat{Z}_T(y,x)$, under the null and the alternative hypothesis.

Lemma A.1. *Given Assumption 1, under $\mathcal{H}_0$ of (2) or $\mathcal{H}_A$ of (3),*

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

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

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

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). □

In the paper, we have a functional parameter $\tau \mapsto \theta(\tau)$, where $\tau \in \mathcal{T}$ and $\theta(\tau) \in \mathcal{B}(\mathcal{T}, \Theta)$, and the true value $\theta_0(\tau)$ solves the moment equations $\Psi(\theta,\tau) = 0$. The following lemma establishes a functional delta method for the empirical analog $\hat{\Psi}_T(\theta,\tau)$ of the previous moment equations and for the estimator of the functional parameter, $\hat{\theta}_T(\bullet)$.

Lemma A.2. *Given Assumptions 1-5, under $\mathcal{H}_0$ of (2) or $\mathcal{H}_A$ of (3), we have*

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

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

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

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

Proof. First, by Lemma E.1 in Chernozhukov et al. (2013), Assumptions 2-5 imply that (i) the inverse of $\Psi(\bullet,\tau)$ defined as $\Psi^{-1}(x,\tau) := \{\theta\in\Theta : \Psi(\theta,\tau) = x\}$ is continuous at $x = 0$ uniformly in $\tau\in\mathcal{T}$ with respect to the Hausdorff distance, (ii) there exists $\dot{\Psi}_{\theta_0,\tau}$ such that $\lim_{t\to 0}\sup_{\tau\in\mathcal{T},||h||=1}|t^{-1}[\Psi(\theta_0(\tau)+th,\tau)-\Psi(\theta_0(\tau),\tau)]-\dot{\Psi}_{\theta_0,\tau}h| = 0$, where $\inf_{\tau\in\mathcal{T}}\inf_{\|h\|=1}||\dot{\Psi}_{\theta_0,\tau}h|| > 0$, (iii) the maps $\tau\mapsto\theta_0(\tau)$ and $\tau\mapsto\dot{\Psi}_{\theta_0,\tau}$ are continuous, and (iv) the mapping $\tau\mapsto\theta_0(\tau)$ is continuously differentiable. Under the previous conditions, Lemma E.2 in Chernozhukov et al. (2013) holds, and the process $r_T(\theta,\tau)$ weakly converges to $\tilde{\mathbb{H}}_2(\theta,\tau)$ in $\ell^\infty(\mathcal{T}\times\Theta)$ and the map $\theta\mapsto\Psi(\theta,\bullet)$ is Hadamard differentiable at θ_0 with continuously invertible derivative $\dot{\Psi}_{\theta_0,\bullet}$. By Hadamard differentiability of the map $\theta\mapsto\Psi(\theta,\bullet)$, it follows the weak convergence of the process $\sqrt{T}(\hat{\theta}_T(\bullet)-\theta_0(\bullet))$ in $\ell^\infty(\mathcal{T})$. □

Lemma A.3. *Given Assumptions 1-5, under* $\mathcal{H}_0$ *of (2) or* $\mathcal{H}_A$ *of (3), 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, $\sqrt{T}(\hat{\theta}_T(\bullet)-\theta_0(\bullet)) \implies -\dot{\Psi}^{-1}_{\theta_0,\bullet}[\tilde{\mathbb{H}}_2(\theta_0(\bullet),\bullet)]$ in $\ell^\infty(\mathcal{T})$, where $\tilde{\mathbb{H}}_2$ is a Gaussian process in $\ell^\infty(\mathcal{T}\times\Theta)$. By the functional delta method, we can rewrite $v_T^{\theta_0}(y,x)$ as

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

By the Hadamard differentiability of the map $\theta\mapsto F(\bullet|\theta(\bullet),\bullet)$ in Assumption 5, we can apply the functional delta method, for fixed y and x, as follows:

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

Similarly to Lemma A.1, under $\mathcal{H}_0$ of (2) or $\mathcal{H}_A$ of (3), given the strong mixing condition of Assumption 1, $\sqrt{T}(\hat{F}_X(\bar{x}) - F_X(\bar{x}))$ weakly converges to a tight mean zero Gaussian process. Now, 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 $(\Gamma, \Pi) \mapsto \int \Gamma(\bullet, x) d\Pi(x)$ - with $\Gamma(\bullet, x) = \mathbb{1}\{\bullet \leq x\} F(\bullet|x)$ and $\Pi = F_X(\bullet)$ - is well defined and Hadamard differentiable at (Γ, Π). Given the Hadamard differentiability of the mapping $(\Gamma, \Pi) \mapsto \int \Gamma(\bullet, x) d\Pi(x)$, the result follows from an application of the functional delta method, where the Gaussian process $\mathbb{H}_2$ is given by

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

where $\mathbb{H}_1$ is the same tight mean zero Gaussian process described in Lemma A.1. □

Lemma A.4. *Under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (11) and Assumptions 1-6,*

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

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

where $F_T^A(y,x) = \int F_T\big(y|\bar{x}\big) \mathbb{1}\{\bar{x} \leq x\} dF_X(\bar{x})$, $\Psi_{F_T}(\theta,\tau) = \mathrm{E}_{F_T}\left[\psi(W_t,\theta,\tau)\right]$, and $(\mathbb{H}_1, \tilde{\mathbb{H}}_2)$ 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 $F\big(y,x,\theta_0\big)$, then the convergence of the process $v_T^{\theta_0}(y,x) := \sqrt{T}(\hat{F}_T(y,x,\hat{\theta}_T) - F_T(y,x,\theta_0))$ on Lemma A.3 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 (11) and Assumptions 1-6, $F_T(y|X_t)$ of (11) is a linear combination of two measures that are VC class with a square integrable envelope. From the convergence of the process $\sqrt{T}(\hat{\Psi}_T(\theta,\tau) - \Psi(\theta,\tau))$ in Lemma A.2 and an application of Lemma 2.8.7 in Van der Vaart and Wellner (2000), we have that $\sqrt{T}(\hat{\Psi}_T(\theta,\tau) - \Psi_{F_T}(\theta,\tau))$ weakly converges to $\tilde{\mathbb{H}}_2(\theta,\tau)$ in $\ell^\infty(\mathcal{T} \times \Theta)$. □

We define weak convergence conditional on the data in probability ($\overset{\mathbb{P}}{\underset{M}{\Longrightarrow}}$ -convergence) in the Hoffmann-Jørgensen sense, i.e., $\hat{X}_n \overset{\mathbb{P}}{\underset{M}{\Longrightarrow}} X$ in a metric space $\mathbb{D}$ denotes conditional bootstrap convergence in probability under $\mathbb{P}$, that is, $\sup_{f\in\ell^\infty(\mathcal{H})} |E_M f(\hat{X}_n) - Ef(X_n)| \xrightarrow{\mathbb{P}} 0$. The subscript M denotes taking the expectation conditional on the data. The following lemma derives the convergence of the block bootstrap of empirical process for dependent observations.

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 $W_1, \ldots, W_T$, let $W_1^*, \ldots, W_T^*$ be generated according to the block bootstrap with block length $\ell := \ell(T)$, with $\ell(T) \to \infty$ as $T \to \infty$. Let $v_T^*(y,x) := \sqrt{T}(\hat{Z}_T^*(y,x) - \hat{Z}_T(y,x))$ be the block bootstrap version of the empirical process $v_T(y,x) = \sqrt{T}(\hat{Z}_T(y,x) - F_{YX}(y,x))$. Suppose that*

$$\limsup_{k\to\infty} k^q \beta(k) < \infty \text{ for some } q > p/(p-2) \text{ and that } P^*(\mathbb{F})^p < \infty \text{ for some } p > 2.$$

Assume that the block length $\ell(T)$ also satisfies

$$\ell(T) = O(T^\rho) \text{ for some } 0 < \rho < (p-2)/[2(p-1)].$$

Then

$$v_T^*(y,x) \underset{M}{\overset{\mathbb{P}}{\Longrightarrow}} \mathbb{H}_1(y,x), \text{ in } \ell^\infty(\mathcal{W}),$$

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

Proof. This result follows directly from an application of Theorem 1 in Radulović (1996) or Theorem 11.26 in Kosorok (2007), slightly modified to address measurability. □

Lemma A.6. *Under Assumptions 2-7, under $\mathcal{H}_0$ of (2), or $\mathcal{H}_A$ of (3), or under the local alternative $\mathcal{H}_{A,T}$ of (11),*

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

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 bootstrap convergence result of 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. □

A.2 Proofs

Proof of Theorem 1. To prove part (i), we consider the empirical processes $v_T(y,x) = \sqrt{T}(\hat{Z}_T(y,x) - F_{YX}(y,x))$ and $v_T^{\theta_0}(y,x) = \sqrt{T}(\hat{F}_T(y,x,\hat{\theta}_T) - F_T(y,x,\theta_0))$ defined in Lemma A.1 and Lemma A.3, respectively. Under $\mathcal{H}_0$ of (2), $F_{YX}(y,x) \equiv F(y,x,\theta_0)$, and we have

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

By Lemma A.1, we have $\sqrt{T}(\hat{Z}_T(y,x) - F_{YX}(y,x)) \implies \mathbb{H}_1(y,x)$ that is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{W})$. Then

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

By 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 an application of the continuous mapping theorem.

In part (ii), under the alternative hypothesis $\mathcal{H}_A$ of (3), $F_{YX}(y,x) \neq F(y,x,\theta_1)$ for some $(y,x) \in \mathcal{W}$ and for all $\theta_1 \in \mathcal{B}(\mathcal{T},\Theta)$. Now the process $v_T^{\theta_0}(y,x)$ becomes $v_T^{\theta_0}(y,x) =$

$\sqrt{T}(\hat{F}_T(y,x,\hat{\theta}_T) - F_T(y,x,\theta_1))$. Then

$$S_T = T\int\left(\hat{Z}_T(y,x) - \hat{F}_T(y,x,\hat{\theta}_T) \pm F_{YX}(y,x) \pm F(y,x,\theta_1)\right)^2 dF_{YX}(y,x)$$

$$= \int\left(v_T(y,x) - v_T^{\theta_0}(y,x) + \sqrt{T}\left(F_{YX}(y,x) - F(y,x,\theta_1)\right)\right)^2 dF_{YX}(y,x) + o_P(1).$$

By 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 $\ell^\infty(\mathcal{W}\times\mathcal{W})$. Therefore, for any fixed constant $\varepsilon > 0$, $\lim_{T\to\infty}\Pr(S_T > \varepsilon) = 1$ and the result follows. □

Proof of Theorem 2. Under the local alternative $\mathcal{H}_{A,T}$ in (11), consider the empirical processes

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

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

where $\Psi_F(\theta,\tau) := \mathrm{E}_F[\psi(W_t,\theta,\tau)]$ as defined in (12). Then

$$\begin{aligned} v_T^1(y,x) =& \sqrt{T}\left(\hat{Z}_T(y,x) - \int F(y|\theta_0,\bar{x})\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right) \\ =& \sqrt{T}\left(\hat{Z}_T(y,x) - \int\left[F_T(y|\bar{x}) + \frac{\delta}{\sqrt{T}}\left(F(y|\theta_0,\bar{x}) - J(y|\bar{x})\right)\right]\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right) \\ =& \sqrt{T}\left(\hat{Z}_T(y,x) - F_T^A(y,x) + \frac{\delta}{\sqrt{T}}\int\left(J(y|\bar{x}) - F(y|\theta_0,\bar{x})\right)\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right). \end{aligned}$$

Thus, it follows from Lemma A.4 that

$$v_T^1(y,x) \implies \mathbb{H}_1(y,x) + \delta\int\left(J(y|\bar{x}) - F(y|\theta_0,\bar{x})\right)\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x}),$$

where $\mathbb{H}_1$ is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{W})$ defined in Lemma A.1. Now we have that

$$
\begin{aligned}
r_T^1(\theta,\tau) =& \sqrt{T}\big(\hat{\Psi}_T(\theta,\tau) - \mathrm{E}_F[\psi\,(W_t,\theta,\tau)]\big) \\
=& \sqrt{T}\big(\hat{\Psi}_T(\theta,\tau) - \{\mathrm{E}_{F_T}[\psi\,(W_t,\theta,\tau)] + \delta\,\mathrm{E}_F[\psi(W_t,\theta,\tau)] - \delta\,\mathrm{E}_J[\psi(W_t,\theta,\tau)]\}\,\big) \\
=& \sqrt{T}\big(\hat{\Psi}_T(\theta,\tau) - \Psi_{F_T}(\theta,\tau) + \delta\,[\mathrm{E}_J[\psi(W_t,\theta,\tau)] - \mathrm{E}_F[\psi(W_t,\theta,\tau)]]\,\big),
\end{aligned}
$$

where $\Psi_J(\theta,\tau) := \mathrm{E}_J[\psi(W_t,\theta,\tau)]$ as defined in (13). Thus, by Lemma A.4, we have

$$
r_T^1(\theta,\tau) \implies \tilde{\mathbb{H}}_2(\theta,\tau) + \delta\big[\,\mathrm{E}_J[\psi(W_t,\theta,\tau)] - \mathrm{E}_F[\psi(W_t,\theta,\tau)]\big],
$$

where $\tilde{\mathbb{H}}_2$ is a tight mean zero Gaussian process in $\ell^\infty(\mathcal{T}\times\Theta)$ defined in Lemma A.2. Now, we consider the empirical process $v_T^{1\theta_0}(y,x)$

$$
v_T^{1\theta_0}(y,x) = \sqrt{T}\left(\int F(y|\hat{\theta}_T,\bar{x})\mathbb{1}\{\bar{x}\le x\}d\hat{F}_X(\bar{x}) - \int F(y|\theta_0,\bar{x})\mathbb{1}\{\bar{x}\le x\}dF_X(\bar{x})\right).
$$

Thus, by Lemma A.3, we have that

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

with $h(\tau) = [\partial\Psi_F(\theta_0,\tau)/\partial\theta]^{-1}\Psi_J(\theta_0,\tau)$. Therefore, under $\mathcal{H}_{A,T}$ of (11), we have

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

then the result follows from the continuous mapping theorem. □

Proof of Theorem 3. For part (i), by Lemma A.6, $\hat{c}_T^*(\alpha) = c(\alpha) + o_P(1)$, where $c(\alpha)$ satisfies $\Pr(S_T > c(\alpha)) = \alpha + o(1)$. Then as $T \to \infty$, $\Pr(S_T > \hat{c}_T^*(\alpha)) = \alpha + o(1)$. For part (ii), there exists a fixed constant $C > 0$ such that

$$\Pr(S_T \le \hat{c}_T^*(\alpha)) = \Pr(S_T \le \hat{c}_T^*(\alpha), S_T \le C) + \Pr(S_T \le \hat{c}_T^*(\alpha), S_T > C)$$
$$\le \Pr(S_T \le C) + \Pr(\hat{c}_T^*(\alpha) > C)$$
$$\le o(1) + \varepsilon + o(1),$$

where the first element of the third line follows from Theorem 1 - $\Pr(S_T \le C) = o(1)$ - and the rest of the third line is due to Lemmas A.5-A.6, that imply the block bootstrap critical value $\hat{c}_T^*(\alpha)$ is bounded in probability under fixed alternatives, i.e., 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

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

Thus, under a sequence of local alternatives, we have $\Pr\left(S_T > c(\alpha)\right) \ge \alpha + o(1)$. Under Assumption 6, the conditional distribution under a local alternative $F_T(\bullet|\bullet)$ implies a sequence of distribution functions $Z_T(y, x)$ that is contiguous to the distribution function $F\big(y, x, \theta_0\big)$ given by $\int F(y|\theta_0, \bar{x}))\mathbb{1}\{\bar{x} \le x\} dF_X(\bar{x})$, under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (11). Since contiguity preserves convergence in probability to constants, under the sequence of local alternatives $\mathcal{H}_{A,T}$ of (11) we have

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

$$= \Pr\left(\int (\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x) + \Delta(y,x))^2 \, dF_{YX}(y,x) \geq c(\alpha)\right) + o(1)$$

$$\geq \Pr\left(\int (\mathbb{H}_1(y,x) - \mathbb{H}_2(y,x))^2 \, dF_{YX}(y,x) \geq c(\alpha)\right)$$

$$= \Pr\left(S_T \geq c(\alpha)\right) \geq \alpha,$$

where equality holds when $\Delta(y,x) \equiv 0$ a.e., with $\Delta(y,x)$ the non-trivial shift function defined in Theorem 2. □

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.

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.

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. (1999), "A simple Nonlinear Time Series Model with Misleading Linear Properties," *Economics Letters*, 62, 161–165.

— (2010), *Modelling Nonlinear Economic Relationships*, Oxford University Press, Oxford.

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. L. (2015), "Specification Tests for Nonlinear Dynamic Models," *The Econometrics Journal*, 18, 67–94.

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 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. L. and Stute, W. (1999), "Nonparametric Model Checks for Time Series," *Annals of Statistics*, 27, 204–236.

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.

Neumann, M. H. and Paparoditis, E. (2008), "Goodness-of-Fit Tests for Markovian Time Series Models: Central Limit Theory and Bootstrap Approximations," *Bernoulli*, 14, 14–46.

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. and Wied, D. (2013), "Misspecification Testing in a Class of Conditional Distributional Models," *Journal of the American Statistical Association*, 108, 314–324.

Rothschild, M. and Stiglitz, J. E. (1970), "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2, 225–243.

Sakov, A. and Bickel, P. (2000), "An Edgeworth Expansion for the m Out of n Bootstrap Median," *Statistics & Probability Letters*, 49, 217–223.

Stute, W. (1997), "Nonparametric Model Checks for Regression," *The Annals of Statistics*, 25, 613–641.

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.

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.