SFB 823

Misspecification testing in a class of conditional distributional models

Christoph Rothe, Dominik Wied

Discussion Paper

Nr. 3/2011

Misspecification Testing in a Class of Conditional Distributional Models

Christoph Rothe and Dominik Wied*

Toulouse School of Economics and TU Dortmund

This Version: January 27, 2011

Abstract

We propose a specification test for a wide range of parametric models for conditional distribution function of an outcome variable given a vector of covariates. The test is based on the Cramer-von Mises distance between an unrestricted estimate of the joint distribution function of the data, and an restricted estimate that imposes the structure implied by the model. The procedure is straightforward to implement, is consistent against fixed alternatives, has non-trivial power against local deviations from the null hypothesis of order $n^{-1/2}$, and does not require the choice of smoothing parameters. We also provide an empirical application using data on wages in the US.

JEL Classification: C12, C14, C31, C52, J31

Keywords: Cramer-von Mises Distance, Quantile Regression, Distributional Regression, Bootstrap

*Christoph Rothe, Toulouse School of Economics, 21 Allée de Brienne, F-31000 Toulouse, France. Email: rothe@cict.fr. Dominik Wied, TU Dortmund, Fakultät Statistik, CDI-Gebäude, D-44221 Dortmund, Germany. Email: wied@statistik.tu-dortmund.de. Financial support by Deutsche Forschungsgemeinschaft (SFB 823) is gratefully acknowledged. We would like to thank Blaise Melly for helpful comments. The usual disclaimer applies.

1. Introduction

In this paper, we propose an omnibus specification test for parametric models for the conditional distribution function of a random variable $Y \in \mathbb{R}$ given a random vector of covariates $X \in \mathbb{R}^k$ in a setting with independent observations. The main innovation of our method is that we allow the unknown parameters to be function-valued. Specifically, our interest is in testing the null hypothesis that

$$\Pr(Y \leq y|X = x) = F(y|x, \theta) \text{ for all } (y, x) \in supp(Y, X), \tag{1.1}$$

where F is a function known up to a possibly function-valued parameter $\theta = \theta(\cdot)$, for which a consistent estimator is available. The alternative hypothesis is that equation (1.1) is violated for at least one pair of values (y, x) contained in the support of (Y, X). This setting covers a wide range of conditional distributional models that are of great importance in empirical applications. For example, the linear quantile regression model (Koenker and Bassett, 1978; Koenker, 2005) implies a linear structure for the inverse of the conditional CDF, namely that $F^{-1}(\tau|x, \theta) = x'\theta(\tau)$, where the functional parameter $\theta : (0, 1) \mapsto \mathbb{R}^k$ is strictly increasing in each of its components; under the distributional regression model (Foresi and Peracchi, 1995), the conditional CDF is specified as $F(y|x, \theta) = \Lambda(x'\theta(y))$, where Λ is a known strictly increasing link function, e.g. the logistic or standard normal distribution function. In both cases, the models are such that the conditional distribution of Y given X cannot be fully described through a finite dimensional parameter.

Our test is an extension of the method proposed by Andrews (1997) in the context of parametric models indexed by finite dimensional parameters. The basic idea is to compare an unrestricted estimate of the joint distribution function of Y and X to a restricted estimate that imposes the structure implied by the conditional distributional model. Our test statistic is a Cramer-von Mises type measure of distance between the restricted and unrestricted estimate, and is therefore called a *Generalized Conditional Cramer-von Mises*

(GCCM) test.[1] We reject the null hypothesis that the parametric model is correctly specified whenever this distance is "too large". Since our test statistic is not asymptotically pivotal, critical values cannot be tabulated, but can be obtained via the bootstrap. While our test is thus computationally somewhat involved, it is straightforward to implement and has a number of attractive theoretical properties: It is consistent against all fixed alternatives, has non-trivial power against local deviations from the null hypothesis of order $n^{-1/2}$ (where n denotes the sample size), and does not require the choice of smoothing parameters.

The correct specification of conditional distributional models of the type considered in this paper is critical in many areas of applied statistics. In economics, there is an extensive literature that uses such specifications in the context of determining the causes of differences in the distribution of wages in different time periods, or different subgroups of a particular population. See e.g. Machado and Mata (2005), Melly (2005), Albrecht, Van Vuuren, and Vroman (2009), Chernozhukov, Fernandez-Val, and Melly (2009) or Rothe (2010b), and Firpo, Fortin, and Lemieux (2010) for an extensive survey. From a statistical point of view, these methods first obtain an estimate of the conditional CDF of Y given X. In a second step, this function is integrated with respect to another CDF, whose exact form depends on the particular application, yielding a new univariate distribution function. As a final step, features of this new distribution function, such as its mean or quantiles, are estimated. Our concern is the implementation of the first step of this procedure. For example, the Machado and Mata (2005) decomposition procedure relies on the assumption that the *entire* conditional distribution of wages given observable individual characteristics can be described by a linear quantile regression model. If this assumption is violated, the method can potentially lead to inappropriate conclusions (see Rothe, 2010a, for some simulation evidence). From a practitioner's point of view, misspecification is a serious concern in this context, as the conditional quantiles of the

[1] We choose to work with the CvM distance instead of other measures, such as e.g. the Kolmogorov distance used by Andrews (1997), since the resulting test statistics turned out to have substantially better power properties in simulations. This is in line with classical findings on the power of specification tests, e.g. Stephens (1974).

wage distribution are e.g. known to be extremely flat in the vicinity of the legal minimum wage, and might thus not be described adequately by a linear specification in this region (Chernozhukov, Fernandez-Val, and Melly, 2009).

As an additional contribution, this paper provides some further empirical evidence on this point: using US data from the Current Population Survey, we show that typical specifications of linear quantile regressions containing a rich set of covariates are frequently rejected by our GCCM test even for small and moderate sample sizes, whereas other classes of models are not. The test we propose in this paper should thus be of great practical importance for researchers applying decomposition methods in practice.

While there exists an extensive literature on specification testing in parametric models for the conditional expectation function (see e.g. Bierens (1990), Härdle and Mammen (1993), Bierens and Ploberger (1997), Horowitz and Spokoiny (2001)), the related problem of testing the validity of a model for the entire conditional distribution function has received much less attention. Andrews (1997) proposes a test for conditional distributional models indexed by finite-dimensional parameters, such as e.g. Generalized Linear Models. As noted above, our GCCM test is an extension of his approach to a substantially more general class of models. While our testing procedure is conceptionally similar, we allow for parametric specifications with possibly function-valued parameters, and thus require somewhat different arguments for our theoretical analysis.

Since our framework covers quantile regression as a special case, this paper also contributes to a growing literature on specification testing in this context. However, while our interest is in testing the validity of a model for the entire conditional distribution, most papers in this area focus on testing the validity at only one particular quantile, such as the median. Examples include Zheng (1998), Bierens and Ginther (2001), Horowitz and Spokoiny (2002), He and Zhu (2003) and Whang (2006). Two notable exceptions are Escanciano and Velasco (2010) and Escanciano and Goh (2010), which both propose testing procedures for the null hypothesis that a conditional quantile restriction is valid over a range of quantiles. The former paper considers specification testing for dynamic quantile regression models in a time series setting using a subsampling approach, whereas the latter paper studies the instrumental variables quantile regression model of e.g. Chernozhukov and Hansen (2005), obtaining critical values via a clever multiplier bootstrap scheme. The settings of those papers are thus very different from ours in general, but they all include the standard quantile regression model with independent observations as a special case. We present some Monte Carlo evidence on the relative merits of these approaches for testing the validity of that particular model below.

The remainder of the paper is structured as follows. In the next section, we describe our testing problem, the test statistic, and a bootstrap procedure to obtain critical values. In Section 3, we establish the theoretical properties of our test under general conditions. Section 4 contains some Monte Carlo evidence on the finite sample properties of our test, and evaluates the appropriateness of various models for the conditional distribution of wages given individual characteristics in the US. Finally, Section 5 concludes.

2. Testing General Conditional Distributional Models

2.1. Testing Problem. We observe an outcome variable Y_i and a vector of explanatory variables X_i for $i = 1, \ldots, n$, where $Z_i = (Y_i, X_i) \sim H$ and $X_i \sim G$. We assume throughout the paper that the data points are independent and identically distributed, although it would also be possible to extend our analysis to certain forms of temporal dependence. Let $\mathcal{F}$ be the class of all conditional distribution functions on the support of Y given X, and define a conditional distributional model $\mathcal{F}^0 \subset \mathcal{F}$ as a parametric family of conditional distribution functions indexed by (potentially) functional parameters. That is, we define

$$\mathcal{F}^0 = \{F(\cdot|\cdot, \theta) : \theta \in \mathcal{B}(T, \Theta)\}, \tag{2.1}$$

where $\mathcal{B}(T, \Theta)$ is a class of mappings $u \mapsto \theta(u)$ such that $\theta(u) \in \Theta \subset \mathbb{R}^p$ for $u \in T \subset \mathbb{R}$. We assume that for every $F \in \mathcal{F}^0$ there exists a unique value $\theta_0 \in \mathcal{B}(T, \Theta)$ of the functional parameter satisfying $F(y|x) = F(y|x, \theta_0)$. Furthermore, we assume that this value is identified as the solution to the equation

$$\Psi(\theta, u) := \mathbb{E}(\psi(Z, \theta, u)) = 0 \tag{2.2}$$

for some uniformly integrable function $\psi : \mathcal{Z} \times \Theta \times T \mapsto \mathbb{R}^p$, whose exact form depends on the specific conditional distributional model $\mathcal{F}^0$. Chernozhukov, Fernandez-Val, and Melly (2009) provide a detailed analysis of this class of conditional distributional models, and our theoretical analysis below uses their results extensively. The framework covers a number of important specifications, including the quantile regression model (Koenker and Bassett, 1978; Koenker, 2005), and the distributional regression model (Foresi and Peracchi, 1995). We discuss these two examples in more detail in Section 3.4. In the special case that the parameter $\theta_0(u) \equiv \bar{\theta}_0$ does not depend on u, the framework also includes many limited dependent variable models and generalized linear models.

Our aim is to test the validity of the conditional distributional model, i.e. whether the conditional distribution function F of Y_i given X_i is contained in $\mathcal{F}^0$. Our testing problem is thus given by

$$\begin{aligned} H_0 &: F(\cdot|\cdot) = F(\cdot|\cdot, \theta_0) \text{ for some } \theta_0 \in \mathcal{B}(T, \Theta) \\ \text{versus } H_1 &: F(\cdot|\cdot) \neq F(\cdot|\cdot, \theta) \text{ for all } \theta \in \mathcal{B}(T, \Theta). \end{aligned} \tag{2.3}$$

Note that this is a challenging problem since it involves testing the value of a function at an infinite number of points.

2.2. Test Statistic. Our test statistic for the problem in (2.3) is based on the distance between an unrestricted estimate of the joint CDF of (Y, X) and an estimate that imposes the structure implied by the conditional distributional model under the null hypothesis. In the present context, the most straightforward unrestricted estimate of H is arguably the empirical distribution function

$$\hat{H}(y,x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{Y_i \leq y\}\mathbb{I}\{X_i \leq x\}.$$

Here $\mathbb{I}\{A\}$ denotes the usual indicator function. To obtain an estimate that imposes the structure implied by the conditional distributional model, note that in the population the function H and the conditional distribution function F are linked via the relationship

$$H(y,x) = \int F(y|t)\mathbb{I}\{t \leq x\}dG(t).$$

Thus a natural choice for the restricted estimator is given by the sample analogue of the last equation, namely

$$\hat{H}^0(y,x) = \int \hat{F}(y|t)\mathbb{I}\{t \leq x\}d\hat{G}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{F}(y|X_i)\mathbb{I}\{X_i \leq x\},$$

where $\hat{G}_n(x) = n^{-1}\sum_{i=1}^{n} \mathbb{I}\{X_i \leq x\}$ is the empirical distribution function of the observed explanatory variables, and

$$\hat{F}(y|x) = F(y|x,\hat{\theta})$$

is a parametric estimate of F based on an estimate $\hat{\theta}$ of θ_0. We take $\hat{\theta}$ to be an approximate Z-estimator satisfying

$$\|\hat{\Psi}_n(\hat{\theta}(u),u)\| = o(n^{-1/2}),$$

where $\hat{\Psi}_n(\theta,u) := n^{-1}\sum_{i=1}^{n} \psi(Z_i,\theta,u)$ is the sample analogue of the moment condition in (2.2). Such an approach is feasible for all examples we consider in this paper. Note that our estimate of F is compatible with the conditional distributional model by construction, i.e. $\hat{F} \in \mathcal{F}^0$ almost surely, irrespectively of whether or not the null hypothesis is true.

Under general conditions described in the following section, both $\hat{H}$ and $\hat{H}^0$ converge to the true joint distribution function H of $Z = (Y,X)$ under the null hypothesis. However, when the alternative is true, the restricted estimate $\hat{H}^0$ has a probability limit $\bar{H}$

that is typically different from H, while $\hat{H}$ remains of course consistent. One can thus construct a specification test for the conditional distributional model $\mathcal{F}^0$ based on the distance between $\hat{H}$ and $\hat{H}^0$. We define our test statistic T_n as a Cramer-von Mises type measure of distance between the restricted and the unrestricted estimate of H at the observed sample points, scaled by the sample size:

$$T_n = n \int \left(\hat{H}(y,x) - \hat{H}^0(y,x) \right)^2 d\hat{H}(y,x) = \sum_{i=1}^{n} \left(\hat{H}(Y_i, X_i) - \hat{H}^0(Y_i, X_i) \right)^2 .$$

With the exception of the use of a different distance measure, our testing principle is analogous to the Conditional Kolmogorov test in Andrews (1997). Since we allow for a much wider class of conditional distributional models, we refer to our test in the following as a Generalized Conditional Cramer-von Mises (GCCM) test. The reason for departing from Andrews (1997) with respect to the distance measure is that our simulation experiments suggest that Cramer-von Mises type statistics have somewhat better power properties than those based on the Kolmogorov distance. They are also simpler to obtain from a computational point of view, since they do not involve a maximization problem over a high-dimensional space.[2]

2.3. Bootstrap Critical Values. As we show in more detail below, for most common conditional distributional models the null distribution of T_n is non-pivotal and depends on the data generating process in a complex fashion. We therefore obtain critical values for our test statistic by the bootstrap, using the restricted estimate $\hat{H}_n^0$ as the bootstrap distribution. Since $\hat{F} \in \mathcal{F}^0$ by construction, this approach ensures that the bootstrap mimics the distribution of the data under the null hypothesis, even though the data might be generated by an alternative distribution. The procedure is implemented as follows. The bootstrap data is given by a random sample $\{(Y_{b,i}, X_{b,i}), 1 \leq i \leq n\}$ drawn from the distribution function $\hat{H}^0$. Such a sample can be obtained by first drawing

[2] It would of course conceptionally be straightforward to consider other measures of distance between the two distribution functions to construct a test statistic for the problem in (2.3). The properties of such a test could be derived in essentially the same way as the one presented in the following subsections.

$\{X_{b,i}, 1 \leq i \leq n\}$ with replacement from the realized values $\{X_i, 1 \leq i \leq n\}$ of the explanatory variables, and then simulating the bootstrap values $Y_{b,i}$ of the dependent variable according to the estimated distribution function $\hat{F}(\cdot|X_{b,i})$. The new data is then used to calculate the restricted and unrestricted bootstrap estimates of H, denoted by $\hat{H}_b^0$ and $\hat{H}_b$, respectively, using the estimators described in the previous subsection. The bootstrap version of our test statistic is then given by

$$T_{b,n} = n \int \left(\hat{H}_b(y, x) - \hat{H}_b^0(y, x) \right)^2 d\hat{H}_b(y, x). \tag{2.4}$$

The distribution of $T_{b,n}$ can be determined through the usual repeated resampling of the data, and, as shown formally below, can then be used as an approximation to the distribution of T_n under the null hypothesis for a wide range of conditional distributional models. In particular, we can obtain an asymptotically valid level α critical value $\hat{c}_n(\alpha)$ for the testing problem in (2.3) by computing the $(1-\alpha)$-quantile of the distribution of $T_{b,n}$, i.e. $\hat{c}_n(\alpha)$ is the smallest constant that satisfies

$$P_b(T_{b,n} \leq \hat{c}_n(\alpha)) \geq 1 - \alpha,$$

where P_b is the probability with respect to bootstrap sampling. The test thus rejects H_0 if $T_n > \hat{c}_n(\alpha)$ for some pre-specified significance level $\alpha \in (0, 1)$.

3. Theoretical Properties

This section shows that the GCCM test has correct asymptotic size, consistency against fixed alternatives, and non-trivial power against local deviations from the null hypothesis of order $n^{-1/2}$. We write "$\xrightarrow{d}$" to denote convergence in distribution of a sequence of random variables, and "$\Rightarrow$" to denote weak convergence of a sequence of random functions. In addition, we write "the data are distributed according to $\tilde{F}$" whenever the joint distribution function of $Z = (Y, X)$ is given by $\tilde{H}(y, x) = \int \tilde{F}(y|t)\mathbb{I}\{t \leq x\}dG(t)$ for some

$\tilde{F} \in \mathcal{F}$, and denote the expectation taken with respect to $\tilde{H}$ by $\mathbb{E}_{\tilde{H}}$. All limits are taken as $n \to \infty$.

3.1. Limiting Distribution of the Test Statistic. The large sample properties of our test statistic require the following assumptions.

Assumption 1. *The set Θ is a compact subset of $\mathbb{R}^p$ and T is either a finite subset or a bounded open subset of $\mathbb{R}$.*

Assumption 2. *For each $u \in T$, there exists a unique value $\theta_0(u)$ in the interior of Θ such that $\Psi(\theta_0(u), u) = 0$.*

Assumption 3. *The mapping $(\theta, u) \mapsto \psi(Z, \theta, u)$ is continuous at each $(\theta, u) \in \Theta \times T$ with probability one, and continuously differentiable at $(\theta_0(u), u)$ with a uniformly bounded derivative on T (where differentiability in u is only required when T is not finite). The function $\dot{\Psi}(\theta, u) := \partial_\theta \Psi(\theta, u)$ is nonsingular at $\theta_0(\cdot)$ uniformly in $u \in T$.*

Assumption 4. *The set of functions $\mathcal{G} = \{\psi(Z, \theta, u), (\theta, u) \in \Theta \times T\}$ is H-Donsker with a square integrable envelope.*

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

Assumption 6. *Under the alternative, $P(\bar{H}(Y, X) \neq H(Y, X)) > 0$, where $\bar{H}(y, x) = \int F(y|t, \theta_0)\mathbb{I}\{t \leq x\}dG(t)$.*

Assumptions 1–4 are standard regularity conditions also imposed by Chernozhukov, Fernandez-Val, and Melly (2009). They allow for establishing a functional central limit theorem for the Z-estimator process $u \mapsto \sqrt{n}(\hat{\theta}(u) - \theta_0(u))$. Assumption 5 is a smoothness condition that can be verified directly in applications. Together with the Functional Delta Method, it implies that the restricted CDF estimator process $(y, x) \mapsto \sqrt{n}(\hat{H}^0(y, x) - \bar{H}(y, x))$ also converges to a Gaussian limit process $\mathbb{G}_1$, with $\bar{H}$ as in Assumption 6. This convergence can be shown to be jointly with that of the empirical CDF process

$(y,x) \mapsto \sqrt{n}(\hat{H}(y,x) - H(y,x))$ to an H-Brownian Bridge $\mathbb{G}_2$. The limiting distribution of our test statistic T_n then follows from the continuous mapping theorem.

Theorem 1. *Under Assumption 1–6 we have*

i) Under the null hypothesis, i.e. when the data are distributed according to some $F \in \mathcal{F}^0$,

$$T_n \xrightarrow{d} \int (\mathbb{G}_1(y,x) - \mathbb{G}_2(y,x))^2 \, dH(y,x), \tag{3.1}$$

where $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2)$ is a bivariate mean zero Gaussian process given in the Appendix.

ii) Under any fixed alternative, i.e. when the data are distributed according to some $F \in \mathcal{F} \setminus \mathcal{F}^0$,

$$\lim_{n\to\infty} P(T_n > c) \to 1 \text{ for all constants } c > 0. \tag{3.2}$$

3.2. Local Alternatives. This section derives the limiting distribution of our test statistic under a sequence of local alternatives that shrink towards an element of $\mathcal{F}^0$ at rate $n^{-1/2}$, where n denotes the sample size. That is, the conditional distribution function of Y given X is given by

$$Q_n(y|x) = (1 - \delta/\sqrt{n})F^*(y|x) + (\delta/\sqrt{n})Q(y|x) \tag{3.3}$$

for some $F^* \in \mathcal{F}^0$, $Q \in \mathcal{F} \setminus \mathcal{F}^0$, a constant δ and $n \geq \delta^2$, satisfying the following assumption.

Assumption 7. *The sequence $H_n(y,x) = \int Q_n(y|t)\mathbb{I}\{t \leq x\}dG(t)$ of distribution functions implied by the local alternative Q_n given in* (3.3) *is contiguous to the distribution function $H^*(y,x) = \int F^*(y|t)\mathbb{I}\{t \leq x\}dG(t)$ implied by F^*.*

The requirement that the local alternatives are contiguous (see Van der Vaart, 2000, Section 6.2 for a formal definition of contiguity) to the limiting distribution function is standard when analyzing local power properties. When the conditional distribution functions F^* and Q admit conditional density functions f^* and q with respect to the same σ-finite measure (e.g. the Lebesgue measure), respectively, a sufficient condition for contiguity is

$$\sup_{(x,y):f^*(y|x)>0} \frac{q(y|x)}{f^*(y|x)} < \infty. \tag{3.4}$$

Intuitively, this would be the case when Q has lighter tails than F^*. This statement is formally proven in Section B in the Appendix.

The following theorem shows that under local alternatives of the form (3.3) the limiting distribution of T_n contains an additional deterministic shift function ensuring nontrivial local power of the test. To describe this function, define $\Psi_Q(\theta, u) = \mathbb{E}_Q(\psi(Z, \theta, u))$ and $\Psi_*(\theta, u) = \mathbb{E}_{F^*}(\psi(Z, \theta, u))$, and let θ_Q and θ_* be the functions satisfying $\Psi_Q(\theta_Q(u), u) = 0$ and $\Psi_*(\theta_*(u), u) = 0$ for all $u \in T$, respectively.

Theorem 2. *Under Assumption 1–5 and 7, and if the data are distributed according to a local alternative Q_n as given in* (3.3),

$$T_n \xrightarrow{d} \int \left(\mathbb{G}_1(y,x) - \mathbb{G}_2(y,x) + \mu(y,x)\right)^2 dH(y,x).$$

where

$$\mu(y,x) = \delta \int (Q(y|t) - F(y|t,\theta_*) + \dot{F}(y|t,\theta_*)\,[h])\mathbb{I}\{t \leq x\} dG(t)$$

with $h(u) = \partial_{\theta'} \Psi_{F^*}(\theta_*(u), u)^{-1} \Psi_Q(\theta_*(u), u)$.

Note that the function F^* to which the local alternative Q_n shrinks can be chosen as $F(\cdot|\cdot, \theta_Q)$, the probability limit of estimator $\hat{F}$ under Q. In this case, we have

$\Psi_Q(\theta_*(u), u) = 0$ for all $u \in T$, and hence the drift term in Theorem 2 simplifies to

$$\mu(y,x) = \delta \int (Q(y|t) - F(y|t,\theta_*))\mathbb{I}\{t \leq x\} dG(t),$$

and is thus proportional to the difference between the joint distribution functions implied by Q and F^*.

3.3. Validity of the Bootstrap. As a final step, we establish asymptotic validity of the critical values obtained via the bootstrap procedure described in Section 2.3. This does not require any further assumptions. Under the null hypothesis, Assumptions 1–5 ensure that the bootstrap consistently estimates the limiting distribution of the test statistic T_n, and hence consistently estimates the true critical values. Under any fixed alternative, the bootstrap critical values can be shown to be bounded in probability. Together with Theorem 1(ii), this implies that the power of our test converges to one in this case. Finally, since contiguity preserves convergence in probability, it follows from Assumption 7 that under any local alternative the bootstrap critical values converge to the same value as under the null hypothesis. We can thus deduce from Anderson's Lemma that our test has non-trivial local power. The following theorem formalizes these arguments.

Theorem 3. *Under Assumption 1–7, the following statements hold for any $\alpha \in (0, 1)$:*

i) Under the null hypothesis, i.e. when the data are distributed according to some $F \in \mathcal{F}^0$, we have that

$$\lim_{n\to\infty} P(T_n > \hat{c}_n(\alpha)) = \alpha.$$

ii) Under any fixed alternative, i.e. when the data are distributed according to some

$F \in \mathcal{F} \setminus \mathcal{F}^0$, *we have that*

$$\lim_{n\to\infty} P(T_n > \hat{c}_n(\alpha)) = 1.$$

iii) Under any local alternative, i.e. when the data are distributed according to some $Q_n \in \mathcal{F} \setminus \mathcal{F}^0$ as described in (3.3), *we have that*

$$\lim_{n\to\infty} P(T_n > \hat{c}_n(\alpha)) \geq \alpha.$$

3.4. Application to Specific Models. In this subsection, we give two examples for classes of conditional distributional models that can be tested in our framework. We provide primitive conditions that imply Assumptions 1–5.

Example 1 (Quantile Regression)**.** Our leading example for a conditional distributional model indexed by function-valued parameters is the linear quantile regression model (Koenker and Bassett, 1978; Koenker, 2005). It postulates that the conditional τ-quantile of Y given $X = x$ is linear in a vector of parameters varying with τ:

$$\mathcal{F}^0 = \{F^{-1}(\tau|x) = x'\theta(\tau) \text{ for some } \theta(\tau) \in \Theta \subset \mathbb{R}^p \text{ and all } \tau \in (0,1)\}.$$

Such a model would be correctly specified if the true data generating process could be represented by the random coefficient model $Y = X'\theta_0(U)$, where $U \sim U[0,1]$ is independent of X and the function $\theta_0(\cdot)$ is strictly increasing in each of its arguments. We consider the usual estimator $\hat{\theta}(u)$ of $\theta_0(u)$, given by

$$\hat{\theta}(u) := \operatorname*{argmin}_{\theta} \frac{1}{n} \sum_{i=1}^{n} \rho_u(Y_i - X_i'\theta), \tag{3.5}$$

where $\rho_u(s) = s(u - \mathbb{I}\{s \leq 0\})$ is the usual "check function". This estimator is contained in the class of approximate Z-estimators we consider in this paper, as it satisfies $\|\hat{\Psi}_n(\hat{\theta}(u), u)\| = o(n^{-1/2})$, where $\hat{\Psi}_n(\theta, u) := n^{-1} \sum_{i=1}^{n} \psi(Z_i, \theta, u)$ and $\psi(Z_i, \theta, u) =$

$(u - \mathbb{I}\{Y_i - X_i'\theta \leq 0\})X_i$. The following theorem gives conditions for the Quantile Regression model to satisfy the "high level" conditions in Section 3.

Theorem 4. *Suppose that (i) the distribution function $F(\cdot|X)$ admits a density function $f(\cdot|X)$ that is continuous, bounded and bounded away from zero at $X'\theta_0(u)$, uniformly over $u \in (0,1)$, almost surely. (ii) The matrix $\mathbb{E}(XX')$ is finite and of full rank, (iii) the parameter $\theta_0(\cdot)$ solving $\mathbb{E}(\psi(Z, \theta_0(u), u)) = 0$ is such that $\theta_0(u)$ is in the interior of the parameter space Θ for every $u \in (0,1)$. Then Assumption 1–5 hold for the Quantile Regression model with $\dot{F}(y|x,\theta)[h(\cdot)] = -f(y|x)x'[h(F(y|x,\theta))]$.*

Example 2 (Distributional Regression)**.** Another interesting class of conditional distributional models is the distributional regression model introduced by Foresi and Peracchi (1995). Here the conditional CDF Y given X is modeled directly by fitting a binary-choice-type model for the event that the dependent variable Y exceeds some threshold y in the support $\mathcal{Y}$ of Y separately for each threshold level:

$$\mathcal{F}^0 = \{F(y|x) = \Lambda(x'\theta(y)) \text{ for some } \theta(y) \in \Theta \subset \mathbb{R}^p\}, \tag{3.6}$$

where $\Lambda(\cdot)$ is a known strictly increasing link function, e.g. the logistic or standard normal distribution function, or simply the identity function. In the context of the analysis of counterfactual distributions, this type of model is e.g. used in Chernozhukov, Fernandez-Val, and Melly (2009) and Rothe (2010b) for the respective empirical applications.

Since for every threshold value $y \in \mathcal{Y}$ the Distributional Regression model resembles a standard binary choice model, a natural estimator for the functional parameter $\theta_0(y)$ is the maximum likelihood estimator $\hat{\theta}(y)$, which solves $\|\hat{\Psi}_n(\hat{\theta}(u), u)\| = 0$, where $\hat{\Psi}_n(\theta, u) := n^{-1}\sum_{i=1}^n \psi(Z_i, \theta, u)$ and $\psi(Z_i, \theta, y) = (\Lambda(X_i'\theta)(1 - \Lambda(X_i'\theta)))^{-1}(\Lambda(X_i'\theta) - \mathbb{I}\{Y_i \leq y\})\lambda(X_i'\theta)X_i$ is the score function. The estimated conditional CDF of Y given X is then given by $\hat{F}(y|x) = \Lambda(x'\hat{\theta}(y))$. The following theorem gives conditions for the Distributional Regression model to satisfy the "high level" conditions in Section 3.

Theorem 5. *Suppose that: (i) The support $\mathcal{Y}$ of Y is a bounded open subset of $\mathbb{R}$ (ii) The distribution function $F(\cdot|X)$ admits a density function $f(\cdot|X)$ that is continuous, bounded and bounded away from zero at all $y \in \mathcal{Y}$, almost surely. (iii) The matrix $E(XX')$ is finite and of full rank. (iv) The parameter $\theta_0(\cdot)$ solving $\mathbb{E}(\psi(Z, \theta_0(y), y)) = 0$ is such that $\theta_0(y)$ is in the interior of the parameter space Θ for every $y \in \mathcal{Y}$. (v) The quantity $\Lambda(X'\theta)$ is bounded away from zero and one uniformly over $\theta \in \Theta$, almost surely. Then Assumption 1–5 hold for the Distributional Regression Model in* (3.6) *with* $\dot{F}(y|x, \theta)[h(\cdot)] = (\partial\Lambda(x, \theta(y))/\partial\theta')[h(y)]$.

4. Numerical Evidence

4.1. Monte Carlo Study. In order to demonstrate the usefulness of our proposed testing procedure, we conduct a number of simulation experiments to assess the size and power properties in finite samples. We consider the same data generating processes as in Escanciano and Goh (2010) to be able to compare the properties of their test to ours for the special case of a quantile regression model, when both methods are applicable. In particular, we simulate a dependent variable Y according to one of the following data generating processes:

$$\text{(DGP1):} \quad Y = X_1 + X_2 + c\sigma^{3/2} + U$$

$$\text{(DGP2):} \quad Y = X_1 + X_2 + (1 + c\sigma^{3/2})U$$

where $\sigma = X_1^2 + X_2^2 + X_1X_2$, and X_1, X_2 and U are taken to be standard normally distributed and mutually independent. We vary the parameter c over the set of values $\{-.3, -.2, -.1, 0, .1, .2, .3\}$ and $\{-1, -.5, 0, .5, 1\}$ for DGP1 and DGP2, respectively, consider the sample sizes $n = 100$ and $n = 250$, and set the number of replications to 1000. In each simulation run, we test for correctness of the specification of the two types of models discussed in Section 3.4: the linear quantile regression model (QR) and the

Table 1: Simulated Rejection Frequencies for DGP1 (left panel: $n = 100$, right panel $n = 250$)

	QR Model		DR Model		QR Model		DR Model	
c	5%	10%	5%	10%	5%	10%	5%	10%
-.3	0.808	0.881	0.897	0.942	1.000	1.000	1.000	1.000
-.2	0.499	0.621	0.662	0.745	0.975	0.996	0.915	0.950
-.1	0.085	0.153	0.167	0.244	0.428	0.565	0.331	0.437
0	0.037	0.064	0.035	0.069	0.043	0.079	0.041	0.073
.1	0.630	0.755	0.320	0.441	0.946	0.986	0.606	0.721
.2	0.899	0.963	0.751	0.834	0.998	1.000	0.966	0.980
.3	0.971	0.996	0.889	0.943	1.000	1.000	1.000	1.000

Table 2: Simulated Rejection Frequencies for DGP2 (left panel: $n = 100$, right panel $n = 250$)

	QR Model		DR Model		QR Model		DR Model	
c	5%	10%	5%	10%	5%	10%	5%	10%
-1	0.393	0.576	0.798	0.918	0.914	0.977	0.991	0.998
-.5	0.160	0.286	0.756	0.856	0.635	0.790	0.982	0.995
0	0.028	0.066	0.026	0.064	0.039	0.086	0.041	0.087
.5	0.379	0.553	0.589	0.761	0.878	0.951	0.914	0.975
1	0.504	0.681	0.649	0.815	0.956	0.981	0.954	0.996

distributional regression model (DR):

$$\text{(QR):} \quad \mathcal{F}^0 = \{F^{-1}(\tau|x) = x'\theta(\tau)\}$$

$$\text{(DR):} \quad \mathcal{F}^0 = \{F(y|x) = \Lambda(x'\theta(y))\}$$

For the DR model, we choose the link function Λ as the standard normal distribution function. Thus both models are correctly specified under DGP1–2 when $c = 0$, and misspecified with $c \neq 0$.

In Table 1–2, we show the empirical rejection probabilities of our GCCM test for the nomial levels 5% and 10%. When setting the parameter to $c = 0$, and thus both models are correctly specified under DGP1–2, the size of the test is slightly below its respective nominal value, but still reasonable close given the small sample sizes. Under misspecification, empirical rejection rates from testing the DR model are mostly higher than those for the QR model, although there is no uniform ordering. The power of our tests is also increasing in the sample size and the absolute value of the parameter c, as expected. Comparing our results for the QR model to those in Escanciano and Goh (2010), we see that the rejection probabilities of our testing procedure are broadly similar to theirs for all parameter constellations in DGP2, and for positive values of c in DGP1. For negative values of c in DGP1, in particular $c = -.1$, their test exhibits substantially higher power.[3] Testing the validity of the DR model is not directly possible with the procedure in Escanciano and Goh (2010), and thus no comparisons can be made for this case.

4.2. Empirical Application. In this subsection, we use our GCCM test to assess the validity of various commonly used models for the conditional distribution of wages given certain individual characteristics. As pointed out in the introduction, such models play an important role in the literature on decomposing counterfactual distributions (Firpo, Fortin, and Lemieux, 2010). There are doubts, however, that simple and parsimonious models are able to capture some important features of such conditional distributions, such as the nonlinearities around th minimum wage. The results in this subsection shed some light on this important empirical issue.

Our data is taken from the 1988 wave of the Current Population Survey (CPS), an extensive survey of US households. The dataset is the same as the one used in DiNardo, Fortin, and Lemieux (1996), to which we refer for details of its construction. The data set contains information on 74,661 males that were employed in the relevant period, including the hourly wage, years of education, years of potential labor market experience, and indicator variables for union coverage, race, marital status, part-time status, living in a Standard Metropolitan Statistical Area (SMSA), type of occupation (2 levels), and industry (20 levels). As in the previous subsection, we consider both the linear quantile regression model (QR), and the distributional regression model (DR) using the normal

[3]Note that Escanciano and Goh (2010) state that for their simulations they experimented with various values for a certain tuning parameter of their bootstrap procedure, reporting only the results that resulted in the best size properties. This is of course not feasible in empirical application, where the true data generating process is unknown. There are no such issues with our test, since it does not require the choice of tuning parameters.

Table 3: Empirical Application: Rejection Probabilities

		Model 1		Model 2		Model 3	
	Nom. level:	.05	.1	.05	.1	.05	.1
DR	n=500	0.515	0.565	0.000	0.015	0.000	0.005
	n=1000	0.655	0.705	0.005	0.015	0.005	0.005
QR	n=500	0.965	0.995	0.875	0.970	0.440	0.655
	n=1000	1.000	1.000	1.000	1.000	0.735	0.915

CDF as a link function. We test the correct specification of each model with hourly wage as the dependent variable, and the following three different subsets of the explanatory variables, respectively:

- **Model 1:** union coverage, education, experience.
- **Model 2:** all variables in Model 1, experience (squared), education interacted with experience, martial status, part-time status, race, SMSA.
- **Model 3:** all variables in Model 2, occupation, industry.

Given the large sample size, we would expect all specifications to be rejected by the data, since every model is at best a reasonable approximation to the true data generating mechanism. However, this would not directly imply that such specifications would result in misleading conclusions, as in large samples our GCCM test should be able to pick up deviations from the null hypothesis even if they are not of economically significant magnitude. On the other hand, we should have much more reason to be concerned if flexible models including many covariates would be rejected even in small samples. We therefore conduct a simulation experiment, where in each run we test the validity of the QR and DR specification of Model 1–3 for random subsamples of the data of size $n = 500$ and $n = 1000$. Results from 200 repetitions are summarized in Table 3.

We can see that there is compelling evidence against the QR model. The specifications in Model 1–2 are rejected in about 90% of all simulation runs for $n = 500$, and in every single one for $n = 1000$. Even the extensively parametrized Model 3, which uses a 31-dimensional parameter vector, is rejected at a nominal level of 5% in about half the simulation runs for $n = 500$, and in three quarters of all runs for $n = 1000$. Given these high rejection rates on small subsamples of the data, practitioners might want to consider a different class of models for the conditional distribution of wages. Our simulation results show that a flexibly specified DR model might be a suitable candidate. While rejection rates are still considerable for the (possibly over-parsimonious) Model 1 for both sample sizes under consideration, the more realistic specifications Model 2 and 3 exhibit empirical rejection probabilities well below the respective nominal testing levels. The simulation results in the previous subsection suggest that this finding should not be due to a lack of power of our test under the DR specification. The class of distributional regression models might thus be more adequate to capture the particular features of conditional wage distributions, such as e.g. the nonlinearities close to the legal minimum wage level.

Remark 1. A particular feature of the CPS data is that the empirical distribution of hourly wages contains a number of mass points, since many workers are paid a "round" amount of dollars, or at least report it in the survey. Since the linear quantile regression model implies a strictly increasing conditional CDF, it is not able to reproduce such patterns. In order to check whether our high rejection rates are simply due to this issue, we repeated the above empirical exercise with the following modification: First, each individual we computed her rank in the distribution of wages, breaking ties at random. Second, we replaced the observed wage by the quantile of a smoothed version of the empirical distribution of wages (obtained by linear interpolation of jump points) corresponding to the individual's rank. That is, we are "spreading" the mass points evenly in order to obtain a smooth distribution of wages. The results remained qualitatively unchanged, and are hence not reported for brevity. There are no theoretical issues related to mass points in the distribution of outcomes when using the class of distributional regression models, which was also confirmed in our simulation.

5. Conclusions

This paper provides a specification test for a general class of conditional distributional models indexed by function-valued parameters. Our method is straightforward to implement and does not require the choice of smoothing parameters. We establish consistency of our testing procedure against fixed alternatives under general conditions, and study its local power properties. We illustrate the usefulness of our procedure via a simulation procedure, highlighting its favorable practical properties compared to a competing approach. In an application to real data, we show that our test is able to detect misspecification of standard linear quantile regression models for conditional distribution of wages in the US even in small samples, while a rich distributional regression model can typically not be rejected.

A. Appendix

A.1. Proofs of Theorems. In this subsection, we collect the proofs of our main theorems. Some auxiliary results needed in the course of the proofs are given in Appendix A.2.

Proof of Theorem 1. To prove part i), define the processes $\nu(y,x) = \sqrt{n}(\hat{H}_n(y,x) - H(y,x))$ and $\nu_0(y,x) = \sqrt{n}(\hat{H}_n^0(y,x) - \bar{H}(y,x))$, and note that $\sqrt{n}(\hat{H}_n(y,x) - \hat{H}_n^0(y,x)) = \nu(y,x) - \nu_0(y,x)$. We can thus rewrite our test statistic as follows:

$$T_n = \int (\nu(y,x) - \nu_0(y,x))^2 dH(y,x) + \int (\nu(y,x) - \nu_0(y,x))^2 d(\hat{H}(y,x) - H(y,x)).$$

From Lemma 2, we know that $(\nu, \nu_0) \Rightarrow \mathbb{G}$, where $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2)$ is a tight bivariate mean zero gaussian process. By the continuous mapping theorem, we thus have that

$$T_n = \int (\mathbb{G}_1(y,x) - \mathbb{G}_2(y,x))^2 dH(y,x) + o_p(1),$$

as claimed. To show part ii), note that under a fixed alternative we have that

$$\sqrt{n}(\hat{H}_n(y,x) - \hat{H}_n^0(y,x)) = (v(y,x) - v_0(y,x)) + \sqrt{n}(H(y,x) - \bar{H}(y,x)),$$

with $\bar{H}$ as defined in Assumption 6. The first summand on the right-hand side of the last equation is bounded in probability by Lemma 2, whereas the second term is of the order $O(n^{1/2})$ for values of (y,x) in a set with positive probability by Assumption 6. Hence $\tilde{T}_n$ becomes arbitrarily large with probability approaching 1 as $n \to \infty$. □

Proof of Theorem 2. To prove the result, we first define the empirical processes $\lambda_1(y,x) = \sqrt{n}((\hat{H}_n(y,x) - \int F^*(y|t)\mathbb{I}\{t \leq x\}dG(t))$ and $\lambda_2(\theta,u) = \sqrt{n}(\hat{\Psi}(\theta,u) - \mathbb{E}_{F^*}(\psi(Z,\theta,u)))$, and denote the joint process by $\lambda(y,x,\theta,u) = (\lambda_1(y,x), \lambda_2(\theta,u))$. It then follows with Lemma 3 that

$$\lambda \Rightarrow \left(\mathbb{G}_1 + \delta\mu_1, \tilde{\mathbb{G}}_1 + \delta\tilde{\mu}_2\right),$$

where $\mu_1(y,x) = \int(Q(y|t) - F^*(y|t))\mathbb{I}\{t \leq x\}dG(t)$ and $\tilde{\mu}_2(\theta,u) = \mathbb{E}_Q(\psi(Z,\theta,u)) - \mathbb{E}_{F^*}(\psi(Z,\theta,u))$. Next, define the empirical processes $\nu^*(y,x) = \sqrt{n}(\hat{H}_n(y,x) - H^*(y,x))$ and $\nu_0^*(y,x) = \sqrt{n}(\hat{H}_n^0(y,x) - H^*(y,x))$, with $H^*(y,x) = \int F^*(y|t)\mathbb{I}\{t \leq x\}dG(t)$. Proceeding in the same way as in the proof of Lemma 2, we find that

$$(\nu, \nu_0) \Rightarrow (\mathbb{G}_1 + \delta\mu_1, \mathbb{G}_2 + \delta\mu_2),$$

where $\mu_2(y,x) = \int \dot{F}(y|t)[h]\mathbb{I}\{t \leq x\}dG(t)$ and $h(u) = \partial_{\theta'}\Psi_{F^*}(\theta_*(u),u)^{-1}\Psi_Q(\theta_*(u),u)$. The statement of the Theorem then follows from the continuous mapping theorem, in the same way as in the proof of Theorem 1. □

Proof of Theorem 3. To prove part i) let $c(\alpha)$ be the "true" critical value satisfying $P(T_n > c(\alpha)) = \alpha + o(1)$. Then it follows from Lemma 4 that $\hat{c}_n(\alpha) = c(\alpha) + o_p(1)$. This implies that T_n and $\tilde{T}_n = T_n - (\hat{c}_n(\alpha) - c(\alpha))$ converge to the same limiting distribution as $n \to \infty$, and hence we have that $P(T_n > \hat{c}_n(\alpha)) = \alpha + o(1)$ as claimed.

To prove part ii), note that by Lemma 4 the bootstrap critical value $\hat{c}(\alpha)$ is bounded in probability under fixed alternatives. Hence for any $\epsilon > 0$ there exists a sufficiently large constant M such that $P(\hat{c}_n(\alpha) > M) < \epsilon + o(1)$. Using elementary inequalities, we also have that

$$P(T_n \leq \hat{c}_n(\alpha)) = P(T_n \leq \hat{c}_n(\alpha), T_n \leq M) + P(T_n \leq \hat{c}_n(\alpha), T_n > M)$$
$$\leq P(T_n \leq M) + P(\hat{c}_n(\alpha) > M).$$

From Theorem 1(ii), we know that $P(T_n \leq M) = o(1)$, and thus $P(T_n \leq \hat{c}_n(\alpha)) < \epsilon + o(1)$, which implies the statement of the theorem since ϵ can be chosen arbitrarily small.

To show part iii), define $c(\alpha)$ as in the proof of part i), i.e. the asymptotic α-quantile of the test statistic T_n under the null hypothesis. Using Anderson's Lemma, we find that

$$P\left(\int (\mathbb{G}_1(y,x) - \mathbb{G}_2(y,x) + \mu(y,x))^2 \, dH(y,x) > c(\alpha)\right)$$
$$\geq P\left(\int (\mathbb{G}_1(y,x) - \mathbb{G}_2(y,x))^2 \, dH(y,x) > c(\alpha)\right) = \alpha,$$

because the Gaussian process $\mathbb{G}_1 - \mathbb{G}_2$ has mean zero (see also Andrews (1997, p. 1114)). Under a local alternative, we therefore have that $P(T_n > c(\alpha)) \geq \alpha + o(1)$. Furthermore, we have already shown in part i) that $P(T_n > \hat{c}_n(\alpha)) = P(T_n > c(\alpha)) + o(1)$ under the null hypothesis. By using contiguity arguments, this can also shown to be true under the local alternative, see e.g. the proof of Corollary 2.1 in Bickel and Ren (2001). □

Proof of Theorem 4–5. This follows by straightforward applications of results in Chernozhukov, Fernandez-Val, and Melly (2009, Appendix F). □

A.2. Auxiliary Results. In this subsection, we collect a number of auxiliary results used in the proofs of our main results above.

Lemma 1. *Define the empirical processes* $\nu(y,x) = \sqrt{n}(\hat{H}_n(y,x) - H(y,x))$ *and* $w(\theta,u) = \sqrt{n}(\hat{\Psi}(\theta,u) - \Psi(\theta,u))$. *Then, under either the null hypothesis or a fixed alternative, and Assumptions 1-6, it holds that* $(v,w) \Rightarrow \tilde{\mathbb{G}}$ *in* $l^\infty(\mathcal{Z} \times \Theta \times T)$, *where* $\tilde{\mathbb{G}} = (\mathbb{G}_1, \tilde{\mathbb{G}}_2)$ *is a* *tight bivariate mean zero Gaussian process. Moreover, the bootstrap procedure in Section 2.3 consistently estimates the law of* $\tilde{\mathbb{G}}$.

Proof. This lemma is a minor generalization of Lemma 13 in Chernozhukov, Fernandez-Val, and Melly (2009). □

Lemma 2. *Let either the null hypothesis or a fixed alternative, and Assumptions 1-6 be true. Define the empirical processes* $\nu(y,x) = \sqrt{n}(\hat{H}_n(y,x) - H(y,x))$ *and* $\nu_0(y,x) = \sqrt{n}(\hat{H}_n^0(y,x) - \bar{H}(y,x))$. *Then it holds that* $(\nu, \nu_0) \Rightarrow \mathbb{G}$ *in* $l^\infty(\mathcal{Z} \times Z)$, *where* $\mathbb{G} = (\mathbb{G}_1, \mathbb{G}_2)$ *is a tight bivariate mean zero gaussian process.*

Proof. Under either the null hypothesis or a fixed alternative, it follows from our Lemma 1 and Lemma 11 in Chernozhukov, Fernandez-Val, and Melly (2009) that

$$\sqrt{n}\begin{pmatrix}\hat{H}_n(\cdot,\cdot) - H(\cdot,\cdot) \\ \hat{\theta}(\cdot) - \theta_0(\cdot)\end{pmatrix} \Rightarrow \begin{pmatrix}\mathbb{G}_1(\cdot,\cdot) \\ -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot}[\tilde{\mathbb{G}}_2(\theta_0(\cdot),\cdot)]\end{pmatrix}$$

in $\ell^\infty(\mathcal{Z}) \times \ell^\infty(T)$. Next, it follows from Assumption 5 that $\sqrt{n}(\hat{F}(y|x) - F(y|x)) \Rightarrow -\dot{F}(y|x,\theta_0)[\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot}[\tilde{\mathbb{G}}_2(\theta_0(\cdot),\cdot)]]$. The statement of the Lemma then follows directly from Hadamard differentiability of the mapping $(G,H) \mapsto \int G(\cdot)\mathbb{I}\{t \leq \cdot\}dH(t)$, and the Functional Delta Method. In particular, for the second component $\mathbb{G}_2$ of the joint limiting process we have that

$$\mathbb{G}_2(y,x) = \int F(y|t)\mathbb{I}\{t \leq x\}d\mathbb{G}_1(\infty,t) + \int \mathbb{G}_1(y,t)\mathbb{I}\{t \leq x\}dG(t),$$

which follows from the form of the Hadamard differential of the mapping $(G,H) \mapsto \int G(\cdot)\mathbb{I}\{t \leq \cdot\}dH(t)$. □

Lemma 3. *Suppose the data are distributed according to a local alternative* Q_n *satisfying Assumption 7. Define the processes* $v_n(y,x) = \sqrt{n}(\hat{H}_n(y,x) - H_n(y,x))$ *and* $w_n(\theta,u) = \sqrt{n}(\hat{\Psi}(\theta,u) - \Psi_n(\theta,u))$, *where* $H_n(y,x) = \int Q_n(y|t)\mathbb{I}\{t \leq x\}dG(t)$ *and*

$\Psi_n(\theta, u) = \mathbb{E}_{Q_n}(\psi(Z, \theta, u))$. *Then it holds* $(v_n, w_n) \Rightarrow \tilde{\mathbb{G}}$ *in* $l^\infty(\mathcal{Z} \times \Theta \times T)$*, where the limiting process is the same as in Lemma 1.*

Proof. This follows by an application of Lemma 2.8.7 in Van der Vaart and Wellner (1996), using the fact that by Assumption 4, Q_n is the linear combination of two measures under which the function class $\mathcal{G}$ is Donsker with a square integrable envelope. □

Lemma 4. *Define the bootstrap empirical processes* $\nu_b(y, x) = \sqrt{n}(\hat{H}_{bn}(y, x) - \hat{H}^0_n(y, x))$ *and* $\nu_{b,0}(y, x) = \sqrt{n}(\hat{H}^0_{bn}(y, x) - \hat{H}^0_n(y, x))$. *Then it holds under either the null hypothesis or a fixed alternative that* $(\nu_b, \nu_{b,0}) \Rightarrow \mathbb{G}_b$*, where* $\mathbb{G}_b = (\mathbb{G}_{b1}, \mathbb{G}_{b2})$ *is a tight bivariate mean zero gaussian process whose distribution coincides with that of the process* $\mathbb{G}$ *in Lemma 1 under the null hypothesis.*

Proof. This follows from Lemma 1 and the Functional Delta Method for the bootstrap (Van der Vaart and Wellner, 1996, Theorem 3.9.11) □

B. Sufficient Condition for Contiguity.

In this section, we show that the condition given in (3.4) is sufficient for contiguity. Our argument is analogous to the one given in Andrews (1997), and stated here only for completeness. For and distribution function H, let P^H be the probability measure induced by H. By definition, P^{H_n} is contiguous to P^{H^*} if $P^{H^*}(A_n) \to 0$ implies $P^{H_n}(A_n) \to 0$ for every sequence of measurable sets A_n. By an application of Le Cam's First Lemma (Van der Vaart, 2000, Theorem 6.4) this is the case if dP^{H_n}/dP^{H^*} converges in distribution to a random variable V with $\mathbb{E}(V) = 1$ under P^{H^*}. We show that $\log(dP^{H_n}/dP^{H^*}) \stackrel{d}{\to} N(-\sigma^2/2, \sigma^2)$ for some value $\sigma^2 > 0$, which directly implies that the aforementioned condition is fulfilled (see also Example 6.5 in Van der Vaart (2000)). Writing $a_n = \delta/\sqrt{n}$, it holds by the definition of dP^{H_n} and dP^{H^*} that

$$\frac{dP^{H_n}}{dP^{H^*}} = \frac{\prod_{i=1}^n \{(1 - a_n) f^*(Y_i|X_i) g(X_i) + a_n q(Y_i|X_i) g(X_i)\}}{\prod_{i=1}^n f^*(Y_i|X_i) g(X_i)},$$

and we thus have that

$$\begin{aligned}\log\left(\frac{dP^{H_n}}{dP^{H^*}}\right) &= \sum_{i=1}^{n}\log\left(\frac{(1-a_n)f^*(Y_i|X_i)g(X_i)+a_nq(Y_i|X_i)g(X_i)}{f^*(Y_i|X_i)g(X_i)}\right)\\ &= \sum_{i=1}^{n}\log\left(1+a_n\frac{q(Y_i|X_i)-f^*(Y_i|X_i)}{f^*(Y_i|X_i)}\right)\\ &= \sum_{i=1}^{n}\log(1+Z_i),\end{aligned}$$

where

$$Z_i = a_n\frac{q(Y_i|X_i)-f^*(Y_i|X_i)}{f^*(Y_i|X_i)}.$$

A Taylor expansion of $\log(1+Z_i)$ around 1 yields

$$\log\left(\frac{dP^{H_n}}{dP^{H^*}}\right) = \sum_{i=1}^{n}\left(Z_i - \frac{1}{2}Z_i^2 + \frac{1}{6}Z_i^3\frac{2}{(1+\xi_i)^3}\right)$$

for some $\xi_i \in [-Z_i, Z_i]$, as $\log(1) = 0$.Next, we show that by the central limit theorem $\sum_{i=1}^{n} Z_i$ converges in distribution to a normally distributed random variable, that by the law of large numbers $\sum_{i=1}^{n} Z_i^2$ converges almost surely to a constant, and that the remaining summand

$$Z_n^* := \sum_{i=1}^{n}\left(\frac{1}{6}Z_i^3\frac{2}{(1+\xi_i)^3}\right)$$

converges to 0 in probability.

First consider the expectation of the random variable $\tilde{Z}_i = \delta\frac{q(Y_i|X_i)-f^*(Y_i|X_i)}{f^*(Y_i|X_i)}$ under P:

$$\begin{aligned}\mathbb{E}(\tilde{Z}_i) &= \delta\int\int\frac{q(y|x)-f^*(y|x)}{f^*(y|x)}f^*(y|x)d\mu(y)dG(x)\\ &= \delta\int\int(q(y|x)-f^*(y|x))d\mu(y)dG(x)\\ &= \delta\int\int q(y|x)d\mu(y)dG(x) - \delta\int\int f^*(y|x)d\mu(y)dG(x) = \delta-\delta = 0.\end{aligned}$$

Furthermore, note that with (3.4) we have that

$$\mathbb{E}(|\tilde{Z}_i|^p) = \mathbb{E}\left(\left|\frac{q(Y_i|X_i)}{f^*(Y_i|X_i)} - 1\right|^p\right) < \infty$$

for all values of p. Since $\sum_{i=1}^n Z_i = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Z}_i$, we can directly apply the standard central limit theorem for i.i.d. random variables and get

$$\sum_{i=1}^n Z_i \xrightarrow{d} N(0, \mathbb{E}((\tilde{Z}_i)^2)).$$

Moreover, we have $\sum_{i=1}^n Z_i^2 = \frac{1}{n} \sum_{i=1}^n (\tilde{Z}_i)^2$. As $\mathbb{E}((\tilde{Z}_i)^2) < \infty$, the usual strong law of large numbers yields

$$\sum_{i=1}^n Z_i^2 \rightarrow_{a.s.} \mathbb{E}((\tilde{Z}_i)^2).$$

To show convergence of the remaining summand to 0, we apply Markov's inequality to Z_n^*. For any $\epsilon > 0$, we have

$$\frac{\mathbb{E}(|Z_n^*|)}{\epsilon} \leq \frac{1}{6} \frac{1}{\sqrt{n}\epsilon} \mathbb{E}(|\tilde{Z}_i^3|) \mathbb{E}\left(\frac{2}{|(1+\xi_i)^3|}\right). \tag{B.1}$$

Since we have $\xi_i \in [-Z_i, Z_i]$, and since (3.3) implies that

$$|Z_i| = \frac{1}{\sqrt{n}} \delta \left|\frac{q(Y_i|X_i)}{f^*(Y_i|X_i)} - 1\right| \leq C_1/\sqrt{n}$$

for a constant C_1, we obtain the bound $|\xi_i| < 1$ for sufficiently large n. Therefore, the expression on the righthand side of (B.1) is bounded, and we find that for arbitrary $\epsilon > 0$, sufficiently large n and a constant C_2 is holds that

$$\mathbb{P}(|Z_n^*| > \epsilon) \leq \frac{\mathbb{E}(|Z_n^*|)}{\epsilon} \leq \frac{C_2}{\sqrt{n}},$$

and thus $Z_n^* \xrightarrow{p} 0$, as claimed. Taken together, we now thus shown that

$$\log\left(\frac{dP^{H_n}}{dP^{H^*}}\right) \xrightarrow{d} N(-\frac{1}{2}\mathbb{E}((\tilde{Z}_i)^2), \mathbb{E}((\tilde{Z}_i)^2)),$$

which completes the argument.

References

Albrecht, J., A. Van Vuuren, and S. Vroman (2009): "Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands," *Labour Economics*, 16(4), 383–396.

Andrews, D. (1997): "A conditional Kolmogorov test," *Econometrica*, 65(5), 1097–1128.

Bickel, P., and J.-J. Ren (2001): "The bootstrap in hypothesis testings," *Lecture Notes-Monograph Series*, 36, 91–112.

Bierens, H. (1990): "A consistent conditional moment test of functional form," *Econometrica*, pp. 1443–1458.

Bierens, H., and D. Ginther (2001): "Integrated conditional moment testing of quantile regression models," *Empirical Economics*, 26(1), 307–324.

Bierens, H., and W. Ploberger (1997): "Asymptotic theory of integrated conditional moment tests," *Econometrica*, 65(5), 1129–1151.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009): "Inference on Counterfactual Distributions," *Working Paper*.

Chernozhukov, V., and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245–261.

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64(5), 1001–1044.

Escanciano, J., and C. Goh (2010): "Specification Analysis of Structural Quantile Regression Models," *Working Papers*.

Escanciano, J., and C. Velasco (2010): "Specification tests of parametric dynamic conditional quantiles," *Journal of Econometrics*, p. forthcoming.

Firpo, S., N. Fortin, and T. Lemieux (2010): "Decomposition Methods in Economics," *NBER Working Paper*.

Foresi, S., and F. Peracchi (1995): "The Conditional Distribution of Excess Returns: An Empirical Analysis.," *Journal of the American Statistical Association*, 90(430), 451–466.

He, X., and L. Zhu (2003): "A lack-of-fit test for quantile regression," *Journal of the American Statistical Association*, 98(464), 1013–1022.

Horowitz, J., and V. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69(3), 599–631.

——— (2002): "An adaptive, rate-optimal test of linearity for median regression models," *Journal of the American Statistical Association*, 97(459), 822–835.

Härdle, W., and E. Mammen (1993): "Comparing nonparametric versus parametric regression fits," *The Annals of Statistics*, 21(4), 1926–1947.

Koenker, R. (2005): *Quantile regression*. Cambridge University Press.

Koenker, R., and G. Bassett (1978): "Regression quantiles," *Econometrica*, 46(1), 33–50.

Machado, J., and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(4), 445–465.

Melly, B. (2005): "Decomposition of differences in distribution using quantile regression," *Labour Economics*, 12(4), 577–590.

Rothe, C. (2010a): "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics*, 155(1), 56–70.

——— (2010b): "Partial Distributional Policy Effects," *TSE Working Paper*.

Stephens, M. (1974): "EDF statistics for goodness of fit and some comparisons," *Journal of the American Statistical Association*, pp. 730–737.

Van der Vaart, A. (2000): *Asymptotic Statistics*. Cambridge University Press.

Van der Vaart, A., and J. Wellner (1996): *Weak convergence and empirical processes: with applications to statistics*. Springer Verlag.

Whang, Y.-J. (2006): "Consistent Specification Testing for Quantile Regression Models," in *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, ed. by D. Corbae, S. Durlauf, and B. Hansen. Cambridge University Press.

Zheng, J. (1998): "A consistent nonparametric test of parametric regression models under conditional quantile restrictions," *Econometric Theory*, 14(1), 123–138.