Misspecification Testing in a Class of Conditional Distributional Models

Christoph Rothe and Dominik Wied*

Columbia University and TU Dortmund

This Version: September 28, 2012

Abstract

We propose a specification test for a wide range of parametric models for the 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 a 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 of order $n^{-1/2}$ from the null hypothesis, and does not require the choice of smoothing parameters. In an empirical application, we use our test to study the validity of various models for the conditional distribution of wages in the US.

Keywords: Cramer-von Mises Distance, Quantile Regression, Distributional Regression, Location-Scale Model, Bootstrap, Wage Distribution

*Christoph Rothe, Columbia University, Department of Economics, 420 W 118th St, New York, NY 10027, USA. Email: cr2690@columbia.edu. Dominik Wied, TU Dortmund, Fakultät Statistik, CDI-Gebäude, D-44221 Dortmund, Germany. Email: wied@statistik.tu-dortmund.de. We would like to thank the associate editor, the referees, Roger Koenker and Blaise Melly for helpful comments. Financial support by Deutsche Forschungsgemeinschaft (SFB 823, project A1) is gratefully acknowledged.

1. Introduction

In this paper, we propose a general principle to construct omnibus specification test for a wide range of parametric models for a conditional distribution function. For $Y \in \mathbb{R}$ an outcome variable and $X \in \mathbb{R}^K$ a vector of covariates, our interest is in testing the validity of a model that asserts that there exists a possibly function-valued parameter θ such that

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

where $F(\cdot|\cdot, \theta)$ is a known function and $\mathcal{Z}$ denotes the support of $Z = (Y, X)$. We refer to any such specification as a *conditional distributional model* (CDM). The alternative hypothesis is that equation (1.1) is violated for at least one value $(y, x) \in \mathcal{Z}$. Allowing unknown parameters to be function-valued is important in this setting, and constitutes one of the main innovations of our approach. We also discuss an extension of our procedure that allows to test the hypothesis in (1.1) for all (y, x) in some set $S \subset \mathcal{Z}$ chosen by the analyst. Through an appropriate choice of S, one can test whether the parametric model provides an adequate fit over a particular range of the conditional distribution function, such as e.g. the area below or above the conditional median.

Our general setting covers a wide range of CDMs that are of great importance in empirical applications. The leading example is certainly the *linear quantile regression model* (Koenker and Bassett, 1978; Koenker, 2005), which implies a linear structure for the inverse of the conditional CDF, namely that $F^{-1}(\tau|x, \theta) = x'\theta(\tau)$ for some functional parameter $\theta(\cdot)$ that is strictly increasing in each of its components. Nonlinear versions of quantile regression could be considered as well. Another example is the *linear location-scale shift model* (Koenker and Xiao, 2002), under which $F^{-1}(\tau|x, \theta) = x'\beta + x'\gamma Q_\epsilon(\tau)$, with Q_ϵ a univariate quantile function and $\theta(\cdot) = (\beta, \gamma, Q_\epsilon(\cdot))$. Our setting also covers the *distributional regression model* (Foresi and Peracchi, 1995), where the conditional CDF is modeled by a series of binary response models with varying "cutoffs". That is, the conditional CDF is specified as $F(y|x, \theta) = \Lambda (x'\theta(y))$, where Λ is a known strictly increasing link function such as e.g. the logistic or standard normal distribution function, and $\theta(\cdot)$ is again a function-valued parameter. This class of models has recently received considerable attention in the econometrics literature (e.g. Chernozhukov et al., 2009; Koenker, 2010; Fortin et al., 2011; Rothe, 2011).

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 CDM. For example, to test the validity of the linear quantile regression model we would first obtain an estimate of the conditional CDF of Y given X by inverting the estimated conditional quantile function, and then transform this object into an estimate of the joint CDF of (Y, X) by "integrating up" the conditioning argument. This restricted CDF estimate can then be compared to the joint empirical distribution function of the data, which is the most natural unrestricted estimate.

Our test statistic is a Cramer-von Mises type measure of distance between the two above-mentioned objects, and is therefore called a *Generalized Conditional Cramer-von Mises* (GCCM) test. 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 CDMs of the type considered in this paper is critical in many areas of applied statistics. In economics, for example, such specifications are employed extensively to study differentials in the distribution of wages between two time periods, or two subgroups of a particular population (see e.g. Machado and Mata, 2005,Chernozhukov et al., 2009 or Rothe, 2011). 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 over the conditioning argument 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 computed. Our concern is the implementation of the first step of this procedure. For example, the Machado and Mata (2005) decomposition 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, 2010, 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 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 et al., 2009). Our testing procedure can be used to formally investigate this issue.

As an additional contribution, our paper provides some empirical evidence on the last point: using US data from the Current Population Survey, we show that typical specifications of linear location-scale models and linear quantile regressions containing a rich set of covariates are frequently rejected by our GCCM test even for small and moderate sample sizes. On the other hand, we find that the distributional regression model, which has thus far received only limited interest in the literature, typically cannot be rejected in such settings.

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, Stute, 1997 and Horowitz and Spokoiny, 2001), and for the conditional quantile function at *one* particular quantile, such as the median (see e.g. Zheng, 1998, Bierens and Ginther, 2001, Horowitz and Spokoiny, 2002, He and Zhu, 2003, and Whang, 2006). In comparison, the related problem of testing the validity of a model for the entire conditional distribution function that we study in this paper has received much less attention. Andrews (1997) proposes a test for CDMs indexed by finite-dimensional parameters. Koenker and Machado (1999) and Koenker and Xiao (2002) consider specification testing in a quantile regression context, and propose tests for e.g. the validity of the location-scale model, but not the validity of the quantile regression model itself. Galvao et al. (2011) test for threshold effects in linear quantile regression in a time series context. Escanciano and Velasco (2010) and Escanciano and Goh (2010) both propose testing procedures for the null hypothesis that a conditional quantile restriction is valid over a range of quantiles in different settings, that both include the usual quantile regression model with independent observations as a special case. We are not aware of any paper that provides a general approach to testing the validity of CDMs indexed by possibly function-valued parameters.

2. Testing General Conditional Distributional Models

2.1. Testing Problem. We observe an outcome variable $Y_i \in \mathbb{R}$ and a vector of explanatory variables $X_i \in \mathbb{R}^K$ for $i = 1, \dots, n$. The cumulative distribution function (CDF) of $Z_i = (Y_i, X_i)$ and X_i are denoted by H and G, respectively. We assume throughout the paper that the data points are independent and identically distributed. Our aim is to test the validity of certain classes of parametric specifications for the conditional CDF F of Y_i given X_i. Let $\mathcal{F}$ be the class of all conditional distribution functions on the support of Y given X that satisfy certain weak regularity conditions given below, and consider a CDM

$$\mathcal{F}^0 = \{F(\cdot|\cdot, \theta) \text{ for some } \theta \in \mathcal{B}(\mathcal{U}, \Theta)\} \subset \mathcal{F},$$

i.e. a family of conditional distribution functions indexed by a (potentially) functional parameter θ taking values in $\mathcal{B}(\mathcal{U}, \Theta)$, the class of mappings $u \mapsto \theta(u)$ such that $\theta(u) \in \Theta \subset \mathbb{R}^p$ for $u \in \mathcal{U} \subset \mathbb{R}$. The hypothesis that we want to test is that F coincides with an element of $\mathcal{F}^0$:

$$\mathcal{H}_0 : F(y|x) = F(y|x,\theta) \text{ for some } \theta \in \mathcal{B}(\mathcal{U},\Theta) \text{ and all } (y,x) \in \mathcal{Z} \tag{2.1}$$

$$\text{vs. } \mathcal{H}_1 : F(y|x) \neq F(y|x,\theta) \text{ for all } \theta \in \mathcal{B}(\mathcal{U},\Theta) \text{ and some } (y,x) \in \mathcal{Z}. \tag{2.2}$$

This paper proposes a testing procedure of the problem in (2.1)–(2.2) for CDMs in which the true value of the functional parameter is identified under the null hypothesis through a moment condition. Specifically, let $\psi : \mathcal{Z} \times \Theta \times \mathcal{U} \mapsto \mathbb{R}^p$ be a uniformly integrable function whose exact form depends on $\mathcal{F}^0$, and suppose that for every $u \in \mathcal{U}$ the equation

$$\Psi(\theta, u) := \mathbb{E}\left(\psi(Z,\theta,u)\right) = 0 \tag{2.3}$$

has a unique solution $\theta_0(u)$. We assume that under the null hypothesis any value $\theta \in \mathcal{B}(\mathcal{U},\Theta)$ of the functional parameter that satisfies $F(y|x) = F(y|x,\theta)$ for all $(y,x) \in \mathcal{Z}$ also satisfies $\theta(u) = \theta_0(u)$ for all $u \in \mathcal{U}$. The moment condition (2.3) thus uniquely determines the value of the "true" functional parameter. Under the alternative, θ_0 remains well-defined as the solution to (2.3), and can thus be thought of as a pseudo-true value of the functional parameter in this case. General results on estimation and inference in this class of models are derived in Chernozhukov et al. (2009). We discuss in Section 4 how a large class of empirically relevant specifications fits into this framework.

2.2. Test Statistic. To motivate a test statistic for the problem in (2.1)–(2.2), note that, following the above discussion, our null hypothesis can be equivalently stated as

$$F(y|x) = F(y|x,\theta_0) \text{ for all } (y,x) \in \mathbb{R}^{K+1}, \tag{2.4}$$

with $\theta_0(u)$ the unique solution to (2.3). This holds because $F(\cdot|\cdot,\theta_0)$ is the only element of $\mathcal{F}^0$ that is a potential candidate value for the true conditional CDF F of Y_i given

X_i by assumption. Since $F(y|x) = \mathbb{E}(\mathbb{I}\{Y \leq y\}|X = x)$, where $\mathbb{I}\{a \leq b\}$ denotes the indicator for the event that each component of a is weakly smaller than the corresponding component of b, equation (2.4) is a restriction on the conditional moments of Y given X, that can be transformed into a restriction on unconditional moments without loss of information by "integrating up" with respect to the marginal distribution G of the conditioning variable X. Writing

$$H(y,x) = \int F(y|t)\mathbb{I}\{t \leq x\}dG(t) \text{ and } H^0(y,x) = \int F(y|t,\theta_0)\mathbb{I}\{t \leq x\}dG(t)$$

for the joint CDF of (Y, X) and another CDF that imposes the structure implied by the CDM, respectively, it follows from Billingsley (1995, Theorem 16.10(iii)) that our testing problem (2.1)–(2.2) can be restated as

$$\begin{aligned} &\mathcal{H}_0 : H(y,x) = H^0(y,x) \text{ for all } (y,x) \in \mathbb{R}^{K+1} \\ \text{vs. } &\mathcal{H}_1 : H(y,x) \neq H^0(y,x) \text{ for some } (y,x) \in \mathbb{R}^{K+1}. \end{aligned}$$

Of course, the idea to transform conditional moment restrictions into unconditional ones is not new in the literature on specification testing. It is used for example by Stute (1997) in the context of testing parametric specifications of conditional expectation functions.

Given the above representation of the testing problem, we propose to construct a specification test for the CDM $\mathcal{F}^0$ based on a Cramer-von Mises type measure of distance between estimates H and H^0, scaled by the sample size. Specifically, the test statistic T_n is defined as

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

where

$$\widehat{H}_n(y,x) = n^{-1}\sum_{i=1}^{n}\mathbb{I}\{Y_i \le y\}\mathbb{I}\{X_i \le x\} \text{ and}$$
$$\widehat{H}_n^0(y,x) = n^{-1}\sum_{i=1}^{n}\widehat{F}_n(y|X_i)\mathbb{I}\{X_i \le x\},$$

with $\widehat{F}_n(y|x) = F(y|x,\widehat{\theta}_n)$ a parametric estimate of F based on an estimate $\widehat{\theta}_n$ of θ_0. Here we take $\widehat{\theta}_n$ to be an approximate Z-estimator satisfying

$$\|\widehat{\Psi}_n(\widehat{\theta}_n(u),u)\| = \inf_{\theta\in\Theta}\|\widehat{\Psi}_n(\theta,u)\| + \eta_n \tag{2.5}$$

for some (possibly random) variable $\eta_n = o_p(n^{-1/2})$ and every $u \in \mathcal{U}$. The function $\widehat{\Psi}_n(\theta,u) := n^{-1}\sum_{i=1}^{n}\psi(Z_i,\theta,u)$ is the sample analogue of the moment condition in (2.3). This approach is feasible for all examples we consider in this paper.

Under general conditions described in the following section, the random process $(y,x) \mapsto \widehat{H}_n(y,x) - \widehat{H}_n^0(y,x)$ converges to zero in probability under the null hypothesis, and to a non-zero probability limit under the alternative. Large realizations of T_n thus indicate a possible violation of the null. Since our testing principle shares some similarities with the Conditional Kolmogorov test in Andrews (1997), 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 suggested that Cramer-von Mises type statistics have somewhat better power properties than those based on the Kolmogorov distance.

2.3. Bootstrap Critical Values. As we show in more detail below, for most common CDMs the asymptotic null distribution of T_n is non-pivotal and depends on the data generating process in a complex fashion. We therefore propose to obtain critical values for our test via a semiparametric bootstrap procedure. Such a procedure is computationally expensive but convenient from a practical point of view, since it avoids the complicated problem of estimating null distribution of T_n directly. While the approach cannot be expected to yield any higher-order improvements over a standard large sample test, one could of course always improve the level accuracy by using a form of the iterated bootstrap (see e.g. Hall, 1986 or Beran, 1988).

The idea of our semiparametric bootstrap is to use the restricted estimate $\widehat{H}_n^0$ as the bootstrap distribution to ensure that the bootstrap mimics the distribution of the data under the null hypothesis, even though the data might be generated by an alternative distribution. A bootstrap realization of our test statistic is computed as follows:

Step 1: Draw a bootstrap sample of covariates $\{X_{b,i}, 1 \leq i \leq n\}$ with replacement from the realized values $\{X_i, 1 \leq i \leq n\}$.

Step 2: For every $1 \leq i \leq n$ put $Y_{b,i} = \widehat{F}_n^{-1}(U_{b,i}|X_{b,i})$, where $\{U_{b,i}, 1 \leq i \leq n\}$ is a simulated i.i.d. sequence of standard uniformly distributed random variables.

Step 3: Use the bootstrap data $\{(Y_{b,i}, X_{b,i}), 1 \leq i \leq n\}$ to compute estimates $\widehat{H}_{b,n}$ and $\widehat{H}_{b,n}^0$ of H and H^0, respectively, exactly as described in the previous subsection, and compute the corresponding bootstrap realization of the test statistic:

$$T_{b,n} = n \int (\widehat{H}_{b,n}(y,x) - \widehat{H}_{b,n}^0(y,x))^2 d\widehat{H}_n(y,x).$$

The distribution of $T_{b,n}$ can be determined through the usual repeated resampling of the data, and, as shown formally below, then be used as an approximation to the distribution of T_n under the null hypothesis for a wide range of CDMs. The test then rejects $\mathcal{H}_0$ if $T_n > \widehat{c}_n(\alpha)$ for some pre-specified significance level $\alpha \in (0,1)$, where the critical value $\widehat{c}_n(\alpha)$ is the smallest constant that satisfies $P_b\left(T_{b,n} \leq \widehat{c}_n(\alpha)\right) \geq 1 - \alpha$, and P_b is the probability with respect to bootstrap sampling.

3. Theoretical Properties

This section shows that the GCCM test has correct asymptotic size, is consistent against fixed alternatives, and has 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 $\widetilde{F}$" whenever the joint distribution function of $Z = (Y, X)$ is given by $\widetilde{H}(y, x) = \int \widetilde{F}(y|t)\mathbb{I}\{t \leq x\}dG(t)$ for some $\widetilde{F} \in \mathcal{F}$, and denote the expectation taken with respect to any such CDF $\widetilde{H}$ by $\mathbb{E}_{\widetilde{H}}$. All limits are taken as $n \to \infty$.

3.1. Limiting Distribution of the Test Statistic. To derive large sample properties of our test statistic we impose the following assumptions.

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

Assumption 2. *For each* $u \in \mathcal{U}$, *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* $(\theta, u) \in \Theta \times \mathcal{U}$ *with probability one, and the mapping* $(\theta, u) \mapsto \Psi(\theta, u)$ *is continuously differentiable at* $(\theta_0(u), u)$ *with a uniformly bounded derivative on* $\mathcal{U}$ *(where differentiability in* u *is only required when* $\mathcal{U}$ *is not finite). The function* $\dot{\Psi}(\theta, u) := \partial_\theta \Psi(\theta, u)$ *is nonsingular at* $\theta_0(\cdot)$ *uniformly over* $u \in \mathcal{U}$.

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

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

Assumptions 1–4 are standard regularity conditions also imposed by Chernozhukov et al. (2009). They ensure that a Functional Central Limit Theorem can be applied to the Z-estimator process $u \mapsto \sqrt{n}(\hat{\theta}_n(u) - \theta_0(u))$. 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 $(y, x) \mapsto \sqrt{n}(\widehat{H}_n^0(y, x) - H^0(y, x))$ also converges to a Gaussian limit process $\mathbb{G}_2$. This convergence can be shown to be jointly with that of the empirical CDF process $(y, x) \mapsto \sqrt{n}(\widehat{H}_n(y, x) - H(y, x))$ to an H-Brownian Bridge $\mathbb{G}_1$. The limiting distribution of our test statistic T_n then follows from an application of the continuous mapping theorem, and from the fact that the functions H and H^0 coincide under the null hypothesis, but differ on a set with positive probability under the alternative.

Theorem 1. *Under Assumption 1–5, the following statements hold.*

i) Under the null hypothesis, i.e. when the data are distributed according to some F that satisfies (2.1),

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

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 that satisfies (2.2),

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

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

$$Q_n(y|x) = (1 - \delta/\sqrt{n})F^*(y|x) + (\delta/\sqrt{n})Q(y|x), \tag{3.1}$$

where F^* is a CDF such that $F^*(y|x) = F(y|x, \theta)$ for some $\theta \in \mathcal{B}(\mathcal{U}, \Theta)$ and all $(y, x) \in \mathcal{Z}$, Q is a CDF such that $Q(y|x) \neq F(y|x, \theta)$ for all $\theta \in \mathcal{B}(\mathcal{U}, \Theta)$ and some $(y, x) \in \mathcal{Z}$, and $\delta \leq n^{1/2}$ is some constant, satisfying the following assumption.

Assumption 6. *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* (3.1) *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 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 that $\sup_{\{(x,y):f^*(y|x)>0\}} q(y|x)/f^*(y|x) < \infty$. Intuitively, this would be the case when Q has lighter tails than F^*. See Rothe and Wied (2012) for a formal proof.

The following theorem shows that under local alternatives of the form (3.1) the limiting distribution of T_n contains an additional deterministic shift function ensuring non-trivial 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 $u \in \mathcal{U}$, respectively.

Theorem 2. *Under Assumption 1–6, and if the data are distributed according to a local alternative Q_n as given in* (3.1),

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

where $\mu(y, x) = \delta \int (Q(y|t) - F(y|t, \theta_*) + \dot{F}(y|t, \theta_*)[h])\mathbb{I}\{t \leq x\}dG(t)$ and the function h is given by $h(u) = \partial_{\theta'}\Psi_{F^*}(\theta_*(u), u)^{-1}\Psi_Q(\theta_*(u), u)$.

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 the estimator $\widehat{F}_n$ under Q. In this case, we have

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

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

which is proportional to the difference between the joint CDFs 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 our test is consistent. Finally, since contiguity preserves convergence in probability, it follows from Assumption 6 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 1–6, 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 CDF F that satisfies (2.1)*, we have that*

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

ii) Under any fixed alternative, i.e. when the data are distributed according to some CDF F that satisfies (2.2)*, we have that*

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

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

CDF Q_n that satisfies (3.1)*, we have that*

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

4. Application to Specific Models

In this subsection, we discuss a number of conditional distributional models whose correct specification can be investigated via our GCCM test. We also provide primitive conditions that imply the "high-level" conditions in Assumption 1–5 that we used to derive asymptotic properties.

4.1. Quantile Regression. Arguably the most important example of a conditional distributional model indexed by function-valued parameters in the sense of this paper is the linear quantile regression model (Koenker and Bassett, 1978; Koenker, 2005). It postulates that the conditional u-quantile of Y given $X = x$ is linear in a vector of parameters that vary with u:

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

Such a model is correctly specified if the true data generating process can 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 $\widehat{\theta}_n(u)$ of $\theta_0(u)$, given by

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

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 (2.5) with $\psi(Z_i, \theta, u) = (u - \mathbb{I}\{Y_i - X_i'\theta \leq 0\})\,X_i$. The conditional distribution function implied by the linear quantile regression model can then be obtained as $F(y|x, \widehat{\theta}_n) = \int_0^1 \mathbb{I}\{x'\widehat{\theta}_n(u) \leq$

$y\}du$. Note that $F(y|x, \widehat{\theta}_n)$ is monotone in y by construction for every x, even if the estimated quantile curve $u \mapsto x'\widehat{\theta}_n(u)$ is not. The test statistic T_n can then be computed in a straightforward fashion. Our asymptotic analysis in Section 3 applies to the linear quantile regression example under the conditions of the following theorem.

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 $u \in \mathcal{U} = (0,1)$, almost surely. (ii) The matrix $\mathbb{E}(XX')$ is finite and of full rank, (iii) the parameter $\theta_0(\cdot)$ solving $\mathbb{E}\left(\psi\left(Z, \theta_0(u), u\right)\right) = 0$ 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 linear quantile regression model with $\dot{F}(y|x,\theta)[h(\cdot)] = -f(y|x)x'\left[h\left(F(y|x,\theta)\right)\right]$.*

The role of the conditions in Theorem 4, which are standard in the literature, is essentially ensure that the moment condition $\mathbb{E}\left(\psi\left(Z, \theta_0(u), u\right)\right) = 0$ has a unique solution $\theta_0(u)$ for every $u \in \mathcal{U}$, and that the process $u \mapsto \sqrt{n}(\widehat{\theta}_n(u) - \theta_0(u)$ converges to a Gaussian limit under both the null hypothesis and the alternative. Note that our Theorem imposes strong conditions on the distribution of Y given X in order to ensure that Assumption 1–5 hold with $\mathcal{U} = (0,1)$. If it is unreasonable to assume that $f(\cdot|X)$ is uniformly bounded away from zero (e.g. because the support of Y is unbounded), one could still test the validity of the linear quantile regression model for $u \in (\epsilon, 1-\epsilon)$ for some $\epsilon > 0$ by using an extension of our test described in Section 5 below.

4.2. Location Shift and Location-Scale Shift Models. The testing procedures proposed in this paper can also be used to assess the validity of various special cases of quantile regression. A leading example is the linear location-scale shift model

$$\begin{aligned} \mathcal{F}^0 = \{ & F^{-1}(u|x) = x'\beta + x'\gamma Q_\epsilon(u) \text{ for some} \\ & \theta(u) = (\beta, \gamma, Q_\epsilon(u)) \in \mathbb{R}^{2p+1} \text{ and all } u \in (0,1)\}, \end{aligned} \tag{4.1}$$

with Q_ϵ some univariate quantile function. In this model, covariates affect both the location and the scale of the conditional distribution of Y given X, but have no influence on its shape. Such a model would e.g. be correctly specified if the data are generated as $Y_i = X_i'\beta_0 + (X_i'\gamma_0)\epsilon_i$ for some random variable $\epsilon_i \sim F_\varepsilon$ that is independent of X_i. An important special case in this class is the linear location shift model, for which $(X_i'\gamma) = 1$:

$$\begin{aligned}\mathcal{F}^0 = \{&F^{-1}(u|x) = x'\beta + Q_\epsilon(u) \text{ for some} \\ &\theta(u) = (\beta, Q_\epsilon(u)) \in \mathbb{R}^{p+1} \text{ and all } u \in (0,1)\}.\end{aligned} \tag{4.2}$$

Location and location-scale shift models can be estimated in a variety of different ways. See for example Rutemiller and Bowers (1968), Harvey (1976) or Koenker and Xiao (2002). For simplicity, we restrict attention to simple two- and three-step methods, respectively. In the pure location shift model (4.2), we can estimate the parameter β_0 by ordinary least squares, and the quantile function Q_ϵ by taking the empirical quantile function of the regression residuals. The corresponding estimator $\widehat{\theta}_n(\cdot) = (\widehat{\beta}_n, \widehat{Q}_{\epsilon,n}(\cdot))$ is contained in the class of approximate Z-estimators we consider in this paper, as it satisfies (2.5) with $\psi(Z_i, (\beta, \alpha), u) = [u - \mathbb{I}\{Y_i - X_i'\beta \leq \alpha\}, \varepsilon_i(\beta)X_i]$ with $\varepsilon_i(\beta) = Y_i - X_i'\beta$. For the location-scale shift model in (4.1), we continue to estimate β_0 by OLS, estimate γ_0 by nonlinear regression of $\varepsilon_i(\widehat{\beta}_n)^2$ on $(X_i'\gamma)^2$, and obtain an estimate $\widehat{Q}_{\epsilon,n}$ via the empirical quantile function of the standardized regression residuals $\varepsilon_i(\widehat{\beta}_n)/(X_i'\widehat{\gamma}_n)$. Again, this is a Z-estimator in the sense of this paper with $\psi(Z_i, (\beta, \gamma, \alpha), u) = [u - \mathbb{I}\{\varepsilon_i(\beta)/X_i'\gamma \leq \alpha\}, \varepsilon_i(\beta)X_i, (\varepsilon_i(\beta)^2 - (X_i'\gamma)^2)X_i'\gamma X_i]$. The following theorem gives conditions for the validity of the "high level" conditions in Section 3 in the location-scale shift case. Conditions for the pure location shift model are similar, with obvious simplifications.

Theorem 5. *Suppose that (i) the residuals $\epsilon_i = (Y_i - X_i'\beta)/(X_i'\gamma)$ are continuously distributed with density function f_ϵ, which is continuous, bounded and bounded away from zero at $Q_\epsilon(u)$, uniformly over $u \in (0,1)$, almost surely, (ii) $P(X_i'\gamma_0 > 0) = 1$, (iii) the* *matrix* $\mathbb{E}(XX')$ *is finite and of full rank, (iv)* $\mathbb{E}(Y^2)$ *is finite, and (v) the parameter* $\theta_0(\cdot) = (\beta_0, \gamma_0, Q_\epsilon(\cdot))$ *solving* $\mathbb{E}\left(\psi\left(Z, \theta_0(u), u\right)\right) = 0$ *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 linear location-scale shift model with* $\dot{F}(y|x,\theta)[h(\cdot)] = -f_\epsilon((y - x'\beta)/(x'\gamma))(x'\beta + x'\gamma Q_\epsilon(h(F(y|x,\theta))))$.

4.3. Distributional Regression. Another class of CDMs covered by our framework are so-called distributional regression models, which were introduced by Foresi and Peracchi (1995). These models have recently received considerable interest in the economics literature, as they conveniently allow to model certain features of conditional wage distributions, such as nonlinearities around the level of the minimum wage (e.g. Chernozhukov et al., 2009; Rothe, 2011). The basic idea is to directly model the conditional CDF of Y given X through a family of binary response models for the event that the dependent variable Y exceeds some threshold $y \in \mathbb{R}$. More specifically the model is given by

$$\mathcal{F}^0 = \{F(y|x) = \Lambda(x'\theta(y)) \text{ for some } \theta(y) \in \Theta \subset \mathbb{R}^p \text{ and all } y \in \mathbb{R}\}, \tag{4.3}$$

where $\Lambda(\cdot)$ is a known strictly increasing link function, e.g. the logistic or standard normal distribution function, or simply the identity function. Distributional regression models can differ substantially from a classical quantile regression model using the same set of covariates, and no class of models is more general than the other. Both classes, however, contain the location shift model as special case. See Koenker (2010) for some theoretical comparisons of the two approaches in this context. An advantage of distributional regression relative to quantile regression is that it does not require the dependent variable to be continuously distributed, which can be an important in certain empirical applications.

Since for every threshold value $y \in \mathbb{R}$ the distributional regression model resembles a standard binary response model, it can be fitted the same way one would e.g. proceed with a logistic regression. A natural estimator for the functional parameter $\theta_0(\cdot)$ is the "pointwise" maximum likelihood estimator $\widehat{\theta}_n(\cdot)$, which solves the equation $\|\widehat{\Psi}_n(\widehat{\theta}_n(y), y)\| = 0$, with $\widehat{\Psi}_n(\theta, y) := n^{-1}\sum_{i=1}^n \psi(Z_i, \theta, y)$ and $\psi(Z_i, \theta, y) = \left(\Lambda(X_i'\theta)\left(1 - \Lambda(X_i'\theta)\right)\right)^{-1} \times$

$(\Lambda(X_i'\theta) - \mathbb{I}\{Y_i \leq y\})\,\lambda(X_i'\theta)X_i$ the usual score function, and λ the derivative of Λ. The estimated conditional CDF of Y given X is then given by $\widehat{F}_n(y|x) = \Lambda(x'\widehat{\theta}_n(y))$, and the test statistic T_n is straightforward to compute from this expression. The following theorem gives conditions for the distributional regression model to satisfy the "high level" conditions in Section 3.

Theorem 6. *Suppose that (i) the support $\mathcal{Y}$ of Y is either a finite set or 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 $\mathbb{E}(XX')$ is finite and of full rank, (iv) the parameter $\theta_0(\cdot)$ solving $\mathbb{E}\left(\psi\left(Z, \theta_0(y), y\right)\right) = 0$ is such that $\theta_0(y)$ is in the interior of the parameter space Θ for every $y \in \mathcal{Y}$, and (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* (4.3) *with* $\dot{F}(y|x,\theta)[h(\cdot)] = \left(\partial\Lambda\left(x, \theta(y)\right)/\partial\theta'\right)[h(y)]$.

5. Extension: Testing over a Subset of the Support

In some applications, it is not only interesting to test the validity of a CDM for the entire conditional CDF, but also its adequacy over some range of the conditional distribution. For example, for models formulated in terms of conditional quantiles, one might be interested in whether the model is correctly specified for all conditional u-quantiles with $u \in (u_L, u_U)$ and $0 < u_L < u_U < 1$. Another question that might be of interest is whether the parametric model correctly describes the conditional CDF on the subset of the support where Y and/or some components of X takes values in a particular interval. To accommodate such settings, we can consider the following generalization of our testing problem (2.1)–(2.2):

$$\mathcal{H}_0 : F(y|x) = F(y|x,\theta) \text{ for some } \theta \in \mathcal{B}(\mathcal{U},\Theta) \text{ and all } (y,x) \in S \tag{5.1}$$

$$\text{vs. } \mathcal{H}_1 : F(y|x) \neq F(y|x,\theta) \text{ for all } \theta \in \mathcal{B}(\mathcal{U},\Theta) \text{ and some } (y,x) \in S \tag{5.2}$$

for some suitably chosen closed and connected set $S \subset \mathcal{Z}$. The two above-mentioned examples correspond to choosing $S = \{(y,x) : F^{-1}(u^L|x) \le y \le F^{-1}(u^U|x)\}$ for $0 < u_L < u_U < 1$, and $S = \{(y,x) : y^L \le y \le y^U, x^L \le x \le x^U\}$ for some $-\infty \le y^L < y^U \le \infty$ and $-\infty \le x^L < x^U \le \infty$, respectively. Of course, other choices of S are possible as well.

Our GCCM test can be adapted to the modified testing problem in (5.1)–(5.2) as follows. First, it might be necessary to modify the function ψ such that $\widehat{\theta}_n$ remains consistent for a population value θ_0 satisfying $F(y|x) = F(y|x,\theta_0)$ for all $(y,x) \in S$ under the null hypothesis. The details of this step critically depend on the type of CDM under consideration, and also the exact form of the set S. It is therefore difficult to give a general recipe. For example, when testing the linear quantile regression specification and $S = \{(y,x) : y \ge y_L\}$ for some constant $y_L \in \mathbb{R}$ one could e.g. work with the censored quantile regression estimator of Powell (1986). Second, one has to redefine the test statistic such that the difference $\widehat{H}_n(y,x) - \widehat{H}_n^0(y,x)$ is only evaluated over S. This can be accomplished by simply putting

$$T_n = n \int_S (\widehat{H}_n(y,x) - \widehat{H}_n^0(y,x))^2 d\widehat{H}_n(y,x) = \sum_{i=1}^n \mathbb{I}\{(Y_i, X_i) \in S\}(\widehat{H}_n(y,x) - \widehat{H}_n^0(y,x))^2.$$

Third, one has to modify the bootstrap sampling scheme in order to impose the new null hypothesis (5.1). To do so, one can obtain a bootstrap data set $\{(Y_{b,i}, X_{b,i}), 1 \le i \le n\}$ by i.i.d. sampling from the CDF $\widehat{H}_n^*$, where

$$\widehat{H}_n^*(y,x) = \begin{cases} \widehat{H}_n(y,x) & \text{if } (y,x) \notin S \\ \widehat{H}_n^0(y,x) & \text{if } (y,x) \in S, \end{cases}$$

and proceed as before with the new data set. Theoretical properties analogous to those derived in Section 3 can be established for the modified testing procedure using the same type of arguments. If the set S is unknown, it can be replaced in the steps outlined above by some consistent estimate $\widehat{S}_n$. It can be shown that this does not affect the test's asymptotic properties as long as $\widehat{S}_n$ satisfies the weak regularity condition that

$n^{-1}\sum_{i=1}^{n}(\mathbb{I}\{(Y_i,X_i)\in\widehat{S}_n\}-\mathbb{I}\{(Y_i,X_i)\in S\})\xrightarrow{p}0.$

6. Simulation Results

6.1. Setup. 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. In particular, we simulate a dependent variable Y according to one of the following data generating processes:

$$
\begin{aligned}
&\text{(DGP1):} \quad Y = X_1 + X_2 + U,\\
&\text{(DGP2):} \quad Y = X_1 + X_2 + V,\\
&\text{(DGP3):} \quad Y = X_1 + X_2 + (.5 + X_1)U,\\
&\text{(DGP4):} \quad Y = X_1 + X_2 + (.5 + X_1 + X_2^2)^{1/2}U,\\
&\text{(DGP5):} \quad Y = X_1 + X_2 + .2(.5 + X_1 + X_2^2)^{3/2} + U.
\end{aligned}
$$

Here $X_1 \sim Binom(1,.5)$, $X_2 \sim N(0,1)$, $U \sim N(0,1)$ and $V = (1-2V_1^*)V_2^*/\sqrt{8}$ with $V_1^* \sim Binom(1,.5)$ and $V_2^* = \chi^2(2)$. The random variables X_1, X_2, U and V are mutually independet. We consider the sample sizes $n = 100$ and $n = 300$, and set the number of replications to 1000. In each simulation run, we use the GCCM test with $B = 199$ bootstrap replications to test the correctness of the four types of models discussed in Section 4: the location shift model (LS), the location-scale shift model (LSS), the linear quantile regression model (QR) and the distributional regression model (DR) with Λ being the standard normal distribution function. For the LS and LSS specification, we also compute the test based on Khmaladzation described in Koenker and Xiao (2002); and for the QR specification we consider the test in Escanciano and Goh (2010)[1]. We are not aware of a competing specification test for the DR model.

Our data generating processes are designed is such a way that a different set of models

[1]Escanciano and Goh (2010) point out that for the case of i.i.d. data we consider in this paper the properties of their procedure are superior to those of tests based on subsampling, as e.g. in Escanciano and Velasco (2010), and hence we do not consider the latter for our simulations

is correctly specified in each of them. DGP1 is a simple location shift model with normally distributed errors, and hence all four specifications are correct in this case. DGP2 is again a simple location shift model, but now the errors follow a mixture of a "positive" and a "negative" χ^2 distribution with 2 degrees of freedom (normalized to have unit variance). We consider this DGP to investigate if our test of the DR specification is able to "pick up" a misspecified link function. DGP3 is a location-scale shift model, and thus the LSS and QR model are correct, whereas the LS and DR specification are not. Finally, under DGP4 and DGP5 all four models considered for this simulation study are misspecified.

6.2. Results. In Table 1 we show the empirical rejection probabilities of our GCCM test and the competing procedures for the nominal levels of 5% and 10%. The GCCM tests generally exhibit good size properties, with rejection rates close to the nominal levels under correct specification. The same is true for Escanciano and Goh's (2010) test of the QR specification. In contrast, the tests for the LS and LSS specification from Koenker and Xiao (2002) seem to be slightly conservative, particularly for $n = 100$. In terms of power, our GCCM-QR test exhibits properties which are roughly on par with those of the test in Escanciano and Goh (2010). The GCCM-LS exhibits rejection rates substantially above those of the corresponding test from Koenker and Xiao (2002). The behavior of the GCCM-LSS test under DGP4 and DGP5 is somewhat peculiar in our simulations, in the sense that it exhibits rejection rates that are substantially below those of the GCCM-QR test, even though one is testing a more restrictive hypothesis in these cases. We conjecture that this is a small sample phenomenon due to the fact that under both DPG4 and DGP5 the error terms tend to contain some large outliers. This causes some instability in the least squares estimates of the scale parameters γ_0, which in turn leads to a loss of power. On the other hand, quantile regression is well-known to be robust against outliers, which seems to be the reason that the corresponding test exhibits better properties in this case. However, our GCCM-LSS still dominates the corresponding test from Koenker and Xiao (2002) in terms of power. Finally, the GCCM-DR test is e.g.

Table 1: Simulation Results: Empirical Rejection Frequencies of the Generalized Conditional Cramer van Mises (GCCM) Test and Several Competing Procedures for Correct Specification of various CDMs.

	GCCM-LS		*GCCM-LSS*		*GCCM-QR*		*GCCM-DR*		*KX-LS*		*KX-LSS*		*EG-QR*	
$n = 100$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DGP1	0.093	0.048	0.095	0.047	0.098	0.045	0.155	0.060	0.067	0.035	0.036	0.017	0.106	0.058
DGP2	0.085	0.033	0.067	0.021	0.087	0.044	0.353	0.207	0.069	0.037	0.045	0.023	0.123	0.066
DGP3	0.829	0.669	0.031	0.010	0.096	0.041	0.908	0.813	0.082	0.047	0.114	0.060	0.101	0.058
DGP4	0.404	0.239	0.136	0.063	0.319	0.199	0.691	0.552	0.097	0.049	0.065	0.031	0.252	0.136
DGP5	0.874	0.746	0.383	0.307	0.913	0.818	0.894	0.813	0.055	0.027	0.052	0.020	0.913	0.827
$n = 300$	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DGP1	0.109	0.056	0.113	0.061	0.099	0.057	0.109	0.057	0.107	0.039	0.064	0.023	0.098	0.039
DGP2	0.096	0.043	0.093	0.042	0.105	0.049	0.501	0.321	0.066	0.024	0.051	0.028	0.106	0.055
DGP3	1.000	0.997	0.033	0.007	0.106	0.052	1.000	0.999	0.336	0.231	0.177	0.109	0.101	0.065
DGP4	0.847	0.679	0.419	0.282	0.753	0.584	0.958	0.885	0.147	0.076	0.109	0.064	0.721	0.504
DGP5	1.000	0.997	0.267	0.261	1.000	1.000	0.999	0.994	0.099	0.050	0.086	0.047	1.000	1.000

GCCM denotes our Generalized Conditional Cramer van Mises test, KX the Koenker and Xiao (2002) specification test based on Khmaladzation, and EG the bootstrap specification test for quantile regression in Escanciano and Goh (2010). Suffixes denote the model being tested: location shift (LS), location-scale shift (LSS), quantile regression (QR) or distributional regression (DR). Left column specifies the true data generating process, as described in the main text.

able to pick up the misspecified link function in DGP2 even for $n = 100$, and produces rejection rates under DGP4 which are substantially higher than that of the GCCM-QR test. In summary, the (certainly limited) simulation evidence suggest that our GCCM tests have good finite sample properties even in relatively small samples, and compare favorably to their respective relevant competitors.

7. Empirical Application

In this section, 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 (Fortin et al., 2011). There are doubts, however, that standard models like linear quantile regression are able to capture some important features of conditional wage distributions, such as e.g. the irregular behavior around the minimum wage. Our results shed some light on this important empirical issue.

We use a data set constructed from the 1988 wave of the Current Population Survey (CPS), an extensive survey of US households. The same data is used in DiNardo et al. (1996), to which we refer for details of its construction. It 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 the industry in which the worker is employed (20 levels). As in the previous subsection, we consider the location shift model (LS), the location-scale sift model (LSS), the linear quantile regression model (QR), and the distributional regression model (DR) using the normal CDF as a link function. We test the correct specification of each model with log hourly wage as the dependent variable, and the following three different subsets of the explanatory variables, respectively:

- **Specification 1:** union coverage, education, experience.

- **Specification 2:** all variables in Covariates 1, experience (squared), education interacted with experience, martial status, part-time status, race, SMSA.

- **Specification 3:** all variables in Covariates 2, occupation, industry.

Given the large sample size, we would expect all specifications to be rejected by the data, since every statistical model is at best a reasonable approximation to the true data generating mechanism. However, this would not directly imply that such specifications 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 be concerned if flexible models using many covariates would be rejected even in small samples. We therefore conduct a simulation experiment, where in each run we test the validity of various conditional distributional models using Specification 1–3, respectively, for random subsamples of the data of size $n = 500$ and $n = 2000$.

In Table 2, we report the empirical rejection probabilities from 1000 replications of the simulation experiment described above. We can see that when using Specification 1 and 2 none of the four models we consider leads to an adequate fit of the conditional wage distribution. All empirical rejection rates are close to one for both sample sizes in case of Specification 1. For Specification 2, we observe rejection rates between about 22% and 66% at the nominal 5% level for $n = 500$, with the lowest rates coming from the DR model. For $n = 2000$, the QR model (or one of its special cases) are rejected in almost all runs at the 5% level, while rejection rates for the DR model are somewhat lower at about 55%. For the most extensive Specification 3, rejection rates for all specifications are around or below the respective nominal level for $n = 500$. When moving to $n = 2000$, rejection rates for the QR model rise to about 37% at the 5% nominal level. The LSS and LS models are rejected at a similar or higher rate, respectively. On the other hand, rejection rates for the DR model remain around the respective nominal level in this case. Our simulation results in the previous subsection suggest that this last finding should

Table 2: Empirical Application: Empirical Rejection Frequencies of the Generalized Conditional Cramer van Mises (GCCM) Test for Correct Specification of various CDMs.

	GCCM-LS		*GCCM-LSS*		*GCCM-QR*		*GCCM-DR*	
$n = 500$	10%	5%	10%	5%	10%	5%	10%	5%
Specification 1	1.000	0.992	0.961	0.936	0.997	0.975	0.992	0.975
Specification 2	0.837	0.661	0.549	0.397	0.691	0.520	0.352	0.211
Specification 3	0.129	0.070	0.070	0.029	0.088	0.029	0.029	0.009
$n = 2000$	10%	5%	10%	5%	10%	5%	10%	5%
Specification 1	1.000	1.000	0.997	0.997	1.000	1.000	1.000	1.000
Specification 2	1.000	0.997	0.947	0.921	0.994	0.983	0.789	0.556
Specification 3	0.760	0.592	0.507	0.323	0.538	0.369	0.136	0.067

GCCM denotes our Generalized Conditional Cramer van Mises test. Suffixes denote the specification being tested: location shift (LS), location-scale shift (LSS), quantile regression (QR) or distributional regression (DR). Left column specifies the set of explanatory variables used, as described in the main text.

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, we computed the rank of each individual 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. The results of our empirical exercise remained qualitatively unchanged using the modified data set, 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.

A. Appendix

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

Proof of Theorem 1. Define the processes $\nu(y,x) = \sqrt{n}(\widehat{H}_n(y,x) - H(y,x))$ and $\nu_0(y,x) = \sqrt{n}(\widehat{H}_n^0(y,x) - H^0(y,x))$. To prove part i), note that $H \equiv H^0$ under $\mathcal{H}_0$, and thus

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

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. Applying the Continuous Mapping Theorem and the Glivenko-Cantelli 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 any fixed alternative $\Pr(H^0(Y,X) \neq H(Y,X)) > 0$ by construction, and thus

$$T_n = \int (\nu(y,x) - \nu_0(y,x) + \sqrt{n}(H(y,x) - H^0(y,x)))^2 dH(y,x) + o_p(1) = O_p(n),$$

which implies that T_n is greater than any fixed constant tends to one as $n \to \infty$. $\square$

Proof of Theorem 2. To prove the result, we first define the empirical processes $\lambda_1(y,x) = \sqrt{n}((\widehat{H}_n(y,x) - \int F^*(y|t)\mathbb{I}\{t \leq x\} dG(t))$ and $\lambda_2(\theta,u) = \sqrt{n}(\widehat{\Psi}_n(\theta,u) - \mathbb{E}_{F^*}(\psi(Z,\theta,u)))$, 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

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

where $\mu_1(y,x) = \int (Q(y|t) - F^*(y|t))\mathbb{I}\{t \leq x\} dG(t)$ and $\widetilde{\mu}_2(\theta,u) = \mathbb{E}_Q(\psi(Z,\theta,u)) - \mathbb{E}_{F^*}(\psi(Z,\theta,u))$. Next, define the empirical processes $\nu^*(y,x) = \sqrt{n}(\widehat{H}_n(y,x) - H^*(y,x))$

and $\nu_0^*(y,x) = \sqrt{n}(\widehat{H}_n^0(y,x) - H^*(y,x))$, 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 $\widehat{c}_n(\alpha) = c(\alpha) + o_p(1)$. This implies that T_n and $\widetilde{T}_n = T_n - (\widehat{c}_n(\alpha) - c(\alpha))$ converge to the same limiting distribution as $n \to \infty$, and hence we have that $P(T_n > \widehat{c}_n(\alpha)) = \alpha + o(1)$ as claimed.

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

$$\begin{aligned} P(T_n \leq \widehat{c}_n(\alpha)) &= P(T_n \leq \widehat{c}_n(\alpha), T_n \leq M) + P(T_n \leq \widehat{c}_n(\alpha), T_n > M) \\ &\leq P(T_n \leq M) + P(\widehat{c}_n(\alpha) > M). \end{aligned}$$

From Theorem 1(ii), we know that $P(T_n \leq M) = o(1)$, and thus $P(T_n \leq \widehat{c}_n(\alpha)) < \epsilon + o(1)$, 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 α-quantile of the limiting distribution of the test statistic T_n under the null hypothesis. Using Anderson's Lemma, we find that

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

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 $P(T_n > \widehat{c}_n(\alpha)) = P(T_n > c(\alpha)) + o(1)$ 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 4–6. This follows by straightforward applications of results in Chernozhukov et al. (2009, Section 5 and Appendix D). □

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

Proof. This lemma is a minor generalization of Lemma 13 in Chernozhukov et al. (2009), and can thus be proven in the same way. □

Lemma 2. *Let either the null hypothesis or a fixed alternative, and Assumptions 1-6 be true. Define the empirical processes* $\nu(y,x) = \sqrt{n}(\widehat{H}_n(y,x) - H(y,x))$ *and* $\nu_0(y,x) = \sqrt{n}(\widehat{H}^0_n(y,x) - H^0(y,x))$*. Then it holds that* $(\nu,\nu_0) \Rightarrow \mathbb{G}$ *in* $l^\infty(\mathcal{Z} \times \mathcal{Z})$*, 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 et al. (2009) that

$$\sqrt{n}(\widehat{H}_n(\cdot,\cdot) - H(\cdot,\cdot), \widehat{\theta}_n(\cdot) - \theta_0(\cdot)) \Rightarrow (\mathbb{G}_1(\cdot,\cdot), -\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot}[\widetilde{\mathbb{G}}_2(\theta_0(\cdot),\cdot)])$$

in $\ell^\infty(\mathcal{Z}) \times \ell^\infty(\mathcal{U})$. Next, it follows from Assumption 5 that

$$\sqrt{n}(\widehat{F}_n(y|x) - F(y|x)) \Rightarrow -\dot{F}(y|x,\theta_0)[\dot{\Psi}^{-1}_{\theta_0(\cdot),\cdot}[\widetilde{\mathbb{G}}_2(\theta_0(\cdot),\cdot)]] =: \mathbb{G}_2^*(y,x).$$

The statement of the Lemma then follows directly from Hadamard differentiability of the mapping $(A,B) \mapsto \int A(\cdot,t)\mathbb{I}\{t \leq \cdot\}dB(t)$, and the Functional Delta Method. In particular, for the second component $\mathbb{G}_2$ of the joint limiting process we have that

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

which follows from the Hadamard differential of $(A,B) \mapsto \int A(\cdot)\mathbb{I}\{t \leq \cdot\}dB(t)$. □

Lemma 3. *Suppose the data are distributed according to a local alternative Q_n satisfying Assumption 7. Define the processes $v_n(y,x) = \sqrt{n}(\widehat{H}_n(y,x) - H_n(y,x))$ and $w_n(\theta,u) = \sqrt{n}(\widehat{\Psi}_n(\theta,u) - \Psi_n(\theta,u))$, 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 $(v_n,w_n) \Rightarrow \widetilde{\mathbb{G}}$ in $l^\infty(\mathcal{Z}\times\Theta\times\mathcal{U})$, where the limiting process $\widetilde{\mathbb{G}}$ has the same properties as the one 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 $\nu_b(y,x) = \sqrt{n}(\widehat{H}_{b,n}(y,x) - \widehat{H}^0_n(y,x))$ and $\nu_{b,0}(y,x) = \sqrt{n}(\widehat{H}^0_{b,n}(y,x) - \widehat{H}^0_n(y,x))$. 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) □

References

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

Beran, R. (1988): "Prepivoting test statistics: a bootstrap view of asymptotic refinements," *Journal of the American Statistical Association*, 687–697.

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*, 1443–1458.

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

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

Billingsley, P. (1995): *Probability and measure*, Wiley.

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

DiNardo, J., N. Fortin, and T. Lemieux (1996): "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," *Econometrica*, 64, 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*, 159, 209–221.

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

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition Methods in Economics," *Handbook of Labor Economics*, 4, 1–102.

Galvao, A., K. Kato, G. Montes-Rojas, and J. Olmo (2011): "Testing linearity against threshold effects: uniform inference in quantile regression," *Working Paper*.

Hall, P. (1986): "On the bootstrap and confidence intervals," *The Annals of Statistics*, 1431–1452.

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

Harvey, A. (1976): "Estimating regression models with multiplicative heteroscedasticity," *Econometrica*, 461–465.

He, X. and L. Zhu (2003): "A lack-of-fit test for quantile regression," *Journal of the American Statistical Association*, 98, 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, 599–631.

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

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

——— (2010): "On Distributional vs. Quantile Regression," *Working Paper*.

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

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

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

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

Powell, J. (1986): "Censored regression quantiles," *Journal of Econometrics*, 32, 143–155.

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

——— (2011): "Partial Distributional Policy Effects," *Econometrica (to appear)*.

Rothe, C. and D. Wied (2012): "Misspecification Testing in a Class of Conditional Distributional Models," *IZA Working Paper*.

Rutemiller, H. and D. Bowers (1968): "Estimation in a heteroscedastic regression model," *Journal of the American Statistical Association*, 552–557.

Stute, W. (1997): "Nonparametric model checks for regression," *The Annals of Statistics*, 613–641.

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, 123–138.