Noname manuscript No.
(will be inserted by the editor)

Specification Testing in Functional Quantile Regression Models with an Application to Income Differences in Germany

Tim Kutzker · **Nadja Klein** · **Dominik Wied**

Abstract We propose a novel consistent specification test for quantile regression models where we allow the covariate effects to be quantile dependent and nonlinear. To achieve this, we parameterize the conditional quantile functions by appropriate basis functions, rather than parametrically and hence allowing to test for functional forms beyond linearity while retaining the linear cases as special cases. Due to the dependence on the quantile itself covariate-quantile relations can differ for distinct quantiles. The induced class of conditional distribution functions can finally be tested with a Cramér-von Mises type test statistic. We derive the theoretical limit distribution and propose a practical bootstrap method. To increase the power of our test, we suggest a modified test statistic using quantile regression splines. A detailed Monte Carlo experiment shows that the test results in a reasonable sized testing procedure with large power. An application to conditional income disparities between East and West Germany over the period 2001 – 2010 indicates that there are still significant differences across the quantiles of the conditional income distributions, when conditioning on age.

Keywords Bootstrap; Cramér-von Mises distance; distributional regression model; quantile regression splines; specification testing

JEL codes: C12, C14, C21, C31, J31

Nadja Klein
Humboldt Universität zu Berlin
Unter den Linden 6, 10099 Berlin, Germany
E-mail: nadja.klein@hu-berlin.de

Tim Kutzker, *corresponding author*
Universität zu Köln
Universitätsstr. 24, 50937 Cologne, Germany
E-mail: tim.kutzker@uni-koeln.de

Dominik Wied
Universität zu Köln
Universitätsstr. 24, 50937 Cologne, Germany
E-mail: dwied@uni-koeln.de

Nadja Klein acknowledges support through the Emmy Noether grant KL 3037/1-1 of the German research foundation (DFG).

1 Introduction

Hypothesis testing plays a central role in many economic research areas. A necessary prerequisite for the statistical validity of the decisions to be made is the correct specification of the underlying model. Specification tests can be used to validate the correctness of theoretical assumptions. Within the framework of linear regression, a whole range of specification tests are available that can test both, parametric and non-parametric approaches. In general, testing misspecification in linear OLS models is well understood and developed. For the parametric setup, e.g., Bierens (1990) showed that any conditional moment test of functional form of nonlinear regression models can be converted into a consistent chi-square test that is consistent against all deviations from the null hypothesis. Härdle and Mammen (1993) suggested a wild bootstrap procedure for regression fits in order to decide whether a parametric model could be justified. Stute (1997) proposed a general method for testing the goodness of fit of a parametric regression model. For the nonparametric case, among others, Gozalo (1993) proposed a general framework for specification testing of the regression function in a nonparametric smoothing estimation context and Stute et al. (1998) suggested a goodness of fit test using a wild bootstrap procedure that checks whether a function belongs to a certain model class.

However, OLS estimates are sensitive to outliers and draw only a part of the whole picture, since they only model the conditional expected value. As it provides more robust estimates compared to OLS and allows a more comprehensive picture and flexible analysis of the economic problem, quantile regression has become increasingly popular since the seminal article by Koenker and Bassett Jr (1978). But it also applies to quantile regressions, that post-estimation inference procedures essentially depend on the validity of the underlying parametric functional form for the quantiles considered (Angrist et al., 2006). For example, assuming the same fixed linear relationship between covariates for all quantiles is the connecting element of the Machado-Mata decomposition (used in particular to describe wage inequalities) by Machado and Mata (2005) and the Khmaladazation (which is based on the Doob-Meyer decomposition of the martingale) by Koenker and Xiao (2002). Since such a linearity assumption considerably limits the number of possible models and hence the hypothesis space, there have recently been successful attempts to weaken the linearity assumption for quantile estimation and inference with independently and identically distributed (i.i.d.) data.

In this context, more general parametric quantile models have been developed that, among others, include works by Hallin et al. (2009) suggesting an estimator for local linear spatial quantile regression and Guerre and Sabbah (2012) investigating the Bahadur representation of a local polynomial estimator of the conditional quantile function and its derivatives. But also nonparametric approaches for estimating conditional quantile functions have attracted much attention. Li and Racine (2008) proposed a nonparametric conditional cumulative distribution function kernel estimator along with an associated nonparametric conditional quantile estimator. Belloni et al. (2019) developed a nonparametric quantile regression-series framework for performing inference on the entire conditional quantile function and its linear functionals and Qu and Yoon (2015) presented estimators for nonparametrically specified condtional quantile processes that are based on local linear regressions. Li et al. (2020) investigated the problem of nonparametrically estimating a conditional quantile function with mixed discrete and continuous covariates suggesting a kernel based approach. But regardless of whether parametric or non-parametric approaches are chosen, the theory concerning the validity of the correct model choice seems to keep up with the rapid development of new estimation methods only to a limited extent. To the best of our knowledge, there does not exist a testing procedure that allows for quantile-specific functional covariate effects.

In a parametric framework, one of the first specification tests for linear location shift and location-scale shift quantile models with i.i.d. data is the test by Koenker and Xiao (2002). Shortly after that, Chernozhukov (2002) proposes a resampling testing procedure avoiding to estimate further objects, such as the score function using the same principle as Koenker and Xiao (2002). However, these two tests proposed do not test the validity of the quantile regression model itself. Escanciano et al. (2010) and Escanciano and Velasco (2010) both tested the validity of the null hypothesis that a conditional quantile restriction is valid over a range of quantiles. Rothe and Wied (2013) proposed a specification test for a larger class of models, including quantile regression models. In case of nonparametric instrumental quantile regression, Breunig (2019) develops a methodology for testing the hypothesis whether the instrumental quantile regression model is correctly specified. However, all models have in common that they require linearity in the regressors. In this paper, we consider a broad approach that tackles the following two challenges simultaneously, hence proposing a general specification test that is an important contribution in the field with potential in a wide range of applied questions. *i*) We suggest a testing procedure for quantile regression models, where the regressors can explicitly depend on the quantile considered, which allows to test for the correct specification of large number of models. *ii*) Due to our general model set up, our proposed methodology does also allow to test for semi-parametric models, e.g., *B*-splines for quantile regressions, where the covariates have a functional form (cf. Cardot et al. (2005) for the estimation procedure of such processes). Such a testing procedure not only increases the range of applications but also offers the advantage that effects can be tested in isolation, depending on the quantile. As such, it extends the literature on quantile regression specification tests for better answering relevant questions in economics and further sciences; wherever specific regressors have a functional, non-linear influence on distinct quantiles may be present.

To illustrate the power and potential of our test, we consider the case of income inequality, with a focus on differences in the conditional income quantiles between East and West Germany in a balanced panel data set. Such disparities have received considerable attention in the economic literature (e.g. Biewen (2000)), and also consistently played a major role in the domestic political debate. Our empirical analysis uses the German socio-economic panel (SOEP) and shows that age has a predominant linear influence on income development in Germany, but for the upper 90% quantile the influence of age is solely quadratic. Such statistically proven statements on income distributions also provide further evidence for investigations on inequalities in income distributions. Importantly, and in line with other studies on this topic, we find through an initial Machado-Mata (Machado and Mata, 2005) decomposition that there are still income differences between East and West Germany, which can be confirmed by our proposed testing procedure.

But also further use cases in, e.g., finance can be addressed with our testing approach. For instance, the correct specification of the left tail of the distribution function is essential to adequately assess risks. Our null hypothesis can be specified in a way that only some parts of the quantile function follow an explicit parametric approach. Thus, we provide a statistical verification procedure for the part of the distribution function that is of relevance for the calculation of the value at risk. Previous procedures usually require a complete characterization of the quantile function, in which the covariates must also be independent of the considered quantiles.

The basic idea of our procedure is based on the principle characterized by Rothe and Wied (2013): We compare an unrestricted estimate of the joint distribution function of the random variable Y and the random vector X with a restricted estimate that imposes the structure implied by the null hypothesis model. Based on a Cramèr-von Mises type measure of distances, the restricted estimate of the joint distribution can then be compared with the unrestricted one. We derive the non-pivotal limiting distribution of our test statistic and show the validity of our suggested parametric bootstrap procedure for the approximation of the critical values. To increase the power of our test, we replace the unrestricted model estimate with a quadratic B-spline, meeting the assumptions of a quantile function. Due to the generality of our test procedure we can subsume previous specification tests for quantile regression models with i.i.d. data as marginal cases of our procedure. The Monte Carlo simulation study shows that the proposed testing procedure has superior power properties than existing methods.

In sum, we believe that the testing procedure proposed in our paper is a useful extension of existing methods for testing the correct specification in quantile regression models, both in terms of the improvement in power, and also in terms of the extension to quantile dependent regressors it offers.

The paper is organized as follows. Section 2 formulates the test problem and derives the test statistics including the parametrized bootstrap approximation. In Section 3, we provide the theoretical properties of the testing procedure and the bootstrap. Section 4 contains an intensive Monte Carlo simulation including comparisons to existing tests and in Section 5 we present the empirical application. The last section concludes. In order to increase the readability of the paper, all proofs are to be found in the Appendix.

2 Quantile regression testing

2.1 Specification test for quantile dependent regressors

We observe an outcome variable $Y_i \in \mathbb{R}$ and a vector of explanatory variables $X_i \in \mathbb{R}^K$ for $i = 1, \ldots, n$, $K \in \mathbb{N}$. We assume the data points to be independent and identically distributed (i.i.d.). Our aim is to test the validity of certain classes of parametric specifications for the conditional cumulative distribution function (cdf) F of Y given X, i.e. $F_{Y|X}$ and with corresponding conditional quantile function (qf) $F^{-1}_{Y|X}$. Since for $y \in \mathbb{R}$ in holds that

$$F_{Y|X}(y\,|\,x) = \int_0^1 \mathbb{1}_{\left\{F^{-1}_{Y|X}(\tau|x) \le y\right\}} d\tau, \tag{2.1}$$

let $\mathscr{F}$ be the set of conditional cdfs $F_{Y|X}$ induced by $F^{-1}_{Y|X}$, i.e.

$$\mathscr{F} := \{F_{Y|X}(y|x,\theta) \mid F^{-1}_{Y|X}(\tau \mid x) = P(x,\tau)'\theta \text{ for some } \theta \in \mathscr{B}(\mathscr{T},\Theta) \text{ and } (y,x) \in \mathscr{S}\}, \tag{2.2}$$

where $\mathscr{S}$ denotes the support of $(y,x) \in \mathbb{R}^{K+1}$ and $\mathscr{B}(\mathscr{T},\Theta)$ the class of functions $\tau \mapsto \theta(\tau) \in \Theta \subset \mathbb{R}^p$ for $\tau \in \mathscr{T} \subseteq [0,1]$ with p the dimension of the parameters. These conditional qfs $F^{-1}_{Y|X}$ are assumed to be of the form $P(X,\tau)'\theta(\tau)$, i.e. $F^{-1}_{Y|X}(\tau|X) = P(X,\tau)'\theta$ for every $\tau \in \mathscr{T}$ and $\theta \in \mathscr{B}(\mathscr{T},\Theta)$. $P(X,\tau)$ is a vector of transformations of the realization X (also known as basis function evaluations in the literature) depending on $\tau \in \mathscr{T}$ such as polynomials or B-splines (which are also called basis functions) evaluated at X that depend on τ and thus may differ for distinct quantiles. In the simplest case, e.g., $P(X,\tau)$ can be equal to X for all $\tau \in \mathscr{T}$, such that the quantiles are linear in X. In this case, the parameterization of the quantile function corresponds to the classical linear quantile regression model, where the vector of transformations $P(X,\tau)$ is constant for all quantiles τ.

The hypothesis we want to test is that the conditional cdf $F_{Y|X}$ coincides with an element of a class of distributions $\mathscr{F}$ of the form (2.2) which corresponding $F^{-1}_{Y|X}$ can be decomposed to a vector of transformations and a functional parameter $\theta \in \mathscr{B}(\mathscr{T},\Theta)$:

$$H_0 : F_{Y|X} \in \mathscr{F} \quad \text{vs.} \quad H_1 : F_{Y|X} \notin \mathscr{F}. \tag{2.3}$$

Again, in comparison to existing parametric quantile regression models we explicitly allow the vector of transformations $P(X,\tau)$ to depend on τ. Hence, this framework allows to model a quantile function that, e.g., contains a linear regressor in the lower fifty percent quantile and a highly non-linear functional regression form in the upper fifty percent quantile. Naturally, models in which the vector of transformations does not depend on τ are captured by our approach as a special case, when $P(X,\tau) \equiv P(X) = const$ for all $\tau \in \mathscr{T}$. Consequently, the aim of this article is to present a testing procedure that is able to give statistical insights if the quantile model assumption (2.3) holds statistically true. We assume that there is a unique $\theta_0 \in \mathscr{B}(\mathscr{T},\Theta)$ under the null hypothesis. Accordingly, we can reformulate our testing problem (2.3) to

$$\begin{aligned} &H_0 : F_{Y|X}(y \mid x) \in \mathscr{F}^0 := \{F_{Y|X}(y \mid x, \theta_0) \mid F^{-1}_{Y|X}(\tau \mid x) = P(x,\tau)'\theta_0 \text{ for some } \theta_0 \in \mathscr{B}(\mathscr{T},\Theta), \text{ for all } (y,x) \in \mathscr{S}\} \\ &H_1 : P\left(F_{Y|X}(y^* \mid x^*) \neq F_{Y|X}(y^* \mid x^*, \theta)\right) > 0 \text{ for all } \theta \in \mathscr{B}(\mathscr{T},\Theta) \text{ and for some } (y^*,x^*) \in \mathscr{S}. \end{aligned} \tag{2.4}$$

In addition, we assume that under the null hypothesis any functional parameter $\theta \in \mathscr{B}(\mathscr{T},\Theta)$ satisfying $F_{Y|X}(y|x) = F^*_{Y|X}(y|x,\theta)$ with $F^*_{Y|X}(y|x,\theta) \in \mathscr{F}^0$ for all $(y,x) \in \mathscr{S}$ also satisfies $\theta(\tau) = \theta_0(\tau)$ for all $\tau \in \mathscr{T}$.

To propose a testing procedure for the problem (2.4) we assume that the true value of the functional parameter, i.e. $\theta_0(\tau)$ for every $\tau \in \mathscr{T}$, is identified under the null hypothesis through a moment condition. Specifically, let

$$g : \mathscr{S} \times \Theta \times \mathscr{T} \to \mathbb{R}^p$$

be a uniformly integrable functions whose exact form depends on $\mathscr{F}^0$, and suppose that for every $\tau \in \mathscr{T}$, the equation

$$G(\theta,\tau) := \mathbb{E}[g(Y,X,\theta,\tau)] = 0 \in \mathbb{R}^p \tag{2.5}$$

has a unique solution $\theta_0(\tau)$. Furthermore, under the alternative, $\theta_0(\tau)$ remains well defined for all $\tau \in \mathscr{T}$ due to Assumption (2.5) and can thus be thought of as a pseudo-true value of the functional parameter in this case.

The null hypothesis can be equivalently stated as

$$F(y|x) = F(y|x,\theta_0) \text{ for all } (y,x) \in \mathbb{R}^{K+1}, \tag{2.6}$$

with $\theta_0(\tau)$ as the unique solution to (2.5) for all $\tau \in \mathscr{T}$. This holds true since $\mathscr{F}^0$ is a singleton containing $F_{\cdot|\cdot}(\cdot|\cdot,\theta_0)$. Since $F_{Y|X}(y|X) = \mathbb{E}[\mathbb{1}_{\{Y\leq y\}}|X]$ we can write the joint cdf of Y and X, $F_{Y,X}$, as[1]

$$F(y,x) = \int_{\mathbb{R}^K} F_{Y|X}(y|x^*)\mathbb{1}_{\{x^*\leq x\}} dF_X(x^*) \tag{2.7}$$

$$F(y,x,\theta_0) = \int_{\mathbb{R}^K} F_{Y|X}(y|x^*,\theta_0)\mathbb{1}_{\{x^*\leq x\}} dF_X(x^*), \tag{2.8}$$

where F_X denotes the marginal cdf of X. From Billingsley (1995) Theorem 16.10 (iii) it follows that the testing problem (2.4) can be restated as

$$\begin{gathered} H_0 : F(y,x) = F(y,x,\theta_0) \text{ for all } (y,x) \in \mathbb{R}^{K+1} \\ \text{versus} \\ H_1 : F(y,x) \neq F(y,x,\theta_0) \text{ for some } (y,x) \in \mathbb{R}^{K+1}. \end{gathered} \tag{2.9}$$

for some $\theta_0 \in \mathscr{B}(\mathscr{T},\Theta)$. With the help of the above representation of the null hypothesis (2.9) we introduce a function $S : \mathbb{R}^{K+1} \times \Theta \to \mathbb{R}$ that measures the difference of the non-parametric $F(y,x)$ and the parametrized cdf $F(y,x,\theta)$ defined as

$$S(y,x,\theta) := F(y,x) - F(y,x,\theta). \tag{2.10}$$

The null hypothesis is true by assumption if $S(y,x,\theta_0) = 0$ for all $(y,x) \in \mathscr{S}$, whereas $S(y,x,\theta) \neq 0$ for all $\theta \neq \theta_0 \in \mathscr{B}(\mathscr{T},\Theta)$ and for some $(y,x) \in \mathscr{S}$. To obtain an applicable test statistic we will replace $F(y,x)$ and $F(y,x,\theta_0)$ by its empirical counterparts $\hat{F}_n(y,x)$ and $\hat{F}_n(y,x,\theta)$. Thus, we have

$$S_n(y,x,\hat{\theta}) := \hat{F}_n(y,x) - \hat{F}_n(y,x,\theta), \tag{2.11}$$

with $\hat{F}_n(y,x,\theta) = F(y,x,\hat{\theta}_n)$, a parametric estimate of F based on a consistent estimate $\hat{\theta}_n$ of θ_0. In order to emphasize that the parametric empirical cdf $\hat{F}_n(y,x,\theta)$ particularly estimates the parameter θ_0 by $\hat{\theta}_n$, we also use the notation $\hat{F}_n(y,x,\hat{\theta}_n)$. Under the null hypothesis, $\hat{F}_n(y,x)$ and $\hat{F}_n(y,x,\hat{\theta}_n)$ are consistent estimators for $F(y,x)$ and $F(y,x,\theta_0)$, respectively. In that case, $S_n(y,x,\theta)$ should be close to zero for all $(y,x) \in \mathscr{S}$. If, however, the alternative holds true, then there is a vector (y,x) for each $\theta \in \mathscr{B}(\mathscr{T},\Theta)$ such that the function S_n is greater than zero.

To obtain an estimate for the parametrized empirical cdf $\hat{F}_n(y,x,\hat{\theta}_n)$ we follow Chernozhukov et al. (2013) and take $\hat{\theta}_n$ to be an approximate Z-estimator satisfying

$$||\hat{G}(\hat{\theta}_n,\tau)|| = \inf_{\theta\in\Theta} ||\hat{G}(\theta,\tau)|| + \eta_n \tag{2.12}$$

[1] Due to readability we will suppress the index Y,X for the joint cdf $F_{Y,X}$ in the following, i.e. $F = F_{Y,X}$.

where the function $\hat{G}(\hat{\theta}_n,\tau) := n^{-1}\sum_{i=1}^{n} g(Y_i,X_i,\theta,\tau)$ is the sample analogue of the moment condition (2.5) for every $\tau \in \mathscr{T}$ and for some possibly random variable $\eta_n = o_p(n^{-1/2})$. For every $\tau \in \mathscr{T}$ and every $(y,x) \in \mathbb{R}^{K+1}$, the estimator based on the testing problem (2.4) takes the form

$$\hat{F}_n(y|x,\hat{\theta}_n) = \delta + \int_{\delta}^{1-\delta} \mathbb{1}_{\{P(x,\tau)'\hat{\theta}_n(\tau)\le y\}} d\tau, \qquad (y,x) \in \mathscr{S}, \tag{2.13}$$

$$\hat{\theta}_n(\tau) = \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{(y,x)\in\mathscr{S}} \left(\tau - \mathbb{1}_{\{y\le P(x,\tau)'\theta\}}\right)\left(y - P(x,\tau)'\theta\right) \tag{2.14}$$

for some arbitrary constant $\delta > 0$. The trimming by δ avoids estimation of tail quantiles (Koenker, 2005) and is valid under the conditions in Theorem 1 in Section 3. Thus, the test statistic (2.11), that is based on the differences of the non-parametric and parametrized empirical distribution functions, can be expressed as

$$S_n(y,x,\hat{\theta}_n) = \hat{F}_n(y,x) - \hat{F}_n(y,x,\hat{\theta}_n) \tag{2.15}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}_{\{Y_i\le y\}}\mathbb{1}_{\{X_i\le x\}}\right) - \int_{\mathbb{R}^K} \mathbb{1}_{\{x^*\le x\}}\left(\delta + \int_{\delta}^{1-\delta}\mathbb{1}_{\{P(x^*,\tau)'\hat{\theta}_n(\tau)\le y\}}d\tau\right) d\hat{F}_X(x^*) \tag{2.16}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}_{\{Y_i\le y\}}\mathbb{1}_{\{X_i\le x\}} - \mathbb{1}_{\{X_i\le x\}}\left[\delta + \int_{\delta}^{1-\delta}\mathbb{1}_{\{P(X_i,\tau)'\hat{\theta}_n(\tau)\le y\}}d\tau\right]\right). \tag{2.17}$$

We propose a Cramér-von Mises type test statistic S_n^{CM} defined as

$$S_n^{CM} := \int ||\sqrt{n}S_n(y,x,\hat{\theta}_n)||^2 d\hat{F}_n(y,x), \tag{2.18}$$

which is a generalization of existing quantile regression tests. However, if the vector of transformations $P(X,\tau)$ in (2.4) is independent of τ then the test statistic coincides with test statistic proposed in Rothe and Wied (2013). Notwithstanding the above, it is also possible to consider a Kolmogorov-Smirnov-type test statistic

$$S_n^{KS} := \sqrt{n} \sup_{(y,x)\in\mathscr{S}} ||S_n(y,x,\hat{\theta}_n)||, \tag{2.19}$$

but the Cramér-von-Mises-type test provides better (power) results, since it is less susceptible to outliers (Chernozhukov, 2002; Rothe and Wied, 2013).

2.2 More powerful testing procedure using splines

In order to obtain better power results, we consider two different test statistics of the form (2.11), using two estimators for the quantile regression model specified under the null hypothesis: one estimator corresponds to the model, the other employs a spline approach, i.e.

$$S_n^*(y,x,\hat{\theta}_n) = \hat{F}(y,x) - \hat{F}_n^S(y,x,\hat{\theta}_n) - \left(\hat{F}(y,x) - \hat{F}_n(y,x,\hat{\theta}_n)\right) \tag{2.20}$$

$$= \hat{F}_n^S(y,x,\hat{\theta}_n) - \hat{F}_n(y,x,\hat{\theta}_n) \tag{2.21}$$

where $\hat{F}_n^S$ is the estimate of the cumulative distribution function by a quantile regression spline that meets some regularity assumptions (cf. Assumptions 2 in Section 3 and Cardot et al. (2005)) and $\hat{F}_n(y,x,\hat{\theta}_n)$ the estimate using the null hypothesis model. In case of non-varying covariates, i.e. $P(X,\tau) = P(X)$ for all $\tau \in \mathscr{T}$, $\hat{F}_n^S$ could be, e.g., estimated by a quadratic B-spline with monotone increasing parameters. Besides standard assumptions, the monotonicity assumption is of central importance for the estimation of the quantile regression function by splines[2]. However, the Monte Carlo simulation clearly shows that these additional assumptions significantly increase the rejection rates in the case of misspecified null hypotheses. Xue and Wang (2010) have shown, e.g., that the estimate of the cumulative distribution function with a smooth monotone polynomial spline has better finite sample properties than the empirical distributional estimate.

However, the goodness and convergence rate of the spline approximation depends, in general, in a complex fashion on the degree of the spline, the number of knots and the position of those knots. He and Shi (1997) have pointed out that if the number of knots $k_n \sim (n/\log n)^{2/5}$ and under some mild assumptions[3] the order of approximation of a quadratic monotone B-spline is $(\log n/n)^{2/5}$ for a quantile regression model with non-varying covariates. Cardot et al. (2005) have generalized the limiting result for quantile regression models with varying covariates. This result is of particular interest since, together with the Donsker-class property[4], it provides the basis for the convergence of the Cramer-von Misès type test statistic that is defined as

$$S_n^{CM^*} := \int ||\sqrt{n} S_n^*(y,x,\hat{\theta}_n)||^2 d\hat{F}_n(y,x). \tag{2.22}$$

2.3 Semiparamteric bootstrap procedure

As we show in more detail in the next section, the asymptotic null distribution of S_n^{CM} and $S_n^{CM^*}$, respectively, depend on the data generating process in a complex fashion. To obtain critical values for our test, we therefore propose a semiparametric bootstrap procedure. This procedure is reasonable from a

[2] There is a whole series of assumptions that guarantee the monotonicity of the quantile function, e.g., derivatives of the (quadratic) spline to be non-negative or estimators are monotonically increasing as quantiles increase assuming static covariates, all having different computational properties. To discuss all these assumptions is beyond the scope of this project. We refer here to the relevant literature, i.e. Koenker et al. (1994), Bondell et al. (2010) among others.

[3] For a detailed description of the requirements we refer to the assumptions $C1-C3$ from He and Shi (1997).

[4] Yu et al. (2017) have shown the Donker-class property for functional linear partial quantile regressions.

practical point of view, since it avoids the complicated problem of estimating the null distribution directly, including the complex covariance structure. The idea of our semiparametric bootstrap is to generate synthetic data that is line with the assumptions under the null hypothesis. Thus, 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 works as follows:

i.) Draw B bootstrap samples of covariates $\{X_{b,i}, 1 \leq i \leq n\}_{b=1,\dots,B}$ of size n with replacement from the realized values $\{X_i, 1 \leq i \leq n\}$.

ii.) Generate independently B n-dimensional vectors U_b with $b = 1,\dots,B$ of standard uniform distributed random variables, i.e. $U_b = (U_{b,i})_{i=1}^n$ with $U_{b,i} \overset{i.i.d.}{\sim} U(0,1)$ for $i = 1,\dots,n$ and $b = 1,\dots,B$, that represent the randomly chosen quantiles.

iii.) For each $b = 1,\dots,B$, estimate the conditional quantile function $\hat{F}^{-1}(U_{b,i} \mid X)$ for every $i = 1,\dots,n$ by the model specified under the null hypothesis using the realized values X and compute n-dimensional estimates $\hat{Y}_b := (\hat{Y}_i)_{b,i=1}^n$ for $b = 1,\dots n$ by means of the bootstrap sample of covariates $\{X_{b,i}, 1 \leq i \leq n\}_{b=1,\dots,B}$, i.e. $\hat{Y}_{b,i} = \hat{F}^{-1}(U_{b,i} \mid X_{b,i})$ for $i = 1,\dots n$ and $b = 1,\dots,B$.

iv.) Calculate B bootstrap versions of the test statistic (2.22), i.e. for $b = 1,\dots,B$ compute

$$S_{n,b}^{CM} := \int ||\sqrt{n} S_{n,B}(\hat{y}, x_b, \hat{\theta}_n)||^2 d\hat{F}_n(\hat{y}, x_b). \tag{2.23}$$

v.) Determine the critical value c such that

$$\frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{\{S_{n,b}^{CM} > c\}} \overset{!}{=} q, \tag{2.24}$$

where $q \in (0,1)$.

With the above described bootstrap procedure we can calculate critical values $c(q)$ for (2.18). Critical values for (2.22) can be obtained in the same manner if the test statistic $S_{n,B}^{CM}$ from (2.23) is replaced by its spline counterpart, i.e. $S_{n,B}^{CM^*}$.

3 Asymptotics

3.1 Theoretical properties for quantile dependent regressors

This section shows that the test statistic S_n^{CM} has the correct asymptotic size which is summarized in Theorem 1 at the end of that subsection. Before we derive large sample properties of our test statistic

(2.18), we need to impose and to discuss some mild assumptions that are in line with Chernozhukov et al. (2013). However, since our proposed test statistic is a generalization of existing tests we need to slightly adjust the standard assumptions. Additionally, we assume that there is a finite compact decomposition of $\mathscr{T} := [\varepsilon, 1-\varepsilon]$, $\varepsilon \in (0, 0.5)$. Hence, we can formulate the assumptions for Θ being an arbitrary subset of $\mathbb{R}^p$ as

Assumption 1

i.) *$P(X,\tau)$ is L^2-bounded in $[0,1]$.*

ii.) *Let $\bigcup_{l=1}^{L} I_l = \mathscr{T}$, $L \in \mathbb{N}$, I_l compact for $l = 1,\dots,L$ and $I_{l_1} \cap I_{l_2}$ a singleton for $l_1 \neq l_2$.*

iii.) *For each $\tau \in I_l$ with $l = 1,\dots,L$, $G(\cdot,\tau) : \Theta \to \mathbb{R}^p$ possesses a unique zero at $\theta_0(\tau) \in interior(\Theta)$ such that $G(\theta_0(\tau),\tau) = 0$, and, for some $\delta > 0$, $\mathscr{B} := \bigcup_{\tau \in \mathscr{I}_l} B_\delta(\theta_0(\tau))$ is a compact subset of $\mathbb{R}^p$ contained in Θ for $l = 1,\dots,L$.*

iv.) *Further, $G(\cdot,\tau)$ has a inverse $G^{-1}(x,\tau) := \{\theta \in \Theta \mid \mathbb{G}(\theta,\tau) = x\}$ that is continuous at $x = 0$ uniformly in $\tau \in I_l$ for all $l = 1,\dots,L$ with respect to the Hausdorff distance.*

v.) *There is a derivative $\dot{G}_{\theta_0(\tau),\tau}$ such that $\lim_{t\to 0} \sup_{\tau \in I_l, ||h||=1} |\frac{G(\theta_0(\tau)+th,\tau) - G(\theta_0(\tau),\tau)}{t} - \dot{G}_{\theta_0(\tau),\tau} h| = 0$, where $\dot{G}_{\theta_0(\tau),\tau}$ is non-singular at $\theta_0(\cdot)$ uniformly over $\tau \in I_l$ with $l = 1,\dots,L$, i.e. $\inf_{\tau \in I_l} \inf_{||h||=1} ||\dot{G}_{\theta_0(\tau),\tau} h|| > 0$ for all $l = 1,\dots,L$.*

vi.) *The maps $\tau \mapsto \theta_0(\tau)$ and $\tau \mapsto \dot{G}_{\theta_0(\tau),\tau}$ are continuous on $\mathscr{T}$.*

vii.) *The function set $\mathscr{G}_l = \{g(Y,X,\theta,\tau) | (\theta,\tau) \in \Theta \times I_l)\}$ is F_{YX}-Donskerfor all $l = 1,\dots,L$ with a square integrable envelope $\tilde{G}$ for $\bigcup_{l=1}^{L} \mathscr{G}_l$. The map $(\theta,\tau) \mapsto g(\cdot,\theta,\tau)$ is continuous at each $(\theta,\tau) \in \Theta \times I_l$ for all $l = 1,\dots,L$.*

viii.) *The mapping $\theta \mapsto F(\cdot|\cdot,\theta)$ is Hadamard differentiable for all $\theta \in \Theta$ with derivative $h \mapsto \dot{F}(\cdot|\cdot,\theta)[h]$*

Assumption 1 *i.)* claiming there is a finite, compact decomposition of the unit interval is required since we consider Donsker classes in the proof. We are using the fact that the union of Donsker classes is also Donsker (see Dudley, 2014, section 3.8). Assumptions 1 *i.)* – *v.)* guarantee the regularity of our estimator $\hat{\theta}_n$ and ensure that a functional central limit theorem can be applied to Z-estimator processes (cf. Corollary 1 in the Appendix 6). Assumption 1 *vi.)* is a smoothness condition, that implies together with the functional delta method that the restricted cdf estimator process

$$(y,x) \mapsto \sqrt{n}\left(\hat{F}_n(y,x,\hat{\theta}) - F(y,x,\theta)\right) \tag{3.1}$$

is F_{YX}-Donsker. The convergence (3.1) can be shown to be jointly with that of the empirical cdf process

$$(y,x) \mapsto \sqrt{n}\left(\hat{F}_n(y,x) - F(y,x)\right) \tag{3.2}$$

to a Brownian Bridge by some standard arguments given in Lemma 1. The limiting distribution of our test statistic S_n^{CM} then follows from an application of the continuous mapping theorem (the proof and further details are shifted to the appendix 6). We are now able to derive our main result:

Theorem 1 *If Assumptions (1) is satisfied then the following statements hold:*

i.) Under the null hypothesis H_0 (2.9),

$$S_n^{CM} \xrightarrow{d} \int ||\mathbb{G}_1(y,x) - \mathbb{G}_2(y,x)||^2 dF_{YX}(y,x), \tag{3.3}$$

where $(\mathbb{G}_1, \mathbb{G}_2)$ are Gaussian processes with zero mean and covariance function

$$\begin{aligned}
\mathbb{C}ov[\mathbb{G}_1(y,x), \mathbb{G}_1(y',x')] &= \sum_{k=-\infty}^{\infty} \mathbb{C}ov[\mathbb{1}_{\{Y_0 \le y\}}\mathbb{1}_{\{X_0 \le x\}}, \mathbb{1}_{\{Y_k \le y'\}}\mathbb{1}_{\{X_k \le x'\}}] \\
\mathbb{G}_2(y,x) &:= \int \mathbb{G}_2^+(y,x^*)\mathbb{1}_{\{x^* \le x\}} dF_X(x^*) + \int F(y\,|\,x^*)\mathbb{1}_{\{x^* \le x\}} d\mathbb{G}_1(\infty, x^*),
\end{aligned}$$

where $\mathbb{G}_2^+(y,x)$ is the limiting Gaussian process of $\sqrt{n}\left(\hat{F}_n(y\,|\,x, \hat{\theta}_n) - F(y\,|\,x)\right) \in \ell^\infty(\mathscr{S})$.

ii.) Under any fixed alternative, i.e., when the data are distributed according to some F that satisfies the alternative hypothesis H_1 in (2.9),

$$\lim_{n\to\infty} P(S_n^{CM} > \varepsilon) = 1 \text{ for all constants } \varepsilon > 0. \tag{3.4}$$

3.2 Theoretical properties for quantile dependent regressors using constrained polynomial spline regression

Theorem 1 represents a generalization of previous tests for quantile regression models, since it allows the covariate to depend on the quantile. Imposing the assumptions from Cardot et al. (2005) on a quantile regression spline enables us to replace the quantile function by an appropriate spline estimator.

Assumption 2

i.) The function $\theta \in \mathscr{B}(\mathscr{T}, \Theta)$ is supposed to have a q'th derivative $\theta^{(q')}$ such that

$$\left|q'(t) - q'(s)\right| \le C_1|t-s|^\nu, \; s,t \in [0,1], \tag{3.5}$$

where $C_1 > 0$ and $\nu \in [0,1]$. In what follows, we set $q = q' + \nu$ and we suppose that $q \ge p \ge m$.

ii.) *The eigenvalues of the covariance of* $\int_0^1 P(X,\tau)d\tau P(X,\tau)$ *are strictly positive*

iii.) *The errors are i.i.d. and have density* $f_{\varepsilon|X=x}$ *given* $X = x$ *that is continuous and bounded below by a strictly positive constant at* 0*, uniformly for* x *bounded and continuous in* 0.

iv.) *The choice of knots* $k_n \sim n^{\frac{1}{2q+1}}$*,* $q > 1/2$ *and quasi-uniform placed.*

Since finite sums of Donkser classes (cf. Assumption 1) are again Donsker, we can now formulate the second theorem

Theorem 2 *If Assumptions 1 and 2 are satisfied then the following statements hold:*

i.) *Under the null hypothesis* H_0 *in* (2.9),

$$S_n^{CM^*} \xrightarrow{d} \int ||\mathbb{G}_2^*(y,x)||^2 dF_{YX}(y,x), \tag{3.6}$$

where $(\mathbb{G}_2^*)$ *is a tight zero mean Gaussian processes with a corresponding covariance structure according to Theorem 1.*

ii.) *Under any fixed alternative, i.e., when the data are distributed according to some* F *that satisfies the alternative hypothesis* H_1 *in* (2.9),

$$\lim_{n\to\infty} P(S_n^{CM^*} > \varepsilon) = 1 \textit{ for all constants } \varepsilon > 0. \tag{3.7}$$

In the simulation part we are using a quadratic B-spline for the test statistic $S_n^{CM^*}$.

3.3 Validity of the bootstrap procedure

Finally, we show that the proposed bootstrap procedure computes the correct critical value for out test statistics (2.18) and (2.22). This does not require any further assumptions. Under the null hypothesis, Assumptions 1 ensure that the bootstrap consistently estimates the limiting distribution for (2.18). For the more powerful test statistic $S_n^{CM^*}$, Assumption 2 has to be additionally fulfilled in order to ensure the Donsker property of the empirical cdf estimator. Under any fixed alternative, the bootstrap critical values can be shown to be bounded in probability. Together with Theorem 1ii.) and Theorem 2ii.), respectively, this implies that the proposed tests (2.18) and (2.22) are consistent.

Theorem 3 *Under Assumption 1, the following statements hold true for every* $\alpha \in (0,1)$

i.) *Under the null hypothesis* H_0 *in* (2.9)*, we have that*

$$\lim_{n\to\infty} P(S_n^{CM} > \hat{c}_n(\alpha)) = \alpha$$

ii.) Under any fixed alternative H_1 in (2.9), we have that

$$\lim_{n\to\infty} P(S_n^{CM} > \hat{c}_n(\alpha)) = 1$$

If additionally Assumption 2 is fulfilled, then the following statements hold true for every $\alpha \in (0,1)$

iii.) Under the null hypothesis H_0 in (2.9), we have that

$$\lim_{n\to\infty} P(S_n^{CM^*} > \hat{c}_n(\alpha)) = \alpha$$

iv.) Under any fixed alternative H_1 in (2.9), we have that

$$\lim_{n\to\infty} P(S_n^{CM^*} > \hat{c}_n(\alpha)) = 1$$

In order to study the behavior of the CS induced Cramér-von Mises type test statistic $S_n^{CM^*}$ in finite samples we perform an extensive Monte Carlo Simulation, presented in the next section.

4 Monte Carlo simulation study

In this section, we show that our test $S_n^{CM^*}$ from (3.6) holds the size level and has superior power properties by means of twelve different data generating processes (DGPs). Thereby, the different DGPs cover location shift models (LS) and location-scale shift models (LSS) including heteroscedastic errors, both, in an univariate and multivariate setting. In order to assess the quality and validity of our proposed test against existing procedures, we will compare the test results (cf. Table 1 - Table 4) with the benchmark tests of Koenker and Xiao (2002), Chernozhukov (2002) and Rothe and Wied (2013) where comparisons are possible. Finally, we consider predominantly linear models and show that our test detects such only weakly misspecified models well.

For the definition of the twelve DGPs we introduce the following variables: Let $x_1 \sim Bin(1,0.5)$, $x_2 \sim N(0,1)$, $x_3 \in U(0,1)$, $x_4 \in \chi^2(1)$, $u \sim N(0,1)$, $w \sim N(0,0.1)$, $v = (1-2x_1)\cdot v_2^* \cdot 8^{-0.5}$ with $v_2^* \sim \chi^2(2)$, where $Bin(\cdot,\cdot)$ describes the Binomial, $N(\cdot,\cdot)$ the normal, $U(\cdot,\cdot)$ the uniform and $\chi^2(\cdot)$ the chi squared distribution. Further, let $x_0 \in [0,2\pi]$ and the variables $x_1, x_2, x_3, x_4, u, v, w$ be ally mutually independent.

$$
\begin{aligned}
&\text{(DGP 1):} && f_1(x_0) := 0.25x_0 + 1 + u \\
&\text{(DGP 2):} && f_2(x_0) := 0.25x_0 + 1 + u \cdot x_0 \qquad (4.1)\\
&\text{(DGP 3):} && f_3(x_0) := 0.25x_0^2 + 1 + u \cdot x_0^2
\end{aligned}
$$

$$
\begin{aligned}
&\text{(DGP 4):} && f_4(x_1,x_2) := x_1 + x_2 + u \\
&\text{(DGP 5):} && f_5(x_1,x_2) := x_1 + x_2 + v \\
&\text{(DGP 6):} && f_6(x_1,x_2) := x_1 + x_2 + (0.5 + x_1)u \qquad (4.2)\\
&\text{(DGP 7):} && f_7(x_1,x_2) := x_1 + x_2 + (0.5 + x_1 + x_2^2)^{0.5}u \\
&\text{(DGP 8):} && f_8(x_1,x_2) := x_1 + x_2 + 0.2(0.5 + x_1 + x_2^2)^{1.5}u
\end{aligned}
$$

$$
\text{(DGP 9):} \qquad f_9(x_2) := x_2 + (1 + \gamma_1 \cdot x_2)u \qquad (4.3)
$$

$$
\begin{aligned}
&\text{(DGP 10):} && f_{10}(x_4) := \begin{cases} 0.25 \cdot x_4^2 + 1 + 0.5 \cdot \varepsilon \cdot x_4^2, & \text{for } \tau \geq 0.5 \\ -0.25 \cdot x_4 + 1 + u \cdot x_4, & \text{otherwise} \end{cases} \\
&\text{(DGP 11):} && f_{11}(x_3) := \sin(-\frac{\pi}{2} + x_3^3) + w \qquad (4.4)\\
&\text{(DGP 12):} && f_{12}(x_3) := e^{f_5(x_3)}
\end{aligned}
$$

DGPs $1-3$ from (4.1) represent the univariate case and serve as preliminary for our empirical application, since they model a linear and quadratic univariate processes. Hereby, DGP 1 describes a simple LS model, DGP 2 a more complex LSS model with a linear regressor and, finally, DPG 3 generates a LSS model with a quadratic influence factor. The multivariate case (cf. (4.2) and (4.3)) is specified by the DGPs $4-8$ that are from Rothe and Wied (2013) and DGP 9 from Chernozhukov (2002). Here, DGP 4 is a simple multivariate LS model with normal distributed errors. DGP 2 is again a simple LS model, but now the errors follow a mixture of a "positive" and "negative" χ^2 distribution with two degrees of freedom (normalized to have unit variance). DGPs $6-8$ are multivariate LSS models where the level of heteroscedasticity increases. DGP 9 is considered in order to compare our proposed testing procedure with those provided in Chernozhukov (2002) and Koenker and Xiao (2002). When $\gamma_1 = 0$ DGP 9 is a LS model, otherwise it is a LSS model. DGPs $10-12$ (cf. (4.4)) are processes in which the functional form appears predominantly linear. DGP 10 is implemented by modeling the lower 50%-quantile linearly, while the upper 50%-quantile is modeled quadratically. Due to the quantile dependence of the regressors, DGP 10 cannot be tested with previous tests and therefore represents an extension of our test. DGP $11-12$ are appearing mainly linear in the interval $[0,1]$ and exhibit non-linear growth only at values close to 1. Assuming a linear model, DGPs of the form $10-12$ often impede the detection of misspecification. In order to illustrate the performance of our test, we draw comparisons to common test procedures in the scope of quantile regression. The test proposed in Koenker and Xiao (2002) (denoted as *KX*), which is based on Khmaladazation, which in turn refers to the Doob-Mayer decomposition of martingales, provides the starting point for quantile regression specification tests. We also consider the enhancement proposed in Chernozhukov (2002) (denoted as *Cher*). Furthermore, we compare our test with Rothe and Wied (2013) (denoted as *RW*) since our test is based on a similar principle but more flexible. The aforementioned tests are characterized by the following properties:

- The *KX*-test models the conditional qf paramterically by assuming a LS or a LSS model. In addition, the regressors are fixed for all quantiles considered and the estimation of non-parameter sparsity and score functions are required.
- In order to avoid such estimation, *Cher* proposes a resampling testing procedure based on *KX* that results in better power and accurate size. However, he still assumes a fully parametrized model under the null hypothesis with non-varying regressors for distinct quantiles.
- *RW* propose a testing procedure for a wide range of parametric models that is based on a Cramèr-von Mises distance between an unrestricted estimate of the joint cdf and the estimate of the joint cdf under the null hypothesis. However, the regressors are assumed to be constant for all quantiles.

To analyze finite sample properties of our testing procedure, we consider different sample sizes n and set the number of Monte Carlo replications to 701, while the number of bootstrap replication is equal to $B = 500$.

Table 1 shows the comparison with *RW* for the univariate DGPs $1-3$. It can be noted that

- compared to *RW* our proposed testing procedure $S_n^{CM^*}$ consistently has better size properties.
- In particular, the test $S_n^{CM^*}$ also manages to maintain the size level when the structure of the error terms is highly heteroscedastic (cf. 5% column of DGP 3 in Table 1).
- In addition, the rejection rate for misspecified models (for DGP 3 we are assuming a linear LSS model in the last column of Table 1) in small samples ($n \leq 300$) is approximately three times higher than for the *RW* test.

Table 1. Size Analysis to the Significance Level 0.10 and 0.05

	*DGP*1		*DGP*2		*DGP*3	
RW	10%	5%	10%	5%	5%	*Power*
$n = 30$	0.077	0.019	0.093	0.039	0.005	**0.032**
$n = 50$	0.061	0.016	0.095	0.038	0.016	**0.045**
$n = 100$	0.056	0.024	0.087	0.033	0.024	**0.075**
$n = 300$	0.055	0.028	0.078	0.032	0.026	**0.312**
$n = 500$	0.056	0.016	0.069	0.029	**0.010**	0.486
$n = 1000$	0.043	0.016	0.069	0.030	0.014	0.883
$n = 2000$	0.064	0.020	0.066	0.030	0.014	1.000
$S_n^{CM^*}$	10%	5%	10%	5%	5%	*Power*
$n = 30$	0.101	0.035	0.089	0.037	0.028	**0.095**
$n = 50$	0.103	0.046	0.074	0.027	0.037	**0.147**
$n = 100$	0.094	0.043	0.112	0.061	0.064	**0.407**
$n = 300$	0.090	0.043	0.159	0.084	0.047	**0.988**
$n = 500$	0.086	0.043	0.111	0.058	**0.050**	1.000
$n = 1000$	0.095	0.048	0.095	0.038	0.056	1.000
$n = 2000$	0.098	0.049	0.092	0.042	0.044	1.000

The number of Monte Carlo repetitions is equal to 701 with 500 bootstrap replications. For the size analysis the wrap speed bootstrap procedure is applied. Here, the quantile is modeled by a B-spline of second order with penalty term $\lambda = 1$ and $\sqrt{n}$ knots, meeting monotonicity assumptions. The 7^{th} and last column named *Power* depicts the power analysis while the quantile function is assumed to follow linear LSS model under the null hypothesis.

Table 2 additionally illustrates the comparison with KX for the DGPs $4-8$, whereby a location shift model is assumed under the null hypothesis. Thus, the results of DGPs 4 and 5 reflect size properties, while DGPs $6-8$ measure the power of our and the benchmark tests RW and KX.

- It can be observed that our test $S_n^{CM^*}$ holds the size for multivariate processes (cf. DGP $4,5$).
- KX has difficulties to detect misspecification when heteroscedasticity prevails (cf. DGP $6-8$).
- RW usually detects misspecification. However, the rejections rate of the test $S_n^{CM^*}$ are clearly higher compared to those from RW even in small samples (cf. $n = 100$ DGP 7 of Table 2).

Table 3 provides a comparison with the standard testing procedure proposed in Koenker and Xiao (2002) and the enhancement from Chernozhukov (2002), where the results of Table 3 of the benchmark tests KX and *Cher* are taken from Chernozhukov (2002).

- Even if the structure of DGP 9 is less complex compared to the other DGPs from (4.1)-(4.4), the test $S_n^{CM^*}$ has consistently better finite sample properties compared to the benchmarks KX and *Cher*.

Table 2. Power Analysis Location Shift (LS)

	RW		*KX*		S_n^{CM*}	
$n = 100$	10%	5%	10%	5%	10%	5%
*DGP*4	0.093	0.048	0.067	0.035	0.122	**0.068**
*DGP*5	0.085	0.033	0.069	0.037	0.114	**0.065**
*DGP*6	0.829	**0.669**	0.082	**0.047**	0.870	**0.838**
*DGP*7	0.404	**0.239**	0.097	**0.049**	0.669	**0.565**
*DGP*8	0.874	**0.746**	0.055	**0.027**	0.970	**0.944**
$n = 300$	10%	5%	10%	5%	10%	5%
*DGP*4	0.109	0.056	0.107	0.039	0.125	**0.068**
*DGP*5	0.096	0.043	0.066	0.024	0.120	**0.056**
*DGP*6	1.000	0.997	0.336	0.231	1.000	1.000
*DGP*7	0.847	0.679	0.147	0.076	0.950	0.908
*DGP*8	1.000	0.997	0.099	0.050	1.000	1.000

All results are one-to-one transferred from Rothe and Wied (2013) (RW). Other details of the set up are as those reported there. The null hypotheses assumes a LS quantile regression model. The number of MC repetitions is equal to 701 with 500 bootstrap replications. Here, the quantile is modeled by a *B*-spline of second order with penalty term $\lambda = 1$ and $\sqrt{n}$ knots, meeting monotonicity assumptions.

- In small samples (cf. $n = 100$) the strong results of *KX* and *Cher* could be improved further.

Table 3. Power Analysis Location Shift (LS) for DGP9

	KX			*Cher*			S_n^{CM*}		
	Size	Power		Size	Power		Size	Power	
$\gamma_1 =$	0	0.2	0.5	0	0.2	0.5	0	0.2	0.5
$n = 100$	0.101	0.264	0.898	0.014	0.348	0.980	**0.0495**	**0.396**	**0.99**
$n = 200$	0.070	0.480	0.988	0.052	0.752	1.000	0.063	0.772	1.000
$n = 300$	0.062	0.622	0.998	0.058	0.910	1.000	0.068	0.930	1.000

All results are one-to-one transferred from Koenker and Xiao (2002) (*KX*) and Chernozhukov (2002) (*Cher*), respectively. Other details of the set up are as those reported there. The null hypotheses assumes a LS quantile regression model. The Monte Carlo study for the proposed test uses 701 replications with $B = 500$ bootstrap replications. The significance level is 0.05.

Finally, Table 4 now examines size and power properties for the DGPSs $10 - 12$.

- In each of the DGPs considered, the test holds the significance level.

- Assuming a linear model, misspecification is detected even in small sample sizes.

- DGP 10 cannot be tested with previous approaches due to the quantile dependent regressors. The slightly lower power for DGP 10 is due to the fact that half of the observations actually follow a linear relationship.

Table 4. Size and Power Analysis under a linear Null Hypothesis

	$DGP10$		$DGP11$		$DGP12$	
S_n^{CM*}	5%	*Power*	5%	*Power*	5%	*Power*
$n = 30$	0.068	0.177	0.014	0.055	0.009	0.069
$n = 50$	0.057	0.189	0.018	0.285	0.013	0.318
$n = 100$	**0.051**	**0.192**	**0.033**	**0.979**	**0.023**	**0.989**
$n = 300$	0.039	0.469	0.040	1.000	0.031	1.000
$n = 500$	0.042	0.519	0.039	1.000	0.029	1.000
$n = 1000$	0.046	0.658	0.034	1.000	0.035	1.000
$n = 2000$	0.042	0.743	0.041	1.000	0.049	1.000

The number of Monte Carlo repetitions is equal to 701 with 500 bootstrap replications. For the size analysis the wrap speed bootstrap procedure is applied. Here, the quantile is modeled by a B-spline of second order with penalty term $\lambda = 1$ and $\sqrt{n}$ knots, meeting monotonicity assumptions. The columns named *Power* depict the power analysis. Under the null hypothesis, the quantile function is modeled as a linear LSS function for DGP 10 and a linear LS function for DGPs $11, 12$, respectively.

In summary, the Monte Carlo study has thus shown that our proposed test procedure holds the significance level and also has superior power properties compared to three benchmark tests, even in small samples. The procedure works for both, univariate and multivariate DGPs and can also test models with quantile-dependent regressors. Even weakly misspecified models are detected in sufficiently large sample sizes.

5 Conditional income disparities between East and West Germany

In this section, we apply the bootstrap version of the specification test for generalized quantile regression models to conditional income distributions in Germany. For this purpose, we utilize information from the German Socio-Economic Panel (SOEP, Wagner et al., 2007). More specifically, we consider real gross annual personal labor income in Germany as defined in Bach et al. (2009) for the years 2001 to 2010. Following the standard literature, we only consider the income of males in full-time employment (see, among others, Card et al. (2013); Dustmann et al. (2009)) in the age range 20-60. This yielded 7220 individuals and is the data set that was also used in Klein et al. (2015). The variables *age*, *origin* (East or West Germany) and *years* are available as covariates (cf. Table 5 for a full description of the data). To obtain an estimate of the quantile function and to take full advantage of the spline approximation, we first regressed the income on the dummy coded variable *years* and then performed a quantile regression using the variables *age* or age^2 on the residuals[5]. This approach takes account of the fact that income grows solely with increasing age. Rather, it can be observed that income increases at the beginning of employment, peaks in middle age and finally decreases (Creedy and Hart, 1979; Klein et al., 2015; Luong and Hébert, 2009) .

In order to gain further insights, we next conduct the Machado-Mata decomposition of the year dummy adjusted data set conditioned on the *origin* according to Melly (2005) and Machado and Mata (2005)[6]. For the decomposition we assume that the quantile function of the income Y can be represented as a function of the form

$$F_{Y|X}^{-1}(\tau|X) = P(X,\tau)'\theta(\tau), \tag{5.1}$$

where X depicts the matrix of covariates, that consists of the variables *age* or age^2 and the quantile $\tau \in (0,1)$. Specifically, we consider here three different linear quantile regression models: The first model describes an entirely linear effect of the regressor *age* on income for all quantiles $\tau \in (0,1)$, i.e. $P(X,\tau) = age$ for all $\tau \in (0,1)$. The second models a quadratic influence of age on income for all quantiles $\tau \in (0,1)$, i.e. $P(X,\tau) = age^2$ for all $\tau \in (0,1)$. And finally, the third model considers the sum of the regressors *age* and age^2 that are constant for all quantiles $\tau \in (0,1)$, i.e. $P(X,\tau) = age + age^2$ for all $\tau \in (0,1)$. Due to the probability integral transform theorem the sequence $P(X,\tau_i)'\hat{\theta}(\tau_i)$ for $\tau_i \overset{i.i.d.}{\sim} Uni(0,1)$, $i = 1,...,n$ constitutes a random sample from the estimated conditional distribution of income Y given the covariates X (Machado and Mata, 2005). In order to obtain the difference between East and West, first, the coefficients for East $\hat{\theta}_E(\tau)$ and West $\hat{\theta}_W(\tau)$ for $\tau \in \{0.1, 0.2, ..., 0.9\}$ are estimated on the basis of the disjoint subsets of the covariates for East X_E and West X_W and the corresponding income in the East Y_E and West Y_W. Second, we draw with replacement B random samples X_E^i and X_W^i for $i = 1,...,B$ from the corresponding covariate subsets X_E and X_W, respectively to obtain a random sample via (5.1) for the distribution of the income Y_l^i, $i = 1,...,B$, $l = E,W$. Thus, the estimated income difference $\hat{\Delta}_y$ for incomes in East Y_E and incomes in West Y_W can now be decomposed according to Machado-Mata as

$$\hat{\Delta}_Y = \hat{F}_{Y_E|X_E}^{-1}(\tau|X_E) - \hat{F}_{Y_W|X_W}^{-1}(\tau|X_W) \tag{5.2}$$

$$= \left(P(X_E^B,\tau) - P(X_W^B,\tau)\right)\hat{\theta}_E(\tau) + \left(\hat{\theta}_E(\tau) - \hat{\theta}_W(\tau)\right)P(X_W^B,\tau), \tag{5.3}$$

[5] We consider this approach justified since four out of six tests did not reject the null hypothesis that there is no correlation between *age* and *year* dummies and age^2 and *year* dummies, respectively.

[6] Another application of the Machado and Mata decomposition for differences in incomes can be found in Landmesser et al. (2016)

where the first summand of (5.3) is the explained while the second summand depicts the unexplained difference.

Table 5. Description of the German labor income data

	Description		
Y	gross market labor income, (continuous $1,257€ \leq Y \leq 280,092€$, average $= 46,641€$)		
origin	indicator for East or West (binary, -1=West (73.8%), 1=East (26.2%))		
age	age of the male in years (continuous, 20 $\leq$ age $\leq$ 60, average = 38)		
years	time in years (categorical, 2001 $\leq$t$\leq$ 2010, 10 years)		
(Sub)sample	Description	Average income (Std.)	Observations
Ger	Entire sample	$51,026$ $(30,569)$	$n = 7220$
West	Subsample with $origin = -1$	$55,141$ $(31,494)$	$n = 5325$
East	Subsample with $origin = 1$	$39,463$ $(24,336)$	$n = 1895$

Table 6 depicts the counterfactual analysis of the effect of *origin* on income. The covariates used for the quantile regressions are *age* (row $4-9$ of Table 6), age^2 (row $11-16$ of Table 6) and the sum of these two variables (row $18-23$ of Table 6). The results in Table 6 suggest that there is a significant income gap between East and West Germany over the period considered, which is particularly striking in the first line, where the observed income differences ranges from 26.21% to 35.49%. However, the income difference between the smallest quantile $\tau = 0.1$ and the largest $\tau = 0.9$ decreases by about eight percent. If income is to be explained by the single covariate *age* or age^2, it cannot be assumed that the model is sufficiently well specified for all quantiles due to high residuals (4.37 for $\tau = 0.1$ and 7.33 for $\tau = 0.9$), indicating misspecification. However, the covariate age^2 seems to be appropriate for the smallest quantile 0.1 while a linear effect of age to income seems to prevail in higher quantiles. In contrast, the additive model $age + age^2$ seems to capture the income effect for all quantiles quite well due to moderate residuals. For all decompositions it holds, that *age* and age^2 contribute a maximum of 16% to the explanation of the income difference between East and West Germany (except highest quantile in age^2, i.e. 25.41). Due to the different residuals and the different explanatory power of the income difference between East and West for the quantile regressions based on *age* or age^2, it seems reasonable to assume that *age* and age^2 have different effects for different quantiles. For example, the residual of the 30% quantile of *age* is about

Table 6. Decomposition of the West/East income differential

$\tau =$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
raw gap	-35.49	-32.4	-33.28	-29.06	-28.38	-26.44	-26.21	-26.93	-27.38
					age				
M-M gap	-39.87	-36.64	-33.52	-31.62	-31.45	-29.37	-29.68	-28.83	-25.89
Explained	-1.84	-3.65	-2.06	-2.99	-2.35	-2.19	-0.85	-0.31	-0.1
Unexplained	-38.02	-32.98	-31.47	-28.63	-29.1	-27.18	-28.83	-28.52	-25.8
% Explained	4.63	9.97	6.13	9.45	7.46	7.46	2.87	1.07	0.38
% Unexplained	95.37	90.03	93.87	90.55	92.54	92.54	97.13	98.93	99.62
Residuals	4.37	4.23	0.24	2.56	3.07	2.93	3.46	1.9	-1.49
					age^2				
M-M gap	-36.41	-38.13	-37.52	-35.49	-31.21	-31.96	-32.26	-31.55	-34.72
Explained	-3.07	-6.08	-4.49	-6.43	-2.81	-2.6	-3.92	-4.35	-8.82
Unexplained	-33.35	-32.05	-33.04	-29.06	-28.4	-29.36	-28.34	-27.2	-25.89
% Explained	8.42	15.94	11.96	18.12	8.99	8.13	12.15	13.8	25.41
% Unexplained	91.58	84.06	88.04	81.88	91.01	91.87	87.85	86.2	74.59
Residuals	0.92	5.73	4.24	6.44	2.83	5.51	6.05	4.62	7.33
					age+age^2				
M-M gap	-33.39	-31.80	-33.16	-30.28	-28.49	-28.90	-27.61	-28.25	-25.69
Explained	2.03	1.55	-1.44	-1.5	0.13	0.31	0.33	-3.09	1.67
Unexplained	-35.42	-33.35	-31.72	-28.78	-28.62	-29.21	-27.94	-25.16	-27.36
% Explained	6.09	4.89	4.34	4.94	0.45	1.09	1.19	10.95	6.49
% Unexplained	93.91	95.11	95.66	95.06	99.55	98.91	98.81	89.05	93.51
Residuals	-2.10	-0.61	-0.12	1.22	0.11	2.45	1.40	1.32	-1.69

The covariates used for the quantile regressions are *age* (row 4-9), *age*^{2} (row 11-16) and the sum of these two variables (row 18-23). The second row *raw gap* depicts the observed income gap between East and West. Remaining rows show three different Machado-Mata decompositions using *age*, *age*^{2} and $age + age^2$ as covariates for the quantile regression models. The rows *M-M gap* are the estimated gap of the income difference depending on the underlying quantile regression model. The quantiles τ range from 0.1 to 0.9. The number of bootstrap replications is equal to 2500. All numbers are in percent. Totals may not sum exactly due to rounding.

18 times smaller than the residual of the corresponding quantile regression using *age*^{2} as explanatory variable. It is therefore suspected that the a linear effect of age dominates in this quantile. The emerging, more general question, at which quantile age has a linear or quadratic effect on income, can be answered with the help of the proposed test.

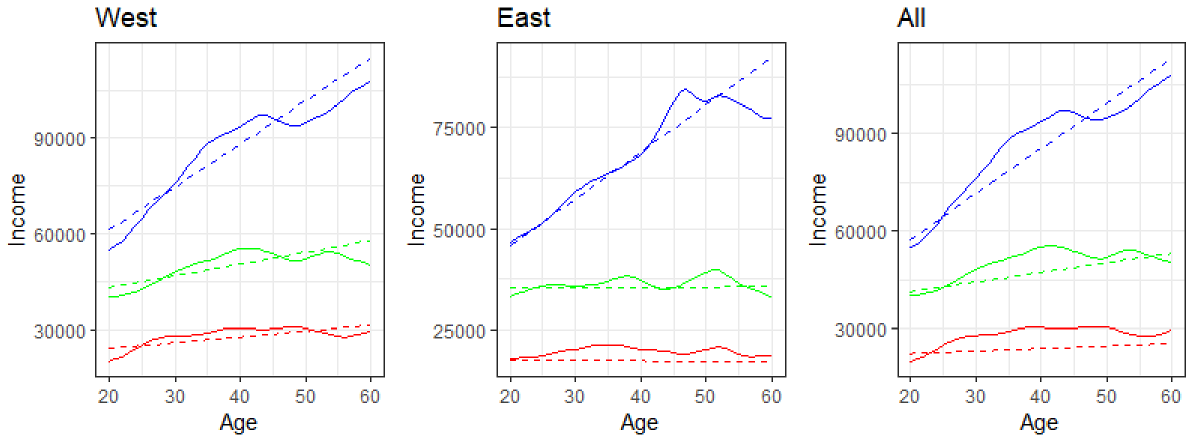
For this purpose, we have defined five different model specifications (5.4), which should take into account the observations of the Machado-Mata decomposition in Table 6. Figure 5.1 visualizes the testing problem and provides further indications of when age might have a quadratic or linear effect. The blue line in Figure 5.1 describes the empirical 90 percent quantile. The corresponding dashed blue line represents the corresponding estimate of the quantile regression. The green (50 percent quantile) and red (10 percent quantile) lines are the equivalent counterparts. In the 5 different quantile regression models considered, the effect of *age* depends on the quantile. Specifications $1-3$ from (5.4) describe quadratic dependencies in the upper or lower quantiles. Specification 4 and 5 model a completely linear and quadratic dependence structure in the covariate, respectively.

The testing procedure is applied to the subsamples East (only individuals from East Germany are considered) and West (only individuals from West Germany are considered) as well as to the complete data set (cf. last column *All* in Table 7).

$$\begin{aligned}
&\text{Specification 1}: && F^{-1}_{Y|X}(\tau|x) = \begin{cases} x^{2\prime}\theta_0, \text{ for } 0 \le \tau \le 0.1 \\ x'\theta_0\ , \text{ for } 0.1 \le \tau \le 0.9 \\ x^{2\prime}\theta_0, \text{ otherwise} \end{cases} \\
&\text{Specification 2}: && F^{-1}_{Y|X}(\tau|x) = \begin{cases} x^{2\prime}\theta_0, \text{ for } 0 \le \tau \le 0.1 \\ x'\theta_0\ , \text{ for } 0.1 \le \tau \le 0.9 \\ x'\theta_0\ , \text{ otherwise} \end{cases} \\
&\text{Specification 3}: && F^{-1}_{Y|X}(\tau|x) = \begin{cases} x'\theta_0\ , \text{ for } 0 \le \tau \le 0.1 \\ x'\theta_0\ , \text{ for } 0.1 \le \tau \le 0.9 \\ x^{2\prime}\theta_0, \text{ otherwise} \end{cases} \\
&\text{Specification 4}: && F^{-1}_{Y|X}(\tau|x) = x'\theta_0 \\
&\text{Specification 5}: && F^{-1}_{Y|X}(\tau|x) = x^{2\prime}\theta_0
\end{aligned} \tag{5.4}$$

Since the sample sizes for East, West and All differ and in order to make the results comparable, we computed the rejection rates of subsamples of East, West and All of size $n = 500, 1000$. We repeated this procedure for every subsample a total of 501 times. The results are listed in Table 7.

Fig. 5.1. Conditional income quantiles for East/West and entire Germany as functions of *age*



Figures show the 0.9 (blue) 0.5 (green) 0.1 (red) smoothed quantiles (using a cubic smoothing spline with smoothing factor 0.5 (cf. R function smooth.spline)) of the income conditioned on the *origin* and the unconditioned data set (*All*). The dashed line depicts the corresponding quantile regression estimate with *age* as covariate.

Table 7. Empirical application: Empirical rejection frequencies of the test statistic $S_n^{CM^*}$

	West	East	All
$n = 500$			
Specification 1	0.023	0.048	0.054
Specification 2	0.122	0.142	0.118
Specification 3	0.010	0.030	0.025
Specification 4	0.080	0.410	0.345
Specification 5	0.066	0.295	0.242
$n = 1000$			
Specification 1	0.014	0.106	0.098
Specification 2	0.242	0.301	0.215
Specification 3	0.019	0.056	0.036
Specification 4	0.128	0.828	0.705
Specification 5	0.082	0.557	0.463

The table depicts the subsample rejection rate of size n of the specification being used from (5.4). The number of subsamplings is 501 and the critical values were calculated at a significance level of 5%.

First, it can be observed that age does not have a completely linear influence on income, as the rejection rates for the 4 specifications are sufficiently high, 0.828 for $n = 1000$, respectively). Assuming a complete quadratic relationship between age and income, this statement cannot be upheld, since the rejection rates for the income distribution in West Germany are below the significance level of 5%. However, the results clearly show that neither in East Germany nor in all of Germany (cf. columns *East* and *All* of Table 7) can the income distribution be adequately described by a quadratic process due to their rejection rates. Second, the Specifications 2 assuming a quadratic structure in the 0.1 and lower quantiles while the remaining quantiles follow a linear model seems for all subsamples considered inappropriate. Third, the Specifications 1 and 3 have the lowest rejection rates for all subsamples indicating that age has a quadratic influence for quantiles 0.9 and higher. In particular, Specification 3 seems to model the income structure sufficiently well for all 3 samples considered. However, as the rejection rates, especially for Specifications 4 and 5, differ sufficiently between East and West, conditional different income distributions between East and West are likely. The test results are in line with the findings of other studies: Based on the different structure of the conditional quantile functions and the corresponding rejections rates for different specifications in Table 7 significant structural differences between East and West Germany can still be assumed (Kluge and Weber, 2018).

6 Conclusion

We believe there are many different areas of application in which the influence of the regressors depends on the quantile linearly or nonlinearly or even in a more complex functional form. A well-known example is the effect of age on the income distribution, which we have taken as illustration. Previous testing procedures of quantile regression are not able to test such influences separately. The present paper proposes a test for generalized quantile regression that addresses these two issues jointly. To improve finite sample properties, we replace quantile regression function by a quadratic monotone B-spline. Our Monte Carlo study illustrates that the proposed method has superior test properties compared to several existing benchmarks from the literature. In addition, a detailed investigation of the conditional income distributions between East and West Germany using the Machado-Mata decomposition reveals that still income differences between the regions in Germany are present, even more than two decades after the reunification. The application of our test could statistically confirm a different functional correlation between the income distributions in East and West Germany.

A Proofs

In order to maintain readability we omit the index $Y|X$ for the conditional cdf F. To prove Theorem 1, we first derive and prove three auxiliary results. Therefore, we define the following three processes for $(y,x) \in \mathbb{R}^{K+1}$ and $(\theta,\tau) \in \Theta \times \mathscr{T}$:

$$\nu_n(y,x) := \sqrt{n}\left(\hat{F}_n(y,x) - F(y,x)\right) \tag{A.1}$$

$$\gamma_n(\theta,\tau) := \sqrt{n}\left(\hat{G}_n(\hat{\theta}_n,\tau) - G(\theta,\tau)\right) \tag{A.2}$$

$$\nu_n^0(y,x) := \sqrt{n}\left(\hat{F}_n(y,x,\hat{\theta}_n) - F(y,x,\theta_0)\right). \tag{A.3}$$

Lemma 1 *Let Assumptions (1) be true. For the processes (A.1) and (A.2) it holds under the null, that*

$$(\nu_n,\gamma_n) \Rightarrow \tilde{\mathbb{G}} := (\mathbb{G}_1,\tilde{\mathbb{G}}_2) \text{ in } \ell^\infty(\mathscr{S} \times \Theta \times \mathscr{T}), \tag{A.4}$$

where $\tilde{\mathbb{G}}$ is a tight bivariate mean zero Gaussian process with

$$\mathbb{G}_1(y,x) := \lim_{n\to\infty} \nu_n(y,x) \tag{A.5}$$

$$\tilde{\mathbb{G}}_2(\theta,\tau) := \lim_{n\to\infty} \gamma_n(\theta,\tau). \tag{A.6}$$

Proof First, we notice that the Donsker property is conserved under the union of Donsker classes. Hence, ν_n and $\gamma_n(\theta,\tau)$ are F_{YX}- Donsker for all $\theta \in \mathscr{B}(\mathscr{T},\Theta)$ and $\tau \in \mathscr{T}$ with limiting process $\mathbb{G}_1(y,x)$ and $\tilde{\mathbb{G}}_2$, respectively. Since arbitrary linear combinations of ν_n and γ_n are Lipschitz and thus Donsker (see Vaart, 1998, Example 29.20), we conclude by the Cramér-Wold theorem that (ν_n,γ_n) converge in distribution to $\tilde{\mathbb{G}}$.

Before we prove the next lemma we slightly generalize Lemma E.3 from Chernozhukov et al. (2013) for our purposes. This modification summarized in the following corollary states conditions under which a Z-estimation process satisfies the functional delta method for Gaussian processes.

Corollary 1 *Let Assumption 1 i.) – iv.) be satisfied and $\sqrt{n}\left(\hat{G}_n - G\right) \Rightarrow \tilde{\mathbb{G}}_2$ in $\ell^\infty(\Theta \times I_l)$ for all $l = 1,\ldots,L$, where $\tilde{\mathbb{G}}_2$ is a Gaussian process with a.s. uniformly continuous paths on $\Theta \times I_l$, $l = 1,\ldots,L$. Further, we assume that the estimator $\hat{\theta}_n(\tau)$ is an approximate Z-estimator ((2.12)) for all $\tau \in I_l$ with $l = 1,\ldots,L$. Then*

$$\sqrt{n}\left(\hat{\theta}_n(\cdot) - \theta_0(\cdot)\right) = -\dot{G}^{-1}_{\theta_0(\cdot),\cdot}\left[\sqrt{n}(\hat{G}_n - G)(\theta_0(\cdot),\cdot)\right] + o_P(1) \tag{A.7}$$

$$\Rightarrow -\dot{G}^{-1}_{\theta_0(\cdot),\cdot}\left[\tilde{\mathbb{G}}_2(\theta_0(\cdot),\cdot)\right] \in \ell^\infty(\mathscr{T}). \tag{A.8}$$

If Assumption 1 v.) also holds, then the paths $\tau \mapsto -\dot{G}^{-1}_{\theta_0(\tau),\tau}\left[\tilde{\mathbb{G}}_2(\theta_0(\tau),\tau)\right]$ *are a.s. uniformly continuous on* $\mathscr{T}$.

Proof The intersection of I_{l_1} and I_{l_2} is a singleton by assumption for $l_1 \neq l_2$. Thus, the set of possible discontinuities is a null set with respect to the Lebesgue measure. Hence, the limiting process $\tilde{\mathbb{G}}_2$ is a.s. continuous on $\Theta \times \mathscr{T}$ with respect to the Euclidean metric. Further we notice, that by assumption the decomposition of the unit interval is finite. Consequently, the property of uniformity is also applicable to the finite union of compact sets. Hence, the conditions of Lemma E.3 in Chernozhukov et al. (2013) are fulfilled.

Lemma 2 *Let either the null hypothesis or a fixed alternative and Assumptions 1 be true. Then it holds that*

$$(\nu_n, \nu_n^0) \Rightarrow \mathbb{G} := (\mathbb{G}_1, \mathbb{G}_2) \text{ in } \ell^\infty(\mathscr{S} \times \mathscr{S}), \tag{A.9}$$

where $\mathbb{G}_1$ *is the limiting tight bivariate mean zero Gaussian process of* ν_n *and*

$$G_2 := \int F(y|x^*)\mathbb{1}_{\{x^* \le x\}} d\mathbb{G}_1(\infty, x^*) + \int \mathbb{G}_2^*(y, x^*)\mathbb{1}_{\{x^* \le x\}} dF_X(x^*). \tag{A.10}$$

Proof Under either the null hypothesis or a fixed alternative, it follows by standard arguments from Lemma 1 and Corollary 1 that

$$\sqrt{n}\left(\hat{F}_n(\cdot,\cdot) - F(\cdot,\cdot), \hat{\theta}_n(\cdot) - \theta_0(\cdot)\right) \Rightarrow \left(\mathbb{G}_1(\cdot,\cdot), -\dot{G}^{-1}_{\theta_0(\cdot),\cdot}(\tilde{\mathbb{G}}_2(\theta_0(\cdot),\cdot)\right) \text{ in } \ell^\infty(\mathscr{S}) \times \ell^\infty(\mathscr{T}). \tag{A.11}$$

Next, it follows from the Hadamard differentiability (cf. Assumption 1 *vii.*)) that

$$\sqrt{n}\left(\hat{F}_n(y|x, \hat{\theta}_n) - F(y|x, \theta_0)\right) \Rightarrow -\dot{F}(y|x, \theta_0)\left[\dot{G}^{-1}_{\theta_0(\cdot),\cdot}(\tilde{\mathbb{G}}_2(\theta_0(\cdot),\cdot)\right] =: \mathbb{G}_2^+(y, x). \tag{A.12}$$

The statement of the lemma then follows directly from the Hadamard derivative $\dot{\phi}$ of the mapping

$$\phi((A,B))[x^*] := \int A(\cdot, x^*)\mathbb{1}_{\{x^* \le \cdot\}} dB(x^*) \tag{A.13}$$

given by

$$\dot{\phi}_{A,B}(\alpha,\beta)[x^*] = \int A(\cdot, x^*)\mathbb{1}_{\{x^* \le \cdot\}} d\beta(x^*) + \int \alpha(\cdot, x^*)\mathbb{1}_{\{x^* \le \cdot\}} dB(\cdot, x^*) \tag{A.14}$$

and the functional delta method. In particular, for the second component $\mathbb{G}_2$ of the joint limiting process, we have

$$\mathbb{G}_2(y,x) = \int \mathbb{G}_2^+(y, x^*)\mathbb{1}_{\{x^* \le x\}} dF_X(x^*) + \int F(y|x^*)\mathbb{1}_{\{x^* \le x\}} d\mathbb{G}_1(\infty, x^*). \tag{A.15}$$

Proof (of Theorem 1) We start with the first statement of Theorem 1. Under the null hypothesis it holds that $\hat{F}_n(y,x) = F(y,x,\theta_0)$ for all $(y,x) \in \mathscr{S}$. By linearity, we have

$$S_n^{CM} = \sqrt{n} \int \left(\hat{F}_n(y,x) - \hat{F}_n(y,x,\hat{\theta})\right) d\hat{F}_n(y,x) \tag{A.16}$$

$$= \int \left(\nu_n(y,x) - \nu_n^0(y,x)\right)^2 dF(y,x) + \int \left(\nu_n(y,x) - \nu_n^0(y,x)\right)^2 d\left(\hat{F}_n(y,x) - F(y,x)\right). \tag{A.17}$$

From Lemma 2 we know that $(\nu, \nu_0) \Rightarrow (\mathbb{G}_1, \mathbb{G}_2) = \mathbb{G}$, where $\mathbb{G}$ is a tight bivariate mean zero Gaussian process. Applying the continuous mapping theorem and the Donsker class property yield

$$S_n^{CM} = \int (\mathbb{G}_1(y,x) - \mathbb{G}_2(y,x)^2 dF(y,x) + o_p(1) \tag{A.18}$$

which claims the statement.

To show part *ii.*), we use the fact that under any fixed alternative $P(F(y,x) \neq F(y,x,\theta_0) > 0$ due to construction of the alternative hypothesis in (2.9). Thus,

$$S_n^{CM} = \int \left(\nu_n(y,x) - \nu_n^0(y,x) + \sqrt{n}(F(y,x) - F(y,x,\theta_0))\right)^2 dF(y,x) + o_P(1) = \mathscr{O}_P(n), \tag{A.19}$$

which implies that S_n^{CM} is greater than any fixed constant $\varepsilon > 0$ and hence, the probability that S_n^{CM} is greater than any $\varepsilon > 0$ tends to 1.

In order to prove Theorem 2 we present the bootstrap version of Lemma 1 as an auxiliary result.

Lemma 3 *Let Assumption 1 be true. We define the bootstrap version of the empirical processes* (A.1) *and* (A.3)

$$\nu_{n,B}(y,x) := \sqrt{n}\left(\hat{F}_{n,B}(y,x) - \hat{F}_n(y,x,\hat{\theta}_n)\right) \tag{A.20}$$

$$\nu_{n,B}^0(y,x) := \sqrt{n}\left(\hat{F}_{n,B}(y,x,\hat{\theta}_n) - \hat{F}_n(y,x,\hat{\theta}_n)\right). \tag{A.21}$$

Then it holds under either the null or a fixed alternative hypothesis that

$$\left(\nu_{n,B}, \nu_{n,B}^0\right) \Rightarrow \mathbb{G}_b, \tag{A.22}$$

where $\mathbb{G}_b := (\mathbb{G}_{b1}, \mathbb{G}_{b2})$ *is a tight bivariate mean zero Gaussian process whose distribution function coincides with that of the process* $\mathbb{G}$ *in Lemma 1.*

Proof This follows from Lemma 1 and the functional delta method for the bootstrap (Rothe and Wied, 2013).

Finally, we can prove the statements of Theorem 3.

Proof (Theorem 3) To prove part *i*.), let $c(\alpha)$ be the true critical value satisfying $P(S_n^{CM} > c(\alpha)) = \alpha + o_P(1)$. Then it follows from Lemma 3 that $\hat{c}_n(\alpha) = c(\alpha) + o_P(1)$. This implies that S_n^{CM} and $\tilde{S}_n := S_n^{CM} - (\hat{c}_n(\alpha) - c(\alpha))$ converge to the same limiting distribution as n tends to infinity. Hence, $P(S_n^{CM} > \hat{c}_n(\alpha)) = \alpha + o_P(1)$ as claimed. To prove part *ii*.), we deduce from Lemma 3 that the bootstrap critical values are bounded in probability under fixed alternatives. Thus, for any $\varepsilon > 0$, there is an $N(\varepsilon)$ such that $P(\hat{c}_n(\alpha) > N(\varepsilon)) < \varepsilon + o_P(1)$. By Kolmogorv axioms we obtain

$$P(S_n^{CM} \leq \hat{c}_n(\alpha)) = P(S_n^{CM} \leq \hat{c}_n(\alpha), S_n^{CM} \leq N(\varepsilon)) + P(S_n^{CM} \leq \hat{c}_n(\alpha), S_n^{CM} > N(\varepsilon)) \tag{A.23}$$

$$\leq P(S_n^{CM} \leq N(\varepsilon)) + P(S_n^{CM} > N(\varepsilon)) \tag{A.24}$$

$$\leq \varepsilon + o_P(1), \tag{A.25}$$

where the last inequality can be deduced from Theorem 1 *ii*.).

Statements *iii*.) and *iv*.) follow in addition with assumption 2 immediately from *i*.) and *ii*.).

References

J. Angrist, V. Chernozhukov, and I. Fernández-Val. Quantile regression under misspecification, with an application to the us wage structure. *Econometrica*, 74(2):539–563, 2006.

S. Bach, G. Corneo, and V. Steiner. From bottom to top: The entire income distribution in germany, 1992 - 2003. *Review of Income and Wealth*, 55(2):303–330, 2009. doi: 10.1111/j.1475-4991.2009.00317.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4991.2009.00317.x.

A. Belloni, V. Chernozhukov, D. Chetverikov, and I. Fernández-Val. Conditional quantile processes based on series or many regressors. *Journal of Econometrics*, 213(1):4–29, 2019.

H. J. Bierens. A consistent conditional moment test of functional form. *Econometrica: Journal of the Econometric Society*, pages 1443–1458, 1990.

M. Biewen. Income inequality in germany during the 1980s and 1990s. *Review of Income and Wealth*, 46 (1):1–19, 2000.

P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1995. ISBN 9780471007104. URL https://books.google.de/books?id=z39jQgAACAAJ.

H. D. Bondell, B. J. Reich, and H. Wang. Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838, 2010.

C. Breunig. Specification testing in nonparametric instrumental quantile regression. *arXiv preprint arXiv:1909.10129*, 2019.

D. Card, J. Heining, and P. Kline. Workplace Heterogeneity and the Rise of West German Wage Inequality*. *The Quarterly Journal of Economics*, 128(3):967–1015, 05 2013. ISSN 0033-5533. doi: 10.1093/qje/qjt006. URL https://doi.org/10.1093/qje/qjt006.

H. Cardot, C. Crambes, and P. Sarda. Quantile regression when the covariates are functions. *Nonparametric Statistics*, 17(7):841–856, 2005.

V. Chernozhukov. Inference on quantile regression process, an alternative. *SSRN Electronic Journal*, 02 2002. doi: 10.2139/ssrn.303187.

V. Chernozhukov, I. Fernández-Val, and B. Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013. doi: 10.3982/ECTA10582. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10582.

J. Creedy and P. E. Hart. Age and the distribution of earnings. *The Economic Journal*, 89(354):280–293, 1979.

R. M. Dudley. *Uniform Central Limit Theorems*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2014. doi: 10.1017/CBO9781139014830.

C. Dustmann, J. Ludsteck, and U. Schönberg. Revisiting the German Wage Structure*. *The Quarterly Journal of Economics*, 124(2):843–881, 05 2009. ISSN 0033-5533. doi: 10.1162/qjec.2009.124.2.843. URL https://doi.org/10.1162/qjec.2009.124.2.843.

J. Escanciano, C. Goh, et al. Specification analysis of structural quantile regression models. Technical report, 2010.

J. C. Escanciano and C. Velasco. Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics*, 159(1):209–221, 2010.

P. L. Gozalo. A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory*, 9(3):451–477, 1993.

E. Guerre and C. Sabbah. Uniform bias study and bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, 28(1):87–129, 2012.

M. Hallin, Z. Lu, K. Yu, et al. Local linear spatial quantile regression. *Bernoulli*, 15(3):659–686, 2009.

X. He and P. Shi. Monotone b-spline smoothing. *Journal of the American Statistical Association*, 93, 10 1997. doi: 10.2307/2670115.

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

N. Klein, T. Kneib, S. Lang, A. Sohn, et al. Bayesian structured additive distributional regression with an application to regional income inequality in germany. *The Annals of Applied Statistics*, 9(2):1024–1052, 2015.

J. Kluge and M. Weber. Decomposing the german east–west wage gap. *Economics of Transition*, 26(1): 91–125, 2018.

R. Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005. doi: 10.1017/CBO9780511754098.

R. Koenker and G. Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

R. Koenker and Z. Xiao. Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612, 2002.

R. Koenker, P. Ng, and S. Portnoy. Quantile smoothing splines. *Biometrika*, 81(4):673–680, 1994.

J. M. Landmesser et al. Decomposition of differences in income distributions using quantile regression. *Statistics in Transition. New Series*, 17(2):331–349, 2016.

D. Li, Q. Li, and Z. Li. Nonparametric quantile regression estimation with mixed discrete and continuous data. *Journal of Business & Economic Statistics*, pages 1–16, 2020.

Q. Li and J. S. Racine. Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, 26(4):423–434, 2008.

M. Luong and B.-P. Hébert. *Age and earnings*. Citeseer, 2009.

J. A. F. Machado and J. Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465, 2005. doi: 10.1002/jae.788. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.788`.

B. Melly. Public-private sector wage differentials in Germany: Evidence from quantile regression. *Empirical Economics*, 30:505–520, 2005. ISSN 0033-5533. doi: https://doi.org/10.1007/s00181-005-0251-y.

Z. Qu and J. Yoon. Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, 185(1):1–19, 2015.

C. Rothe and D. Wied. Misspecification testing in a class of conditional distributional models. *Journal of the American Statistical Association*, 108(501):314–324, 2013. doi: 10.1080/01621459.2012.736903. URL `https://doi.org/10.1080/01621459.2012.736903`.

W. Stute. Nonparametric model checks for regression. *The Annals of Statistics*, pages 613–641, 1997.

W. Stute, W. G. Manteiga, and M. P. Quindimil. Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93(441):141–149, 1998.

A. W. v. d. Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

G. Wagner, J. Frick, and J. Schupp. The german socio-economic panel study (soep) - scope, evolution and enhancements. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 127(1):139 – 169, 2007. URL `https://EconPapers.repec.org/RePEc:aeq:aeqsjb:v127_y2007_i1_q1_p139-169`.

L. Xue and J. Wang. Distribution function estimation by constrained polynomial spline regression. *Journal of Nonparametric Statistics*, 22(4):443–457, 2010.

D. Yu, L. Kong, and I. Mizera. An alternative approach to functional linear partial quantile regression. *arXiv preprint arXiv:1709.02069*, 2017.