

# Flexible specification testing in quantile regression models

Tim Kutzker<sup>1,2</sup> | Nadja Klein<sup>1,3</sup>  | Dominik Wied<sup>2</sup>

<sup>1</sup>Emmy Noether Research Group, Humboldt-Universität zu Berlin, Berlin, Germany

<sup>2</sup>Institute for Econometrics and Statistics, Universität zu Köln, Cologne, Germany

<sup>3</sup>Research Center for Trustworthy Data Science and Security (UA Ruhr) and Department of Statistics (Technische Universität Dortmund), Dortmund, Germany

## Correspondence

Tim Kutzker, Emmy Noether Research Group, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany.  
Email: [tim.kutzker@gmail.com](mailto:tim.kutzker@gmail.com)

## Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Numbers: Emmy Noether grant KL 3037/1-1, INST 216/512/1FUGG

## Abstract

We propose three novel consistent specification tests for quantile regression models which generalize former tests in three ways. First, we allow the covariate effects to be quantile-dependent and nonlinear. Second, we allow parameterizing the conditional quantile functions by appropriate basis functions, rather than parametrically. We are thereby able to test for general functional forms, while retaining linear effects as special cases. In both cases, the induced class of conditional distribution functions is tested with a Cramér–von Mises type test statistic for which we derive the theoretical limit distribution and propose a bootstrap method. Third, a modified test statistic is derived to increase the power of the tests. We highlight the merits of our tests in a detailed MC study and two real data examples. Our first application to conditional income distributions in Germany indicates that there are not only still significant differences between East and West but also across the quantiles of the conditional income distributions, when conditioning on age and year. The second application to data from the Australian national electricity market reveals the importance of using interaction effects for modeling the highly skewed and heavy-tailed distributions of energy prices conditional on day, time of day and demand.

## KEYWORDS

B-splines, Cramér–von Mises test statistic, distribution regression, series terms, specification tests

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

## 1 | INTRODUCTION

Hypothesis testing plays a central role in many 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. For linear regression, a whole range of specification tests are available for both, parametric and nonparametric approaches. In general, testing misspecification in linear ordinary least squares (OLS) models is well understood and developed.

In parametric models, for example, Bierens (1990) showed that any conditional moment test of functional form of nonlinear regression models can be converted into a consistent chi-squared 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, while Stute (1997) proposed a more 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 class.

However, OLS estimates are sensitive to outliers and draw only a part of the whole picture since they only model the mean. In contrast, quantile regression provides more robust estimates and allows a more comprehensive picture of the entire conditional distribution. Due to these advantages, quantile regression has become increasingly popular since the seminal article by Koenker and Bassett (1978). However, post-estimation inference procedures for quantile regression models 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 (M–M) decomposition in order to describe wage inequalities (Machado & Mata, 2005) and the Khmaladze transformation (Koenker & Xiao, 2002). Thus, testing the validity of the imposed structure remains one of the key tasks associated with challenges for valid posterior inference.

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 thereafter, Chernozhukov (2002) proposes a resampling test procedure that avoids the estimation of additional objects, such as the score function, while building on the principles stated in Koenker and Xiao (2002). However, these two tests do not test the validity of the quantile regression model itself, they only test for restrictions on the parameters. Escanciano and Velasco (2010) and Escanciano and Goh (2014) 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. They consider classes of conditional distribution functions with function-valued parameters and test whether the underlying cdf of the sample lies in one of these classes. This principle was extended to dynamic models by Troster and Wied (2021).

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. Hallin et al. (2009) suggests 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 (qf) and its derivatives. Li and Racine (2008) propose

a nonparametric conditional cumulative distribution function (cdf) kernel estimator along with an associated nonparametric conditional quantile estimator. Belloni et al. (2019) develop nonparametric quantile regression for performing inference on the entire conditional qf and its linear functionals and Qu and Yoon (2015) presented estimators for nonparametrically specified conditional quantile processes that are based on local linear regressions. Li et al. (2020) investigated the problem of nonparametrically estimating a conditional qf with discrete and continuous covariates suggesting a kernel based approach.

But regardless of whether parametric or nonparametric 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. For example, to the best of our knowledge, there does not exist a testing procedure that allows to test for general quantile-specific functional (such as nonlinear) covariate effects. While nonlinear regressors can be considered in a standard parametric quantile regression framework, testing for presence of such nonlinear effects at certain quantiles usually requires estimating them for all quantiles. We propose a method that does not require this. Moreover, we incorporate spline-based estimation approaches in order to allow for more flexible nonlinear effects than in a parametric setting. On the one hand, we are able to test if a particular spline approximation is appropriate, on the other hand, we show that the spline approach can be used to obtain more powerful tests.

Our contributions are as follows. First, we suggest a general procedure for quantile regression models, where the regressors can explicitly depend on quantiles. This allows to test for the correct specification of large number of parametric models. Second, due to our general model set-up, our proposed methodology also allows to test for finite semi-parametric models. One of such examples are B-splines for quantile regressions, where the finite number of covariates have a general functional form depending on the quantile (Cardot et al., 2005). Additionally, our second test allows to test for the order and the correct number of knots of the B-spline specifications. Third, a test is developed in the framework of quantile regression models with an increasing number of knots. This third test can also be applied to (semi-)parametric quantile regression models which turns out to be a more powerful testing procedure. Last, we derive a valid bootstrap procedure is a practical easy-to-implement algorithm to calculate critical values of the limiting distributions. Overall, our framework therefore extends the literature on quantile regression specification tests to situations where specific regressors may have a complex functional impact and/or the respective effects may vary over quantiles.

The key idea of our framework 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 vector  $X$  with a restricted estimate that imposes the structure implied by the null hypothesis model. Based on a Cramér–von Mises type measure of distances, the restricted estimate of the joint distribution can then be compared with the unrestricted one. We derive the nonpivotal 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. Due to the generality of our test procedure we can subsume previous specification tests for quantile regression models with i.i.d. data as special cases of our procedure. Our extensive Monte Carlo (MC) simulation study in the Supplement shows that our testing procedures are consistent and have superior power properties than existing benchmark methods, where comparisons are possible.

Finally, to illustrate the power and potential of our tests, we consider two real data applications. First, the case of income inequality is treated, 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. Importantly, and in line with other studies on this topic, we find through an initial M–M decomposition that there are still income differences between East and West Germany, which can be confirmed by our proposed testing procedure. The second application arises from energy economics. Following recent work in Smith and Klein (2021), we consider spot prices from the Australian national electricity market from 2019 and analyze in which sense its conditional quantiles can be explained by different covariates. These authors have shown that the distribution is heavily skewed and far from Gaussian with complex interactions of the three covariates day of the year, the time of day and the demand. We statistically confirm that interaction effects have a substantial impact on the electricity price, especially for the lower quantiles.

The paper is organized as follows. Section 2 formulates the test problem for the finite-dimensional parametric and semi-parametric models. In Section 3, we provide the theoretical properties of the testing procedures and derive their limiting distributions. Section 4 describes a practical and easy-to-implement bootstrap procedure, which provides valid coverage. In Section 6 we present the two empirical applications. The last Section 7 concludes. Supplement contains all proofs of our theoretical results, as well as an extensive MC study including comparisons to existing tests and further results on the second application.

## 2 | QUANTILE REGRESSION TESTING

In this section, we introduce three specification tests for (semi-)parametric quantile regression models comparing the empirical conditional cumulative distribution function (ecdf) with the (semi-)parametric joint cdf that is based on the estimated conditional qf. We denote these tests by  $S_n^{CM}$ ,  $S_n^{CM,S}$  and  $S_n^{CM*}$ . In Section 2.1, we derive the general test principle along the lines of parametric models. In contrast to existing approaches, the test for parametric quantile regression models  $S_n^{CM}$  allows the covariates  $X$  to be quantile-dependent. This feature is important in many applications as we illustrate in our first application where the effect of age has distinct functional forms depending on the quantile of interest of the conditional income distribution. For the case of quantile-independent covariates, the test reduces to the test statistic from Rothe and Wied (2013). Section 2.2 applies the general test principle to finite-dimensional semi-parametric models with the specification test denoted by  $S_n^{CM,S}$ . As an illustrative example, we consider B-splines, where the degree of the spline and the dimension of the vector of knots is known and finite. We choose the wording “semi-parametric” because we consider a (possibly penalized) spline with a fixed dimension, whereas the total number of parameters is (possibly much) higher than in the first test of Section 2.1. Similar to the first test, it is of high relevance in practice, as we show in our second application on electricity price distributions. It should be stated, however, that the computational complexity increases in the degree of the spline and in the dimension of the vector of knots, compare Toraichi et al. (1987) for precise results for splines. In Section 2.3, we introduce a more powerful model specification testing procedure  $S_n^{CM*}$ , which is illustrated on the class of parametric quantile regression models. To do so, we replace the empirical conditional cdf in the test statistic  $S_n^{CM}$  with an appropriate spline representation that approximates the true joint cdf faster. The price of the higher power is that the class of true cdfs is restricted more strongly.

For the approach, we need (in contrast to the case in Section 2.2) splines whose dimension grows as a function of the number of observations, that is, the degree of the spline is fixed while the dimension of the knot vector (and that the dimension of the vector of regression coefficients) diverges at an appropriate rate. This of course increases the computational complexity even more. Finally, we note that it would in principle also be possible to do this extension for  $S_n^{CM,S}$  with some additional assumptions but doing so in detail is beyond the scope of this paper.

## 2.1 | Quantile regression and the general test principle

Let  $Y_i \in \mathbb{R}$  denote the outcome variable and  $X_i \in \mathbb{R}^K$  the vector of explanatory variables of i.i.d. data points for  $i = 1, \dots, n$  and  $K \in \mathbb{N}$ . Our aim is to test the validity of certain model specifications for quantile regression. Specifically, we consider models of the form

$$F_{Y|X}^{-1}(\tau | x) = P(x, \tau)^\top \theta(\tau), \quad (1)$$

where  $F_{Y|X}^{-1}(\tau | x)$  denotes the qf of  $Y$  conditional on  $X = x \in \mathbb{R}^K$  at quantile  $\tau$ ,  $P(x, \tau) \in \mathbb{R}^{p_\tau}$  is a transformation vector of  $x$  with  $p_\tau \in \mathbb{N}$  and  $\theta(\tau) \in \mathbb{R}^{p_\tau}$  is the parameter vector depending on  $\tau$  for all  $\tau \in \mathcal{T} \subset (0, 1)$ . Naturally, models in which the vector of transformations does not depend on  $\tau$  are captured by our approach as a special case ( $P(x, \tau) \equiv x$  is the familiar case of linear quantile regression). As noted by Belloni et al. (2019) for  $P(x, \tau) \equiv P(x)$ , the above framework incorporates a variety of models such as parametric (Koenker, 2005) and semi-parametric (He & Shi, 1998) ones. However, since we allow the transformation vector  $P(x, \tau)$  to depend on the quantile  $\tau$ , models of the form (1) are generalizations. In parametric quantile regression models,  $P(x, \tau)$  could for instance represent a linear covariate in the lower 50% quantile and a highly nonlinear functional form in the upper 50% quantile, for example,  $P(x, \tau) = x$  if  $\tau \leq 0.5$  and  $P(x, \tau) = \sin(x)x^2$  otherwise. In semi-parametric models,  $P(x, \tau)$  could represent the knot vector for cubic B-splines that differs for distinct quantiles as in our second application in Section 6.2. For ease of notation and without loss of generality, we assume  $p_\tau =: p \in \mathbb{N}$  for all  $\tau \in \mathcal{T}$ . In the remainder of this subsection we assume the qf according to (1) to be specified by a parametric model, while generalizations are treated thereafter.

Our test principle is designed for the comparison of the nonparametric with the parametric joint cdf, where the latter can be expressed by means of the parametric conditional cdf. In general, the conditional cdf  $F$  of  $Y$  conditioned on  $X$ , denoted as  $F_{Y|X}$ , in turn is induced by its corresponding (generalized) conditional qf  $F_{Y|X}^{-1}$  through the following equation

$$F_{Y|X}(y | x) = \int_0^1 \mathbb{1}_{\{F_{Y|X}^{-1}(\tau | x) \leq y\}} d\tau \quad \forall y \in \mathbb{R}. \quad (2)$$

In the following, we consider the set of all conditional distribution functions satisfying (2) given the model specification (1), which we denote by  $\mathcal{F}$ , that is

$$\mathcal{F} := \{F_{Y|X}(y | x, \theta) \mid F_{Y|X}^{-1}(\tau | x) = P(x, \tau)^\top \theta(\tau) \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta), (y, x) \in \mathcal{S}\}, \quad (3)$$

where  $\mathcal{S}$  denotes the support of  $(y, x) \in \mathbb{R}^{K+1}$  and  $\mathcal{B}(\mathcal{T}, \Theta)$  the class of functions  $\tau \mapsto \theta(\tau) \in \Theta \subset \mathbb{R}^p$ . The specification testing problem of whether our model (1) is correctly specified for all  $\tau \in \mathcal{T}$

transfers by means of (3) to hypotheses of the form

$$H_0 : F_{Y|X} \in \mathcal{F} \quad \text{vs.} \quad H_1 : F_{Y|X} \notin \mathcal{F}. \quad (4)$$

Thus, we want to test if the conditional cdf  $F_{Y|X}$  coincides with an element of  $\mathcal{F}$  from (3). For this testing problem, we assume a unique  $\theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$  under the null hypothesis, such that  $\theta(\tau) = \theta_0(\tau)$  for all  $\tau \in \mathcal{T}$ . This yields  $\mathcal{F}^0 := \{F_{Y|X}(y|x, \theta_0) \mid F_{Y|X}^{-1}(\tau|x) = P(x, \tau)^\top \theta_0(\tau) \text{ for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \forall (y, x) \in S\}$ . Hence, we can reformulate (4) as

$$H_0 : F_{Y|X}(y|x) = F_{Y|X}(y|x, \theta_0) \text{ for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \text{ for all } (y, x) \in S \quad (5)$$

$$\text{vs. } H_1 : F_{Y|X}(y|x) \neq F_{Y|X}(y|x, \theta) \text{ for all } \theta \in \mathcal{B}(\mathcal{T}, \Theta) \text{ for some } (y, x) \in S. \quad (6)$$

Additionally we assume that  $\theta_0$  is identified under the null hypothesis through a moment condition. Specifically, let  $g : S \times \Theta \times \mathcal{T} \rightarrow \mathbb{R}^p$  be a uniformly integrable function whose exact form depends on  $\mathcal{F}^0$ , and suppose that for every  $\tau \in \mathcal{T}$

$$G(\theta, \tau) := \mathbb{E}[g(Y, X, \theta, \tau)] = 0 \in \mathbb{R}^p \quad (7)$$

has a unique solution  $\theta_0(\tau)$ . Furthermore, under the alternative  $H_1$ , we assume  $\theta_0(\tau)$  to be uniquely defined as the solution to  $\inf_{\theta \in \Theta} \|G(\theta, \tau)\|$  from (7) for all  $\tau \in \mathcal{T}$  and thus can be regarded as a pseudo-true value of the functional parameter in this case. Incorporating the moment condition, we can now rewrite the null hypothesis of (4) as

$$F_{Y|X}(y|x) = F_{Y|X}(y|x, \theta_0) \text{ for all } (y, x) \in \mathbb{R}^{K+1},$$

with  $\theta_0(\tau)$  as the unique solution to (7) for all  $\tau \in \mathcal{T}$ . This holds true since  $\mathcal{F}^0$  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  $F$  of  $Y$  and  $X$  as

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

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

where  $F_X$  denotes the marginal cdf of  $X$ . From Theorem 16.10(iii) of Billingsley (1995) it follows that the testing problem (5) can be restated as

$$H_0 : F(y, x) = F(y, x, \theta_0) \text{ for all } (y, x) \in \mathbb{R}^{K+1}$$

$$\text{vs. } H_1 : F(y, x) \neq F(y, x, \theta_0) \text{ for some } (y, x) \in \mathbb{R}^{K+1}. \quad (8)$$

Furthermore, let  $S : \mathbb{R}^{K+1} \times \Theta \rightarrow \mathbb{R}$  be a function that measures the difference of the nonparametric  $F(y, x)$  and the parametrized cdf  $F(y, x, \theta)$  defined as

$$S(y, x, \theta) := F(y, x) - F(y, x, \theta). \quad (9)$$

The null hypothesis is true if  $S(y, x, \theta_0) = 0$  for all  $(y, x) \in S$ , whereas  $S(y, x, \theta) \neq 0$  for all  $\theta \neq \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta)$  and for some  $(y, x) \in S$ . The sample analog is

$$S_n(y, x, \hat{\theta}_n) := \hat{F}_n(y, x) - \hat{F}_n(y, x, \hat{\theta}_n), \quad (10)$$

where  $\hat{F}_n(y, x)$  is the empirical cdf and  $\hat{F}_n(y, x, \hat{\theta})$  a parametric estimate of  $F$  based on a consistent estimate  $\hat{\theta}_n(\tau)$  of  $\theta_0(\tau)$  for all  $\tau \in \mathcal{T}$  corresponding to the underlying model assumption (1). Under the null hypothesis,  $\hat{F}_n(y, x, \hat{\theta}_n)$  is a consistent estimator for  $F(y, x, \theta_0)$ , whereas  $\hat{F}_n(y, x)$  consistently estimates  $F(y, x)$ . In that case,  $S_n(y, x, \hat{\theta}_n)$  should be close to zero for all  $(y, x) \in S$ . If, however, the alternative holds true, then there is a vector  $(y, x) \in S$  for each  $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$  such that the absolute value of the function  $S_n$  from (10) 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 the function  $\hat{\theta}_n$  to be an approximate  $Z$ -estimator satisfying

$$\|\hat{G}_n(\hat{\theta}_n, \tau)\| = \inf_{\theta \in \Theta} \|\hat{G}_n(\theta, \tau)\| + \eta_n, \quad (11)$$

where the function  $\hat{G}_n(\hat{\theta}_n, \tau) := n^{-1} \sum_{i=1}^n g(Y_i, X_i, \theta, \tau)$  is the sample analog of the moment condition (7) for every  $\tau \in \mathcal{T}$  and for some possibly random variable  $\eta_n = o_p(n^{-1/2})$ . For every  $\tau \in \mathcal{T}$  and every  $(y, x) \in S$ , the estimator based on the testing problem (5) is

$$\begin{aligned} \hat{F}_n(y | x, \hat{\theta}_n) &= \int_{\mathcal{T}} \mathbb{1}_{\{P(x, \tau)^\top \hat{\theta}_n(\tau) \leq y\}} d\tau + \int_{(0,1) \setminus \mathcal{T}} \mathbb{1}_{\{\hat{F}_{Y|X}^{-1}(\tau | X) \leq y\}} d\tau, \\ \hat{\theta}_n(\tau) &= \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^n (\tau - \mathbb{1}_{\{y_i \leq P(x_i, \tau)^\top \theta\}})(y_i - P(x_i, \tau)^\top \theta). \end{aligned} \quad (12)$$

For  $\tau \notin \mathcal{T}$ , the conditional qf  $\hat{F}_{Y|X}^{-1}(\tau | X)$  is some estimator for the conditional qf which is consistent both under the null and the alternative hypothesis (e.g., Soni et al., 2012; Takeuchi et al., 2006). If  $\mathcal{T} = [\epsilon, 1 - \epsilon]$  for a small  $\epsilon > 0$ , this term is negligible in practice. The integral in (12) can be computed by means of standard numerical integration techniques (i.e., averaging over a fine equidistant grid of  $\tau$  with 49 supporting points starting at 0.02 using the trapezoidal rule) and corresponds to the canonical quantile regression approach, that is, the loss function  $g$  from (7) is given by  $g(Y, X, \theta, \tau) = (\tau - \mathbb{1}\{Y \leq P(X, \tau)^\top \theta\})P(X, \tau)$  (cf. Lemma 14 of Chernozhukov et al., 2013). Additionally, (12) and other typical estimation methods fit the estimated conditional qf  $\hat{F}_n^{-1}(\tau | x)$  pointwise in  $\tau \in \mathcal{T}$ , which might induce the problem that the estimated quantile curve  $\tau \mapsto \hat{F}_n^{-1}(\tau | x)$  violates the monotonicity constraint. This in turn may cause crossing quantile curves. However, a violation of the monotonicity constraint does not affect the validity of the test statistic, since it is based on transformations of  $\hat{F}_n(y | x, \hat{\theta}_n)$  which is monotone in  $y$  by construction for every  $x$ . Hence, a valid test statistic can be based on the differences of the nonparametric and parametric ecdfs  $\hat{F}_n(y, x)$  and  $\hat{F}_n(y, x, \hat{\theta}_n)$  and thus expressed as

$$\begin{aligned} S_n(y, x, \hat{\theta}_n) &= \hat{F}_n(y, x) - \hat{F}_n(y, x, \hat{\theta}_n) \\ &= \frac{1}{n} \sum_{i=1}^n (\mathbb{1}_{\{Y_i \leq y\}} \mathbb{1}_{\{X_i \leq x\}}) - \int_{\mathbb{R}^K} \mathbb{1}_{\{x^* \leq x\}} \hat{F}_n(y | x, \hat{\theta}_n) d\hat{F}_X(x^*) \\ &= \frac{1}{n} \sum_{i=1}^n (\mathbb{1}_{\{Y_i \leq y\}} \mathbb{1}_{\{X_i \leq x\}} - \mathbb{1}_{\{X_i \leq x\}} \hat{F}_n(y | x, \hat{\theta}_n)), \end{aligned} \quad (13)$$

where the third line exploits the definition of the integral with respect to the ecdf  $\hat{F}_X$ . We propose a Cramér–von Mises type (CM) test statistic  $S_n^{CM}$  defined as

$$S_n^{CM} := \int \left( \sqrt{n} S_n(y, x, \hat{\theta}_n) \right)^2 d\hat{F}_n(y, x), \quad (14)$$

which is due to the quantile dependence of the covariates a generalization of existing quantile regression tests. However, if the vector of transformations  $P(x, \tau)$  in (1) is independent of  $\tau$  then the test statistic coincides with test statistic proposed in Rothe and Wied (2013). It is also possible to consider a Kolmogorov–Smirnov-type test statistic

$$S_n^{KS} := \sqrt{n} \sup_{(y,x) \in S^1} \left| S_n(y, x, \hat{\theta}_n) \right|,$$

but the CM test yields better (power) results (Chernozhukov, 2002; Rothe & Wied, 2013).

## 2.2 | Specification test for semi-parametric quantile regression

Since we introduced the general testing principle of (1) by means of the parametric model, this subsection briefly demonstrates that the general test principle is also applicable to finite dimensional semi-parametric models. This particularly addresses the fact that parametric models are often too restrictive and implausible from an applied perspective, since, among others, the constantly increasing complexity of data sets also makes modeling by simple functional relationships more difficult.

In the following, we identify the vector of transformations  $P(x, \tau)$  as basis functions (often referred to as series terms; Belloni et al., 2019; Chao et al., 2017; Chernozhukov et al., 2013). To distinguish such basis functions from the vector of transformations in the previous subsection, we use the notation  $B$  instead. Due to their widespread use, we will derive the semi-parametric test for B-splines bases, although our general test principle also allows for other semi-parametric forms such as penalized splines, Fourier series or compactly supported wavelets (Chao et al., 2017). For ease of well-defined expression and readability, we assume w.l.o.g. that the vector of covariates  $X \in \mathbb{R}^K$  is properly scaled and centered and that  $M \in \mathbb{N}$  uniformly spaced knots  $0 = t_1 < \dots < t_M = 1$  in the interval  $[0, 1]$  are given. For  $x = (x_1, \dots, x_K)^\top \in [0, 1]^K$  with  $K \in \mathbb{N}$ , we identify the B-spline quantile regression model in the spirit of (1) as

$$F_{Y|X}^{-1}(\tau | X = x) = \sum_{j=1}^K B(x_j | d_\tau)^\top \theta_j(\tau) \quad (15)$$

with  $B(x_j | d_\tau) := (B_1(x_j | d_\tau), \dots, B_{M+d_\tau-1}(x_j | d_\tau))^\top$  being the vector of  $M + d_\tau - 1$  basis functions of degree  $d_\tau$  that are defined recursively on the vector of knots on  $[0, 1]$  and evaluated at  $x_j$  for  $j = 1, \dots, K$  (cf. De Boor, 1978, for the recursive Definition). For every  $\tau \in \mathcal{T}$  and  $j = 1, \dots, K$ ,  $\theta_j(\tau) = (\theta_{j,1}(\tau), \dots, \theta_{j,M+d_\tau-1}(\tau))^\top$  defines the corresponding functional coefficient vectors. Although both  $M$  and  $d_\tau$  can be conceived to depend on  $j$  for  $j = 1, \dots, K$  and additionally  $M$  on  $\tau$ , we suppress these dependencies at this point due to readability and clearness. Note that for distinct quantiles  $\tau$  the degree of the B-spline might differ. If  $d_\tau \equiv d \in \mathbb{N}$  we refer to (15) as B-spline quantile regression model of degree  $d$ . Since our general quantile regression model in (1) conceptually

allows for multivariate covariates, we make (15) more flexible by adding  $q^* \in \mathbb{N}$  arbitrary product interaction effects of the form  $\pi_i(x) = \prod_{j \in J_i} f_j(x_j)$ , where  $J_i$  is an arbitrary subset of  $\{1, \dots, K\}$  for  $i = 1, \dots, q^*$  and  $f_j$  an arbitrary continuous function for  $j \in J_i$ . Thus, (15) generalizes to

$$F_{Y|X}^{-1}(\tau | X = x) = \sum_{j=1}^q B(\pi_j(x) | d_\tau)^\top \theta_j(\tau) \quad (16)$$

with  $q = K + q^*$ . For  $\pi_j(x) = x_j$  and  $J_i$  singletons for  $j = 1, \dots, q$  with  $q = K$  we receive our initial B-spline model (15). The estimator for models of the form (16) is given by

$$\hat{\theta}_n(\tau) = \operatorname{argmin}_{\theta \in \mathbb{R}^{q \cdot M}} \left\{ \sum_{i=1}^n \rho_\tau \left( y_i - \sum_{j=1}^q B(\pi_j(x) | d_\tau)^\top \theta_j \right) \right\} \quad (17)$$

where  $\rho_\tau(u) = u(\tau - \mathbb{1}(u < 0))$  is the check function (Koenker & Bassett, 1978) for  $\tau \in \mathcal{T}$ ,  $u \in \mathbb{R}$ . In case of no misspecification, Bondell et al. (2010) have shown that the unconstrained estimator (17) has the same limiting distribution as the classical constrained quantile regression estimator. Hence, in accordance with the discussion on monotonicity in Section 2.1,  $\theta_0$  can be estimated consistently based on unconstrained methods, noting that possible quantile curves crossing of the conditional qf estimator does not affect the validity of the CM test statistic. In our MC simulation II.1 and empirical application 6.2, we will also consider penalized splines, which can be captured by (15) with an additional penalty term, but are omitted at this point for ease of notation (compare the discussion after Corollary 1). Besides the well-known estimation method (17), there are other consistent approaches. A prominent and easy to implement algorithm is the *divide and conquer* algorithm at fixed  $\tau$ . The *quantile projection* algorithm, in contrast, is used to construct an estimator for the quantile process (cf. Volgushev et al., 2019, for further details).

To develop a CM test for null hypotheses of the form (5), we replace the estimator of the conditional qf in (12) with our estimator (17). This yields a new conditional distribution function  $\hat{F}_n^S(y | x, \hat{\theta}_n)$ . Integrating over  $x$  leads to the function  $S_n^S(y, x, \hat{\theta}_n) := \hat{F}_n(y, x) - \hat{F}_n^S(y, x, \hat{\theta}_n)$ , where  $\hat{F}_n^S(y, x, \hat{\theta}_n)$  is the spline based estimate of the cdf in the spirit of (13) for fixed and finite  $q$  and  $M$ . We then define the CM test statistic for finite-dimensional semi-parametric quantile regression models

$$S_n^{CM,S} := \int \left( \sqrt{n} S_n^S(y, x, \hat{\theta}_n) \right)^2 d\hat{F}_n(y, x). \quad (18)$$

The second test  $S_n^{CM,S}$  is thus able to test, for instance, whether a spline is correctly specified with respect to its predefined fixed degree  $d$ . In contrast to  $S_n^{CM}$ , the semi-parametric test  $S_n^{CM,S}$  also allows for possible penalization. Consequently, questions of the form whether linear splines characterize a data set similarly well as cubic splines (with possible penalization) can be addressed by means of  $S_n^{CM,S}$ .

## 2.3 | A more powerful testing procedure using splines

The underlying principle of the test  $S_n^{CM}$  is to compare the parametric cdf induced by (5) with the nonparametric cdf. As the class of alternative hypotheses in (4) gets smaller, the power

of  $S_n^{CM}$  can be improved if a spline is used for the nonparametric part of the test, that is, modeling the conditional qf with an appropriate spline function used to estimate the ecdf  $\hat{F}_n$ . Xue and Wang (2010) have shown, for instance, that the estimate of the cdf with a smooth monotone polynomial spline has better finite sample properties than the empirical distributional estimate. Cardot et al. (2005) have generalized limiting results for quantile regression models with quantile-dependent covariates. 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 the knots which may change for increasing  $n$ . For a quantile regression model with quantile-independent covariates He and Shi (1998) have pointed out that if the number of knots  $k_n \sim (n/\log n)^{2/5}$  and under some mild assumptions (cf. He & Shi, 1998, assumptions C1–C3), the order of approximation of a quadratic monotone B-spline is  $(\log n/n)^{2/5}$ .

However, in order to approximate the theoretical cdf sufficiently well, it is necessary that  $M$  grows as a function of  $n$ , that is,  $M$  diverges at a proper rate. This is known as nonparametric quantile regression, that is, a linear model with increasing dimension in the regression coefficients. Note that in this framework the true functional parameter vector  $\theta_0$  also depends on  $n$  (Belloni et al., 2019). Consequently, for nonparametric quantile regression, (1) expands to

$$F_{Y|X}^{-1}(\tau | x) = P_n(x_n, \tau)^\top \theta_0_n(\tau). \quad (19)$$

We discuss the theoretical framework of (19) in Section 3.2. Let  $\hat{F}_n^{S^M}$  be the spline based estimate of the cdf via the nonparametric conditional qf according to (19). Thus in these models, the test statistic that is based on the difference of the parametric and semi-parametric ecdf reads

$$S_n^*(y, x, \hat{\theta}_n) = \frac{1}{a_n^*} \hat{F}_n^{S^M}(y, x, \hat{\theta}_n) - \frac{1}{a_n} \hat{F}_n(y, x, \hat{\theta}_n),$$

where  $a_n, a_n^*$  are scaling factors defined in Section 3.2. This yields the new test statistic

$$S_n^{CM^*} := \int \left( \sqrt{n} S_n^*(y, x, \hat{\theta}_n) \right)^2 d\hat{F}_n(y, x). \quad (20)$$

In comparison to  $S_n^{CM}$ , the test statistic  $S_n^{CM^*}$  replaces the estimate of the ecdf  $\hat{F}_n(y, x)$  in (13) by an appropriate spline estimate of the conditional qf via (19), which is then transformed to estimate  $\hat{F}_n^{S^M}(y, x, \hat{\theta}_n)$ . Note that finite-dimensional parametric models can also be tested with  $S_n^{CM^*}$ . In our MC simulation study, we will therefore compare the two tests  $S_n^{CM}$  and  $S_n^{CM^*}$ , as they address questions of similar kind. It turns out that  $S_n^{CM^*}$  is a more powerful testing procedure than  $S_n^{CM}$ , particularly in small samples.

### 3 | ASYMPTOTICS

In this section, we first derive theoretical properties of the parametric test statistic  $S_n^{CM}$  in Section 3.1 before generalizing the statements to the semi-parametric test statistic  $S_n^{CM,S}$  and the more powerful test statistic  $S_n^{CM^*}$  in Section 3.2.

### 3.1 | Theoretical properties for testing (semi-)parametric quantile regression models

In Theorem 1 below we show that the test statistic  $S_n^{CM}$  has correct asymptotic size. To be able to derive large sample properties of  $S_n^{CM}$ , we make and discuss the following mild assumptions. Since our proposed test statistic is a generalization of existing tests, these assumptions modify those previously made (Chernozhukov et al., 2013; Rothe & Wied, 2013). For this purpose, we restate the assumptions on compact subsets on  $\mathcal{T}$ . Let  $\Theta$  be an arbitrary subset of  $\mathbb{R}^p$  and  $\mathcal{T} := [\varepsilon, 1 - \varepsilon]$  with  $\varepsilon \in (0, 0.5)$ .

#### Assumption 1.

- (i)  $P(X, \tau)$  is  $L_2$ -bounded in  $[0, 1]$  and continuous in  $X$ .
- (ii) Let  $\bigcup_{l=1}^L I_l = \mathcal{T}$ ,  $L \in \mathbb{N}$ ,  $I_l$  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  $l = 1, \dots, L$ ,  $G(\cdot, \tau) : \Theta \rightarrow \mathbb{R}^p$  possesses a unique zero at  $\theta_0 \in \text{interior}(\Theta)$  such that  $G(\theta_0, \tau) = 0$  for all  $\tau \in \mathcal{T}$  and for some  $\delta > 0$ ,  $\mathcal{B} := \bigcup_{\tau \in I_l} B_\delta(\theta_0)$  is a compact subset of  $\mathbb{R}^p$  contained in  $\Theta$  for  $l = 1, \dots, L$ .
- (iv) Furthermore,  $G(\cdot, \tau)$  has an inverse  $G^{-1}(x, \tau) := \{\theta \in \Theta \mid G(\theta, \tau) = x\}$  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) The mapping  $(\theta, \tau) \mapsto g(\cdot, \theta, \tau)$  is continuous at each  $(\theta(\tau), \tau) \in \Theta \times I_l$  for all  $l = 1, \dots, L$  with probability one and  $(\theta, \tau) \mapsto G(\theta, \tau)$  is continuously differentiable at  $(\theta_0(\tau), \tau)$  with uniformly bounded derivative on  $\mathcal{T}$ .
- (vi) The function  $\dot{G}(\theta, \tau) := \partial_\theta G(\theta, \tau)$  is nonsingular at  $\theta_0(\cdot)$  uniformly over  $\tau \in I_l$  with  $l = 1, \dots, L$ .
- (vii) The function set  $\mathcal{G}_l = \{g(Y, X, \theta, \tau) \mid (\theta, \tau) \in \Theta \times I_l\}$  is  $F_{YX}$ -Donsker for all  $l = 1, \dots, L$  with a square integrable envelope  $\tilde{G}$  for  $\bigcup_{l=1}^L \mathcal{G}_l$ .
- (viii) The mapping  $\theta \mapsto F(\cdot \mid \cdot, \theta)$  is Hadamard differentiable for all  $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$  with derivative  $h \mapsto \dot{F}(\cdot \mid \cdot, \theta)[h]$

Due to quantile dependence of the regressors  $X$ , we further require continuity of the function  $P(X, \tau)$  in  $X$ , which is provided by Assumption 1(i). Assumption 1(ii) ensures that there is a finite and compact decomposition of the unit interval. This is required since we consider Donsker classes in the proof of Theorem 1. We are using the fact that the union of Donsker classes is also Donsker (see Dudley, 2014, sec. 3.8). Assumptions 1(ii)–(vii) guarantee the regularity of our estimator  $\hat{\theta}_n$  and ensure that a functional central limit theorem can be applied to  $Z$ -estimator processes (see Corollary 2 in Supplement I.1). Assumption 1(viii) is a smoothness condition. Together with the functional delta method it implies that the restricted cdf estimator process

$$(y, x) \mapsto \sqrt{n}(\hat{F}_n(y, x, \hat{\theta}_n) - F(y, x, \theta)) \quad (21)$$

is  $F_{YX}$ -Donsker. This convergence can be shown to be jointly with that of the ecdf process  $(y, x) \mapsto \sqrt{n}(\hat{F}_n(y, x) - F(y, x))$  to a Brownian bridge by some standard arguments given in Lemma 2 in Supplement I.1. Applying the continuous mapping theorem yields the following proposition.

**Theorem 1.** *If Assumption 1 is satisfied, then the following statements hold:*

(i) *Under the null hypothesis  $H_0$  in (8),*

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

where  $(\mathbb{G}_1, \mathbb{G}_2)$  is a bivariate zero mean Gaussian processes with

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

where  $\mathbb{G}_2^+(y, x)$  is the limiting Gaussian process of  $\sqrt{n}(\hat{F}_n(y | x, \hat{\theta}_n) - F(y | x, \theta)) \in \ell^\infty(S)$  defined in Lemma 2. Moreover,

$$\begin{aligned} & \text{Cov}(\mathbb{G}_1(y_1, x_1), \mathbb{G}_2(y_2, x_2)) \\ &= \lim_{n \rightarrow \infty} n \text{Cov}(\hat{F}_n(y_1, x_1) - F(y_1, x_1), \hat{F}_n(y_2, x_2, \hat{\theta}_n) - F(y_2, x_2, \theta)). \end{aligned}$$

(ii) *Under any fixed alternative, that is, when the data are distributed according to some  $F$  that satisfies the alternative hypothesis  $H_1$  in (8),*

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

Theorem 1 ensures distributional convergence of the test statistic  $S_n^{CM}$  and further shows that the nonparametric ecdf  $\hat{F}_n(y, x)$  and the parametric ecdf  $\hat{F}_n(y, x, \hat{\theta}_n)$  differ with probability one under the alternative hypothesis. Hence in case of misspecification, the power of the test statistic  $S_n^{CM}$  converges to one as  $n$  approaches infinity. Based on the generality of Assumption 1 and the proof structure in Supplement I.1, the statements from Theorem 1 can also be extended to semi-parametric quantile regressions models with fixed  $q, M$  as discussed in Section 2.2. Thus, we have

**Corollary 1.** *If Assumption 1 is satisfied, then the following statements hold:*

(i) *Under the null hypothesis  $H_0$  in (8),*

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

where  $(\mathbb{G}_1, \mathbb{G}_2^S)$  is a bivariate zero mean Gaussian processes with

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

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

$$\begin{aligned} & \text{Cov}(\mathbb{G}_1(y_1, x_1), \mathbb{G}_2^S(y_2, x_2)) = \\ & \lim_{n \rightarrow \infty} n \text{Cov}(\hat{F}_n(y_1, x_1) - F(y_1, x_1), \hat{F}_n^S(y_2, x_2, \hat{\theta}_n) - F(y_2, x_2, \theta)). \end{aligned}$$

- (ii) Under any fixed alternative, that is, when the data are distributed according to some cdf  $F$  that satisfies the alternative hypothesis  $H_1$  in (8),

$$\lim_{n \rightarrow \infty} P\left(S_n^{CM,S} > \varepsilon\right) = 1 \text{ for all constants } \varepsilon > 0.$$

In empirical applications, however, it is often common to estimate series terms with smoothing penalty parameters  $\lambda_j$  for  $j = 1, \dots, p$ , since this avoids overfitting the data. Imposing the assumption that the penalty parameters are  $\lambda_j = o(n^{1/2})$  for  $j = 1, \dots, p$  the penalties can be asymptotically ignored. This indicates that Corollary 1 is also valid in case of penalized quantile regression (Lian et al., 2015).

### 3.2 | Theoretical properties for the more powerful test

In the context of nonparametric quantile regression, that is, the number of knots diverges at a proper rate, we need to introduce some additional notation: Since the dimension  $K$  and the true distribution  $F$  of i.i.d. samples  $(X_i, Y_i) \in \mathbb{R}^{K+1}$  for  $i = 1, \dots, n$  can depend on  $n$ , we consider triangular arrays. For brevity of notation, we omit the index  $n$  in the following and we write  $P_i = P(X_i, \tau)$  and  $P = P(X, \tau)$ . Let  $\lambda_{\min}(A)$  and  $\lambda_{\max}(A)$  be the smallest and largest eigenvalue of a matrix  $A$ . By  $\|b\|$  we denote the  $L^2$ -norm of a vector  $b$ . Moreover, we set  $a_n := \sqrt{n}/\|P(x)\|$  and  $\alpha_n^* := \sqrt{n}/\|B(x)\|$ , respectively.

Imposing the assumptions from Chao et al. (2017) adapted to quantile regression with quantile-dependent series terms enables us to replace the conditional qf by an appropriate (spline) estimator and thus to derive large sample properties for our third test statistic  $S_n^{CM^*}$ :

#### Assumption 2.

- (i) For  $p := qM$ , assume that  $\|P_i\| \leq \xi_p = O(n^a)$  almost surely with  $a > 0$ , and that  $\frac{1}{M^*} \leq \lambda_{\min}(\mathbb{E}[PP^T]) \leq \lambda_{\max}(\mathbb{E}[PP^T]) \leq M^*$  holds uniformly in  $n$  and  $\tau \in \mathcal{T}$  for some fixed  $M^* > 0$ .
- (ii) The conditional distribution  $F_{Y|X}(y|x)$  is twice differentiable w.r.t.  $y$ . We denote the corresponding derivatives by  $f_{Y|X}(y|x)$  and  $f_{Y|X}^\top(y|x)$ . Assume that  $\bar{f} := \left| \sup_{y,x} f_{Y|X}(y|x) \right| < \infty$  and  $\bar{f}^\top := \sup_{y,x} \left| f_{Y|X}^\top(y|x) \right| < \infty$  uniformly in  $n$ .
- (iii) Assume there exists a constant  $f_{\min} > 0$  such that  $\inf_{\tau \in \mathcal{T}} \inf_x f_{Y|X}(F_{Y|X}^{-1}(\tau|x)|x) \geq f_{\min}$ .
- (iv) For each  $x$ , the basis vector  $P$  has zeroes in all but at most  $r$  consecutive entries, where  $r$  is fixed. Moreover,  $\sup_{\tau,x} \mathbb{E}[|P^\top \tilde{J}_m(\tau)^{-1} P|] = O(1)$ , where  $\tilde{J}_m(\tau) := \mathbb{E}[PP^\top f_{Y|X}(F_{Y|X}^{-1}(\tau|x)|X)]$ .
- (v) Assume that  $\xi_p^4 (\log n)^6 = o(n)$  and letting  $c_n := \sup_{\tau,x} \left| F_{Y|X}^{-1}(\tau|X) - P^\top \hat{\theta}_n(\tau) \right|$  with  $c_n^2 = o(n^{-1/2})$ .

As mentioned in Chao et al. (2017), Assumption 2(i) claims rescaling in case of B-splines and for linear models with increasing dimension  $P(X, \tau)$  to be bounded for all  $\tau$ . Assumption 2(ii) and (iii) is fairly standard. Assumption 2(iv) and (v) implies that for any sequence satisfying  $c_n = o(1)$  and that the smallest eigenvalues of the matrix  $J_m(\tau)$  are bounded away from zero uniformly in  $\tau$  for all  $n$ . Using Theorem 2.4 of Chao et al. (2017) showing that a standardized version of the quantile series terms process converges to a centered Gaussian process we have

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

(i) *Under the null hypothesis  $H_0$  in (8),*

$$S_n^{CM^*} \xrightarrow{d} \int \left( \mathbb{G}_2(y, x) - \mathbb{G}_2^{S_M}(y, x) \right)^2 dF_{YX}(y, x),$$

*where  $(\mathbb{G}_2, \mathbb{G}_2^{S_M})$  are Gaussian processes with zero mean given in Supplement I.2.*

(ii) *Under any fixed alternative, that is, when the data are distributed according to some cdf  $F$  that satisfies the alternative hypothesis  $H_1$  in (8),*

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

Theorem 2 ensures that the test  $S_n^{CM^*}$  is asymptotically normal and has power in case of misspecification. The convergence statements from Corollary 1 and Theorem 2 hold for additive univariate series terms, including, for instance, univariate B-splines with product interacting covariates. In line with Chao et al. (2017), we further conjecture that such arguments as those given in the proofs (cf. Supplement 1.1) can also be applied to multivariate splines and thus in particular to tensor product B-splines considered later in Section 6.2. Therefore, convergence statements from Corollary 1 and Theorem 2 can be extended to a more general class of (multivariate) splines. Inspired by this observation and our second application, we show empirically that the test statistic  $S_n^{CM^*}$  based on tensor product B-splines also yields a reasonable sized testing procedure with large power (cf. Tables I and II, Supplement III). However, a detailed theoretical investigation of this interesting topic is beyond the scope of this paper and left for future research.

## 4 | BOOTSTRAP

To obtain critical values for our test  $S_n^{CM}$ , we therefore propose a semi-parametric bootstrap procedure. This procedure is reasonable from a practical point of view, since it avoids to estimate the null distribution directly, including a complex covariance structure.

### 4.1 | Semi-parametric bootstrap procedure

The idea of our semi-parametric bootstrap is to generate synthetic data that is consistent with the assumptions under the null hypothesis. Since the conditional qf is already known according to our null hypothesis, our bootstrap procedure is based on the principle of inverse sampling transformation, which provides a method to generate samples from arbitrary distributions. 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. Let  $B$  be the number of bootstrap samples. Then

- (i) Draw  $B$  independent bootstrap samples of covariates  $\{X_{b,i} \mid 1 \leq i \leq n\}_{b=1, \dots, B}$  of size  $n$  with replacement from  $\{X_i \mid 1 \leq i \leq n\}$ ,
- (ii) For every  $i = 1, \dots, n$  put  $Y_{b,i} = \hat{F}_n^{-1}(U_{b,i} \mid X_{b,i}, \hat{\theta}_n)$ , where  $\{U_{b,i} \mid 1 \leq i \leq n\}$  is a simulated i.i.d. sequence of standard uniformly distributed random variables.

(iii) Use the bootstrap data  $\{(Y_{b,i}, X_{b,i}) \mid 1 \leq i \leq n\}_{b=1, \dots, B}$  to calculate  $B$  bootstrap versions of the test statistic  $S_n^{CM}$  from (14), that is, for  $b = 1, \dots, B$  compute

$$S_{n,b}^{CM} := \int \left( \sqrt{n} S_{n,b}(y_b, x_b, \hat{\theta}_n) \right)^2 d\hat{F}_n(y_b, x_b).$$

(iv) For  $q \in (0, 1)$ , determine the critical value  $\hat{c}_n(q)$  such that

$$\frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{S_{n,b}^{CM} > \hat{c}_n(q)\}} = q.$$

With the bootstrap procedure described above, we can calculate critical values  $\hat{c}_n(q)$  for (14) and the null hypothesis is rejected on the level of significance  $q$  if  $S_n^{CM} > \hat{c}_n(q)$ . Critical values for (18) and (20) can be obtained in the same manner if the test statistic  $S_{n,B}^{CM}$  is replaced by its counterparts, that is,  $S_{n,B}^{CM,S}$  or  $S_{n,B}^{CM,*}$ .

## 4.2 | Validity of the bootstrap procedure

Finally, according to Rothe and Wied (2013), we show that the proposed bootstrap procedure computes the correct critical value for our test statistic (14). This does not require any further assumptions. Assumption 1 ensures that the bootstrap consistently estimates the limiting distribution for (14). Under the null hypothesis and any fixed alternative (5), the bootstrap critical values can be shown to be bounded in probability. Thus,

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

(i) *Under the null hypothesis  $H_0$  in (8), we have that*

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

(ii) *Under any fixed alternative  $H_1$  in (8), we have that*

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

In order to study the behavior of the Cramér–von Mises type test statistics  $S_n^{CM}$ ,  $S_n^{CM,S}$  and  $S_n^{CM,*}$  in finite samples, we conducted an extensive MC study, whose results are reported in Supplement II. Overall, the MC study has shown that our proposed testing procedures are also consistent based on critical values obtained via the bootstrap procedure described in Section 4.1 and have superior power properties compared with three benchmark tests (cf. Supplement 5.1), even in small samples. The testing procedures works for both, univariate and multivariate DGPs (including product interacting or more complex tensor product covariates) and can also test models with quantile-dependent regressors. Even weakly misspecified models are detected in sufficiently large sample sizes.

## 5 | MONTE CARLO SIMULATION STUDY

This contains a comprehensive MC simulation study for the test statistics  $S_n^{CM}$  and  $S_n^{CM*}$ , where the spline part in the latter test statistic is modeled by a penalized B-spline. Wherever it is possible, we also compare our results with existing benchmark tests, for instance, those given in Koenker and Xiao (2002) (*KX*), Chernozhukov (2002) (*CH*) and Rothe and Wied (2013) (*RW*). Note, in case of quantile-independent covariates, *RW* is a special case of our proposed test  $S_n^{CM}$ . In Section II.1 in the Appendix, we examine power and size properties for the semi-parametric model specification test  $S_n^{CM,S}$  where possible interacting covariates are modeled by a tensor product. In Section II.2 in the Appendix, we examine power and size properties for the semi-parametric model specification test  $S_n^{CM,S}$  with univariate product interacting covariates.

### 5.1 | MC simulation study for $S_n^{CM*}$

In this subsection, we show that our test  $S_n^{CM*}$  holds the size level and has superior power properties compared with  $S_n^{CM}$  by means of the twelve different data generating processes (DGPs) based on i.i.d. data  $\{(y_i, x_i) \mid 1 \leq i \leq n\}$  for  $n \in \{30, 50, 100, 300, 500, 1000, 2000\}$ . The different DGPs cover location shift models (LS) and location-scale shift models (LSS) including heteroscedastic errors, both, in a univariate and multivariate setting. In order to assess the quality and validity of our proposed test against existing procedures, we benchmark against the tests of Koenker and Xiao (2002), Chernozhukov (2002) and Rothe and Wied (2013) where comparisons are possible (DGPs 1–9). Finally, we also consider linear models and show that our test detects even weakly misspecified models well.

For the definition of the twelve DGPs we introduce the following variables: Let  $x_0 \in U(0, 2\pi)$ ,  $x_1 \sim \text{Bin}(1, 0.5)$ ,  $x_2 \sim N(0, 1)$ ,  $x_3 \in U(0, 1)$ ,  $x_4 \in \chi^2(1)$ ,  $u \sim N(0, 1)$ ,  $w \sim N(0, 0.1)$ ,  $v = (1 - 2x_1) \cdot v_2^* \cdot 8^{-0.5}$  with  $v_2^* \sim \chi^2(2)$ , where  $\text{Bin}(\cdot, \cdot)$ ,  $N(\cdot, \cdot)$ ,  $U(\cdot, \cdot)$  and  $\chi^2(\cdot)$  are Binomial, Gaussian, uniform and chi-square distributions, respectively.

#### Data generating processes

DGPs 1–3 represent the univariate case with one covariate and additive noise. Hereby, DGP 1 describes a simple LS model, DGP 2 a more complex LSS model with a linear regressor and, finally, DGP 3 generates a quadratic LSS model. The multivariate case 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 normally distributed errors. DGP 2 is again a simple LS model, but now the errors follow a mixture of a “positive” and “negative”  $\chi^2$  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 are processes in which the functional form appears predominantly nonlinear. DGP 10 is implemented by modeling different functional forms for quantiles below and above the 0.5 threshold. Due to the quantile dependence of the regressors, DGP 10 cannot be correctly tested with previous tests but with our test  $S_n^{CM*}$  it can. DGPs 11–12 are appearing mainly linear in the interval  $[0, 1]$  and exhibit nonlinear growth only at values close to 1. Both DGPs require flexible modeling beyond the

parametric framework of simple polynomials. Assuming a linear model, DGPs of the form 10–12 often impede the detection of misspecification.

## Estimation and further settings

Computations have been carried out using the R package `cobs` (Ng & Maechler, 2007, 2020). In what follows,  $\hat{F}_n^S$  is modeled by a B-spline of second order with penalty term  $\lambda = 1$  and  $\sqrt{n}$  knots evaluated for  $\tau \in \{0.1, 0.2, \dots, 0.9\}$ , meeting monotonicity assumptions. The number of MC repetitions is equal to 701 with 500 bootstrap replications. The significance level is 0.05.

$$\begin{aligned}
 \text{DGP 1: } y_1(x_0) &:= \frac{x_0}{4} + 1 + u, & \text{DGP 2: } y_2(x_0) &:= \frac{x_0}{4} + 1 + u \cdot x_0 \\
 \text{DGP 3: } y_3(x_0) &:= \frac{x_0^2}{4} + 1 + u \cdot x_0^2, & \text{DGP 4: } y_4(x_1, x_2) &:= x_1 + x_2 + u \\
 \text{DGP 5: } y_5(x_1, x_2) &:= x_1 + x_2 + v, & \text{DGP 6: } y_6(x_1, x_2) &:= x_1 + x_2 + \left(\frac{1}{2} + x_1\right)u \\
 \text{DGP 7: } y_7(x_1, x_2) &:= x_1 + x_2 + \left(\frac{1}{2} + x_1 + x_2^2\right)^{0.5} u & (22) \\
 \text{DGP 8: } y_8(x_1, x_2) &:= x_1 + x_2 + \frac{1}{5} \left(\frac{1}{2} + x_1 + x_2^2\right)^{1.5} u \\
 \text{DGP 9: } y_9(x_3) &:= x_3 + (1 + \gamma_1 \cdot x_2)u \\
 \text{DGP 10: } y_{10}(x_3) &:= \begin{cases} \frac{x_3^2}{4} + 1 + \frac{u \cdot x_3^2}{2}, & \text{if } \tau \geq 0.5 \\ \frac{-x_3^2}{4} + 1 + u \cdot x_3, & \text{otherwise} \end{cases} \\
 \text{DGP 11: } y_{11}(x_3) &:= \sin\left(-\frac{\pi}{2} + x_3^3\right) + w, & \text{DGP 12: } y_{12}(x_3) &:= e^{y_5(x_3)}
 \end{aligned}$$

## Benchmark tests

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) (*KX*), which is based on the Khmaladze transformation, 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) (*CH*) and compare our test with *RW*. The aforementioned tests are characterized as follows:

- The *KX*-test models the conditional qf parametrically by assuming a LS or a LSS model. The regressors are fixed for all quantiles considered and the estimation of nonparametric sparsity and score functions are required (Chernozhukov, 2002).
- In order to avoid the latter, *CH* employs a resampling testing procedure based on *KX* that results in better power and accurate size. However, this tests still assumes a fully parametrized model under the null hypothesis with quantile-independent regressors.
- *RW* propose a testing procedure for a wide range of parametric models that is based on a Cramér–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. Thus, the  $RW$  test approach equals  $S_n^{CM}$  in case that the vector of transformations  $P(X, \tau)$  is constant for all  $\tau$ .

## Results

Table 1 shows the comparison with  $RW$  for all  $n$  in the univariate DGPs 1–3 in terms of size and power of the statistics at 10%, 5% levels and a 5% level, respectively. We make three observations. First, compared with  $RW/S_n^{CM}$  our proposed testing procedure  $S_n^{CM*}$  consistently has better size properties. For example, in DGP 3,  $RW/S_n^{CM}$  is way too conservative, which is not true for  $S_n^{CM*}$ . Second, the test  $S_n^{CM*}$  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). Last, 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 2 illustrates the comparison with  $KX$  for the DGPs 4–8 for  $n = 100, 300$ , whereby a LS model is assumed under the null hypothesis. Thus, the results of DGPs 4 and 5 reflect size properties, while DGPs 6–8 illustrate the power of  $S_n^{CM*}$  compared with the benchmark tests  $RW/S_n^{CM}$  and  $KX$  at significance levels 10% and 5% each. We again make three observations. First, our test  $S_n^{CM*}$  holds the size for multivariate models (cf. DGPs 4 and 5 in Table 2). Second,  $KX$  has difficulties to detect misspecification when heteroscedasticity is present (cf. DGP 6–8 in Table 2).

TABLE 1 Size and power for  $RW/S_n^{CM}$  and  $S_n^{CM*}$ .

$RW/S_n^{CM}$	DGP 1		DGP 2		DGP 3	
	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

Note: MC Study. The table compares the test statistics  $S_n^{CM*}$  and  $S_n^{CM}$  in terms of size (significance levels 10% and 5%) and power (at a 5% level), where the latter coincides with test statistic of Rothe and Wied (2013) ( $RW$ ) in case of quantile-independent covariates. The last column named *Power* shows the power analysis while the  $qf$  is assumed to follow a linear LSS model under the null hypothesis.

TABLE 2 Size and power for DGPs 4–8.

	$RW/S_n^{CM}$		$KX$		$S_n^{CM*}$	
	10%	5%	10%	5%	10%	5%
<b><math>n = 100</math></b>						
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
<b><math>n = 300</math></b>	<b>10%</b>	<b>5%</b>	<b>10%</b>	<b>5%</b>	<b>10%</b>	<b>5%</b>
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

Note: MC Study. The table compares size and power (at significance level 10% and 5%) of the test statistics  $RW/S_n^{CM}$ ,  $KX$  and  $S_n^{CM*}$ . All results are one-to-one transferred from Rothe and Wied (2013). The results of DGPs 4 and 5 reflect size properties, while DGPs 6–8 illustrate the power of  $S_n^{CM*}$  compared with the benchmark tests  $RW/S_n^{CM}$  and  $KX$ .

TABLE 3 Size and power for DGP 9.

	$KX$			$CH$			$S_n^{CM*}$		
	Size	Power		Size	Power		Size	Power	
$\gamma_1 =$	0	0.2	0.5	0	0.2	0.5	0	0.2	0.5
<b><math>n = 100</math></b>	0.101	0.264	0.898	0.014	0.348	0.980	0.050	0.396	0.990
<b><math>n = 200</math></b>	0.070	0.480	0.988	0.052	0.752	1.000	0.063	0.772	1.000
<b><math>n = 300</math></b>	0.062	0.622	0.998	0.058	0.910	1.000	0.068	0.930	1.000

Note: MC Study. The table compares size and power (at significance level 5%) of the test statistics  $KX$ ,  $CH$  and  $S_n^{CM*}$ .  $KX$  refers to the specification test suggested by Koenker and Xiao (2002). The more powerful test of Chernozhukov (2002) is abbreviated by  $CH$ . All results are one-to-one transferred from Chernozhukov (2002). The null hypothesis assumes a LS quantile regression model, i.e.  $\gamma_1 = 0$ .

Third,  $RW/S_n^{CM}$  usually detects misspecification. However, the rejection rates of the test  $S_n^{CM*}$  are clearly higher compared with 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) using  $n = 100, 200, 300$  and DGP 9. Results of Table 3 of the benchmark tests  $KX$  and  $CH$  are taken from Chernozhukov (2002). This table indicates that the test  $S_n^{CM*}$  has consistently better size and power properties compared with the benchmarks  $KX$  and  $CH$  for small sample sizes. Finally, Table 4 examines size and power properties for the DGPs 10–12. Here, in each of the DGPs 10–12, the test  $S_n^{CM*}$  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 and are thus in line with the null hypothesis.

TABLE 4 Size and power for DGPs 10–12.

$S_n^{CM^*}$	DGP 10		DGP 11		DGP 12	
	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

Note: MC Study. The table reports size and power of the test statistic  $S_n^{CM^*}$  at a significance level 5%.

## 6 | EMPIRICAL ILLUSTRATIONS

### 6.1 | Income disparities between East and West Germany

In this section, we apply the bootstrap version of the specification test  $S_n^{CM^*}$  to conditional income distributions in Germany. 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) from 2001 to 2010 as our response  $Y$ . We deflate the incomes by the consumer price index (Statistisches Bundesamt, 2012), setting 2010 as our base year. Thus, all incomes are expressed in real-valued 2010 Euros from here on. Following the standard literature, we focus on incomes 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* (dummy for East/West Germany) and *year* are available as covariates, see Table 5 for a full description of the data.

To obtain an estimate of the qf, we first regress income on the dummy coded variable *year* and then performed a linear quantile regression using the variables *age* or  $age^2$  on the residuals. 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. This approach takes into account that income increases at the beginning of employment, peaks in middle age and finally decreases (Creedy & Hart, 1979; Klein et al., 2015; Luong & Hébert, 2009). We next conduct a M–M decomposition (Landmesser et al., 2016; Machado & Mata, 2005), of the *year*-adjusted dataset conditional on *origin*. For the decomposition we assume that the conditional qf of the income  $Y$  can be represented as a function of the form  $F_{Y|X}^{-1}(\tau | X) = P(X, \tau)^\top \theta(\tau)$  with  $X$  consisting of the variables *age* or  $age^2$ . 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  $\tau \in (0, 1)$ , that is,  $P(X, \tau) = age$  for all  $\tau \in (0, 1)$ . The second models a quadratic influence of age on income for all quantiles  $\tau \in (0, 1)$ , that is,  $P(X, \tau) = age^2$  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  $P(X, \tau_i)^\top \hat{\theta}_n(\tau_i)$  for  $\tau_i \stackrel{i.i.d.}{\sim} U(0, 1)$ ,  $i = 1, \dots, n$  constitutes a random sample from the estimated conditional distribution of  $Y$  given the covariates  $X$  (Machado & Mata, 2005). In order to obtain the difference between East and West, first,

TABLE 5 Description of the German labor income data from 2001 to 2010.

	Description		
<b>Y</b>	Gross market labor income (in €) (continuous $1257 \leq Y \leq 280,092$ , average = 46,641)		
<b>origin</b>	indicator for East or West (binary, $-1 = \text{West}$ (73.8%), $1 = \text{East}$ [26.2%])		
<b>age</b>	age of the male in year (continuous, $20 \leq \text{age} \leq 60$ , average = 38)		
<b>years</b>	time in years (categorical, $2001 \leq \text{year} \leq 2010$ , 10 years)		
Sample	Description	Average (std.) income	Observations
<b>Ger</b>	Complete sample	51,026€ (30,569€)	$n = 7220$
<b>West</b>	Subsample ( $\text{origin} = -1$ )	55,141€ (31,494€)	$n = 5325$
<b>East</b>	Subsample ( $\text{origin} = 1$ )	39,463€ (24,336€)	$n = 1895$

Note: Incomes. The table summarizes the descriptive statistics of the German labor income data.

the coefficients for East ( $\hat{\theta}_E(\tau)$ ) and West ( $\hat{\theta}_W(\tau)$ ) for  $\tau \in \{0.1, 0.2, \dots, 0.9\}$  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  $B \in \mathbb{N}$  random samples  $X_E^i$  and  $X_W^i$  for  $i = 1, \dots, B$  with replacement from the corresponding covariate subsets  $X_E$  and  $X_W$ , respectively to obtain a random sample via the probability integral transform for the distribution of the income  $Y_l^i, i = 1, \dots, B, l = E, W$ . Thus, the estimated income difference  $\hat{\Delta}_Y$  for incomes in the East  $Y_E$ /West  $Y_W$  can be decomposed according to M–M into

$$\begin{aligned} \hat{\Delta}_Y &= \frac{1}{B} \sum_{b=1}^B \left( (P(X_E^b, \tau) - P(X_W^b, \tau)) \hat{\theta}_E(\tau) + (\hat{\theta}_E(\tau) - \hat{\theta}_W(\tau)) P(X_W^b, \tau) \right) \\ &\approx F_{Y_E|X_E}^{-1}(\tau | X_E) - F_{Y_W|X_W}^{-1}(\tau | X_W), \end{aligned} \quad (23)$$

where the first summand of (23) is the explained, while the second summand depicts the unexplained difference.

Table 6 summarizes results from the counterfactual analysis described above. The covariates used for the quantile regressions are *age* (rows 4–9), *age*<sup>2</sup> (rows 11–16) and the sum of these two variables (rows 18–23). The results 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. It cannot be assumed that the model is sufficiently well specified by a single covariate *age* or *age*<sup>2</sup> 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*<sup>2</sup> seems to be appropriate for the smallest quantile 0.1 (residual of 0.92 in Table 6), while a linear effect of age to income seems to prevail in higher quantiles (−1.49 in Table 6). In contrast, the additive model *age* + *age*<sup>2</sup> seems to capture the income effect for all quantiles quite well due to moderate residuals (cf. last row *Residuals* in Table 6). For all decompositions it holds, that *age* and *age*<sup>2</sup> contribute a maximum of 16% to the explanation of the income difference between East and West Germany (except highest quantile in *age*<sup>2</sup>, i.e., 25.41). Due to the different residuals and the different explanatory power of the income gap between East and West for the quantile regressions based on *age* or *age*<sup>2</sup>, it seems reasonable to assume that *age* and *age*<sup>2</sup> have different effects for different quantiles. For example, the residual of the 30% quantile

TABLE 6 Decomposition of the West/East income differential.

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

Note: Incomes. The covariates used for the quantile regressions are *age* (rows 4–9), *age<sup>2</sup>* (rows 11–16) and the sum of these two variables (rows 18–23). The second row *raw gap* depicts the observed income gap between East and West. Remaining rows show three different M–M decompositions using *age*, *age<sup>2</sup>* and *age + age<sup>2</sup>* as covariates for the quantile regression models. The rows *M–M gap* are the estimated gap of the income difference. The quantiles  $\tau$  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 to 100% due to rounding.

of *age* (0.24 in Table 6) is about 18 times smaller than the residual of the corresponding quantile regression using *age<sup>2</sup>* as explanatory variable (4.24 in Table 6). It is therefore reasonable that the a linear effect of age dominates in the  $\tau = 0.3$  quantile. The emerging, more general question is at which quantiles *age* has a linear or quadratic effect on incomes. This can be answered with the help of our proposed test  $S_n^{CM^*}$ .

$$\begin{aligned}
 S1 : F_{Y|X}^{-1}(\tau | x) &= \begin{cases} x^\top \theta_0, & \text{if } 0.1 \leq \tau \leq 0.9 \\ (x^2)^\top \theta_0, & \text{otherwise} \end{cases} & S3 : F_{Y|X}^{-1}(\tau | x) &= \begin{cases} x^\top \theta_0, & \text{if } 0 \leq \tau \leq 0.9 \\ (x^2)^\top \theta_0, & \text{otherwise} \end{cases} \\
 S2 : F_{Y|X}^{-1}(\tau | x) &= \begin{cases} (x^2)^\top \theta_0, & \text{if } 0 \leq \tau \leq 0.1 \\ x^\top \theta_0, & \text{otherwise} \end{cases} & S4 : F_{Y|X}^{-1}(\tau | x) &= x^\top \theta_0 \\
 & & S5 : F_{Y|X}^{-1}(\tau | x) &= (x^2)^\top \theta_0
 \end{aligned}$$

For this purpose, we have defined five different model specifications S1–S5, which should take into account the observations made in Table 6. Specifications S1–S3 describe quadratic dependencies in the upper or lower quantiles. Specification S4 and S5 model a completely linear and quadratic dependence structure in the covariate, respectively.

The testing procedure is applied to the subsamples East and West as well as to the complete data set. We estimate the function  $\hat{F}_n^S(y, x, \theta)$  in (20) by a cubic spline with second order difference penalty, setting the basis dimension to 20 using the R package `qgam`. The smoothing parameter  $\lambda$  is estimated using the restricted maximum likelihood (REML) procedure within the package. We then reestimate the models with optimized smoothing parameter and compute our test statistic  $S_n^{CM^*}$  for  $\tau \in \{0.1, 0.2, \dots, 0.9\}$ . 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  $n = 500, 1000, 1500$ . The reason for considering different samples is, similarly to Rothe and Wied (2013), that consistent specification tests detect also small deviations from the null hypothesis in large samples, so that smaller samples are more appropriate for model comparisons. We repeated this procedure for every subsample a total of 501 times and refer to it as *subsamplings* in the following. Table 7 summarizes the resulting rejection rates of the test statistic  $S_n^{CM^*}$ .

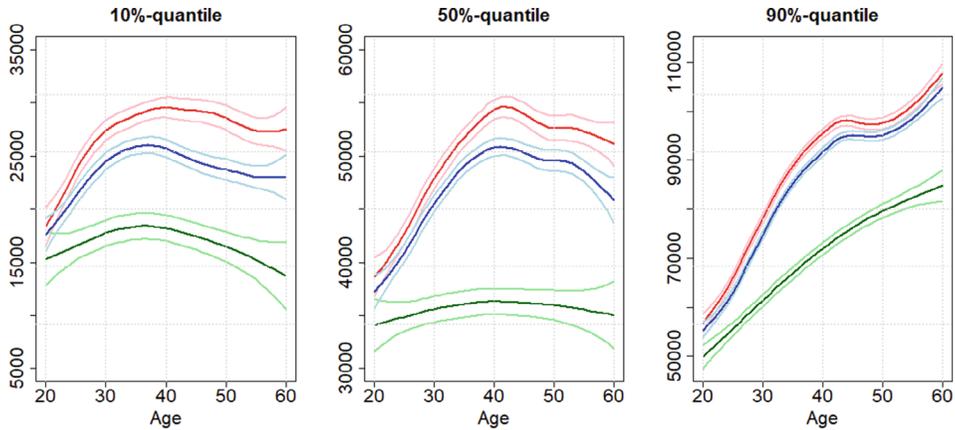
From this table we make two observations: First, it can be observed that the model in which age has a completely quadratic influence on income (S5) provides the worst fit. Also the models with either a completely linear influence or a linear influence in the upper quantiles (S2 and S4) are worse than the models in which the influence is quadratic in the upper quantiles and linear in the lower ones (S1 and S3). Second, the model fits are in general much better for East Germany than for West Germany and for the whole country, whose rejection rates can be interpreted as the weighted average of the two rejection rates. For example, for  $n = 1500$ , the rejection rate of S1 and S3 are even lower than 5% for East Germany, whereas they are larger than 50% for West Germany. This indicates that the conditional income distributions differ significantly between East and West Germany.

Finally, Figure 1 visualizes the estimated quantiles at  $\tau = 0.1, 0.5, 0.9$  (from left to the right) and provides further indications of when age might have a quadratic or linear effect. Shown are the results for West (red), East (green) and entire Germany (blue). Overall, our results are in line with the findings of other studies. Based on the different structure of the conditional qfs and rejection rates for different specifications significant structural differences between East and West Germany can still be assumed (Kluge & Weber, 2018).

TABLE 7 Rejection frequencies of the test statistic  $S_n^{CM^*}$ .

	$n = 500$			$n = 1000$			$n = 1500$		
	East	West	Ger	East	West	Ger	East	West	Ger
<b>S1</b>	0.034	0.134	0.132	0.038	0.329	0.303	0.026	0.553	0.535
<b>S2</b>	0.063	0.204	0.164	0.090	0.517	0.479	0.099	0.755	0.673
<b>S3</b>	0.050	0.136	0.094	0.026	0.353	0.339	0.030	0.551	0.529
<b>S4</b>	0.092	0.198	0.158	0.086	0.449	0.461	0.104	0.745	0.661
<b>S5</b>	0.089	0.429	0.387	0.276	0.880	0.775	0.507	0.966	0.948

Note: Incomes. Shown are the rejection rates of size  $n$  of the specification S1–S5. The number of subsamplings is 501 and the critical values are calculated at a significance level of 5% and for  $\tau \in \{0.1, 0.2, \dots, 0.9\}$ .



**FIGURE 1** Income quantiles for East/West and entire Germany as functions of age. Figures show the penalized conditional quantile estimates for West (red), East (green) and entire Germany (blue) at  $\tau = 0.1, 0.5, 0.9$ . The lines shown in lighter colors represent the 95% confidence intervals.

## 6.2 | Interaction effects in modeling Australian electricity prices

In this section, we apply the specification test  $S_n^{CM,S}$  to electricity data from the Australian national electricity market (NEM) in 2019. The NEM is a wholesale market, where generators, distributors and third party participants bid for sale and purchase of electricity one day ahead of transmission (Ignatieva & Trück, 2016; Shively & Smith, 2018). We consider hourly market-wide price  $P_i$  from January 1, 2019 to December, 31, 2019, which yields  $n = 8760$  observations. The market-wide price  $P_i$  is the demand-weighted average price across the five regions (<http://www.aemo.com.au>). We correct for the three main drivers of the electricity spot price distribution, namely day of the year  $x_1$ , time of day  $x_2$  and total market demand  $x_3$ , which is the sum of demand across the five regions in the NEM. Following Smith and Klein (2021) we thus choose a regression approach for the electricity data from the Australian NEM even if the problem could be addressed by a time series approach. For convenience, we scale each covariate to the unit interval.

Our main purposes are to identify (i) potential interactions between the covariates on different quantiles of the electricity spot price distributions (ii) to statistically investigate if the impact of the covariates  $x_1, x_2$  and  $x_3$  varies for distinct quantiles and (iii) to test which (interaction) effects are statistically significant. In contrast to the previous application in Section 6.1, it is not clear a priori how to optimally determine a functional relationship between the three covariates  $x_1, x_2$  and  $x_3$  for distinct quantiles  $\tau \in (0, 1)$ . Therefore, the functional relationship for different quantiles is modeled very flexibly by a spline approach. We employ trivariate P-splines (tensor product B-splines) as proposed by Eilers et al. (1996) which combine a multivariate B-spline basis, with a discrete penalty on the basis coefficients.

In order to investigate our main purposes (i)-(iii), we assume that the data generating process can be represented by one of the eight different specifications S6–S13. To increase the readability, the notation is geared to the implementation in R, that is,  $s(\cdot, \tau)$  models the marginal P-spline and  $ti(\cdot, \tau)$  solely the interaction effect at the quantile  $\tau$ . For example, S6 describes a P-spline for the three covariates  $x_1, x_2, x_3$  represented by the marginal main effects  $s(x_1, \cdot), s(x_2, \cdot), s(x_3, \cdot)$ , their mutual bivariate interactions  $ti(x_1, x_2, \cdot), ti(x_1, x_3, \cdot), ti(x_2, x_3, \cdot)$  and their mutual trivariate interaction  $ti(x_1, x_2, x_3, \cdot)$ . In contrast, specification S7 does not incorporate any interactions between

the covariates and thus models the marginal effects only. Specification S12 describes a P-spline that models the marginals and bivariate interaction effects within the 0.25 and 0.75-quantile. Specifically, we define

$$\begin{aligned}
 \text{S6: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= s(x_1, \tau) + s(x_2, \tau) + s(x_3, \tau) \\
 &\quad + ti(x_1, x_2, \tau) + ti(x_1, x_3, \tau) + ti(x_2, x_3, \tau) + ti(x_1, x_2, x_3, \tau) \\
 \text{S7: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= s(x_1, \tau) + s(x_2, \tau) + s(x_3, \tau) \\
 \text{S8: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= s(x_1, \tau) + s(x_2, \tau) + s(x_3, \tau) + ti(x_1, x_3, \tau) + ti(x_2, x_3, \tau) \\
 \text{S9: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= \begin{cases} \text{S6,} & \text{if } 0.25 < \tau < 0.75 \\ \text{S7,} & \text{otherwise} \end{cases} \\
 \text{S10: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= \begin{cases} \text{S6,} & \text{if } \tau \leq 0.25 \\ \text{S7,} & \text{otherwise} \end{cases} \\
 \text{S11: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= \begin{cases} s(x, \tau), & \text{if } 0.25 < \tau \\ s(x, \tau) + ti(x_1, x_2, \tau) + ti(x_2, x_3, \tau), & \text{otherwise} \end{cases} \\
 \text{S12: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= \begin{cases} \text{S6} - ti(x_1, x_2, x_3, \tau), & \text{if } 0.25 < \tau < 0.75 \\ \text{S7,} & \text{otherwise} \end{cases} \\
 \text{S13: } F_{Y|X}^{-1}(\tau | x_1, x_2, x_3) &:= \begin{cases} \text{S6} - ti(x_1, x_2, \tau) - ti(x_1, x_2, x_3, \tau), & \text{if } 0.25 < \tau < 0.75 \\ \text{S7,} & \text{otherwise} \end{cases}
 \end{aligned}$$

Similar to the previous section, all estimations are carried out using the `qgam` package in R. As before we use REML to optimize the smoothing parameters. Due to the extreme skew in electricity prices, we follow previous authors and set  $Y_i = \log(P_i + 17)$ . This avoids a negative dependent variable  $Y_i$  since the minimum observed price in our data is  $-\$15.78$ . For the application of our test, we set  $n \in \{500, 1000, 2000\}$  and  $\tau \in \{0.02, 0.04, \dots, 0.98\}$ . The number of subsamples is 501 and the critical values are calculated at a significance level of 5%. To ensure comparability of rejection rates for different  $n$  and since we include multivariate interaction effects (cf. S6 and S8–S13), we set the number of knots to 5. The rejection rates of the specification test  $S_n^{CM*}$  are listed in Table 8. From this table we make four observations. First, it can be observed that the rejection rates increase as  $n$  increases, which is plausible as our specification test is consistent and also small deviations from the null hypothesis are detected for large sample sizes. In addition, an increase in the rejection rates as  $n$  increases could be due to possible structural breaks. Second, based on the rejection rates for S8–S13 at  $n = 500$ , interaction effects seem to have a significant impact, especially in the lower quantile, i.e. at  $\tau \leq 0.25$ . This is particularly reflected in the comparison of the specifications S9 and S10, which differ in the modeling of the upper quantile ( $\tau \geq 0.75$ ) but show similar rejection rates. Third, we can conclude from the specifications S12 and S13 that the interaction between the day of the year ( $x_1$ ) and the daytime ( $x_2$ ) has no significant impact to the log electricity prices. Fourth, specification S7, however, which does not incorporate interaction effects, is rejected at all sample sizes.

Figure 2 shows the decomposition of the main and interaction effects at the 90% quantile at 6:00 p.m. using S6. Since the contour lines in the second and third panel show the presence of interactions between demand and day, we conclude that the relation between the three covariates cannot be fully captured by product interactions based on univariate splines. In addition, different

TABLE 8 Rejection rates of the test statistic  $S_n^{CM^*}$ .

	S6	S7	S8	S9	S10	S11	S12	S13
$n = 500$	0.050	0.115	0.065	0.086	0.043	0.086	0.058	0.079
$n = 1000$	0.089	0.338	0.300	0.178	0.185	0.135	0.218	0.224
$n = 2000$	0.228	0.811	0.748	0.256	0.237	0.445	0.713	0.764

Note: Electricity prices. The table shows the subsample rejection rates of size  $n$  of the specification S6–S13.

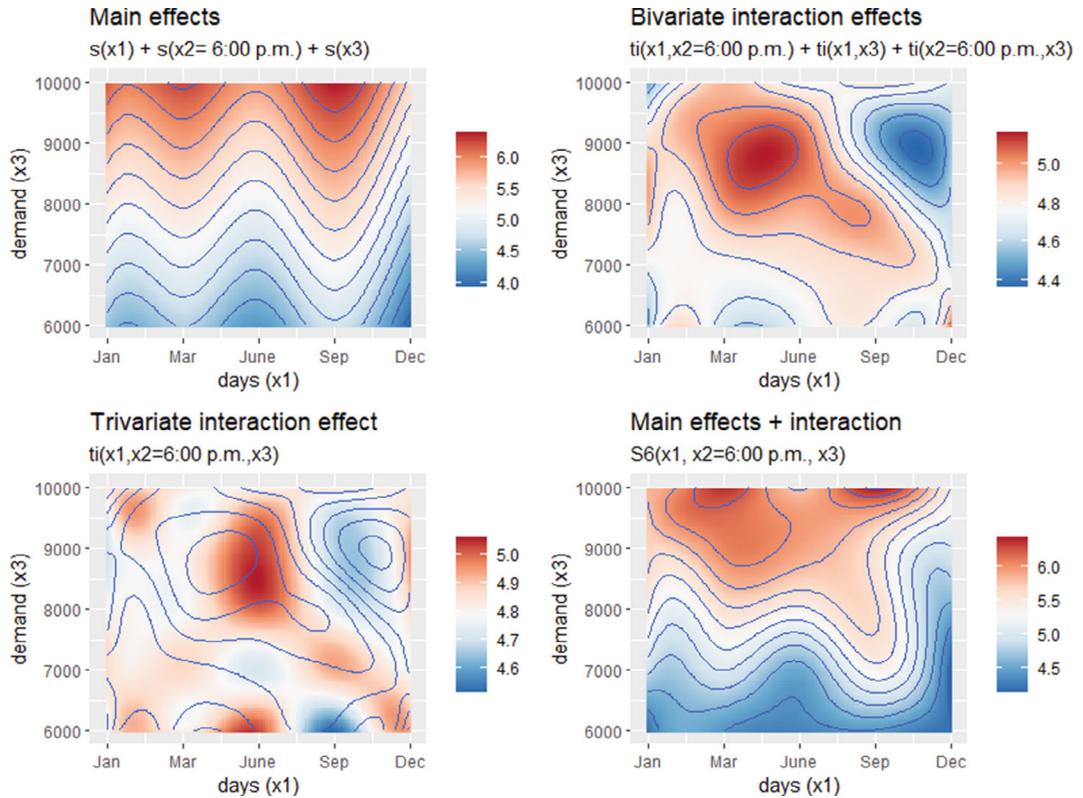


FIGURE 2 Estimated main and interaction effects at the 90% quantile at 6:00 p.m. Electricity prices.

Figures depict the estimated effects of the three covariates on the 90% quantile of the Australian NEM hourly electricity price distribution for 2019, where the time of the day ( $x_2$ ) is set to 06:00 p.m. The estimation is based on the model specification S6. The first panel (upper left) shows the sum of the univariate main effects of days ( $x_1$ ),  $x_2$  and total market demand ( $x_3$ ). The second and third panel illustrate the bivariate and trivariate interaction effects. The overall effect is depicted in the last panel.

day-demand combinations have a different impact on the market wide price  $P_t$ . A similar graphical analysis additionally reveals this behavior for the 10% quantile (see Supplement III). This observation is also confirmed by the higher rejection rates of our specification test when using univariate splines rather than bivariate tensor product B-splines (cf. Tables 8 and IV in Supplement III). Overall, we conclude that for a thorough specification of the Australian NEM mutual interaction effects are important. Particularly, there seems to be a complex dependence structure in the lower quantile ( $\tau \leq 0.25$ ) of log electricity prices, which can be captured by means of

the mutual interaction effects. However, the interaction between day of the year and daytime is negligible here. This might be important for risk management purposes.

## 7 | CONCLUSION

In this paper, we derive and test new specification tests for parametric and semi-parametric quantile regression models, which allow the covariates to vary over quantiles in a flexible nonlinear way. To improve finite sample properties in the parametric model framework, we replace the nonparametric ecdf by an estimator based on an estimate of the conditional qf using penalized splines. Our MC study illustrates that the proposed method has superior test properties compared to several existing benchmarks from the literature. We illustrate this in two relevant examples on income inequality and electricity spot prices. In the former, the (nonlinear) effect of age on the income distribution is known in the literature. A detailed investigation of the conditional income distributions between East and West Germany using the M–M decomposition reveals that still income differences between the regions in Germany are present, even more than two decades after the reunification. Similarly, modeling and predicting electricity spot prices is a common issue in economics. We treat the problem in a semi-parametric framework and reveal the importance of interaction effects between demand and time variables, particularly for lower quantiles of the price distributions.

We believe our test statistics make an important contribution in the specification testing literature since nonlinear or even more complex functional forms of covariates are omnipresent in many applications.

## ACKNOWLEDGMENTS

Tim Kutzker (*corresponding author*) and Nadja Klein acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German research foundation) through the Emmy Noether grant KL 3037/1-1. We thank Matteo Fasiolo for helpful comments on the `qgam` (Fasiolo et al., 2021) package in R. The authors acknowledge computational facilities by the Regional Computing Center of the University of Cologne via the DFG-funded High Performance Computing system CHEOPS (grant INST 216/512/1FUGG). Open Access funding enabled and organized by Projekt DEAL.

## ORCID

Nadja Klein  <https://orcid.org/0000-0002-5072-5347>

## REFERENCES

- Angrist, J., Chernozhukov, V., & Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structure. *Econometrica*, 74, 539–563.
- Bach, S., Corneo, G., & Steiner, V. (2009). From bottom to top: The entire income distribution in Germany, 1992–2003. *Review of Income and Wealth*, 55, 303–330.
- Belloni, A., Chernozhukov, V., Chetverikov, D., & Fernández-Val, I. (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics*, 213, 4–29.
- Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58, 1443–1458.
- Biewen, M. (2000). Income inequality in Germany during the 1980s and 1990s. *Review of Income and Wealth*, 46, 1–19.
- Billingsley, P. (1995). *Probability and measure. Wiley series in probability and statistics*. Wiley.

- Bondell, H. D., Reich, B. J., & Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, *97*, 825–838.
- Breunig, C. (2019). *Specification testing in nonparametric instrumental quantile regression*. arXiv preprint arXiv:1909.10129.
- Statistisches Bundesamt. (2012). *Periodensterbetafeln für deutschland: allgemeine sterbetafeln, abgekürzte sterbetafeln und sterbetafeln*. Statistisches Bundesamt.
- Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, *128*, 967–1015.
- Cardot, H., Crambes, C., & Sarda, P. (2005). Quantile regression when the covariates are functions. *Nonparametric Statistics*, *17*, 841–856.
- Chao, S.-K., Volgushev, S., & Cheng, G. (2017). Quantile processes for semi and nonparametric regression. *Electronic Journal of Statistics*, *11*, 3272–3331.
- Chernozhukov, V. (2002). *Inference on quantile regression process, an alternative*. SSRN.
- Chernozhukov, V., Fernández-Val, I., & Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, *81*, 2205–2268.
- Creedy, J., & Hart, P. E. (1979). Age and the distribution of earnings. *Economic Journal*, *89*, 280–293.
- De Boor, C. (1978). *A practical guide to splines* (Vol. 27). Springer-Verlag.
- Dudley, R. M. (2014). *Uniform central limit theorems*. *Cambridge Studies in Advanced Mathematics* (2nd ed.). Cambridge University Press.
- Dustmann, C., Ludsteck, J., & Schönberg, U. (2009). Revisiting the German wage structure. *The Quarterly Journal of Economics*, *124*, 843–881.
- Eilers, P. H., Marx, B. D., et al. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, *11*, 89–121.
- Escanciano, J. C., & Goh, S.-C. (2014). Specification analysis of linear quantile models. *Journal of Econometrics*, *178*, 495–507.
- Escanciano, J. C., & Velasco, C. (2010). Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics*, *159*, 209–221.
- Fasiolo, M., Wood, S. N., Zaffran, M., Nedellec, R., & Goude, Y. (2021). Fast calibrated additive quantile regression. *Journal of the American Statistical Association*, *116*, 1402–1412.
- Gozalo, P. L. (1993). A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory*, *9*, 451–477.
- Guerre, E., & Sabbah, C. (2012). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, *28*, 87–129.
- Hallin, M., Lu, Z., Yu, K., et al. (2009). Local linear spatial quantile regression. *Bernoulli*, *15*, 659–686.
- Härdle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, *21*, 1926–1947.
- He, X., & Shi, P. (1998). Monotone B-Spline smoothing. *Journal of the American Statistical Association*, *93*, 643–650.
- Ignatieva, K., & Trück, S. (2016). Modeling spot price dependence in Australian electricity markets with applications to risk management. *Computers & Operations Research*, *66*, 415–433.
- Klein, N., Kneib, T., Lang, S., Sohn, A., et al. (2015). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *The Annals of Applied Statistics*, *9*, 1024–1052.
- Kluge, J., & Weber, M. (2018). Decomposing the German East–West wage gap. *Economics of Transition*, *26*, 91–125.
- Koenker, R. (2005). *Quantile regression*. *Econometric society monographs*. Cambridge University Press.
- Koenker, R., & Bassett, G., Jr. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.
- Koenker, R., & Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, *70*, 1583–1612.
- Landmesser, J. M., et al. (2016). Decomposition of differences in income distributions using quantile regression. *Statistics in Transition. New Series*, *17*, 331–349.
- Li, D., Li, Q., & Li, Z. (2020). Nonparametric quantile regression estimation with mixed discrete and continuous Data. *Journal of Business & Economic Statistics*, *39*(3), 741–756.
- Li, Q., & Racine, J. S. (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, *26*, 423–434.
- Lian, H., Meng, J., & Fan, Z. (2015). Simultaneous estimation of linear conditional quantiles with penalized splines. *Journal of Multivariate Analysis*, *141*, 1–21.

- Luong, M., & Hébert, B.-P. (2009). *Age and earnings*. Citeseer.
- Machado, J. A. F., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20, 445–465.
- Ng, P., & Maechler, M. (2007). A Fast and efficient implementation of qualitatively constrained quantile smoothing splines. *Statistical Modelling*, 7, 315–328.
- Ng, P. T., & Maechler, M. (2020). COBS—Constrained B-splines (Sparse matrix based). *R package version 1.3-4*.
- Qu, Z., & Yoon, J. (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, 185, 1–19.
- Rothe, C., & Wied, D. (2013). Misspecification testing in a class of conditional distributional models. *Journal of the American Statistical Association*, 108, 314–324.
- Shively, T. S., & Smith, M. S. (2018). Econometric modeling of regional electricity spot prices in the Australian market. *Energy Economics*, 74, 886–903.
- Smith, M. S., & Klein, N. (2021). Bayesian inference for regression copulas. *Journal of Business & Economic Statistics*, 39, 712–728.
- Soni, P., Dewan, I., & Jain, K. (2012). Nonparametric estimation of quantile density function. *Computational Statistics & Data Analysis*, 56, 3876–3886.
- Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, 25, 613–641.
- Stute, W., Manteiga, W. G., & Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93, 141–149.
- Takeuchi, I., Le, Q., Sears, T., & Smola, A. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7, 1231–1264.
- Toraichi, K., Katagishi, K., Sekita, I., & Mori, R. (1987). Computational complexity of spline interpolation. *International Journal of Systems Science*, 18, 945–954.
- Troster, V., & Wied, D. (2021). A specification test of dynamic conditional distributions. *Econometric Review*, 40, 109–127.
- Volgushev, S., Chao, S.-K., & Cheng, G. (2019). Distributed inference for quantile regression processes. *The Annals of Statistics*, 47, 1634–1662.
- Wagner, G., Frick, J., & Schupp, J. (2007). 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, 139–169.
- Xue, L., & Wang, J. (2010). Distribution function estimation by constrained polynomial spline regression. *Journal of Nonparametric Statistics*, 22, 443–457.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kutzker, T., Klein, N., & Wied, D. (2024). Flexible specification testing in quantile regression models. *Scandinavian Journal of Statistics*, 51(1), 355–383. <https://doi.org/10.1111/sjos.12671>