Flexible Specification Testing in Quantile Regression Models

Tim Kutzker[1,3,*] and Nadja Klein[1,2,3*] and Dominik Wied[3,*]

[1]Chair of Statistics and Data Science; Emmy Noether Research Group, Humboldt-Universität zu Berlin
[2]Chair of Uncertainty Quantification and Statistical Learning, Research Center for Trustworthy Data Science and Security (UA Ruhr) and Department of Statistics (Technische Universität Dortmund)
[3]Institute for Econometrics and Statistics, Universität zu Köln

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 non-linear. 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 modelling 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;

---

*E-mails: tim.kutzker@gmail.com, nadja.klein@.tu-dortmund.de, dwied@uni-koeln.de. Addresses: [1] Unter den Linden 6, 10099 Berlin, Germany. [2] Joseph-von-Fraunhofer-Str. 25, 44227 Dortmund, Germany. [3] Albertus Magnus Platz, 50923 Cologne, Germany. 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., 2020) 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).

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 non-parametric approaches. In general, testing misspecification in linear ordinary least squares (OLS) models is well understood and developed.

In parametric models, e.g. , Bierens (1990) showed that any conditional moment test of functional form of non-linear 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 non-parametric case, amongst others, Gozalo (1993) proposed a general framework for specification testing of the regression function in a non-parametric 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 Jr (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 and Mata, 2005) and the Khmaladze transformation (Koenker and Xiao, 2002). Thus, testing the validity of the imposed structure remains one of the key taks 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 theses classes. This principle was extended to dynamic models by Troster and Wied (2021).

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

But regardless of whether parametric or non-parametric approaches are chosen, the theory concerning the validity of the correct model choice seems to keep up with the rapid development of new estimation methods only to a limited extent. 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 non-linear) covariate effects. While non-linear regressors can be considered in a standard paraemtric quantile regression framework, testing for presence of such non-linear 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 non-linear 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 non-pivotal limiting distribution of our test statistic and show the validity of our suggested parametric bootstrap procedure for the approximation of the critical values. To increase the power of our test, we replace the unrestricted model estimate with a quadratic B-spline. 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 (2020), 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. Sec. 2 formulates the test problem for the finite-dimensional parametric and semi-parametric models. In Sec. 3, we provide the theoretical properties of the testing procedures and derive their limiting distributions. Sec. 4 describes a practical and easy-to-implement bootstrap procedure, which provides valid coverage. In Sec. 6 we present the two empirical applications. The last Sec. 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 Sec. 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). Sec. 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 Sec. 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 Sec. 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 Sec. 2.2) splines whose dimension grows as a function of the number of observations, i.e., 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 $i = 1, \ldots, n$ 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

$$F^{-1}_{Y|X}(\tau \mid x) = P(x, \tau)^\top \theta(\tau), \tag{2.1}$$

where $F^{-1}_{Y|X}(\tau \mid x)$ denotes the qf of Y conditional on $X = x \in \mathbb{R}^K$ at quantile τ, $P(x, \tau) \in \mathbb{R}^{p_\tau}$ 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 τ for all $\tau \in \mathcal{T} \subset (0, 1)$. Naturally, models in which the vector of transformations does not depend on τ 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. (2019b) for $P(x, \tau) \equiv P(x)$, the above framework incorporates a variety of models such as parametric (Koenker, 2005) and semi-parametric (He and Shi, 1997) ones. However, since we allow the transformation vector $P(x, \tau)$ to depend on the quantile τ, models of the form (2.1) 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 non-linear functional form in the upper 50% quantile, e.g. $P(x, \tau) = x$ 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 Sec. 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 (2.1) to be specified by a parametric model, while generalizations are treated thereafter.

Our test principle is designed for the comparison of the non-parametric 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 $F^{-1}_{Y|X}$ through the following equation

$$F_{Y|X}(y \mid x) = \int_0^1 \mathbb{1}_{\left\{F^{-1}_{Y|X}(\tau|x) \leq y\right\}} d\tau \qquad \forall\, y \in \mathbb{R}. \tag{2.2}$$

In the following, we consider the set of all conditional distribution functions satisfying (2.2) given the model specification (2.1), which we denote by $\mathcal{F}$, i.e.

$$\mathcal{F} := \{F_{Y|X}(y \mid x, \theta) \mid F^{-1}_{Y|X}(\tau \mid x) = P(x,\tau)^\top \theta(\tau) \text{ for some } \theta \in \mathcal{B}(\mathcal{T}, \Theta),\ (y,x) \in \mathcal{S}\}, \tag{2.3}$$

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 (2.1) is correctly specified for all $\tau \in \mathcal{T}$ transfers by means of (2.3) to hypotheses of the form

$$H_0 : F_{Y|X} \in \mathcal{F} \quad \text{vs.} \quad H_1 : F_{Y|X} \notin \mathcal{F}. \tag{2.4}$$

Thus, we want to test if the conditional cdf $F_{Y|X}$ coincides with an element of $\mathcal{F}$ from (2.3). 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 $\mathcal{F}^0 := \{F_{Y|X}(y \mid x, \theta_0) \mid F^{-1}_{Y|X}(\tau \mid x) = P(x,\tau)^\top \theta_0(\tau) \text{ for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \forall (y,x) \in \mathcal{S}\}$. Hence, we can reformulate (2.4) as

$$\begin{aligned} H_0 &: F_{Y|X}(y \mid x) = F_{Y|X}(y \mid x, \theta_0) \text{ for some } \theta_0 \in \mathcal{B}(\mathcal{T}, \Theta) \text{ for all } (y,x) \in \mathcal{S} \\ \text{vs. } H_1 &: F_{Y|X}(y \mid x) \neq F_{Y|X}(y \mid x, \theta) \text{ for all } \theta \in \mathcal{B}(\mathcal{T}, \Theta) \text{ for some } (y,x) \in \mathcal{S}. \end{aligned} \tag{2.5}$$

Additionally we assume that θ_0 is identified under the null hypothesis through a moment condition. Specifically, let $g : \mathcal{S} \times \Theta \times \mathcal{T} \to \mathbb{R}^p$ be a uniformly integrable function whose exact form depends on $\mathcal{F}^0$, and suppose that for every $\tau \in \mathcal{T}$

$$G(\theta, \tau) := \mathbb{E}[g(Y, X, \theta, \tau)] = 0 \in \mathbb{R}^p \tag{2.6}$$

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 (2.6) 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 (2.4) as

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

with $\theta_0(\tau)$ as the unique solution to (2.6) for all $\tau \in \mathcal{T}$. This holds true since $\mathcal{F}^0$ is a singleton containing $F_{\cdot|\cdot}(\cdot \mid \cdot, \theta_0)$. Since $F_{Y|X}(y \mid X) = \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \mid X]$, we can write the joint cdf F of Y and X as

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

where F_X denotes the marginal cdf of X. From Theorem 16.10 (iii) of Billingsley (1995) it follows that the testing problem (2.5) can be restated as

$$\begin{aligned} &H_0 : F(y, x) = F(y, x, \theta_0) \text{ for all } (y, x) \in \mathbb{R}^{K+1} \\ \text{vs. } &H_1 : F(y, x) \neq F(y, x, \theta_0) \text{ for some } (y, x) \in \mathbb{R}^{K+1}. \end{aligned} \tag{2.7}$$

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

$$S(y, x, \theta) := F(y, x) - F(y, x, \theta). \tag{2.8}$$

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

$$S_n(y, x, \hat{\theta}_n) := \hat{F}_n(y, x) - \hat{F}_n(y, x, \hat{\theta}_n), \tag{2.9}$$

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 (2.1). 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 $(y, x) \in \mathcal{S}$. If, however, the alternative holds true, then there is a vector $(y, x) \in \mathcal{S}$ for each $\theta \in \mathcal{B}(\mathcal{T}, \Theta)$ such that the absolute value of the function S_n from (2.9) 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

$$\left\|\hat{G}_n(\hat{\theta}_n, \tau)\right\| = \inf_{\theta \in \Theta} \left\|\hat{G}_n(\theta, \tau)\right\| + \eta_n, \tag{2.10}$$

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 analogue of the moment condition (2.6) 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 $(y, x) \in \mathcal{S}$, the estimator based on the testing problem (2.5) is

$$\begin{aligned} \hat{F}_n(y \mid x, \hat{\theta}_n) &= \int_{\mathcal{T}} \mathbb{1}_{\{P(x,\tau)^\top \hat{\theta}_n(\tau) \leq y\}} d\tau + \int_{(0,1)\setminus\mathcal{T}} \mathbb{1}_{\{\hat{F}^{-1}_{Y|X}(\tau|X) \leq y\}} d\tau, \\ \hat{\theta}_n(\tau) &= \operatorname*{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} \left(\tau - \mathbb{1}_{\{y_i \leq P(x_i,\tau)^\top \theta\}}\right) \left(y_i - P(x_i, \tau)^\top \theta\right). \end{aligned} \tag{2.11}$$

For $\tau \notin T$, the conditional qf $\hat{F}^{-1}_{Y|X}(\tau \mid X)$ is some estimator for the conditional qf which is consistent both under the null and the alternative hypothesis (e.g. Takeuchi et al., 2006; Soni et al., 2012). If $\mathcal{T} = [\epsilon, 1-\epsilon]$ for a small $\epsilon > 0$, this term is negligible in practice. The integral in (2.11) can be computed by means of standard numerical integration techniques (i.e. averaging over a fine equidistant grid of τ with 49 supporting points starting at 0.02 using the trapezoidal rule) and corresponds to the canonical quantile regression approach, i.e. the loss function g from (2.6) is given by $g(Y, X, \theta, \tau) = (\tau - \mathbb{1}\{Y \leq P(X, \tau)^\top \theta(\tau)\}) P(X, \tau)$ (compare Lemma 14 of Chernozhukov et al., 2013). Additionally, (2.11) and other typical estimation methods fit the estimated conditional qf $\hat{F}_n^{-1}(\tau \mid x)$ pointwise in $\tau \in \mathcal{T}$, which might induce the problem that the estimated quantile curve $\tau \mapsto \hat{F}_n^{-1}(\tau \mid x)$ 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 $\hat{F}_n(y \mid x, \hat{\theta}_n)$ which is monotone in y by construction for every x. Hence, a valid test statistic can be based on the differences of the non-parametric and parametric ecdfs $\hat{F}_n(y,x)$ and $\hat{F}_n(y,x,\hat{\theta}_n)$ and thus expressed as

$$\begin{aligned} S_n(y,x,\hat{\theta}_n) &= \hat{F}_n(y,x) - \hat{F}_n(y,x,\hat{\theta}_n) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}_{\{Y_i \le y\}}\mathbb{1}_{\{X_i \le x\}}\right) - \int_{\mathbb{R}^K} \mathbb{1}_{\{x^* \le x\}} \hat{F}_n(y \mid x, \hat{\theta}_n)\, d\hat{F}_X(x^*) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}_{\{Y_i \le y\}}\mathbb{1}_{\{X_i \le x\}} - \mathbb{1}_{\{X_i \le x\}}\hat{F}_n(y \mid x, \hat{\theta}_n)\right), \end{aligned} \tag{2.12}$$

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

$$S_n^{CM} := \int \left(\sqrt{n} S_n(y,x,\hat{\theta}_n)\right)^2 d\hat{F}_n(y,x), \tag{2.13}$$

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 (2.1) is independent of τ 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

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

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

2.2 Specification Test for Semi-Parametric Quantile Regression

Since we introduced the general testing principle of (2.1) 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, amongst 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; Chao et al., 2017; Belloni et al., 2019b; Chernozhukov et al., 2013). To distinguish such basis functions from the vector of transformations in the previous subsection, we use the notation $B^{\cdot}$ 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 $0 = t_1 < \ldots < t_M = 1$ in the interval $[0, 1]$ are given. For $x = (x_1, \ldots, x_K)^\top \in [0, 1]^K$ with $K \in \mathbb{N}$, we identify the B-spline quantile regression model in the spirit of (2.1) as

$$F^{-1}_{Y|X}(\tau \mid X = x) = \sum_{j=1}^{K} B(x_j \mid d_\tau)^\top \theta_j(\tau) \tag{2.14}$$

with $B(x_j \mid d_\tau) := (B_1(x_j \mid d_\tau), \ldots, B_{M+d_\tau-1}(x_j \mid d_\tau))^\top$ being the vector of $M + d_\tau - 1$ basis functions of degree d_τ that are defined recursively on the vector of knots on $[0, 1]$ and evaluated at x_j for $j = 1, \ldots, K$ (cf. De Boor, 1978, for the recursive Definition). For every $\tau \in \mathcal{T}$ and $j = 1, \ldots, K$, $\theta_j(\tau) = (\theta_{j,1}(\tau), \ldots, \theta_{j,M+d_\tau-1}(\tau))^\top$ defines the corresponding functional coefficient vectors. Although both M and d_τ can be conceived to depend on j for $j = 1, \ldots, K$ and additionally M on τ, we suppress these dependencies at this point due to readability and clearness. Note that for distinct quantiles τ the degree of the B-spline might differ. If $d_\tau \equiv d \in \mathbb{N}$ we refer to (2.14) as B-spline quantile regression model of degree d. Since our general quantile regression model in (2.1) conceptually allows for multivariate covariates, we make (2.14) 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 $\{1, \ldots, K\}$ for $i = 1, \ldots, q^*$ and f_j an arbitrary continuous function for $j \in J_i$. Thus, (2.14) generalizes to

$$F^{-1}_{Y|X}(\tau \mid X = x) = \sum_{j=1}^{q} B(\pi_j(x) \mid d_\tau)^\top \theta_j(\tau) \tag{2.15}$$

with $q = K + q^*$. For $\pi_j(x) = x_j$ and J_i singletons for $j = 1, \ldots, q$ with $q = K$ we receive our initial B-spline model (2.14). The estimator for models of the form (2.15) is given by

$$\hat{\theta}_n(\tau) = \underset{\theta \in \mathbb{R}^{q \cdot M}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{n} \rho_\tau \left(y_i - \sum_{j=1}^{q} B(\pi_j(x) \mid d_\tau)^\top \theta_j \right) \right\} \tag{2.16}$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}(u < 0))$ is the check function (Koenker and Bassett Jr, 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 (2.16) has the same limiting distribution as the classical constrained quantile regression estimator. Hence, in accordance with the discussion on monotonicity in Sec. 2.1, θ_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 (2.14) 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 (2.16), there are other consistent approaches. A prominent and easy to implement algorithm is the *divide and conquer* algorithm at fixed τ. 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 (2.5), we replace the estimator of the conditional qf in (2.11) with our estimator (2.16). This yields a new conditional distribution function $\hat{F}_n^S(y \mid x, \hat{\theta}_n)$. 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 (2.12) for fixed and finite q and M. We then define the CM test statistic for finite-dimensional semi-parametric quantile regression models

$$S_n^{CM,S} := \int \left(\sqrt{n} S_n^S(y, x, \hat{\theta}_n) \right)^2 d\hat{F}_n(y, x). \tag{2.17}$$

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 (2.5) with the non-parametric cdf. As the class of alternative hypotheses in (2.4) gets smaller, the power of S_n^{CM} can be improved if a spline is used for the non-parametric part of the test, i.e., 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 (1997) have pointed out that if the number of knots $k_n \sim (n/\log n)^{2/5}$ and under some mild assumptions (cf. He and Shi, 1997, 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, i.e. M diverges at a proper rate. This is known as non-parametric quantile regression, i.e. a linear model with increasing dimension in the regression coefficients. Note that in this framework the true functional parameter vector θ_0 also depends on n (Belloni et al., 2019b). Consequently, for non-parametric quantile regression, (2.1) expands to

$$F_{Y|X}^{-1}(\tau \mid x) = P_n(x_n, \tau)^\top \theta_{0_n}(\tau). \tag{2.18}$$

We discuss the theoretical framework of (2.18) in Sec. 3.2. Let $\hat{F}_n^{S_M}$ be the spline based estimate of the cdf via the non-parametric conditional qf according to (2.18). 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 Sec. 3.2. This yields the new test statistic

$$S_n^{CM^*} := \int \left(\sqrt{n} S_n^*(y, x, \hat{\theta}_n)\right)^2 d\hat{F}_n(y, x). \tag{2.19}$$

In comparison to S_n^{CM}, the test statistic $S_n^{CM^*}$ replaces the estimate of the ecdf $\hat{F}_n(y, x)$ in (2.12) by an appropriate spline estimate of the conditional qf via (2.18), 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 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 Sec. 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 and Wied, 2013). For this purpose, we restate the assumptions on compact subsets on $\mathcal{T}$. Let Θ 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 $l = 1,\dots,L$, $G(\cdot,\tau) : \Theta \to \mathbb{R}^p$ possesses a unique zero at $\theta_0 \in$ interior(Θ) such that $G(\theta_0,\tau) = 0$ for all $\tau \in \mathcal{T}$ and for some $\delta > 0$, $\mathcal{B} := \bigcup_{\tau \in \mathcal{I}_l} B_\delta(\theta_0)$ is a compact subset of $\mathbb{R}^p$ contained in Θ for $l = 1,\dots,L$.*

iv.) *Further, $G(\cdot,\tau)$ 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 non-singular 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 1i. Assumption 1ii 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 1ii–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

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

is F_{YX}-Donsker. This convergence can be shown to be jointly with that of the ecdf process $(y, x) \mapsto \sqrt{n}\left(\hat{F}_n(y, x) - F(y, x)\right)$ 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 (2.7),

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

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

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

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

$$Cov(\mathbb{G}_1(y_1, x_1), \mathbb{G}_2(y_2, x_2)) = \lim_{n \to \infty} n \; Cov\left(\hat{F}_n(y_1, x_1) - F(y_1, x_1), \hat{F}_n(y_2, x_2, \hat{\theta}_n) - F(y_2, x_2, \theta)\right).$$

ii.) Under any fixed alternative, i.e., when the data are distributed according to some F that satisfies the alternative hypothesis H_1 in (2.7),

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

Theorem 1 ensures distributional convergence of the test statistic S_n^{CM} and further shows that the non-parametric 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 Sec. 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 (2.7),

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

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

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

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

$$Cov(\mathbb{G}_1(y_1,x_1),\mathbb{G}_2^S(y_2,x_2)) = \lim_{n\to\infty} n \; Cov\left(\hat{F}_n(y_1,x_1) - F(y_1,x_1), \hat{F}_n^S(y_2,x_2,\hat{\theta}_n) - F(y_2,x_2,\theta)\right).$$

ii.) Under any fixed alternative, i.e., when the data are distributed according to some cdf F that satisfies the alternative hypothesis H_1 in (2.7),

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

In empirical applications, however, it is often common to estimate series terms with smoothing penalty parameters λ_j for $j = 1, \ldots, p$, since this avoids overfitting the data. Imposing the assumption that the penalty parameters are $\lambda_j = o(n^{1/2})$ for $j = 1, \ldots, p$ 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 non-parametric quantile regression, i.e. 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 $i = 1, \ldots, n$ 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 $a_n^* := \sqrt{n}/\|B(x)\|$, 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 $\frac{1}{M^*} \leq \lambda_{min}(\mathbb{E}[PP^\top]) \leq \lambda_{max}(\mathbb{E}[PP^\top]) \leq M^*$ holds uniformly in n and $\tau \in \mathcal{T}$ for some fixed $M^* > 0$.*

ii.) *The conditional distribution $F_{Y|X}(y \mid x)$ is twice differentiable w.r.t. y. We denote the corresponding derivatives by $f_{Y|X}(y \mid x)$ and $f^\top_{Y|X}(y \mid x)$. Assume that $\bar{f} := \left|\sup_{y,x} f_{Y|X}(y \mid x)\right| < \infty$ and $\bar{f^\top} := \sup_{y,x}\left|f^\top_{Y|X}(y \mid x)\right| < \infty$ uniformly in n.*

iii.) Assume there exists a constant $f_{min} > 0$ *such that* $\inf_{\tau \in \mathcal{T}} \inf_{x} f_{Y|X}(F^{-1}_{Y|X}(\tau \mid x) \mid x) \geq f_{min}$.

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* $\tilde{J}_m(\tau) := \mathbb{E}[PP^\top f_{Y|X}(F^{-1}_{Y|X}(\tau \mid x) \mid X)]$.

v.) Assume that $\xi_p^4 (\log n)^6 = o(n)$ *and letting* $c_n := \sup_{\tau,x} \left| F^{-1}_{Y|X}(\tau \mid X) - P^\top \hat{\theta}_n(\tau) \right|$ *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 τ. Assumptions 2 ii.)–iii.) are fairly standard. Assumptions 2 iv.) and v.) imply 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 τ 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* (2.7),

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

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, i.e., when the data are distributed according to some cdf F *that satisfies the alternative hypothesis* H_1 *in* (2.7),

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

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 I.1) can also be applied to multivariate splines and thus in particular to tensor product B-splines considered later in Sec. 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 II). 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 $\{X_{b,i} \mid 1 \leq i \leq n\}_{b=1,\ldots,B}$ 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 (2.13), i.e. 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 (2.13) and the null hypothesis is rejected on the level of significance q if $S_n^{CM} > \hat{c}_n(q)$. Critical values for (2.17) and (2.19) can be obtained in the same manner if the test statistic $S_{n,B}^{CM}$ is replaced by its counterparts, i.e. $S_{n,B}^{CM,S}$ 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 (2.13). This does not require any further assumptions. Assumption 1 ensures that the bootstrap consistently estimates the limiting distribution for (2.13). Under the null hypothesis and any fixed alternative (2.5), 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 (2.7), we have that

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

ii.) Under any fixed alternative H_1 in (2.7), we have that

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

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 Sec. 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 modelled 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 Sec. 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 modelled by a tensor product. In Sec. 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 $x_0 \in U(0, 2\pi)$, $x_1 \sim Bin(1, 0.5)$, $x_2 \sim N(0, 1)$, $x_3 \in U(0, 1)$, $x_4 \in \chi^2(1)$, $u \sim N(0, 1)$, $w \sim N(0, 0.1)$, $v = (1 - 2x_1) \cdot v_2^* \cdot 8^{-0.5}$ with $v_2^* \sim \chi^2(2)$, where $Bin(\cdot, \cdot)$, $N(\cdot, \cdot)$, $U(\cdot, \cdot)$ 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, DPG 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" χ^2 distribution with two degrees of freedom (normalized to have unit variance). DGPs 6–8 are multivariate LSS models where the level of heteroscedasticity increases. DGP 9 is considered in order to compare our proposed testing procedure with those provided in Chernozhukov (2002) and Koenker and Xiao (2002). When $\gamma_1 = 0$ DGP 9 is a LS model, otherwise it is a LSS model. DGPs 10–12 are processes in which the functional form appears predominantly non-linear. DGP 10 is implemented by modeling different functional forms for quantiles below and 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 non-linear growth only at values close to 1. Both DGPs require flexible modelling 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 and Maechler, 2020, 2007). 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 $\tau \in \{0.1, 0.2, \ldots, 0.9\}$, meeting monotonicity assumptions. The number of MC repetitions is equal to 701 with 500 bootstrap replications. The significance level is 0.05.

$$
\begin{aligned}
&\text{DGP 1: } y_1(x_0) := \frac{x_0}{4} + 1 + u, && \text{DGP 2: } y_2(x_0) := \frac{x_0}{4} + 1 + u \cdot x_0 \\
&\text{DGP 3: } y_3(x_0) := \frac{x_0^2}{4} + 1 + u \cdot x_0^2, && \text{DGP 4: } y_4(x_1, x_2) := x_1 + x_2 + u \\
&\text{DGP 5: } y_5(x_1, x_2) := x_1 + x_2 + v, && \text{DGP 6: } y_6(x_1, x_2) := x_1 + x_2 + (\frac{1}{2} + x_1)u \\
&\text{DGP 7: } y_7(x_1, x_2) := x_1 + x_2 + (\frac{1}{2} + x_1 + x_2^2)^{0.5} u && \qquad (5.1) \\
&\text{DGP 8: } y_8(x_1, x_2) := x_1 + x_2 + \frac{1}{5}(\frac{1}{2} + x_1 + x_2^2)^{1.5} u \\
&\text{DGP 9: } y_9(x_3) := x_3 + (1 + \gamma_1 \cdot x_2)u \\
&\text{DGP 10: } y_{10}(x_3) := \begin{cases} \frac{x_3^2}{4} + 1 + \frac{u \cdot x_3^2}{2}, & \text{if } \tau \geq 0.5 \\ \frac{-x_3^2}{4} + 1 + u \cdot x_3, & \text{otherwise} \end{cases} \\
&\text{DGP 11: } y_{11}(x_3) := \sin\left(-\frac{\pi}{2} + x_3^3\right) + w, && \text{DGP 12: } y_{12}(x_3) := e^{y_5(x_3)}
\end{aligned}
$$

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 non-parameteric 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 τ.

Results Tab. 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

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

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

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.

of DGP 3 in Tab. 1). Last, the rejection rate for misspecified models (for DGP 3 we are assuming a linear LSS model in the last column of Tab. 1) in small samples ($n \leq 300$) is approximately three times higher than for the RW test.

Tab. 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 Tab. 2). Second, KX has difficulties to detect misspecification when heteroscedasticity is present (cf. DGP $6-8$ in Tab. 2). 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 Tab. 2). Tab. 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 Tab. 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, Tab. 4 examines size and power properties for the DGPSs $10 - -12$. 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 2. Size and power for DGPs 4–8

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

MC Study. The table compares size and power (at significance level 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 at significance levels 10% and 5% each.

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
$n = 100$	0.101	0.264	0.898	0.014	0.348	0.980	**0.050**	**0.396**	**0.99**
$n = 200$	0.070	0.480	0.988	0.052	0.752	1.000	0.063	0.772	1.000
$n = 300$	0.062	0.622	0.998	0.058	0.910	1.000	0.068	0.930	1.000

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. $\gamma_1 = 0$.

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

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, Dustmann et al., 2009; Card et al., 2013) 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 Tab. 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 and Hart, 1979; Luong and Hébert, 2009; Klein et al., 2015) . We next conduct a M-M decomposition (Machado and Mata, 2005; Landmesser et al., 2016), of the *year*-adjusted dataset conditional on *origin*. For the decomposition we assume that the conditional qf of the income Y

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

	Description		
Y	gross market labor income (in €), (continuous $1257 \leq Y \leq 280092$, average = 46641)		
origin	indicator for East or West (binary, -1=West (73.8%), 1=East (26.2%))		
age	age of the male in year (continuous, $20 \leq age \leq 60$, average = 38)		
years	time in years (categorical, $2001 \leq year \leq 2010$, 10 years)		
Sample	**Description**	**average (std.) income**	**observations**
Ger	complete sample	51026€ (30569€)	$n = 7220$
West	sub-sample ($origin = -1$)	55141€ (31494€)	$n = 5325$
East	sub-sample ($origin = 1$)	39463€ (24336€)	$n = 1895$

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

can be represented as a function of the form $F^{-1}_{Y|X}(\tau \mid X) = P(X, \tau)^\top \theta(\tau)$ 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 $\tau \in (0, 1)$, i.e. $P(X, \tau) = age$ for all $\tau \in (0, 1)$. The second models a quadratic influence of age on income for all quantiles $\tau \in (0, 1)$, i.e. $P(X, \tau) = age^2$ for all $\tau \in (0, 1)$. And finally, the third model considers the sum of the regressors *age* and age^2 that are constant for all quantiles $\tau \in (0, 1)$, i.e. $P(X, \tau) = age + age^2$ for all $\tau \in (0, 1)$. Due to the probability integral transform theorem the sequence $P(X, \tau_i)^\top \hat{\theta}_n(\tau_i)$

for $\tau_i \overset{i.i.d.}{\sim} U(0,1)$, $i = 1, \ldots, n$ constitutes a random sample from the estimated conditional distribution of Y given the covariates X (Machado and Mata, 2005). In order to obtain the difference between East and West, first, the coefficients for East ($\hat{\theta}_E(\tau)$) and West ($\hat{\theta}_W(\tau)$) for $\tau \in \{0.1, 0.2, \ldots, 0.9\}$ 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 $i = 1, \ldots, B$ 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, $i = 1, \ldots, B$, $l = E, W$. 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

$$\begin{aligned} \hat{\Delta}_Y &= \frac{1}{B} \sum_{b=1}^{B} \left(\left(P(X_E^b, \tau) - P(X_W^b, \tau)\right) \hat{\theta}_E(\tau) + \left(\hat{\theta}_E(\tau) - \hat{\theta}_W(\tau)\right) P(X_W^b, \tau) \right) \\ &\approx F^{-1}_{Y_E|X_E}(\tau \mid X_E) - F^{-1}_{Y_W|X_W}(\tau \mid X_W), \end{aligned} \tag{6.1}$$

where the first summand of (6.1) is the explained, while the second summand depicts the unexplained difference.

Tab. 6 summarizes results from the counterfactual analysis described above. The covariates used for the quantile regressions are *age* (rows 4–9), *age*^{2} (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*^{2} for all quantiles due to high residuals (4.37 for $\tau = 0.1$ and 7.33 for $\tau = 0.9$), indicating misspecification. However, the covariate *age*^{2} seems to be appropriate for the smallest quantile 0.1 (residual of 0.92 in Tab. 6), while a linear effect of age to income seems to prevail in higher quantiles (-1.49

Table 6. Decomposition of the West/East income differential

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

Incomes. The covariates used for the quantile regressions are *age* (rows 4-9), age^2 (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^2 and $age + age^2$ as covariates for the quantile regression models. The rows *M-M gap* are the estimated gap of the income difference. The quantiles τ range from 0.1 to 0.9. The number of bootstrap replications is equal to 2500. All numbers are in percent. Totals may not sum exactly to 100% due to rounding.

in Tab. 6). In contrast, the additive model $age + age^2$ seems to capture the income effect for all quantiles quite well due to moderate residuals (cf. last row *Residuals* in Tab. 6). For all decompositions it holds, that *age* and age^2 contribute a maximum of 16% to the explanation of the income difference between East and West Germany (except highest quantile in age^2, i.e. 25.41). Due to the different residuals and the different explanatory power of the income gap between East and West for the quantile regressions based on *age* or age^2, it seems reasonable to assume that *age* and age^2 have different effects for different quantiles. For example, the residual of the 30% quantile of *age* (0.24 in Tab. 6) is about 18 times smaller than the residual of the corresponding quantile regression using age^2 as explanatory variable (4.24 in Tab. 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 $S_n^{CM^*}$.

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

$$\text{S2} : F_{Y|X}^{-1}(\tau \mid x) = \begin{cases} (x^2)^\top\theta_0, \ \text{if } 0 \leq \tau \leq 0.1 \\ x^\top\theta_0 \ , \ \text{otherwise} \end{cases} \qquad \begin{array}{l} \text{S4} : F_{Y|X}^{-1}(\tau \mid x) = x^\top\theta_0 \\ \text{S5} : F_{Y|X}^{-1}(\tau \mid x) = (x^2)^\top\theta_0 \end{array}$$

For this purpose, we have defined five different model specifications S1–S5, which should take into account the observations made in Tab. 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 sub-samples East and West as well as to the complete data set. We estimate the function $\hat{F}_n^S(y, x, \theta)$ in (2.19) by a cubic spline with second order difference penalty, setting the basis dimension to 20 using the R package `qgam`. The smoothing parameter λ is estimated using the restricted maximum likelihood (REML) procedure within the package. We then re-estimate the models with optimized smoothing parameter and compute our test statistic $S_n^{CM^*}$ for $\tau \in \{0.1, 0.2, \ldots, 0.9\}$. Since the sample sizes for East, West and All differ and in order to make the results comparable, we computed the rejection rates of sub-samples 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 sub-sample a total of 501 times and refer to it as *sub-samplings* in the following. Tab. 7 summarizes the resulting rejection rates of the test statistic $S_n^{CM^*}$.

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

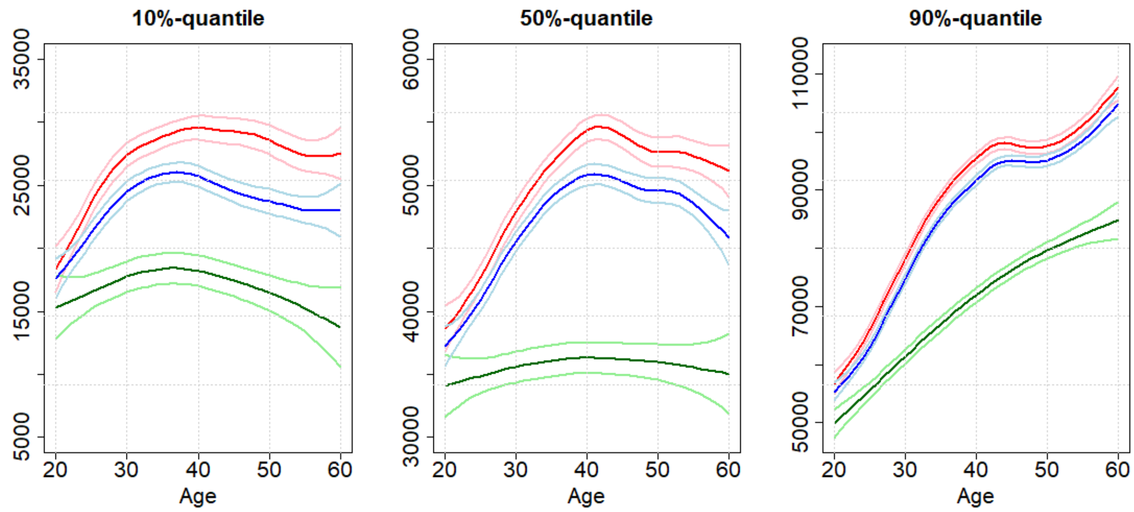
	$n = 500$			$n = 1000$			$n = 1500$		
	East	West	Ger	East	West	Ger	East	West	Ger
S 1	0.034	0.134	0.132	0.038	0.329	0.303	0.026	0.553	0.535
S 2	0.063	0.204	0.164	0.090	0.517	0.479	0.099	0.755	0.673
S 3	0.050	0.136	0.094	0.026	0.353	0.339	0.030	0.551	0.529
S 4	0.092	0.198	0.158	0.086	0.449	0.461	0.104	0.745	0.661
S 5	0.089	0.429	0.387	0.276	0.880	0.775	0.507	0.966	0.948

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

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, Fig. 6.1 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 rejections rates for different specifications significant structural differences between East and West Germany can still be assumed (Kluge and Weber, 2018).

Figure 6.1. Income quantiles for East/West and entire Germany as functions of *age*



Incomes. 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 Modelling 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 and Trück, 2016; Shively and 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 (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 (2020) 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 Sec. 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, i.e. $s(\cdot, \tau)$ models the marginal P-spline and $ti(\cdot, \tau)$ solely the interaction effect at the quantile τ. 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 marginals 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

$$\text{S6:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := s(x_1,\tau) + s(x_2,\tau) + s(x_3,\tau) + ti(x_1,x_2,\tau) + ti(x_1,x_3,\tau) + ti(x_2,x_3,\tau) + ti(x_1,x_2,x_3,\tau)$$

$$\text{S7:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := s(x_1,\tau) + s(x_2,\tau) + s(x_3,\tau)$$

$$\text{S8:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := s(x_1,\tau) + s(x_2,\tau) + s(x_3,\tau) + ti(x_1,x_3,\tau) + ti(x_2,x_3,\tau)$$

$$\text{S9:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} S6 \ , & \text{if } 0.25 < \tau < 0.75 \\ S7 \ , & \text{otherwise} \end{cases}$$

$$\text{S10:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} S6 \ , & \text{if } \tau \leq 0.25 \\ S7 \ , & \text{otherwise} \end{cases}$$

$$\text{S11:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} s(x,\tau) \ , & \text{if } 0.25 < \tau \\ s(x,\tau) + ti(x_1,x_2,\tau) + ti(x_2,x_3,\tau) \ , & \text{otherwise} \end{cases}$$

$$\text{S12:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} S6 - ti(x_1,x_2,x_3,\tau) \ , & \text{if } 0.25 < \tau < 0.75 \\ S7 \ , & \text{otherwise} \end{cases}$$

$$\text{S13:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} S6 - ti(x_1,x_2,\tau) - ti(x_1,x_2,x_3,\tau), & \text{if } 0.25 < \tau < 0.75 \\ S7 \ , & \text{otherwise} \end{cases}$$

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 sub-samples 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 Tab. 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

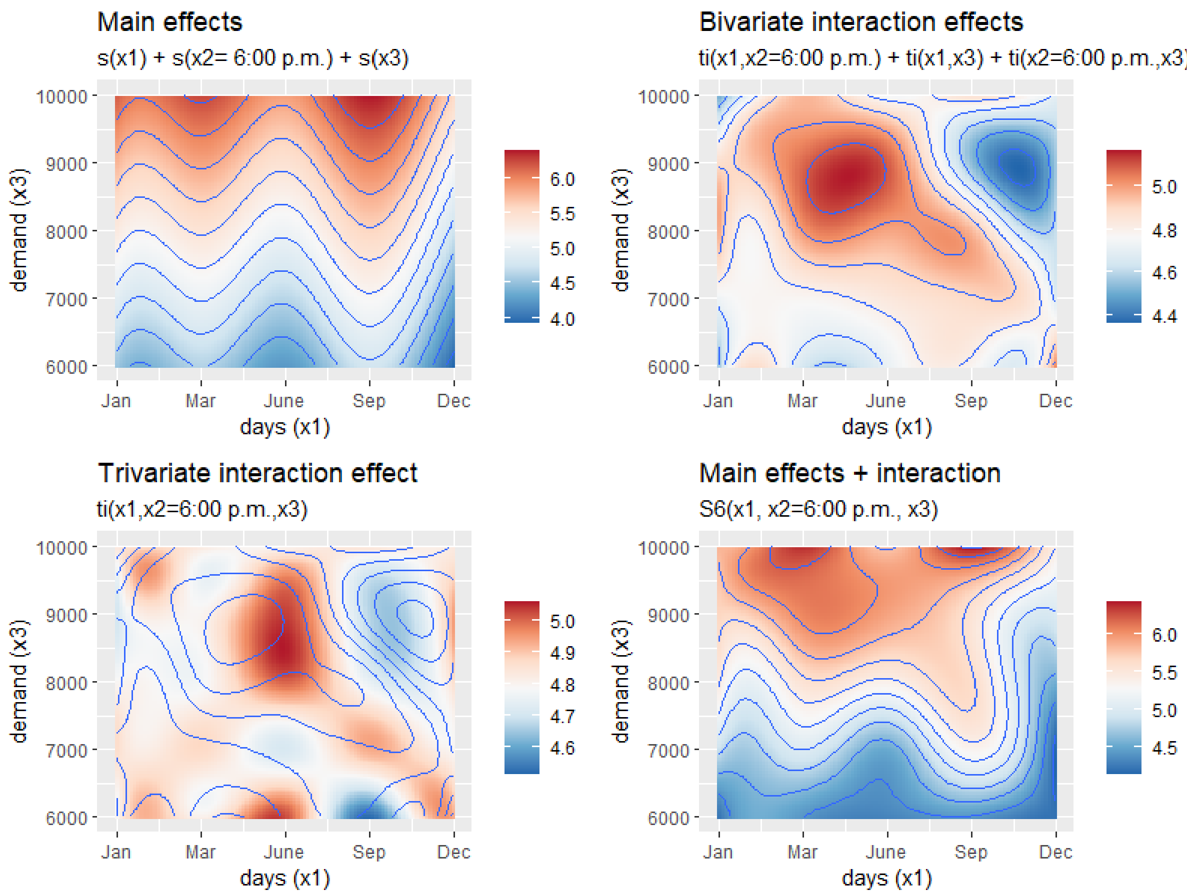
Table 8. Rejection rates of the test statistic $S_n^{CM^*}$

	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

Electricity prices. The table shows the sub-sample rejection rates of size n of the specification S6–S13.

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. Fig. 6.2 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 day-demand combinations have a different impact on the market wide price P_i. 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. Tab. 8 and Tab. IV in Supplement III). Overall, we conclude that for a thorough specification of the Australien 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.

Figure 6.2. 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 (x2) 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 (x1), x2 and total market demand (x3). The second and third panel illustrate the bivariate and trivariate interaction effects. The overall effect is depicted in the last panel.

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 non-linear way. To improve finite sample properties in the parametric model framework, we replace the non-parametric 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 (non-linear) 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, modelling 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 non-linear or even more complex functional forms of covariates are omnipresent in many applications.

References

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

Bach, S., G. Corneo, and V. Steiner (2009): "From bottom to top: The entire income distribution in Germany, 1992 – 2003," *Review of Income and Wealth*, 55, 303–330.

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

——— (2019b): "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*, 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., B. J. Reich, and H. Wang (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*.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly Journal of Economics*, 128, 967–1015.

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

Chao, S.-K., S. Volgushev, and G. Cheng (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., I. Fernández-Val, and B. Melly (2013): "Inference on counterfactual distributions," *Econometrica*, 81, 2205–2268.

Creedy, J. and P. E. Hart (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 New York.

Dudley, R. M. (2014): *Uniform Central Limit Theorems*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2 ed.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009): "Revisiting the German wage structure*," *The Quarterly Journal of Economics*, 124, 843–881.

Eilers, P. H., B. D. Marx, et al. (1996): "Flexible smoothing with B-splines and penalties," *Statistical Science*, 11, 89–121.

Escanciano, J. C. and S.-C. Goh (2014): "Specification analysis of linear quantile models," *Journal of Econometrics*, 178, 495–507.

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

Fasiolo, M., S. N. Wood, M. Zaffran, R. Nedellec, and Y. Goude (2020): "Fast calibrated additive quantile regression," *Journal of the American Statistical Association*, 1–11.

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

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

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

He, X. and P. Shi (1997): "Monotone B-Spline smoothing," *Journal of the American Statistical Association*, 93.

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

Ignatieva, K. and S. Trück (2016): "Modeling spot price dependence in Australian electricity markets with applications to risk management," *Computers & Operations Research*, 66, 415–433.

Klein, N., T. Kneib, S. Lang, A. Sohn, 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. and M. Weber (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. and G. Bassett Jr (1978): "Regression quantiles," *Econometrica*, 33–50.

Koenker, R. and Z. Xiao (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., Q. Li, and Z. Li (2020): "Nonparametric quantile regression estimation with mixed discrete and continuous Data," *Journal of Business & Economic Statistics*, 1–16.

Li, Q. and J. S. Racine (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., J. Meng, and Z. Fan (2015): "Simultaneous estimation of linear conditional quantiles with penalized splines," *Journal of Multivariate Analysis*, 141, 1–21.

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

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

Ng, P. and M. Maechler (2007): "A Fast and efficient implementation of qualitatively constrained quantile smoothing splines," *Statistical Modelling*, 7, 315–328.

Ng, P. T. and M. Maechler (2020): "COBS – Constrained B-splines (Sparse matrix based)," *R package version 1.3-4*.

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

Rothe, C. and D. Wied (2013): "Misspecification testing in a class of conditional distributional models," *Journal of the American Statistical Association*, 108, 314–324.

Shively, T. S. and M. S. Smith (2018): "Econometric modeling of regional electricity spot prices in the Australian market," *Energy Economics*, 74, 886–903.

Smith, M. S. and N. Klein (2020): "Bayesian inference for regression copulas," *Journal of Business & Economic Statistics*.

Soni, P., I. Dewan, and K. Jain (2012): "Nonparametric estimation of quantile density function," *Computational Statistics & Data Analysis*, 56, 3876–3886.

Statistisches Bundesamt (2012): *Periodensterbetafeln für Deutschland: Allgemeine Sterbetafeln, abgekürzte Sterbetafeln und Sterbetafeln*, Wiesbaden: Statistisches Bundesamt.

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

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

Takeuchi, I., Q. Le, T. Sears, and A. Smola (2006): "Nonparametric Quantile Estimation," *Journal of Machine Learning Research*, 7, 1231–1264.

Toraichi, K., K. Katagishi, I. Sekita, and R. Mori (1987): "Computational complexity of spline interpolation," *International Journal of Systems Science*, 18, 945–954.

Troster, V. and D. Wied (2021): "A specification test of dynamic conditional distributions," *Econometric Review*, 40, 109–127.

Vaart, A. W. v. d. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

Volgushev, S., S.-K. Chao, and G. Cheng (2019): "Distributed inference for quantile regression processes," *The Annals of Statistics*, 47, 1634–1662.

Wagner, G., J. Frick, and J. Schupp (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. and J. Wang (2010): "Distribution function estimation by constrained polynomial spline regression," *Journal of Nonparametric Statistics*, 22, 443–457.

SUPPLEMENTARY MATERIAL

Part I: Proofs and derivations.

Part II: Monte Carlo simulation study for S_n^{CM}, $S_n^{CM^*}$ and $S_n^{CM,S}$.

Part III: Additional figures and results to the empirical application on electricity prices of the Australian NEM of the manuscript.

I Proofs

I.1 Proof of Theorem 1

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

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

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

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

Let ℓ^∞ denote the set of all uniformly bounded real functions.

Lemma 1. *Assume Assumption 1 holds. For the processes (I.1) and (I.2) it holds under the null, that*

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

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

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

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

Corollary 2. *Let Assumption 1 i.)–iv.) be satisfied and* $\sqrt{n}\left(\hat{G}_n - G\right) \Rightarrow \tilde{\mathbb{G}}_2$ *in* $\ell^\infty(\Theta \times I_l)$ *for all* $l = 1, \dots, L$, *where* $\tilde{\mathbb{G}}_2$ *is a Gaussian process with a.s. uniformly continuous paths on* $\Theta \times I_l$, $l = 1, \dots, L$. *Further, we assume that the estimator* $\hat{\theta}_n(\tau)$ *is an approximate* Z*-estimator (2.10) for all* $\tau \in I_l$ *with* $l = 1, \dots, L$. *Then*

$$\sqrt{n}\left(\hat{\theta}_n(\cdot) - \theta_0(\cdot)\right) = -\dot{G}^{-1}_{\theta_0(\cdot),\cdot}\left[\sqrt{n}(\hat{G}_n - G)(\theta_0(\cdot), \cdot)\right] + o_P(1)$$
$$\Rightarrow -\dot{G}^{-1}_{\theta_0(\cdot),\cdot}\left[\tilde{\mathbb{G}}_2(\theta_0(\cdot), \cdot)\right] \in \ell^\infty(\mathcal{T}).$$

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

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

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

$$(\nu_n, \nu_n^0) \Rightarrow \mathbb{G} := (\mathbb{G}_1, \mathbb{G}_2) \text{ in } \ell^\infty(\mathcal{S} \times \mathcal{S}),$$

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

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

with $\mathbb{G}_2^+ := -\dot{F}(y \mid x, \theta_0) \left[\left[\dot{G}(\theta_0(\cdot), (\cdot) \right]^{-1} \tilde{\mathbb{G}}_2(\cdot) \right].$

Proof. Under either the null hypothesis or a fixed alternative, it follows by standard arguments from Lemma 1 and Corollary 2 that

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

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

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

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

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

given by

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

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

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

□

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

$$\begin{aligned} S_n^{CM} &= \sqrt{n} \int \left(\hat{F}_n(y, x) - \hat{F}_n(y, x, \hat{\theta}) \right) d\hat{F}_n(y, x) \\ &= \int \left(\nu_n(y, x) - \nu_n^0(y, x) \right)^2 dF(y, x) + \int \left(\nu_n(y, x) - \nu_n^0(y, x) \right)^2 d\left(\hat{F}_n(y, x) - F(y, x) \right). \end{aligned}$$

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

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

which claims the statement.

To show part $ii.$), we use the fact that under any fixed alternative $P(F(y,x) \neq F(y,x,\theta_0) > 0$ due to construction of the alternative hypothesis in (2.7). Thus,

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

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

I.2 Proof of Theorem 2

The proof is shown for $P(X,\tau) = P(X)$. In case of quantile dependent regressors, standard arguments those as given in the proof of Theorem 1 apply. To prove Theorem 2.i), we consider the parametric and the semi-parametric model with increasing dimension. The steps from the proof of Theorem 1 are applied analogously replacing $v_n(y,x)$ and $v_n^0(y,x)$ from (I.1) and (I.3) by $v_n^0 := a_n\left(\hat{F}_n(y,x,\hat{\theta}_n) - F(y,x,\theta_0)\right)$ and $v_n^{0,S} := a_n^*\left(\hat{F}_n^S(y,x,\hat{\theta}_n) - F(y,x,\theta_0)\right)$ with $a_n = \sqrt{n}/\|P(x)\|$ and $a_n^* = \sqrt{n}/\|B(x)\|$, respectively. By Theorem 1, Corollary 1 in Belloni et al. (2019b) and Theorem 1 we have that

$$\left(v_n^0, a_n(\hat{\theta}_n(\cdot) - \theta_0(\cdot))\right) \Rightarrow \left(\mathbb{G}_2(\cdot,\cdot), -\mathbb{G}_2^{S_M}\right),$$

where $\mathbb{G}_2^{S_M} := -\dot{G}^{-1}_{\theta_0(\cdot),\cdot}\left[\tilde{\mathbb{G}}_2^{S_M}(\theta_0(\cdot),\cdot)\right]$. Together with the Hadamard differentiablity in Assumption 1 and if

$$H(\tau_1,\tau_2,P(x)) := \lim_{n\to\infty} \|P(x)\|^{-2} P(x)^\top J_m^{-1}(\tau_1)\mathbb{E}[P(x)P(x)^\top] J_m^{-1}(\tau_2)P(x)(\tau_1 \wedge \tau_2 - \tau_1\tau_2)$$

exists for any $\tau_1, \tau_2 \in \mathcal{T}$, we know from Chao et al. (2017) Theorem 2.1 and Corollary 4.1 that for any fixed x and initial estimator $\hat{F}_n^{-1}(\cdot \mid x)$ the expression $a_n^*(\hat{F}_n(\cdot \mid x) - F_{Y|X}(\cdot \mid x)) \Rightarrow -f_{Y|X}(\cdot \mid x)\tilde{\mathbb{G}}_2^{S_M}$, where $\tilde{\mathbb{G}}_2^{S_M}$ is a centered Gaussian process with covariance function $H(\tau_1, \tau_2, P(x))$.

In order to show that $(v_n^0, v_n^{0,S}) \Rightarrow \mathbb{G}_S := (\mathbb{G}_2, \mathbb{G}_2^{S_M})$ we use the Hadamard differentiability of the mapping $\phi(A,B)[x^*] := \int A(\cdot, x^*)\mathbb{1}_{\{x^* \leq \cdot\}} dB(x^*)$ the functional delta method as stated. The continuous mapping theorem completes the proof. Part *ii.*) can be proved analogously to part *ii.*) of Theorem 2. Moreover,

$$Cov(\mathbb{G}_2(y_1, x_1), \mathbb{G}_2^{S_M}(y_2, x_2)) = \\ \lim_{n\to\infty} n\, Cov\left(\hat{F}_n(y_1, x_1, \hat{\theta}_n) - F(y_1, x_1, \theta_n), \hat{F}_n^{S_M}(y_2, x_2, \hat{\theta}_n) - F(y_2, x_2, \theta)\right),$$

where the true functional vector $\theta(\tau)$ depends on n for all $\tau \in \mathcal{T}$.

I.3 Proof of Theorem 3

In order to prove Theorem 3 we present the bootstrap version of Lemma 1 as an auxiliary result.

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

$$\begin{aligned} \nu_{n,B}(y,x) &:= \sqrt{n}\left(\hat{F}_{n,B}(y,x) - \hat{F}_n(y,x,\hat{\theta}_n)\right) \\ \nu_{n,B}^0(y,x) &:= \sqrt{n}\left(\hat{F}_{n,B}(y,x,\hat{\theta}_n) - \hat{F}_n(y,x,\hat{\theta}_n)\right). \end{aligned} \tag{I.4}$$

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

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

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

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

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

$$\begin{aligned} P(S_n^{CM} \leq \hat{c}_n(\alpha)) &= P(S_n^{CM} \leq \hat{c}_n(\alpha), S_n^{CM} \leq N(\epsilon)) + P(S_n^{CM} \leq \hat{c}_n(\alpha), S_n^{CM} > N(\epsilon)) \\ &\leq P(S_n^{CM} \leq N(\epsilon)) + P(S_n^{CM} > N(\epsilon)) \\ &\leq \varepsilon + o_P(1), \end{aligned}$$

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

II Monte Carlo Simulation Study

II.1 MC Simulation Study for $S_n^{CM,S}$

In this section, we show that our test $S_n^{CM,S}$ holds the size level and has considerable power properties using five different functional forms of complex interacting covariate effects, which we denote by DGP 13–17. Motivated by our second application, we use more flexible product tensor B-splines instead of product interactions using univariate B-splines to model the interacting covariate effects. We consider this approach reasonable for two reasons. First, our second application reveals that the covariates indeed interact in a very complex way. Second, we show empirically in this section that multivariate tensor product B-splines yield satisfactorily testing results. For this, we couple DGPs 13–17 with various model specifications, which we refer to as B1–B6 (for two covariates) and T1–T5 (for three covariates). Overall, these settings attempt to mimic the situation of interaction effects as seen in our second real data illustration on electricity prices in Sec. 6.2 of our manuscript and are based on tensor product B-splines. For a detailed MC simulation study investigating univariate product interacting covariate effects, we refer to II.2. Here, we demonstrate for two-dimensional functions that our test S_n^{CM*} holds the size level and has good power properties in case of product interacting covariate effects.

Data Generating Processes DGPs 13–17 contain two or three interacting covariates and are defined as follows:

$$\begin{aligned}
&\text{DGP 13:} \quad y_{13}(x_1, x_2) := 7 \cdot \sin(x_1 \cdot x_2) + x_1 + u \\
&\text{DGP 14:} \quad y_{14}(x_1, x_2) := \sin(x_1 \cdot x_2) + x_1 \cdot x_2^2 + z_1(x_1, x_2) + u \\
&\text{DGP 15:} \quad y_{15}(x_3, x_4) := 1 + 2x_3 + 4x_4 + 70\cos(x_3 \cdot x_4) + u \\
&\text{DGP 16:} \quad y_{16}(x_2, x_5) := x_2^2 \cdot x_5 + x_2 \cdot u + \cos(x_2 u) + z_1(x_2, x_5) \\
&\text{DGP 17:} \quad y_{17}(x_1, x_2, x_3) := x_1 + \sin(x_2) \cdot x_3 + z_2(x_1, x_2, x_3) + u
\end{aligned} \tag{II.1}$$

Above, let $x_1 \sim U(-4, -4)$, $x_2 \sim N(5, 1)$, $x_3, x_4 \sim U(0, 1)$, $x_5 \sim U(-10, 10)$, $z_1(x_1, x_2) \sim SN(x_1 + x_2^2, 2 + \sin(2x_1), x_1/4)$, $z_2(x_1, x_2, x_3) \sim SN(x_2 + x_3^2, 5 + \sin(x_1)x_3, x_3)$, $u \sim N(0, 1)$, where $U(\cdot, \dot{)}$, $N(\cdot, \cdot)$ and $SN(\cdot, \cdot, \cdot)$ denote uniform, Gaussian and skew normal distributions, respectively. To estimate the functional forms in DGPs 13–17, we use cubic P-splines.

Model Specifications To increase the readability, the notation is geared to the implementation in R, i.e. $s(\cdot, \tau)$ models the marginal P-spline and $ti(\cdot, \cdot, \tau)$ for the interaction effect at the quantile τ excluding the basis functions associated with the lower dimensional marginal effects of the marginal smooths. For the case of two covariates, we define the following specifications:

$$\text{B1:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2) := s(x_1, \tau)$$

$$\text{B2:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2) := s(x_1, \tau) + s(x_2, \tau)$$

$$\text{B3:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2) := s(x_1, \tau) + s(x_2, \tau) + ti(x_1, x_2, \tau)$$

$$\text{B4:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2) := \begin{cases} s(x_1, \tau) + s(x_2, \tau), & \text{if } 0.25 < \tau < 0.75 \\ s(x_1, \tau) + s(x_2, \tau) + ti(x_1, x_2, \tau), & \text{otherwise} \end{cases}$$

$$\text{B5:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2) := \begin{cases} s(x_1, \tau) + s(x_2, \tau), & \text{if } 0.25 < \tau \\ s(x_1, \tau) + s(x_2, \tau) + ti(x_1, x_2, \tau), & \text{otherwise} \end{cases}$$

$$\text{B6:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2) := \begin{cases} s(x_1, \tau) + s(x_2, \tau), & \text{if } \tau < 0.75 \\ s(x_1, \tau) + s(x_2, \tau) + ti(x_1, x_2, \tau), & \text{otherwise} \end{cases}$$

For the case of three covariates, let $s(x, \tau) := s(x_1, \tau) + s(x_2, \tau) + s(x_3, \tau)$. We define the following specifications:

$$\text{T1:} \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := s(x, \tau)$$

$$\text{T2}: \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := s(x, \tau) + ti(x_1, x_2, \tau) + ti(x_2, x_3, \tau)$$

$$\text{T3}: \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} s(x, \tau)\ , & \text{if } 0.25 < \tau < 0.75 \\ s(x, \tau) + ti(x_1, x_2, \tau) + ti(x_2, x_3, \tau)\ , & \text{otherwise} \end{cases}$$

$$\text{T4}: \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} s(x, \tau)\ , & \text{if } 0.25 < \tau \\ s(x, \tau) + ti(x_1, x_2, \tau) + ti(x_2, x_3, \tau)\ , & \text{otherwise} \end{cases}$$

$$\text{T5}: \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) := \begin{cases} s(x, \tau)\ , & \text{if } \tau < 0.75 \\ s(x, \tau) + ti(x_1, x_2, \tau) + ti(x_2, x_3, \tau)\ , & \text{otherwise} \end{cases}$$

Estimation and Further Settings The estimation is carried out in the R-package `qgam` by Fasiolo et al. (2020). To keep the computational costs for our MC simulation study in reasonable limits and to make the results comparable, we use cubic P-splines with second order difference penalty and set the number of knots to 5. We evaluate the test statistic for the quantiles $u \in \{0.02, 0.04, \ldots, 0.96, 0.98\}$. The number of overall replications is equal to 301 and the significance level is set to 0.05.

Table I. $S_n^{CM,S}$: Size and power for DGPs 13–15 and B1–B6

	DGP 13						DGP 14					
	B1	B2	**B3**	**B4**	B5	B6	B1	B2	**B3**	**B4**	B5	B6
n=500	1.000	0.050	**0.033**	**0.033**	0.058	0.033	1.000	0.106	**0.057**	**0.076**	0.0764	0.089
n=1000	1.000	0.088	**0.058**	**0.050**	0.045	0.067	1.000	0.103	**0.043**	**0.040**	0.0565	0.057
n=2000	1.000	0.150	**0.079**	**0.046**	0.070	0.121	1.000	0.237	**0.050**	**0.057**	0.1296	0.156
n=3000	1.000	0.392	**0.083**	**0.096**	0.187	0.292	1.000	0.445	**0.060**	**0.073**	0.1761	0.199
n=5000	1.000	0.655	**0.046**	**0.067**	0.241	0.492	1.000	0.694	**0.066**	**0.080**	0.2126	0.269
n=6000	1.000	0.867	**0.076**	**0.053**	0.279	0.613	1.000	0.764	**0.089**	**0.050**	0.1927	0.316
	DGP 15						Electricity Data					
	B1	B2	**B3**	B4	B5	B6	B1	B2	**B3**	B4	B5	B6
n=500	1.000	1.000	**0.027**	1.000	1.000	1.000	0.661	0.110	**0.086**	0.096	0.126	0.086
n=1000	1.000	1.000	**0.019**	1.000	1.000	1.000	0.924	0.099	**0.059**	0.076	0.086	0.077
n=2000	1.000	1.000	**0.039**	1.000	1.000	1.000	1.000	0.199	**0.063**	0.239	0.206	0.226
n=3000	1.000	1.000	**0.019**	1.000	1.000	1.000	1.000	0.329	**0.057**	0.435	0.336	0.422
n=5000	1.000	1.000	**0.029**	1.000	1.000	1.000	1.000	0.688	**0.040**	0.824	0.688	0.804
n=6000	1.000	1.000	**0.049**	1.000	1.000	1.000	1.000	0.864	**0.057**	0.917	0.794	0.920

MC Study. Shown are the size and power properties for the test statistic $S_n^{CM,S}$. The columns with bold numbers depict the size of the specification test $S_n^{CM,S}$. The remaining columns represent the power of the test.

Results Considering the bivariate case (cf. Tables I) we make four observations. First, the test $S_n^{CM,S}$ holds the size level for the DGPs 13–15 (cf. Tables I, bold columns). Second, generally, power properties depend on the degree of misspecification. In the case of a moderately misspecified model (the difference between $B4$ and $B5$ or $B6$ is only that the lower 25% or upper 75% quantile contains interaction effects), the test $S_n^{CM,S}$ shows reasonable power properties. The higher the degree of misspecification the higher the rejection rates. This is particularly evident on DGP 15, where the test always detects misspecification. Third, consistent with our theoretical investigations in Sec. 3, the rejection rate for misspecified models increases with increasing sample size. Fourth, omitted variable bias is always detected (cf. Tables I, column B1 in DGP 13–15). Considering the multivariate case

Table II. $S_n^{CM,S}$: Size and power for DGPs 16–17 and B2, T1–T5

	DGP 16						DGP 17					
	B2	T1	**T2**	**T3**	T4	T5	B2	T1	**T2**	**T3**	T4	T5
$n = 500$	0.476	0.179	**0.034**	**0.037**	0.186	0.033	0.691	0.073	**0.063**	**0.073**	0.060	0.076
$n = 1000$	0.754	0.332	**0.017**	**0.043**	0.303	0.040	0.940	0.083	**0.073**	**0.073**	0.083	0.073
$n = 2000$	0.898	0.472	**0.008**	**0.055**	0.458	0.063	0.998	0.083	**0.043**	**0.069**	0.080	0.070
$n = 3000$	0.984	0.780	**0.032**	**0.055**	0.764	0.063	1.000	0.149	**0.063**	**0.099**	0.163	0.123
$n = 5000$	1.000	0.852	**0.012**	**0.066**	0.835	0.070	1.000	0.179	**0.069**	**0.093**	0.163	0.096
$n = 6000$	1.000	0.878	**0.017**	**0.086**	0.889	0.104	1.000	0.183	**0.053**	**0.086**	0.179	0.089

MC Study. Shown are the size and power properties for the test statistic $S_n^{CM,S}$. The columns with bold numbers depict the size of the specification test $S_n^{CM,S}$. The remaining columns represent the power of the test.

with three covariates (cf. DGPs 16–17 and B2, T1–T5 in Tab. II), we make the following observations: First, the test $S_n^{CM,S}$ holds the size level (cf. Tab. II, bold columns). Second, as in the bivariate cases, power properties depend on the degree of misspecification. For a moderately misspecified model ($T4$ and $T5$ contain interaction effects only in the lower

25% or upper 75% quantile), the test shows reasonable power properties. The higher the degree of misspecification the higher the rejection rates (columns B2 and T1). Third, due to the curse of dimensions, however, the multivariate case with three covariates requires a larger number of observations n to obtain similar properties as the multivariate case with two covariates. Last, the omitted variable bias is sufficiently well detected for $n \geq 2000$ (cf. Tab. II, column B2).

II.2 Further Results from the MC Study using Product Interacting Covariates

In this section, we show that our test $S_n^{CM,S}$ holds the size level and has good power properties in case of product interacting covariates. For this, we consider the complex multivariate case with three covariates for the test statistic $S_n^{CM,S}$ (cf. DGPs 16–17 and B2, T1–T5 in Tab. III), where we replace the tensor interaction in T2–T5 by univariate product interactions, i.e. we replace ti by $s(x_1 \cdot x_2)$, $s(x_2 \cdot x_3)$ and $s(x_1 \cdot x_2 \cdot x_3)$, respectively. We make the following observations:

First, the test $S_n^{CM,S}$ holds the size level (cf. Tab. III, bold columns). Second, as in the tensor product cases in Sec. II.1, power properties depend on the degree of misspecification. For a moderately misspecified model ($T4$ and $T5$ contain interaction effects only in the lower 25% or upper 75% quantile), the test shows reasonable power properties. The higher the degree of misspecification the higher the rejection rates (cf. columns B2 and T1). Third, due to the curse of dimensions, however, the multivariate case with three covariates requires a large number of observations n to obtain a powerful testing procedure as discussed in Sec. II.1. Last, the omitted variable bias is sufficiently well detected for $n \geq 2000$ (cf. Tab. III, column B2).

Table III. Product interaction $S_n^{CM,S}$: Size and power for DGPs 16–17 and B2, T2–T5

	DGP 16						DGP 17					
	B2	T1	**T2**	**T3**	T4	T5	B2	T1	**T2**	**T3**	T4	T5
$n = 500$	0.194	0.043	**0.016**	**0.010**	0.054	0.010	0.623	0.041	**0.050**	**0.050**	0.040	0.054
$n = 1000$	0.461	0.107	**0.026**	**0.029**	0.131	0.030	0.884	0.038	**0.024**	**0.030**	0.031	0.030
$n = 2000$	0.854	0.273	**0.036**	**0.054**	0.287	0.060	0.999	0.059	**0.043**	**0.061**	0.061	0.059
$n = 3000$	0.984	0.673	**0.039**	**0.069**	0.666	0.067	1.000	0.088	**0.063**	**0.087**	0.091	0.089
$n = 5000$	0.999	0.900	**0.027**	**0.098**	0.886	0.126	1.000	0.128	**0.059**	**0.066**	0.131	0.067
$n = 6000$	1.000	0.927	**0.033**	**0.087**	0.924	0.146	1.000	0.108	**0.054**	**0.063**	0.113	0.077

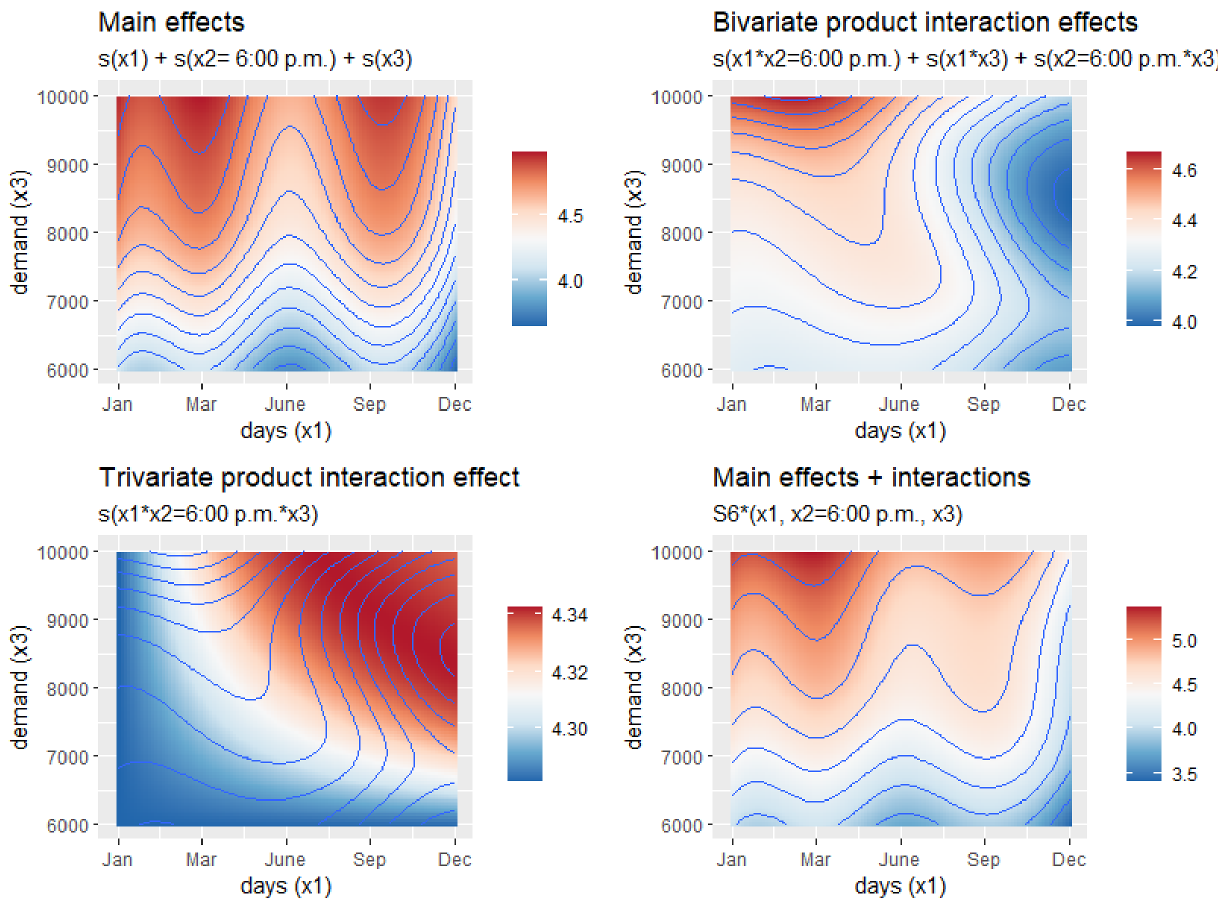
MC Study on product interactions. Shown are the size and power properties for the test statistic $S_n^{CM,S}$ using product interaction effects, i.e. the tensor product splines in T2–T5 $ti(x_1, x_2), ti(x_2, x_3)$ are replace by product interactions $s(x_1 \cdot x_2)$ and $s(x_2 \cdot x_3)$, respectively. The columns with bold numbers depict the size of the specification test $S_n^{CM,S}$. The remaining columns represent the power of the test.

III Further Results from Modelling Australian Electricity Prices

Tensor product interacting covariates In addition to Sec. 6.2 of our manuscript, Fig. II shows the decomposition of the main and interaction effects at the 10% quantile at 6:00 p.m. using specification S6. Since the contour lines in the second and third panel (upper right and lower left) 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 day-demand combinations have a different impact on the market wide price P_i.

Univariate product interacting covariates Fig. I shows the decomposition of the main and interaction effects at the 10% quantile at 6:00 p.m. using specification S6, where

Figure I. Estimated main and product interaction effects at the 90% quantile at 6:00 p.m.



Electricity prices. Figures depict the estimated effects of the three covariates on the 10% quantile of the Australian NEM hourly electricity price distribution for 2019. The time of the day is set to 06:00 p.m. The estimation of the conditional qf is based on the sum of the main effects and bi- and trivariate product interaction effects, i.e. $F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) = s(x_1, \tau) + s(x_2, \tau) + s(x_3, \tau) + s(x_1 \cdot x_2, \tau) + s(x_1 \cdot x_3, \tau) + s(x_2 \cdot x_3, \tau) + s(x_1 \cdot x_2 \cdot x_3, \tau)$. The first panel (upper left) shows the sum of the univariate main effects of days (x1), time of day (x2) and total market demand (x3). The second and third panel illustrate the bivariate and trivariate product interaction effects. The overall effect is depicted in the last panel (lower right).

the tensor product interaction effects are replaced by univariate interacting covariates, i.e. specification S6 is modified to

$$\text{S6}^*: \quad F^{-1}_{Y|X}(\tau \mid x_1, x_2, x_3) = s(x_1, \tau) + s(x_2, \tau) + s(x_3, \tau) + s(x_1 \cdot x_2, \tau) + s(x_1 \cdot x_3, \tau) + s(x_2 \cdot x_3, \tau) + s(x_1 \cdot x_2 \cdot x_3, \tau).$$

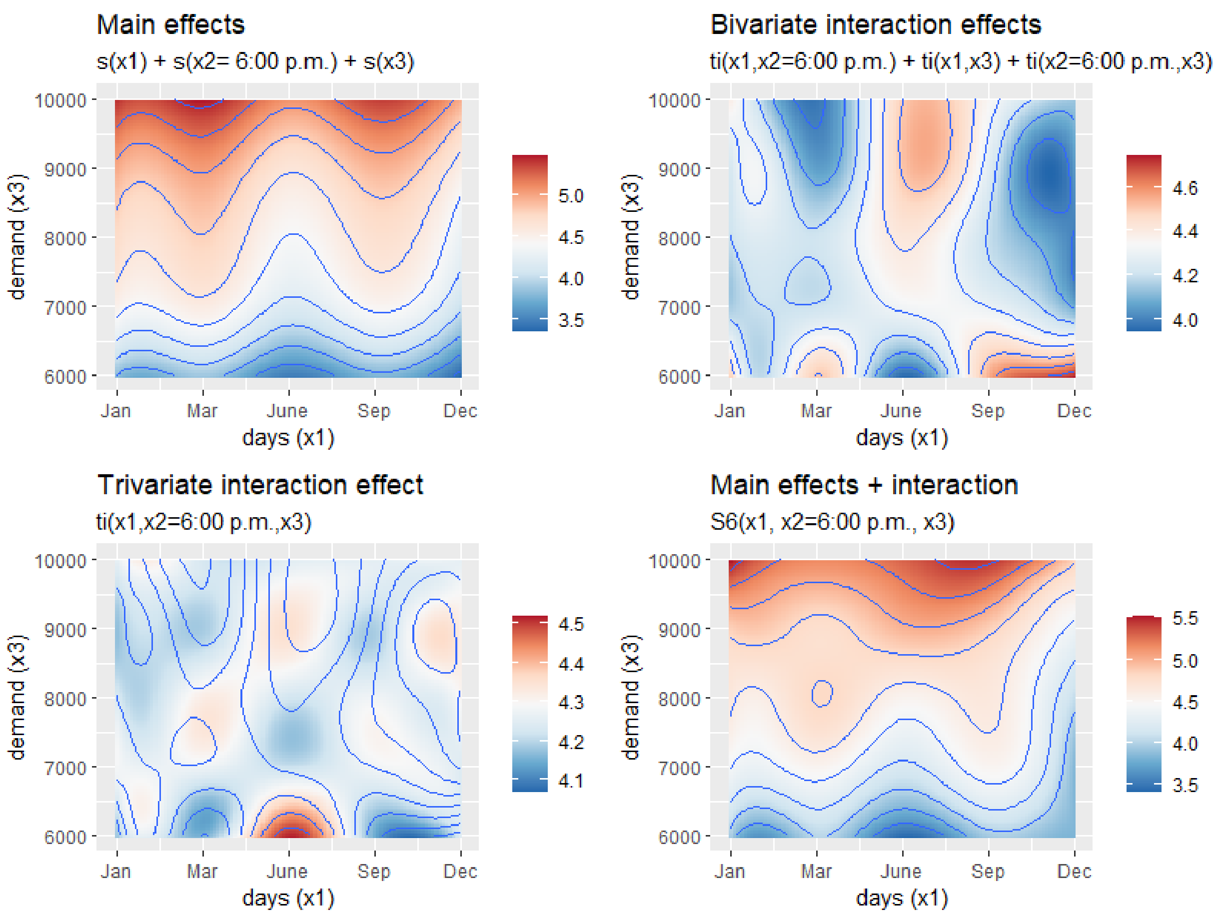
For the application of our test $S_n^{CM^*}$ with univariate product interacting effects to the Australian NEM, we consider a rolling window and set $n \in \{500, 1000, 2000\}$ and $\tau \in \{0.02, 0.04, \ldots, 0.98\}$. The number of sub-samples is 501 and the critical values are calculated at a significance level of 5%. To ensure comparability of rejection rates for different n and since we replaced the multivariate tensor by univariate 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 Tab. IV. For further details on the application we refer to Sec. 6.2.

Table IV. Product Interaction: Rejection rates of the test statistic $S_n^{CM^*}$

	S6	S7	S8	S9	S10	S11	S12	S13
n=500	0.058	0.072	0.068	0.058	0.072	0.056	0.062	0.081
n=1000	0.173	0.235	0.212	0.183	0.217	0.164	0.171	0.192
n=2000	0.646	0.874	0.773	0.661	0.739	0.670	0.685	0.784

Electricity prices. The table shows the sub-sample rejection rates of size n of the specification S6–S13, where the interaction effects modeled by tensor product splines ti are replaced by product interaction effects, i.e. $s(x_1 \cdot x_2)$, $s(x_1 \cdot x_3)$, $s(x_2 \cdot x_3)$ and $s(x_1 \cdot x_2 \cdot x_3)$.

Figure II. Estimated main and product interaction effects at the 10% quantile at 6:00 p.m.



Electricity prices. Figures depict the estimated effects of the three covariates on the 10% quantile of the Australian NEM hourly electricity price distribution for 2019. The time of the day 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 (x1), time of day (x2) and total market demand (x3), where $x2$ is set to 6:00 p.m. The second and third panel illustrate the bivariate and trivariate interaction effects. The overall effect is depicted in the last panel (lower right).