Checking the Adequacy of Quantile Regression Processes*

Shengtao Dai† Xiaojun Song‡
Boston College Peking University

Zixin Yang§
Shanghai University of Finance and Economics

Abstract

We develop a class of omnibus Cramér–von Mises tests for assessing the adequacy of parametric quantile regression models over a continuum of quantile levels. The proposed tests are based on a vector-weighted cumulative sum process that incorporates the gradient of the quantile regression objective function. To address the curse of dimensionality induced by multivariate covariates, we employ projection-based weight functions. An orthogonal projection onto the tangent space of nuisance parameters removes the effect of parameter estimation and yields a limiting null distribution that is invariant to the choice of estimator. This feature enables a simple multiplier bootstrap for critical values approximation. We establish the asymptotic properties of the tests under the null hypothesis, fixed alternatives, and sequences of local alternatives. Monte Carlo simulations and an empirical application illustrate the finite-sample performance of the proposed procedure.

Keywords: Specification tests, Multiplier bootstrap, Projection, Empirical processes

*The authors contributed equally to this work and are listed in alphabetical order.

†Department of Economics, Boston College, Chestnut Hill, MA 02467 USA. Email: shengtao.dai@bc.edu.

‡Guanghua School of Management, Peking University, Beijing, 100871, China. Email: sxj@gsm.pku.edu.cn.

§Corresponding author. School of Statistics and Data Science, Shanghai University of Finance and Economics, Shanghai, 200433, China. Email: yangzixin@mail.sufe.edu.cn.

1 Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression has emerged as a key tool in empirical economics for analyzing the entire conditional distribution of a response variable. In practice, parametric forms of quantile regression are often adopted to mitigate the curse of dimensionality and facilitate interpretation. The validity of such specifications, however, is paramount, as an inadequate model can lead to biased estimates and misleading inferences (Kim and White, 2003; Angrist et al., 2006). In this paper, we address the critical problem of testing the adequacy of parametric quantile regression process models. Formally, let Y be a response variable and X a vector of covariates. We consider testing the continuum of conditional probability restrictions given by

$$H_0 : \mathbb{P}\left(Y \leq m(X, \theta_0(\tau))|X\right) = \tau \text{ a.s. for some } \theta_0 \in \mathcal{B} \text{ and for all } \tau \in \mathcal{T}, \tag{1}$$

where $\mathcal{T}$ is a compact subinterval of $[0, 1]$, $\mathcal{B}$ is a family of uniformly bounded functions from $\mathcal{T}$ to $\Theta \subset \mathbb{R}^p$, and $\{m(\cdot, \theta) : \theta \in \Theta\}$ represents the prespecified parametric family of quantile functions. The alternative hypothesis is the negation of H_0.

The importance of this testing problem cannot be overstated. Reliable inference in quantile regression, whether for assessing wage inequality, value-at-risk in finance, or heterogeneous treatment effects, depends on the correct specification of the model (Engle and Manganelli, 2004; Machado and Mata, 2005). A test applied only to a single prespecified quantile, such as the median, provides an incomplete diagnostic, as misspecification may occur elsewhere in the conditional distribution. Consequently, developing powerful tests that are consistent against misspecification at any quantile in a continuum $\mathcal{T}$ is essential for rigorous empirical practice.

A rich literature has developed tests for parametric quantile models, yet the analysis has been mostly limited to a single, prespecified quantile (e.g., Zheng, 1998; Bierens and Ginther, 2001; Horowitz and Spokoiny, 2002; He and Zhu, 2003; Horowitz and Lee, 2009; Conde-Amboage et al., 2015; and Horvath et al., 2022). To the best of our knowledge, the only omnibus proposals for testing over a continuum of quantiles are those given by Escanciano and Velasco (2010) for dynamic models and Escanciano and Goh (2014) for linear models. These authors developed tests based on a cumulative sum (CUSUM) process of the quantile residuals. While pioneering, the tests are susceptible to low power when the dimension of the covariate X is high or even moderate. This is due to the nature of the weight function or integrating measure employed. Additionally, as Eubank and Hart (1993) and He and Zhu (2003) have noted, tests based on processes with unidimensional residuals can be insensitive to departures in the form of oscillation around zero. Recently, Feng et al. (2023) proposed a lack-of-fit test for a linear quantile regression process, extending the idea of Dong et al. (2019), which was designed for a fixed quantile. Their test is robust to the presence of multivariate covariates; however, its implementation can be challenging due to the necessity of a complex bootstrap procedure to obtain the critical values.

To overcome the above limitations, we propose a new class of lack-of-fit tests for parametric quantile regression process models that are powerful against a broad range of alternatives. Our approach is founded on a vector-weighted projection-based empirical process. A key and distinctive feature is that the process is marked not only by the quantile residual, but also by $\dot{m}(X, \theta(\tau))$, the gradient of the quantile function $m(X, \theta(\tau))$. This creates a process that directly incorporates the first-order condition (or "score") of the quantile regression estimator. This vector-weighting strategy makes the test more sensitive to misspecifications that violate these estimating equations, thereby improving its power (He and Zhu, 2003). It also improves the test's size performance in relatively high-dimensional contexts, as evidenced by our simulation studies.

Furthermore, to mitigate the curse of dimensionality associated with multivariate covariates, we adopt a class of projection-based weight functions to achieve dimension reduction. Finally, we employ an orthogonal projection to eliminate the "parameter estimation effect", a major technical hurdle that complicates the asymptotic properties of the test process. This projection renders the limiting null distribution invariant to the choice of estimators and facilitates a simple multiplier bootstrap procedure for computing the critical values.

Our work is most closely related to He and Zhu (2003) and Escanciano and Goh (2014), but it introduces critical innovations that distinguish it from both. Relative to He and Zhu (2003), our primary contribution is the extension from a single quantile to a continuum of quantiles, a generalization that is nontrivial due to the functional nature of the parameter $\theta(\cdot)$. We also introduce a projection device to eliminate the effect of parameter estimation. The most significant distinction from Escanciano and Goh (2014) lies in the construction of the underlying empirical process. While these authors base their test on a process marked by the quantile residual, our process is marked by its product with the gradient $\dot{m}(X, \theta(\tau))$. The incorporation of the gradient vector effectively enhances power, particularly against alternatives involving oscillations around 0, which are poorly detected by the standard CUSUM process. We also avoid using the standard indicator function to construct the CUSUM process, thereby alleviating the curse of dimensionality associated with multivariate covariates. Consequently, our test not only shares the desirable properties of Escanciano and Goh (2014), such as the orthogonal projection that enables a simple multiplier bootstrap, but also delivers superior power.

The remainder of this paper is organized as follows. Section 2 presents our testing framework and constructs the test statistics. Section 3 illustrates the asymptotic null distributions and the asymptotic power properties of our tests under the fixed and a sequence of local alternatives. Section 4 introduces a multiplier bootstrap procedure for the computation of critical values. Section 5 contains a set of Monte Carlo simulations and an empirical application on checking the validity of quantile regression models. Section 6 concludes. All proofs and additional simulation results are gathered into the Supplementary Appendix.

The following notation is adopted throughout this paper. For a generic matrix A, let $A^\top$, $\|A\|_1$, and $\|A\|$ denote the transpose, L_1 norm, and Frobenius norm of A, respectively. Denote $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. Let $\mathbb{1}(A)$ denote the indicator function for the event A, that is, $\mathbb{1}(A) = 1$ if A occurs, and zero otherwise.

2 Testing setup

2.1 A modified empirical process

To test the null hypothesis in (1), we adopt an integrated approach which is based on the observation that (1) can be equivalently characterized by an infinite number of parametric unconditional moment restrictions:

$$\mathbb{E}\left\{\left[\tau - \mathbb{1}\big(Y - m(X, \theta_0(\tau)) \le 0\big)\right] w(X, x)\right\} = 0 \text{ for all } x \in \Pi \text{ and } \tau \in \mathcal{T},$$

where $\{w(\cdot, x) : x \in \Pi\}$ is a properly chosen parametric family to make the above unconditional moment restrictions equivalent to (1). A commonly used weight function is $w(X, x) = \mathbb{1}(X \le x)$ with $\Pi = \mathbb{R}^d$; see, e.g., Koul and Stute (1999) and Whang (2006). Given a sample $\{(Y_i, X_i^\top)^\top : 1 \le i \le n\}$, we consider the residual-marked empirical process

$$R_n(x, \tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\tau - \mathbb{1}\left(Y_i \le m(X_i, \hat{\theta}_n(\tau))\right)\right) \mathbb{1}(X_i \le x) \tag{2}$$

indexed by $x \in \mathbb{R}^d$ and $\tau \in \mathcal{T}$. Here, $\hat{\theta}_n(\tau)$ denotes a $\sqrt{n}$-consistent estimator of $\theta_0(\tau)$. An obvious example is the regression τ-quantile estimator developed by Koenker and Bassett (1978). This estimator is defined as

$$\hat{\theta}_{n,KB}(\tau) = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \rho_\tau(Y_i - m(X_i, \theta)), \tag{3}$$

where $\rho_\tau(\epsilon) = \epsilon(\tau - \mathbb{1}(\epsilon \le 0))$. For the linear case where $m(X, \theta(\tau)) = X^\top \theta(\tau)$, the estimator $\hat{\theta}_{n,KB}(\tau)$ can be solved via linear programming (Koenker and d'Orey, 1987). For the generalized cases of nonlinear regression, an algorithm is given in Koenker and Park (1996). The null hypothesis is likely to hold when $R_n(x, \tau)$ is close to zero for almost all $(x^\top, \tau)^\top \in \mathbb{R}^d \times \mathcal{T}$.

The above test is straightforward. However, as noted by Eubank and Hart (1993), the test based on the unidimensional process (2) could be sensitive to certain alternatives but have low power against departures in the form of oscillation around zero. To avoid this problem, He and Zhu (2003) put forward a modified version of (2) that incorporates all components in the gradient equation. Specifically, let $\dot{m}(X, \theta) := \partial m(X, \theta)/\partial \theta$ be the gradient of $m(X, \theta)$

and denote

$$\psi_\tau(Y, X, \theta) := \dot{m}(X, \theta(\tau)) \left(\tau - \mathbb{1}\left(Y \leq m(X, \theta(\tau))\right)\right).$$

These authors proposed to build the test on the vector empirical process

$$\tilde{R}_n(x, \tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(Y_i, X_i, \hat{\theta}_n) \mathbb{1}\left(X_i \leq x\right). \tag{4}$$

Note that the residuals are only one component of the gradient equation. Hence, the process (4) is expected to be more sensitive to departures from the null hypothesis than (2). This idea has been widely adopted in the literature; for example, Qu (2008) constructed the subgradient test statistic in quantile regression to test for structural changes. Zhang et al. (2014) proposed a score function-based test for checking the jumping threshold effect in the threshold models.

The above tests use the indicator weight function $\mathbb{1}(X \leq x)$ to form the residual-marked empirical process. As mentioned by Bierens (1990), Bierens and Wang (2012), and Escanciano (2006a), the power of tests based on the indicator weight may increase slowly when the data-generating process gradually departs from the null model. This issue is exacerbated when the dimension of the covariates X is large or even moderate. To address this problem, we propose a new empirical process that uses weight functions $w(X, x)$ with a nice dimension-reduction property. We denote such a process as

$$\breve{R}_n(x, \tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(Y_i, X_i, \hat{\theta}_n) w(X_i, x), \tag{5}$$

where $w(X, x)$ represents one of the following two types of weight functions: (i) the exponential weight function $w(X, x) = \exp(\mathrm{i}x^\top X)$ proposed by Bierens (1982), where $\mathrm{i} = \sqrt{-1}$ denotes the imaginary unit, and (ii) the projection-based indicator weight function $w(X, x) = 1(\beta^\top X \leq u)$ by Escanciano (2006a), where $x = (\beta^\top, u)^\top \in \mathbb{S}^d \times [-\infty, +\infty]$, and $\mathbb{S}^d = \{\beta \in \mathbb{R}^d : \|\beta\| = 1\}$. Different families w yield different power properties of the associated tests, providing flexibility to the integrated methodology. We will discuss these weight functions in a more detailed way in Section 2.3.

The process (5) is expected to have better power properties when high-dimensional covariates are present than the process (2). However, implementing a test based on (5) is still not straightforward since the asymptotic null distribution of $\breve{R}_n$ depends on the estimator $\hat{\theta}_n$. Specifically, under some mild conditions, we can show that $\breve{R}_n$ admits the asymptotic decomposition

$$\breve{R}_n(x, \tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(Y_i, X_i, \theta_0) w(X_i, x)$$

$$- \mathbb{E}\left[f(m(X,\theta_0(\tau))|X)\dot{m}(X,\theta_0(\tau))\dot{m}^\top(X,\theta_0(\tau))w(X,x)\right]\sqrt{n}(\hat{\theta}_n(\tau)-\theta_0(\tau)) + o_p(1) \tag{6}$$

uniformly in $(x,\tau) \in \Pi \times \mathcal{T}$, where $f(y|X)$ denotes the conditional density function of Y given X, evaluated at $Y = y$. The second term on the right-hand side of (6) represents the estimation effect of $\theta_0(\tau)$, which is non-negligible since $\sqrt{n}\|\hat{\theta}_n(\tau) - \theta_0(\tau)\| = \mathcal{O}_p(1)$. The presence of this term complicates the limiting null distribution of $\breve{R}_n(x,\tau)$ and the approximation of the critical values. In practice, one may choose an estimator $\hat{\theta}_n(\tau)$ and construct a test based on (5) using that estimator. However, this method requires that the estimator admits a uniform asymptotic linear representation, and also necessitates a subsampling or an estimator-dependent bootstrap procedure to simulate the null distribution of the test statistic; see, e.g., Whang (2006) and Escanciano and Velasco (2010). These features make implementing the test based on (5) difficult.

Instead of proceeding with a particular estimator, we acknowledge our lack of knowledge of $\theta_0(\cdot)$ in testing for (1), which can be accounted for by an orthogonal projection of the weight function into the so-called tangent space of nuisance parameters at each fixed quantile. Specifically, we consider a transformation on the weight function $w(X,x)$ so that the transferred weight is "orthogonal" to the matrix $f(m(X,\theta_0(\tau))|X)\dot{m}(X,\theta_0(\tau))\dot{m}^\top(X,\theta_0(\tau))$.[1] As such, the expectation component in the second term of (6) would be zero, which means that the estimation effect is eliminated. To fix the idea, we consider the infeasible projection operator

$$\mathcal{P}_\tau w(X,x) = w(X,x)I_p - G^\top(x,\theta_0(\tau))\Delta^{-1}(\theta_0(\tau))g(X,\theta_0(\tau)), \tag{7}$$

where I_p is the identity matrix of dimension p,

$$g(X,\theta_0(\tau)) = f(m(X,\theta_0(\tau))|X)\dot{m}(X,\theta_0(\tau))\dot{m}^\top(X,\theta_0(\tau)), \tag{8}$$

and

$$\Delta(\theta_0(\tau)) = \mathbb{E}[g(X,\theta_0(\tau))g^\top(X,\theta_0(\tau))],\ G(x,\theta_0(\tau)) = \mathbb{E}[g(X,\theta_0(\tau))w(X,x)]. \tag{9}$$

We can see that $G^\top(x,\theta_0(\tau))\Delta^{-1}(\theta_0(\tau))g(X,\theta_0(\tau))$ is the best linear predictor of $w(X,x)$ given $g(X,\theta_0(\tau))$, and thus, for all x and τ,

$$\mathbb{E}\left[\mathcal{P}_\tau w(X,x)g^\top(X,\theta_0(\tau))\right] = G^\top(x,\theta_0(\tau)) - G^\top(x,\theta_0(\tau))\Delta^{-1}(\theta_0(\tau))\Delta(\theta_0(\tau)) = 0.$$

This implies that with the projected weight function, the second term on the right-hand side

[1] This idea was first proposed by Neyman (1959) in the context of parametric models, and subsequently adapted to the context of quantile regression model checking by Escanciano and Goh (2014). However, these authors considered only a scalar process.

of (6) is eliminated. Hence, the resulting test is free from the estimation effect of $\theta_0(\cdot)$.

Therefore, we turn to a vector-weighted projection-based empirical process

$$\hat{R}_n(x,\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{P}_{n,\tau} w(X_i, x) \psi_\tau(Y_i, X_i, \hat{\theta}_n), \tag{10}$$

where $\mathcal{P}_{n,\tau} w(X,x)$ is the sample analog of $\mathcal{P}_\tau w(X,x)$ defined as

$$\mathcal{P}_{n,\tau} w(X_i, x) = w(X_i, x) I_p - \hat{G}_n^\top(x, \hat{\theta}_n(\tau)) \hat{\Delta}_n^{-1}(\hat{\theta}_n(\tau)) \hat{g}_n(X_i, \hat{\theta}_n(\tau)), \tag{11}$$

where

$$\hat{g}_n(X_i, \hat{\theta}_n(\tau)) = \hat{f}_h(m(X_i, \hat{\theta}_n(\tau))|X_i) \dot{m}(X_i, \hat{\theta}_n(\tau)) \dot{m}^\top(X_i, \hat{\theta}_n(\tau)),$$

and

$$\hat{\Delta}_n(\hat{\theta}_n(\tau)) = \frac{1}{n} \sum_{i=1}^{n} \hat{g}_n(X_i, \hat{\theta}_n(\tau)) \hat{g}_n^\top(X_i, \hat{\theta}_n(\tau)), \quad \hat{G}_n(x, \hat{\theta}_n(\tau)) = \frac{1}{n} \sum_{i=1}^{n} \hat{g}_n(X_i, \hat{\theta}_n(\tau)) w(X_i, x).$$

Here, $\hat{f}_h(m(X_i, \hat{\theta}_n(\tau))|X_i)$ denotes a consistent estimator of $f(m(X_i, \theta_0(\tau))|X_i)$, which will be discussed in Section 2.2 below in details. As shown in Theorem 3.1 below, the projection-based empirical process (10) satisfies

$$\sup_{(x,\tau)\in\Pi\times\mathcal{T}} \left| \hat{R}_n(x,\tau) - R_{n0}(x,\tau) \right| = o_p(1),$$

where

$$R_{n0}(x,\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{P}_\tau w(X_i, x) \psi_\tau(Y_i, X_i, \theta_0). \tag{12}$$

This implies that the process $\hat{R}_n(x,\tau)$ is asymptotically invariant to the choice of $\hat{\theta}_n(\tau)$. Such property enables a convenient multiplier bootstrap procedure for computing the critical values, see Section 4 for more details.

Remark 2.1 *In the above discussion, we use the full gradient $\dot{m}(X, \theta_0(\tau))$ to construct the empirical process. It is also possible to use only a subvector of $\dot{m}(X, \theta_0(\tau))$. Denote $\dot{m}_s(X, \theta_0(\tau))$ the subvector, whose dimension is p_s with $1 \le p_s \le p$. In this case, similar to (7), the infeasible projection operator is given by*

$$\mathcal{P}_{\tau,s} w(X, x) = w(X, x) I_{p_s} - G_s^\top(x, \theta_0(\tau)) \Delta_s^{-1}(\theta_0(\tau)) g_s(X, \theta_0(\tau)),$$

where $g_s(X, \theta_0(\tau)) = f(m(X, \theta_0(\tau))|X) \dot{m}(X, \theta_0(\tau)) \dot{m}_s^\top(X, \theta_0(\tau))$, and $\Delta_s(\theta_0(\tau))$ and $G_s(x, \theta_0(\tau))$ are defined as in (9) except that $g(X, \theta_0(\tau))$ is replaced by $g_s(X, \theta_0(\tau))$. Similar *to* (11)*,* $\mathcal{P}_{n,\tau,s}w(X,x)$*, the feasible version of* $\mathcal{P}_{\tau,s}w(X,x)$*, can be obtained. Then the corresponding test is based on the following process*

$$\hat{R}_{n,s}(x,\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{P}_{n,\tau,s}w(X_i,x)\dot{m}_s(X_i,\hat{\theta}_n(\tau)) \left(\tau - \mathbb{1}\left(Y_i \leq m(X_i,\hat{\theta}_n(\tau))\right)\right),$$

which may direct power to specific alternatives. For simplicity, we do not focus on $\hat{R}_{n,s}(x,\tau)$*.*

2.2 Conditional density estimation

We consider a consistent estimator for $f(m(X,\theta(\tau))|X)$ following the idea in Escanciano and Goh (2019). Instead of directly estimating the conditional density function, these authors exploited the behavior of fitted conditional U_j-quantiles $m(X,\theta_0(U_j))$ over a range of quantiles $U_1,\ldots,U_m$ which are uniformly distributed on $\mathcal{T}$. For illustration, let $\mathcal{T} = [\tau_1,\tau_2] \subset (0,1)$. Note the following identity of the distribution function

$$F(y|X_i) = \tau_1 + \int_{\tau_1}^{\tau_2} \mathbb{1}\left(F^{-1}(\tau|X_i) \leq y\right) d\tau, \;\; \text{for } \tau_1 \leq F(y|X_i) \leq \tau_2.$$

This suggests using a smooth approximation to the indicator function and then differentiating to obtain an estimator of the density $f(y|X_i)$. The key quantity is given by

$$\frac{|\mathcal{T}|}{h}\mathbb{E}\left[K\left(\frac{y - F^{-1}(U|X_i)}{h}\right)\Big|X_i\right], \tag{13}$$

where $K(\cdot)$ is a smoothing kernel satisfying some mild regularity conditions, $U|X_i$ is uniformly distributed on $[\tau_1,\tau_2]$, and $|\mathcal{T}| = \tau_2 - \tau_1$. We expect (13) to be a good approximation of $f(y|X_i)$ as $h \to 0$, where $h > 0$ is a scalar smoothing parameter. To avoid numerical integration, we approximate the integral by a finite sum with m terms. Both m and h are allowed to depend on the sample size n, with $m \to \infty$ and $h \to 0$ as $n \to \infty$.

Under the null hypothesis, the conditional quantile function $F^{-1}(U|X) = m(X,\theta_0(U))$. Based on the above discussion, our final estimator of $f(m(X_i,\theta_0(\tau))|X_i)$ is

$$\hat{f}_h(m(X_i,\hat{\theta}_n(\tau))|X_i) = \frac{|\mathcal{T}|}{mh} \sum_{r=1}^{m} K\left(\frac{m(X_i,\hat{\theta}_n(\tau_r)) - m(X_i,\hat{\theta}_n(\tau))}{h}\right), \tag{14}$$

where $\{\tau_j\}_{j=1}^{m}$ is a random sample from the uniform distribution in $\mathcal{T}$. Unlike the classical conditional density estimators in Rosenblatt (1969), the above estimator converges to the true conditional density at a rate free of dimensionality. Moreover, it does not have random denominators, which makes it robust to the choice of bandwidth.

2.3 Test statistics

Following (1), our test statistics are based on a distance from the residual-marked process (10) to zero, i.e., on a norm of $\hat{R}_n$. A popular norm is the Cramér–von Mises (CvM) functional $\int_{\Pi\times\mathcal{T}} \hat{R}_n(x,\tau)\hat{R}_n^\top(x,\tau)\psi(dx,d\tau)$, where ψ denotes some integrating measures on $\Pi\times\mathcal{T}$. Since $\hat{R}_n$ is in a vector form, directly using the above functional is inconvenient. We consider test statistics of the form

$$CvM_n^{\mathrm{E}} = \lambda_{\max}\left(\int_{\Pi\times\mathcal{T}} \hat{R}_n(x,\tau)\hat{R}_n^\top(x,\tau)\psi(dx,d\tau)\right) \tag{15}$$

and

$$CvM_n^{\mathrm{F}} = \mathrm{trace}\left(\int_{\Pi\times\mathcal{T}} \hat{R}_n(x,\tau)\hat{R}_n^\top(x,\tau)\psi(dx,d\tau)\right), \tag{16}$$

where $\lambda_{\max}(A)$ and $\mathrm{trace}(A)$ denote the largest eigenvalue and trace of matrix A, respectively.

A key element for the CvM statistics is the choice of the integrating measure $\psi(\cdot)$. A desired choice of $\psi(\cdot)$ delivers a closed-form expression of the test statistic and eases the computation. For integration with respect to τ, we consider a uniform discrete distribution over a grid of $\mathcal{T}$ as in Escanciano and Velasco (2010) and Escanciano and Goh (2014). This leads to test statistics of form $\Gamma\big(m^{-1}\sum_{r=1}^m \int_\Pi \hat{R}_n(x,\tau_r)\hat{R}_n^\top(x,\tau_r)\Psi(dx)\big)$, where $\{\tau_r\}_{r=1}^m$ are m equidistributed points on $\mathcal{T}$ with $m\to\infty$ as $n\to\infty$, and $\Gamma(\cdot)$ denotes some matrix norm (e.g., the spectral or nuclear norm). For integration with respect to x, we consider below two types of weight functions $w(X,x)$, present the corresponding integrating measure $\Psi(dx)$, and derive the closed form of the test statistics.

First, we consider the linear indicator weight function $w(X,x) = \mathbb{1}\left(\beta^\top X \le u\right)$, where $x = (\beta^\top, u)^\top$, and $\beta \in \mathbb{S}^d := \left\{\beta\in\mathbb{R}^d : \|\beta\| = 1\right\}$. Compared with the standard indicator weight $w(X,x) = \mathbb{1}(X\le x)$, this weight uses one-dimensional projections of X and is less sensitive to the dimension d (Escanciano, 2006a). A choice of the integrating measure in this case is $\Psi^{\mathrm{lin}}(dx) := F_{n,\beta}(du)d\beta$, where $F_{n,\beta}(u)$ is the empirical distribution function of the projected regressors $\{\beta^\top X_i\}_{i=1}^n$ and $d\beta$ the uniform density on $\mathbb{S}^d$. The integration is taken over $\Pi = \mathbb{S}^d \times [-\infty,+\infty]$. We show in Lemma A.9 in the Appendix that CvM test statistics equipped with such weight functions have the closed-form expressions as

$$CvM_n^{\mathrm{lin,E}} = \lambda_{\max}(\mathcal{B}_n) \quad \text{and} \quad CvM_n^{\mathrm{lin,F}} = \mathrm{trace}(\mathcal{B}_n), \tag{17}$$

where $\mathcal{B}_n$ is a p-dimensional matrix defined as

$$\mathcal{B}_n := \frac{1}{m}\sum_{r=1}^m \int_\Pi \hat{R}_n(x,\tau_r)\hat{R}_n^\top(x,\tau_r)F_{n,\beta}(du)d\beta = \frac{1}{mn^2}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n B_{ij}\xi_i(\tau_r)\xi_j^\top(\tau_r).$$

Here, $\xi_i(\tau_r)$ denotes the "residuals" of regressing $\psi_{\tau_r}(Y_i, X_i, \hat{\theta}_n)$ on $\hat{g}_n(X_i, \hat{\theta}_n(\tau_r))$:

$$\xi_i(\tau_r) = \psi_{\tau_r}(Y_i, X_i, \hat{\theta}_n) - \hat{g}_n^\top(X_i, \hat{\theta}_n(\tau_r))\hat{\Delta}_n^{-1}(\hat{\theta}_n(\tau_r))\frac{1}{n}\sum_{s=1}^{n}\hat{g}_n(X_s, \hat{\theta}_n(\tau_r))\psi_{\tau_r}(Y_s, X_s, \hat{\theta}_n), \quad (18)$$

and

$$B_{ij} = \sum_{v=1}^{n}\left(\pi - \arccos\left(\frac{(X_i - X_v)^\top(X_j - X_v)}{\|X_i - X_v\|\|X_j - X_v\|}\right)\right)\frac{\pi^{\frac{d}{2}-1}}{\Gamma(\frac{d}{2}+1)}. \quad (19)$$

Note that the first component on the right-hand side is the complementary angle between vectors $X_i - X_v$ and $X_j - X_v$ measured in radians.

Another frequently used weight function is the exponential function $w(X, x) = \exp(\mathrm{i}x^\top X)$, where $x \in \Pi$, a compact subset of $\mathbb{R}^d$ containing the origin. This type of weight function was initially proposed by Bierens (1982) and has since gained widespread popularity in integrated-approach-based model specification tests (Escanciano and Velasco, 2010). Commonly used integrating measures for exponential weight function include the Gaussian-type measure $\Psi^{\text{gau}}(dx) = \exp(-\|x\|^2/2)dx$, and the Laplace-type measure $\Psi^{\text{lap}}(dx) = \exp(-\|x\|_1)dx$. We show in the Appendix that for Gaussian-type integrating measures, the CvM statistics have a closed form

$$CvM_n^{\text{gau,E}} = \lambda_{\max}(\mathcal{U}_n) \text{ and } CvM_n^{\text{gau,F}} = \text{trace}(\mathcal{U}_n), \quad (20)$$

where

$$\mathcal{U}_n := \frac{1}{m}\sum_{r=1}^{m}\int_{\Pi}\hat{R}_n(x,\tau_r)\hat{R}_n^\top(x,\tau_r)\Psi^{\text{gau}}(dx) = \frac{1}{mn^2}\sum_{r=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n}U_{ij}\xi_i(\tau_r)\xi_j^\top(\tau_r),$$

with $\xi_i(\tau_r)$ defined in (18) and $U_{ij} = \exp\left(-\|X_i - X_j\|^2/2\right)$. For the CvM statistics with a Laplace-type integrating measure, we have

$$CvM_n^{\text{lap,E}} = \lambda_{\max}(\mathcal{L}_n) \text{ and } CvM_n^{\text{lap,F}} = \text{trace}(\mathcal{L}_n), \quad (21)$$

where

$$\mathcal{L}_n := \frac{1}{m}\sum_{r=1}^{m}\int_{\Pi}\hat{R}_n(x,\tau_r)\hat{R}_n^\top(x,\tau_r)\Psi^{\text{lap}}(dx) = \frac{1}{mn^2}\sum_{r=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n}L_{ij}\xi_i(\tau_r)\xi_j^\top(\tau_r)$$

with $L_{ij} = 1/(1+\|X_i - X_j\|^2)$.[2] Our test rejects H_0 if the realized values of (15) or (16) appear in the right tail of its asymptotic null distributions, which is developed below in Section 3.2.

[2]In practice, we compute the CvM test statistics by integrating in the whole $\mathbb{R}^d$ space. Strictly speaking, our present theory does not allow this treatment, but our theory can be easily adapted, see e.g. the Hilbert space approach in Escanciano (2006b), to allow for such definition of the CvM test.

3 Large sample properties

We now investigate the limiting behavior of $\hat{R}_n$ under the null hypothesis, the fixed alternative, and sequences of local alternatives that approach H_0 in an appropriate sense. The asymptotic distributions of the CvM test statistics can be established by a combination of weak convergence of the process (10) and the continuous mapping theorem (CMT).

3.1 Regularity conditions

To derive asymptotic properties, we consider the following notations and assumptions. For a generic function $g : \mathcal{Z} \to \mathbb{R}$, define $\|g\|_{\mathcal{Z}} := \sup_{z\in\mathcal{Z}} |g(z)|$. We study the weak convergence of $\hat{R}_n(\cdot,\cdot)$ as an element of $l^\infty(\mathscr{A})$, the space of real-valued functions from $\mathscr{A}$ to $\mathbb{R}^p$ with each element uniformly bounded on $\mathscr{A}$, where $\mathscr{A} := \Pi \times \mathcal{T}$. The space $l^\infty(\mathscr{A})$ is equipped with the supremum norm $\|\cdot\|_{\mathscr{A}}$. Let "$\Rightarrow$" denote weak convergence on $(l^\infty(\mathscr{A}), \mathcal{B}_{d_\infty})$ in the sense of Hoffmann–Jørgensen, where $\mathcal{B}_{d_\infty}$ is the corresponding Borel σ-algebra; see, e.g., Definition 1.3.3 of Van Der Vaart and Wellner (2023). Let $\mathcal{X}$ denote the support of X, and let $\mathcal{Y} := \{m(x, \theta_0(\tau)) : x \in \mathcal{X}, \tau \in \mathcal{T}\}$.

Assumption 3.1 *(i) $\{(Y_i, X_i^\top)^\top\}_{i=1}^n$ is a sequence of independently and identically distributed (i.i.d.) random variables, and $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$. (ii) $|\mathcal{T}| > 0$.*

Assumption 3.2 *(i) The conditional density $f(y|x)$ is uniformly bounded from above and below on $y \in \mathcal{Y}$ and $x \in \mathcal{X}$; and (ii) $\partial f(y|x)/\partial y$ is uniformly bounded with respect to $y \in \mathcal{Y}$.*

Assumption 3.3 *The function $m(\mathbf{x}, \boldsymbol{\theta})$ satisfies the following: (i) For all $\mathbf{x} \in \mathcal{X}$, $m(\mathbf{x}, \cdot)$ is twice continuously differentiable in $\boldsymbol{\theta}$ on Θ. (ii) There exist functions $M(\cdot) : \mathcal{X} \to \mathbb{R}$ and $N(\cdot) : \mathcal{X} \to \mathbb{R}$ such that $\sup_{\theta\in\Theta} \|\dot{m}(x, \theta)\| \le M(x)$, and $\|\dot{m}(x, \theta_1) - \dot{m}(x, \theta_2)\| \le N(x)\|\theta_1 - \theta_2\|$ for all $x \in \mathcal{X}$, $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$. In addition, $\mathbb{E}|M(X)|^6 < \infty$ and $\mathbb{E}|N(X)|^4 < \infty$. (iii) For all $\tau \in \mathcal{T}$, $\Delta(\theta(\tau)) = \mathbb{E}[g(X, \theta(\tau))g^\top(X, \theta(\tau))]$ is nonsingular in a neighborhood of $\theta(\tau) = \theta_0(\tau)$.*

Assumption 3.4 *(i) The parameter space Θ is compact in $\mathbb{R}^p$, and the true parameter $\theta_0(\tau)$ belongs to its interior for all $\tau \in \mathcal{T}$. (ii) A consistent estimator $\hat{\theta}_n(\tau)$ of $\theta_0(\tau)$ is available such that $\sup_{\tau\in\mathcal{T}} \|\hat{\theta}_n(\tau) - \theta_0(\tau)\| = \mathcal{O}_p(n^{-1/2})$.*

Assumption 3.5 *(i) The kernel $K(\cdot)$ is of bounded variation and $\sup_u |K(u)| < \infty$; (ii) $K(\cdot)$ satisfies a Lipschitz condition on $\mathbb{R}$; and (iii) $\int_{-\infty}^{\infty} K(u)du = 1$, $\int_{-\infty}^{\infty} uK(u)du = 0$, $\int_{-\infty}^{\infty} u^2K(u)du = \mu_2$, and $\int_{-\infty}^{\infty}(K(u))^2du = \nu_2$ for some $\mu_2, \nu_2 \in (0, \infty)$.*

Assumption 3.6 *The bandwidth h satisfies $\mathbb{P}(a_m \le h \le b_m) \to 1$ for deterministic sequences of positive numbers a_m and b_m such that $ma_m/\log m \to \infty$ and $b_m \to 0$ as $n \to \infty$.*

Assumption 3.7 *$\psi(\cdot,\cdot)$ is absolutely continuous with respect to the Lebesgue measure on $\Pi \times \mathcal{T}$.*

Assumption 3.1 is standard in the model checking literature. It excludes the case where the quantile set $\mathcal{T}$ is a singleton, as this would render the conditional density estimator $\hat{f}_h$ in (14) invalid. Assumption 3.2 is used to prove the asymptotic tightness of the process $\hat{R}_n$. It also plays a key role in showing the continuity of $\theta_0(\tau)$ with respect to τ on $\mathcal{T}$, which is essential for the proof of asymptotic equicontinuity of R_{n0}. Assumption 3.3 concerns the smoothness of the parametric model $m(x, \theta(\tau))$. It is classical in inferences about nonlinear models, see, e.g., He and Zhu (2003). Assumption 3.4 imposes regularity conditions on the parameter space and the estimator $\hat{\theta}_n$. The uniform consistency condition is satisfied by the regression quantile estimator $\hat{\theta}_{n,KB}$, as shown by Mukherjee (1999) and Angrist et al. (2006), among others. Finally, Assumptions 3.5 and 3.6 allow for various smoothing kernels with data-driven bandwidths, see Escanciano et al. (2014) for a detailed discussion. Assumption 3.7 is imposed for the consistency of the CvM test statistics.

3.2 Asymptotic null distribution

Our first result concerns the asymptotic behavior of the empirical process (10) and its related functionals under the null hypothesis.

Theorem 3.1 *Suppose that Assumptions 3.1–3.6 hold. Under the null hypothesis H_0,*

$$\hat{R}_n \Rightarrow R_\infty,$$

where R_∞ is a Gaussian process with zero mean and covariance function

$$K_\infty(v_1, v_2) = (\tau_1 \wedge \tau_2 - \tau_1 \cdot \tau_2)\, \mathbb{E}\left[\mathcal{P}_{\tau_1} w(X, x_1)\dot{m}(X, \theta_0(\tau_1))\dot{m}^\top(X, \theta_0(\tau_2))\mathcal{P}_{\tau_2}^\top w(X, x_2)\right]$$

with $v_1 = (x_1^\top, \tau_1)^\top$ and $v_2 = (x_2^\top, \tau_2)^\top$ representing two generic elements of $\Pi \times \mathcal{T}$.

From Theorem 3.1, we see that the limiting distribution of $\hat{R}_n$ does not depend on $\hat{\theta}_n$ or the procedure used to obtain $\hat{\theta}_n$. This renders the proposed test more analytically tractable than the test based on the non-projection process (5). The proof of the theorem proceeds in two steps. First, we show that the process $\hat{R}_n$ is asymptotically equivalent, in the supremum norm, to R_{n0} in (12). Then, we establish the weak convergence of R_{n0} under the null. Details are provided in Appendix B.

The above convergence result has important implications on the asymptotic null distribution of the CvM test statistics defined in Section 2.3. For example, we have $CvM_n^{\text{lin,E}} \xrightarrow{d} \lambda_{\max}(|\mathcal{T}|^{-1} \int_{\Pi\times\mathcal{T}} R_\infty(x,\tau)R_\infty^\top(x,\tau))F_\beta(du)d\beta d\tau)$, as a consequence of Theorem 3.1, the CMT, and the Glivenko–Cantelli Theorem. Here, $F_\beta(u)$ is the cumulative distribution function of $\beta^\top X$. The justification for substituting the empirical distributions of $\{\beta^\top X_i\}_{i=1}^n$ and $\{\tau_j\}_{j=1}^m$ with the limiting distributions follows from the arguments in Chang (1990, Lemma 3.1). The CvM test statistic with the Frobenius norm can be similarly established.

3.3 Asymptotic power

In this subsection, we investigate the power properties of the CvM tests. The first theorem shows that these tests are consistent against all fixed alternatives provided that a mild condition is satisfied.

Assumption 3.8 *Under the alternative hypothesis H_1, there exists a $\theta^* \in \mathcal{B}$ such that* $\sup_{\tau\in\mathcal{T}} \|\hat{\theta}_n(\tau) - \theta^*(\tau)\| = o_p(1)$.

See Kim and White (2003) for conditions on $\hat{\theta}_{n,KB}$ to satisfy Assumption 3.8; see also Section 3 in Angrist et al. (2006).

Under the fixed alternative, the conditional density estimator $\hat{f}_h(y|X)$ in (14) is not a consistent estimator for $f(y|X)$. To characterize its asymptotic property, we introduce the following notations. For a vector $x \in \mathbb{R}^d$, let $f^*(\cdot|x)$ denote the density function of the random variable $m(x, \theta^*(U))$, where U is uniformly distributed on $\mathcal{T}$. Note that under H_1, $f^*(y|X)$ is generally different from the true conditional density function of Y given X, namely, $f(y|X)$. Let $g^*(X, \theta(\tau)) = f^*(m(X, \theta(\tau))|X)\dot{m}(X, \theta(\tau))\dot{m}^\top(X, \theta(\tau))$, $\Delta^*(\theta(\tau)) = \mathbb{E}\big[g^*(X, \theta(\tau))g^{*\top}(X, \theta(\tau))\big]$, and $G^*(x, \theta(\tau)) = \mathbb{E}\big[g^*(X, \theta(\tau))w(X, x)\big]$.

Theorem 3.2 *Under the alternative hypothesis H_1 and Assumptions 3.1–3.6 and 3.8,*

$$\sup_{(x,\tau)\in\Pi\times\mathcal{T}} \left| \frac{\hat{R}_n(x,\tau)}{\sqrt{n}} - \mathbb{E}\left[\mathcal{P}_\tau^* w(X,x)\psi_\tau(Y,X,\theta^*)\right] \right| = o_p(1), \tag{22}$$

where $\mathcal{P}_\tau^* w(X,x) := w(X,x)I_p - G^{*\top}(x,\theta^*(\tau))\Delta^{*-1}(\theta^*(\tau))g^*(X,\theta^*(\tau))$.

A consequence of Theorem 3.2 and the continuous mapping theorem is that

$$\int_{\Pi\times\mathcal{T}} \left\| n^{-1/2}\hat{R}_n(x,\tau) \right\|^2 \psi(dx, d\tau) \xrightarrow{p} \int_{\Pi\times\mathcal{T}} \left\| \mathbb{E}\left[\mathcal{P}_\tau^* w(X,x)\psi_\tau(Y,X,\theta^*)\right] \right\|^2 \psi(dx, d\tau) > 0,$$

provided that Assumption 3.7 holds and $\|\mathbb{E}[\mathcal{P}_\tau^* w(X,x)\psi_\tau(Y,X,\theta^*)]\|$ is different from zero in a subset with positive Lebesgue measure on $\Pi \times \mathcal{T}$. In such a case, the CvM test statistic will diverge to infinity under any fixed alternative, and the test will be consistent against all directions in the alternative hypothesis.

Next, we derive the asymptotic distribution of $\hat{R}_n$ under a certain sequence of local alternatives converging to the null at a parametric rate of $n^{-1/2}$. Specifically, we consider the data-generating process for the sequence of local alternatives given by

$$H_{1n} : \mathbb{E}\left[\mathbb{1}\left(Y \le m(X,\theta_0(\tau))\right) - \tau \middle| X\right] = \frac{d(X,\tau)}{\sqrt{n}} \text{ a.s. for some } \theta_0 \in \mathcal{B} \text{ and all } \tau \in \mathcal{T}, \tag{23}$$

where the function $d(\cdot, \tau) : \mathbb{R}^d \to \mathbb{R}$ satisfies the following assumption.

Assumption 3.9 *The function $d(X,\cdot)$ is such that $\mathbb{E}[\sup_{\tau\in\mathcal{T}}|d(X,\tau)|] < \infty$. There exists a square integrable random variable $\mathcal{C}(X)$ such that for all τ_1, $\tau_2 \in \mathcal{T}$, $|d(X,\tau_1) - d(X,\tau_2)| \leq \mathcal{C}(X)|\tau_1 - \tau_2|$ a.s.*

Theorem 3.3 *Suppose that Assumptions 3.1–3.6 and 3.9 hold. Then, under the sequence of local alternatives H_{1n}, we have*

$$\hat{R}_n \Rightarrow R_\infty + D,$$

where R_∞ is defined the same as in Theorem 3.1, and $D(x,\tau) := \mathbb{E}\left[\mathcal{P}_\tau w(X,x)\dot{m}(X,\theta_0(\tau))d(X,\tau)\right]$ is a deterministic shift term.

Theorem 3.3 implies that the proposed test possesses non-trivial power against local alternatives (23) in most cases. A situation where our tests cannot lack local power is when the direction $d(X,\tau)$ is collinear with the score vector $f(m(X,\theta_0(\tau))|X)\dot{m}(X,\theta_0(\tau))$, i.e., $d(X,\tau) = f(m(X,\theta_0(\tau))|X)\dot{m}^\top(X,\theta_0(\tau))a$ a.s. for some vector $a \in \mathbb{R}^p$. Under this case, the local shift $D(x,\tau)$ is equal to zero, so that the derivation from the null cannot be detected.

Remark 3.1 *Note that the lack of power against the directions that are parallel to the score is not a serious limitation. In fact, all ICM tests have trivial local power against these alternatives. Once we acknowledge the lack of power in these directions, we may obtain more power against other alternatives; see more detailed discussions in Strasser (1990), Escanciano (2009), and Sant'Anna and Song (2019). As a result, our tests outperform the non-projection-based tests in most cases. In practice, an advantage of our test is that it allows a simple multiplier bootstrap to simulate its null distribution, which is another convenience provided by the projection device.*

4 Multiplier bootstrap procedure

As shown in Theorem 3.1, the limiting distribution of the test statistics depends in a complex manner on the underlying data-generating process (DGP). Therefore, the critical values for the test statistics based on continuous functionals of $\hat{R}$ cannot be tabulated in general. In this section, we address this problem using a multiplier bootstrap. This particular resampling scheme has the practical advantage of not requiring the estimation of nuisance parameters at each bootstrap replication. Furthermore, it is fully data-driven and does not utilize any pre-selected tuning parameters.

It is also noteworthy that when testing the correct specification of a quantile regression model over a continuum of quantiles, other commonly used bootstrap procedures, such as the wild bootstrap and the parametric bootstrap, are rendered invalid by the quantile-dependent nature of the newly generated samples. We will expand on this subject in greater detail subsequently, highlighting the efficacy of the multiplier bootstrap.

4.1 Validity of multiplier bootstrap

We propose to approximate the asymptotic behavior of $\hat{R}_n(x,\tau)$ by that of

$$\hat{R}_n^*(x,\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \mathcal{P}_{n,\tau} w(X_i, x) \psi_\tau(Y_i, X_i, \hat{\theta}_n), \tag{24}$$

where $\{V_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with mean zero, unit variance, and bounded support. They are also independent of the original sample $\{(Y_i, X_i^\top)^\top\}_{i=1}^n$. A popular example is i.i.d.Bernoulli variates $\{V_i\}$ with $\mathbb{P}(V = 1-\kappa) = \kappa/\sqrt{5}$ and $\mathbb{P}(V=\kappa) = 1-\kappa/\sqrt{5}$, where $\kappa = (\sqrt{5}+1)/2$, see Mammen (1993).

With $\hat{R}_n^*$ at hand, the bootstrap version of our test statistics is given by

$$CvM_n^* := \Gamma\left(\int_{\Pi\times\mathcal{T}} \hat{R}_n^*(x,\tau) \hat{R}_n^{*\top}(x,\tau) \psi(dx, d\tau) \right), \tag{25}$$

where Γ denotes the spectral or Frobenius norm. For example, the bootstrap version of the test statistics (17) is constructed as $CvM_n^{*,\text{lin,E}} = \lambda_{\max}(\mathcal{B}_n^*)$ and $CvM_n^{*,\text{lin,F}} = \text{trace}(\mathcal{B}_n^*)$, where

$$\mathcal{B}_n^* := \frac{1}{m} \sum_{r=1}^{m} \int_{\Pi} \hat{R}_n^*(x,\tau_r) \hat{R}_n^{*\top}(x,\tau_r) F_{n,\beta}(du) d\beta = \frac{1}{mn^2} \sum_{r=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} B_{ij} \xi_i^*(\tau_r) \xi_j^{*\top}(\tau_r)$$

and

$$\xi_i^*(\tau_r) = \psi_{\tau_r}(Y_i, X_i, \hat{\theta}_n) V_i - \hat{g}_n^\top(X_i, \hat{\theta}_n(\tau_r)) \hat{\Delta}^{-1}(\hat{\theta}_n(\tau_r)) \frac{1}{n} \sum_{s=1}^{n} \hat{g}_n(X_s, \hat{\theta}_n(\tau_r)) \psi_{\tau_r}(Y_s, X_s, \hat{\theta}_n) V_s.$$

Bootstrap version of the CvM test statistics with exponential weight function can be obtained similarly, simply by replacing B_{ij} with U_{ij} or L_{ij} defined in (20) and (21).

The bootstrap empirical distribution of CvM_n^*, i.e., $\hat{F}_n^*(t \,|\, \{(Y_i, X_i)\}_{i=1}^n) = \mathbb{P}(CvM_n^* \leq t \,|\, \{(Y_i, X_i)\}_{i=1}^n)$ is shown to be a consistent estimate of the asymptotic null distribution function of the original test statistic CvM_n given in (15) and (16). Hence, the null hypothesis will be rejected at the α-level of significance when $CvM_n \geq c_{n,\alpha}^*$, where $c_{n,\alpha}^*$ is such that $\hat{F}_n^*(c_{n,\alpha}^* | \{(Y_i, X_i)\}_{i=1}^n) = 1-\alpha$. We can also use bootstrap p-values in this context. In this case, the null hypothesis could be rejected if $p_n^* < \alpha$, where $p_n^* = \mathbb{P}\big(CvM_n^* \geq CvM_n | \{(Y_i, X_i)\}_{i=1}^n\big)$.

Denote by "$\overset{*}{\Rightarrow}$ in probability" the weak convergence in probability under the bootstrap law, i.e., conditional on the original sample $\{(Y_i, X_i^\top)\}_{i=1}^n$, see Van Der Vaart and Wellner (2023).

Theorem 4.1 *Suppose Assumptions 3.1–3.9 hold.*

(i) Under H_0 and H_{1n}, $\hat{R}_n^* \overset{*}{\Rightarrow} R_\infty$ in probability, where R_∞ is defined in Theorem 3.1. For the CvM test statistics, we have $CvM_n^* \xrightarrow{d} \Gamma\big(\int_{\Pi\times\mathcal{T}} R_\infty(x,\tau) R_\infty^\top(x,\tau) \psi(dx, d\tau)\big)$ in probability.

(ii) Under H_1, $\hat{R}_n^* \overset{*}{\Rightarrow} R_\infty^1$ in probability, where R_∞^1 is a Gaussian process with mean zero and

covariance function

$$K^1_\infty(v_1, v_2) = \mathbb{E}\big[\mathcal{P}^*_{\tau_1} w(X, x_1)\psi_{\tau_1}(Y, X, \theta^*)\psi^\top_{\tau_2}(Y, X, \theta^*)\mathcal{P}^{*\top}_{\tau_2} w(X, x_2)\big],$$

and $\mathcal{P}^*_\tau w(X, x)$ *is defined in Theorem 3.2. For the CvM test statistics, we have* $CvM^*_n \xrightarrow{d} \Gamma\big(\int_{\Pi\times\mathcal{T}} R^1_\infty(x, \tau)(R^1_\infty(x, \tau))^\top \psi(dx, d\tau)\big)$ *in probability.*

Theorem 4.1(i) implies that under the null, both the original CvM_n statistic and the bootstrap CvM^*_n statistic converge in distribution to $\Gamma\big(\int_{\Pi\times\mathcal{T}} R_\infty(x, \tau)R^\top_\infty(x, \tau)\psi(dx, d\tau)\big)$. Hence, our test based on the bootstrap p-value would yield an asymptotically correct level. On the other hand, under the alternative H_1, the CvM_n statistic diverges to infinity in probability by Theorem 3.2, whereas CvM^*_n is stochastically bounded conditional on the original sample by Theorem 4.1(ii). It follows that the test based on the bootstrap p-values is consistent against all such fixed alternatives. Therefore, the multiplier bootstrap test is asymptotically valid. Finally, under H_{1n}, the bootstrap statistic CvM^*_n still converges to the asymptotic null process, while the original statistic CvM_n converges to the asymptotic null process coupled with the deterministic drift $D(x, \tau)$. This difference guarantees our test to distinguish H_{1n} from H_0, as long as the drift $D(x, \tau)$ is not equal to zero for some (x, τ).

4.2 Invalidity of wild bootstrap

In this subsection, we explain why the standard wild bootstrap, where $Y^*_i(\tau)$ are resampled independently for each τ, is not valid for simulating critical values in the context of a continuum of quantiles.

Fix a quantile level τ. After estimating the quantile model and obtaining fitted conditional quantiles $\hat{q}(X_i, \tau)$ and residuals $\hat{u}_i(\tau) = Y_i - \hat{q}(X_i, \tau)$, the usual wild bootstrap for that τ generates pseudo-outcomes by multiplying residuals with i.i.d. mean-zero, variance-one multipliers:

$$Y^*_i(\tau) \;=\; \hat{q}(X_i, \tau) + \omega_i\, \hat{u}_i(\tau), \qquad i = 1, \dots, n,$$

where $\{\omega_i\}$ are iid with $\mathbb{E}(\omega_i) = 0,\ Var(\omega_i) = 1$ (typical choices are Rademacher, normal, or Mammen's two-point distribution). One then refits the quantile estimator (or, for some tests) on $\{(Y^*_i(\tau), X_i)\}$ and computes the bootstrap analog of the test statistic. This method works well for a fixed τ because the multiplier resamples preserve the first-order properties of the model (Escanciano, 2006a).

However, when applying this to a continuum of quantiles $\tau \in \mathcal{T}$, resampling $Y^*_i(\tau)$ independently for each τ_r introduces a crucial problem. The resampling is no longer consistent across quantiles, meaning that the generated paths $\{Y^*_i(\tau_r)\}$ may not be monotone in τ, and more importantly, the dependence structure across τ is broken. Specifically, the joint behavior of the process $\{\hat{R}_n(x, \tau)\}_{\tau\in\mathcal{T}}$ is distorted because the random components $\omega_{i,r}$ are independently drawn for each τ_r, which alters the correlation structure across τ. The resulting bootstrap statistic, expressed as

$$CvM_n^* = \frac{1}{m}\sum_{r=1}^{m}\int_{\Pi}\left\|\hat{R}_n^*(x,\tau_r)\right\|^2\psi(dx),$$

is not asymptotically equivalent to the original statistic, which leads to inflated rejection rates under the null hypothesis. Thus, this approach fails to preserve the process's global properties across the entire continuum of quantiles.

In contrast, the multiplier bootstrap applied to the summands, in which the same multiplier V_i is used across all quantiles, preserves the joint structure and yields consistent critical values. Therefore, we rely on the multiplier bootstrap method, which avoids the dependence distortion caused by independent resampling at each quantile.

5 Simulation and empirical study

5.1 Comparison to the tests without projection

In this section, we examine the finite-sample performance of the proposed tests through Monte Carlo experiments. We compare our omnibus test in Section 2.3 with two related tests. The first is CvM tests based on the nonprojected process $\breve{R}_n$ in (5), that is, $CvM_n^{\text{np}} = \Gamma\big(\int_{\Pi\times\mathcal{T}}\breve{R}_n(x,\tau)\breve{R}_n^\top(x,\tau)\psi(dx,d\tau)\big)$, where $\Gamma(\cdot)$ denotes either the spectral or nuclear (trace) norm. Such test statistics admit closed-form expressions which can be derived similarly to those in Section 2.3. For example, when $\breve{R}_n$ is equipped with the indicator weight function, the test statistics have the form

$$CvM_n^{\text{lin,E,np}} = \lambda_{\max}(\mathcal{B}_n^{\text{np}}) \text{ and } CvM_n^{\text{lin,F,np}} = \text{trace}(\mathcal{B}_n^{\text{np}}), \tag{26}$$

where $\mathcal{B}_n^{\text{np}}$ is a p-dimensional matrix defined as

$$\mathcal{B}_n^{\text{np}} = \frac{1}{mn^2}\sum_{r=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n}B_{ij}\psi_{\tau_r}(Y_i,X_i,\hat{\theta}_n)\psi_{\tau_r}^\top(Y_j,X_j,\hat{\theta}_n)$$

and B_{ij} is given by (19). The closed form of nonprojected CvM tests with an exponential weight function can be similarly derived. The only difference is that we replace the weight B_{ij} with U_{ij} and L_{ij} for the Gaussian and Laplace integrating measure, respectively.

The asymptotic null distribution of CvM_n^{np} depends on the choice of $\hat{\theta}_n(\cdot)$. To simulate the critical values, we consider a multiplier bootstrap similar to He and Zhu (2003), which exploits the asymptotic linear representation of $\hat{\theta}_n(\cdot)$. Specifically, we take $\hat{\theta}_n(\cdot)$ as the regression quantile estimator $\hat{\theta}_{n,KB}(\cdot)$ in (3). Using the asymptotic linear representation of this type of estimator, we have the bootstrap version of $\breve{R}_n$ as

$$\breve{R}_n^*(x,\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n V_i\big(\tau - \mathbb{1}\big(Y_i \le m(X_i,\hat{\theta}_n(\tau))\big)\big)\big(w(X_i,x)I_p - \hat{\mathbb{W}}_n\big)\dot{m}(X_i,\hat{\theta}_n(\tau)), \tag{27}$$

where $\hat{\mathbb{W}}_n := \hat{G}_n(x,\hat{\theta}_n(\tau))^\top\big(n^{-1}\sum_{i=1}^n \hat{g}_n(X_i,\hat{\theta}_n(\tau))\big)^{-1}$ and $\hat{G}_n(x,\hat{\theta}_n(\tau))$ is defined below (11). With $\breve{R}_n^*$ at hands, the bootstrap version of the test statistic CvM_n^{np} is simply given by $CvM_n^{*,\text{np}} := \Gamma\big(\int_{\Pi\times\mathcal{T}}\breve{R}_n^*(x,\tau)\breve{R}_n^{*\top}(x,\tau)\psi(dx,d\tau)\big)$. Equipped with the linear indicator weight function, such a quantity admits the closed-form expression as

$$CvM_n^{*,\text{lin,E,np}} = \lambda_{\max}(\mathcal{B}_n^{*,\text{np}}) \text{ and } CvM_n^{*,\text{lin,F,np}} = \text{trace}(\mathcal{B}_n^{*,\text{np}}),$$

where

$$\mathcal{B}_n^{*,\text{np}} = \frac{1}{mn^2}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n B_{ij}\xi_i^{*,\text{np}}(\tau_r)\xi_j^{*,\text{np}\top}(\tau_r), \tag{28}$$

and

$$\xi_i^{*,\text{np}}(\tau_r) = V_i\psi_{\tau_r}(Y_i,X_i,\hat{\theta}_n) - \hat{g}_n^\top(X_i,\hat{\theta}_n(\tau_r))\left(\frac{1}{n}\sum_{i=1}^n\hat{g}_n(X_i,\hat{\theta}_n(\tau_r))\right)^{-1}\frac{1}{n}\sum_{s=1}^n V_s\psi_{\tau_r}(Y_s,X_s,\hat{\theta}_n).$$

The bootstrap test statistics with an exponential weight function can be derived analogously. The null hypothesis is rejected at α significance level if $CvM_n^{\text{np}} \ge c_{n,\alpha}^{*,\text{np}}$, where $c_{n,\alpha}^{*,\text{np}} = \inf\{c_\alpha \in [0,+\infty) : \lim_{n\to\infty}\mathbb{P}(CvM_n^{*,\text{np}} > c_\alpha|\{(Y_i,X_i)\}_{i=1}^n) = \alpha\}$.

The second type of test that we compare our test with is the CvM test without the gradient component in the residual-marked process. Such a test is based on a unidimensional empirical process

$$\bar{R}_n(x,\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\big(\tau - \mathbb{1}(Y_i \le m(X_i,\hat{\theta}_n(\tau)))\big)\hat{\mathcal{P}}_{n,\tau}w(X_i,x), \tag{29}$$

where

$$\hat{\mathcal{P}}_{n,\tau}w(X_i,x) = w(X_i,x) - s_n^\top(X_i,\hat{\theta}_n(\tau))\hat{\Xi}_n^{-1}(\hat{\theta}_n(\tau))\hat{S}_n(x,\hat{\theta}_n(\tau)),$$

with $s_n(X_i,\hat{\theta}_n(\tau)) := \hat{f}_h(m(X_i,\hat{\theta}_n(\tau))|X_i)\dot{m}(X_i,\hat{\theta}_n(\tau))$,

$$\hat{\Xi}_n(\hat{\theta}_n(\tau)) = \frac{1}{n}\sum_{i=1}^n s_n(X_i,\hat{\theta}_n(\tau))s_n^\top(X_i,\hat{\theta}_n(\tau)), \text{ and } \hat{S}_n(x,\hat{\theta}_n(\tau)) = \frac{1}{n}\sum_{i=1}^n s_n(X_i,\hat{\theta}_n(\tau))w(X_i,x).$$

The CvM test statistic based on (29) is given by

$$CvM_n^{\mathrm{ng}} = \int_{\Pi\times\mathcal{T}} |\bar{R}_n(x,\tau)|^2\psi(dx,d\tau) = \frac{1}{mn^2}\sum_{r=1}^{m}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\xi_i^{\mathrm{ng}}(\tau_r)\xi_j^{\mathrm{ng}}(\tau_r),$$

where W_{ij} can be B_{ij}, U_{ij} or L_{ij}, and

$$\begin{aligned}\xi_i^{\mathrm{ng}}(\tau_r) = &\tau_r - \mathbb{1}(Y_i \le m(X_i,\hat{\theta}_n(\tau_r))) - s_n^\top(X_i,\hat{\theta}_n(\tau_r))\hat{\Xi}_n^{-1}(\hat{\theta}_n(\tau)) \\ &\times \frac{1}{n}\sum_{\ell=1}^{n} s_n(X_\ell,\hat{\theta}_n(\tau_r))\big(\tau_r - \mathbb{1}(Y_\ell \le m(X_\ell,\hat{\theta}_n(\tau_r)))\big).\end{aligned}$$

The multiplier bootstrap version of CvM_n^{ng} can be derived similarly to CvM_n. The details are omitted here for brevity.

5.2 Numerical evidence

This section presents the results of a series of Monte Carlo experiments to evaluate the finite-sample performance of the CvM tests. The performance of the proposed test, which is based on the empirical process $\hat{R}_n$ in (10), is compared to the same test based on the unidimensional process $\bar{R}_n$ in (29), and the non-projection process $\breve{R}_n$ in (5). For each case, we consider three types of weight functions and integrating measures: (i) the linear indicator weight $w(X_i,x) = \mathbb{1}(\beta^\top X_i \le u)$ with integrating measure $\Psi^{\mathrm{lin}}(\cdot)$, (ii) the exponential weight $w(X_i,x) = \exp(\mathrm{i}x^\top X_i)$ with integrating measure $\Psi^{\mathrm{gau}}(\cdot)$, and (iii) the exponential weight with integrating measure $\Psi^{\mathrm{lap}}(\cdot)$. For gradient CUSUM tests, we consider two types of functionals: one based on the nuclear norm and the other on the spectral norm. The critical values of the tests are simulated by the multiplier bootstrap described in Sections 4.1 and 5.1.

Our first exercise concerns testing the correct specification of a linear quantile regression model. The DGPs are given by

- DGP 1: $Y_i = 1 + X_{1i} + X_{2i} + \epsilon_i$,
- DGP 2: $Y_i = 1 + X_{1i} + X_{2i} + e_i$,
- DGP 3: $Y_i = 1 + X_{1i} + X_{2i} + 0.25(X_{1i}^2 + X_{2i}^2 + X_{1i} * X_{2i}) + \epsilon_i$,
- DGP 4: $Y_i = 1 + X_{1i} + X_{2i} + \exp(-0.5(1 + X_{1i} + X_{2i})) + \epsilon_i$,
- DGP 5: $Y_i = 1 + X_{1i} + X_{2i} + \sqrt{|X_{1i} + X_{2i}|}\,\mathrm{sgn}(X_{1i} + X_{2i}) + \epsilon_i$,
- DGP 6: $Y_i = 1 + X_{1i} + X_{2i} + \sin(X_{1i} + X_{2i}) + \epsilon_i$,

where X_{1i} and X_{2i} are taken to be iid $N(0,1)$ and mutually independent. The random variables ϵ_i and e_i are i.i.d. error terms with standard normal and centered exponential distributions, respectively. They are also independent of $\{(X_{1i},X_{2i})\}_{i=1}^n$. For each case, we run 1000

Monte Carlo experiments and record the rejection frequencies of the CvM tests for H_0 with $m(X, \theta(\tau)) = \tilde{X}^\top \theta(\tau)$ and $\tilde{X} = (1, X^\top)^\top$. The null hypothesis is true under DGPs 1 and 2, but not true under DGPs 3–6.

We consider the sample size $n \in \{50, 100, 200\}$ and the quantile interval $\mathcal{T} = [0.1, 0.9]$. The estimator $\theta_n(\tau)$ of the null parameter vector is taken as the regression τ-quantile given by (3). The CvM tests are computed as an approximation over a grid of $m = 30$ evenly spaced points in $\mathcal{T}$. The bandwidth for the conditional density estimator (14) is chosen as $h = n^{-1/5}$ and the kernel function $K(\cdot)$ is a Gaussian kernel. The number of bootstrap multipliers generated for each Monte Carlo replication is 200.

Table 1 displays the rejection frequencies of the CvM tests with the exponential weight function and a Gaussian-type integrating measure. The significance levels are 10%, 5%, and 1%. Our findings are summarized as follows. First, all types of test statistics exhibit good empirical size for sample sizes of 200. The empirical sizes for DGP 2 are also close to the nominal levels, showing that the tests are robust to fat-tailed error distributions. Second, the empirical power of all test statistics increases with the sample size n. The projection-based unidimensional CvM test CvM_n^{ng} has the best empirical power for DGPs 3 and 4, while the gradient tests CvM_n and CvM_n^{np} have the best power for DGPs 5-6. This is consistent with the theory that the gradient test is more powerful in detecting departures in the form of oscillations around 0. Finally, the CvM functional with nuclear norm has better power performance than that with spectral norm in most cases.

To examine the robustness of the tests in relatively high-dimensional settings, we conduct an additional set of Monte Carlo experiments in which we set X_i to a six-dimensional random vector. The results show that the proposed test (17) has significant advantages over the competitive tests in controlling the Type I error; see Section C.2 in the Appendix for more details.

Next, we examine the performance of the tests when the null model involves a nonlinear conditional quantile function. The null hypothesis we consider is (1), where $m(X, \theta(\tau)) = \theta_0(\tau) + \exp(X^\top \theta_1(\tau))$, and $\theta(\tau) = (\theta_0(\tau), \theta_1(\tau)^\top)^\top$ denotes the unknown parameter. The DGPs are given by

- DGP 7: $Y_i = \exp(X_{1i} + X_{2i}) + \epsilon_i$,
- DGP 8: $Y_i = \exp(-(X_{1i} + X_{2i})) + \epsilon_i$,
- DGP 9: $Y_i = 1 + 0.2 * (X_{1i} + X_{2i}) + \exp(X_{1i} + X_{2i}) + \epsilon_i$,
- DGP 10: $Y_i = 1 + 0.2 * (X_{1i} + X_{2i}) + \exp(-(X_{1i} + X_{2i})) + \epsilon_i$,

with X_{1i}, X_{2i} and ϵ_i being i.i.d. $N(0, 1)$ random variables and mutually independent. The remaining Monte Carlo parameters are as before. The null hypothesis corresponds to DGPs 7 and 8, under which the conditional quantile function satisfies $F^{-1}_{Y_i|X_i}(\tau) = \Phi^{-1}(\tau) + \exp(X_{1i}\theta_1 +$

Table 1: Rejection frequencies of the CvM tests with exponential weights under DGPs 1–6.

	$n=50$			$n=100$			$n=200$		
DGP 1	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{gau,F}}$	0.092	0.046	0.007	0.112	0.053	0.012	0.108	0.069	0.022
$CvM_n^{\text{gau,E}}$	0.080	0.038	0.011	0.110	0.052	0.012	0.116	0.065	0.021
$CvM_n^{\text{gau,F,np}}$	0.117	0.063	0.010	0.117	0.059	0.016	0.114	0.066	0.022
$CvM_n^{\text{gau,E,np}}$	0.099	0.046	0.014	0.116	0.055	0.014	0.112	0.067	0.019
$CvM_n^{\text{gau,ng}}$	0.090	0.046	0.009	0.100	0.055	0.011	0.111	0.060	0.020
DGP 2									
$CvM_n^{\text{gau,F}}$	0.123	0.066	0.014	0.114	0.057	0.013	0.106	0.052	0.013
$CvM_n^{\text{gau,E}}$	0.113	0.066	0.022	0.117	0.061	0.015	0.110	0.053	0.017
$CvM_n^{\text{gau,F,np}}$	0.148	0.078	0.021	0.123	0.065	0.015	0.115	0.054	0.011
$CvM_n^{\text{gau,E,np}}$	0.131	0.075	0.023	0.122	0.060	0.016	0.102	0.053	0.016
$CvM_n^{\text{gau,ng}}$	0.129	0.064	0.017	0.121	0.060	0.017	0.103	0.053	0.012
DGP 3									
$CvM_n^{\text{gau,F}}$	0.612	0.474	0.235	0.913	0.854	0.662	0.999	0.994	0.973
$CvM_n^{\text{gau,E}}$	0.655	0.521	0.270	0.935	0.868	0.679	1.000	0.998	0.976
$CvM_n^{\text{gau,F,np}}$	0.677	0.531	0.279	0.928	0.873	0.699	0.999	0.995	0.975
$CvM_n^{\text{gau,E,np}}$	0.687	0.558	0.308	0.942	0.894	0.703	1.000	0.997	0.979
$CvM_n^{\text{gau,ng}}$	0.697	0.580	0.307	0.951	0.908	0.772	1.000	1.000	0.987
DGP 4									
$CvM_n^{\text{gau,F}}$	0.211	0.120	0.027	0.351	0.238	0.089	0.603	0.481	0.279
$CvM_n^{\text{gau,E}}$	0.212	0.116	0.038	0.337	0.222	0.091	0.584	0.454	0.246
$CvM_n^{\text{gau,F,np}}$	0.259	0.149	0.042	0.374	0.260	0.115	0.629	0.499	0.297
$CvM_n^{\text{gau,E,np}}$	0.225	0.130	0.040	0.352	0.232	0.099	0.603	0.461	0.247
$CvM_n^{\text{gau,ng}}$	0.240	0.143	0.036	0.396	0.266	0.115	0.659	0.553	0.339
DGP 5									
$CvM_n^{\text{gau,F}}$	0.230	0.125	0.041	0.403	0.290	0.116	0.720	0.600	0.366
$CvM_n^{\text{gau,E}}$	0.201	0.122	0.035	0.357	0.230	0.094	0.655	0.510	0.243
$CvM_n^{\text{gau,F,np}}$	0.270	0.146	0.047	0.424	0.298	0.125	0.734	0.606	0.364
$CvM_n^{\text{gau,E,np}}$	0.222	0.129	0.039	0.365	0.232	0.093	0.665	0.514	0.248
$CvM_n^{\text{gau,ng}}$	0.206	0.108	0.030	0.355	0.229	0.089	0.660	0.518	0.272
DGP 6									
$CvM_n^{\text{gau,F}}$	0.410	0.286	0.093	0.729	0.588	0.361	0.963	0.936	0.823
$CvM_n^{\text{gau,E}}$	0.369	0.233	0.080	0.654	0.498	0.256	0.938	0.880	0.694
$CvM_n^{\text{gau,F,np}}$	0.459	0.312	0.105	0.751	0.609	0.366	0.963	0.934	0.834
$CvM_n^{\text{gau,E,np}}$	0.390	0.257	0.084	0.660	0.499	0.253	0.933	0.882	0.691
$CvM_n^{\text{gau,ng}}$	0.390	0.258	0.080	0.705	0.557	0.301	0.965	0.924	0.787

$X_{1i}\theta_2)$. Here, $\Phi^{-1}(\cdot)$ denotes the quantile function of a standard normal random variable, $\theta_1 = \theta_2 = 1$ for DGP 7 and $\theta_1 = \theta_2 = -1$ for DGP 8.

The rejection frequencies of the CvM tests with exponential weights associated with DGPs 7–10 are shown in Table 2. We see that the test statistics are slightly oversized when the sample size n is small, but they tend to the nominal level as n increases. The empirical levels of the non-projection-based test $CvM_n^{\text{gau,np}}$ are less accurate than the other two tests, but are reasonable. From the results of DGPs 9 and 10, we observe that the empirical power increases with the sample size for all test statistics. The test $CvM_n^{\text{gau,ng}}$ has the best empirical power under DGP 9, while the test CvM_n^{gau} and the non-projection-based test $CvM_n^{\text{gau,np}}$ have the best power performance under DGP 10.

Finally, we also conduct simulations for the tests with the indicator weight function and the exponential weight function with a Laplace-type integrating measure. The results are presented in Tables C.1–C.4 in Appendix C. We find that the empirical size performance of the proposed tests with these weight functions is similar to that of the tests presented above. However, the power performance of the tests varies with the choice of weight functions and also depends on the null hypotheses considered. For example, when testing the linearity hypothesis, the test statistic with the exponential weight function possesses higher empirical power than the test with the indicator weight function under DGPs 5 and 6. However, for testing the nonlinear null hypothesis as illustrated above, the test statistic with the indicator weight function outperforms its counterparts under both DGPs 9 and 10.

5.3 Real data application

We apply the proposed method to assess the fit of a linear quantile regression model to the salaries of major league baseball (MLB) players. The data is taken from the April 20, 1987 issue of *Sports Illustrated*,[3] which consists of the base salaries of $n = 176$ pitchers at the beginning of the 1987 season. It also records the number of years of experience and several performance measures of these pitchers in 1986. In this section, we analyze the relationship between years of experience and players' salaries, a subject that has attracted considerable attention in the field of sports statistics (Hoaglin and Velleman, 1995). Existing methods have relied on quantile/expectile regression to explore the heterogeneous effects of performance and years of experience on salaries, see, e.g., Zhang and Li (2017). However, the analysis has been mostly limited to a single or some fixed quantiles.

In what follows, we examine whether the widely adopted parametric specification of the salaries-experience relationship is supported across a broad range of quantiles in the distribution of salaries. First, we apply the testing approach to assess the fit of the quantile regression model

$$F_{Y|X}^{-1}(\tau) = \theta_0(\tau) + \theta_1(\tau)X + \theta_2(\tau)X^2, \quad \tau \in \mathcal{T},$$

[3] The data is now available at https://lib.stat.cmu.edu/datasets/baseball.data.

Table 2: Rejection frequencies of the CvM tests with exponential weights under DGPs 7–10.

	$n = 50$			$n = 100$			$n = 200$		
DGP 7	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{gau,F}}$	0.101	0.048	0.007	0.115	0.063	0.015	0.105	0.055	0.012
$CvM_n^{\text{gau,E}}$	0.098	0.038	0.012	0.120	0.059	0.014	0.106	0.052	0.016
$CvM_n^{\text{gau,F,np}}$	0.128	0.066	0.012	0.129	0.071	0.021	0.109	0.058	0.013
$CvM_n^{\text{gau,E,np}}$	0.107	0.044	0.011	0.128	0.064	0.016	0.111	0.054	0.014
$CvM_n^{\text{gau,ng}}$	0.105	0.053	0.010	0.124	0.063	0.013	0.103	0.052	0.011
DGP 8									
$CvM_n^{\text{gau,F}}$	0.110	0.051	0.009	0.121	0.053	0.013	0.105	0.046	0.010
$CvM_n^{\text{gau,E}}$	0.110	0.049	0.008	0.121	0.061	0.014	0.114	0.052	0.018
$CvM_n^{\text{gau,F,np}}$	0.135	0.072	0.011	0.129	0.061	0.014	0.110	0.044	0.009
$CvM_n^{\text{gau,E,np}}$	0.108	0.056	0.011	0.125	0.067	0.013	0.116	0.054	0.017
$CvM_n^{\text{gau,ng}}$	0.111	0.051	0.010	0.116	0.064	0.016	0.097	0.051	0.010
DGP 9									
$CvM_n^{\text{gau,F}}$	0.229	0.130	0.040	0.428	0.297	0.132	0.707	0.598	0.391
$CvM_n^{\text{gau,E}}$	0.203	0.115	0.032	0.378	0.239	0.100	0.642	0.519	0.295
$CvM_n^{\text{gau,F,np}}$	0.276	0.163	0.053	0.460	0.327	0.156	0.720	0.616	0.408
$CvM_n^{\text{gau,E,np}}$	0.225	0.124	0.043	0.396	0.262	0.110	0.642	0.532	0.308
$CvM_n^{\text{gau,ng}}$	0.252	0.158	0.059	0.468	0.332	0.159	0.764	0.644	0.435
DGP 10									
$CvM_n^{\text{gau,F}}$	0.191	0.116	0.037	0.293	0.176	0.071	0.480	0.363	0.191
$CvM_n^{\text{gau,E}}$	0.157	0.100	0.022	0.238	0.149	0.050	0.395	0.283	0.138
$CvM_n^{\text{gau,F,np}}$	0.230	0.140	0.050	0.341	0.224	0.086	0.545	0.434	0.247
$CvM_n^{\text{gau,E,np}}$	0.176	0.110	0.039	0.284	0.172	0.066	0.475	0.359	0.193
$CvM_n^{\text{gau,ng}}$	0.189	0.119	0.038	0.286	0.175	0.074	0.483	0.355	0.197

where Y and X denote the log of the base salary in dollars and the log of number of years experience, respectively, $\mathcal{T}$ denotes one of $[0.1, 0.9]$, $[0.1, 0.5]$, and $[0.5, 0.9]$, and $\theta_0(\cdot)$, $\theta_1(\cdot)$, and $\theta_2(\cdot)$ are unknown parameters.

We also use the same approach to evaluate the fit of the cubic specification

$$F_{Y|X}^{-1}(\tau) = \theta_0(\tau) + \theta_1(\tau)X + \theta_2(\tau)X^2 + \theta_3(\tau)X^3, \quad \tau \in \mathcal{T},$$

where $\theta_0(\cdot)$, $\theta_1(\cdot)$, $\theta_2(\cdot)$, and $\theta_3(\cdot)$ are unknown parameters.

The results are shown in Table 3, which displays the bootstrapped p-values of the $CvM_n^{\text{gau,F}}$ test of the quadratic and cubic log-wage equations. The test is implemented in the same way as the simulation experiments in Section 5.2. In particular, the unknown parameter $\theta_0(\cdot)$ is estimated using the regression τ-quantiles in (3), and the continuum $\mathcal{T}$ is approximated using a grid of $m = 30$ evenly spaced points between the corresponding endpoints. The distribution of the CvM statistic is approximated using 500 sequences of bootstrap multipliers, where the multipliers are generated in accordance with the Bernoulli distribution summarized below (24). The conditional density estimate in the expression for CvM_n is computed using (14) with a Gaussian kernel. This density estimate is implemented using three different settings of bandwidth form $h = \kappa n^{-1/5}$, where $\kappa \in \{0.5, 1.0, 1.5\}$.

Table 3: Linearity tests over various ranges of quantiles

Range of quantiles	$h = 0.5n^{-1/5}$	$h = n^{-1/5}$	$h = 1.5n^{-1/5}$
		p-values H_0 : quadratic	
[0.1,0.5]	<0.001	<0.001	<0.001
[0.5,0.9]	<0.001	<0.001	<0.001
[0.1,0.9]	<0.001	<0.001	<0.001
		p-values H_0 : cubic	
[0.1,0.5]	0.720	0.740	0.692
[0.5,0.9]	0.858	0.866	0.886
[0.1,0.9]	0.796	0.866	0.836

It is clear from Table 3 that the quadratic specification provides a poor fit for all salary quantiles in the range $[0.1, 0.9]$. In contrast, the cubic specification cannot be rejected at the 10% level across any range of quantiles considered. The fit is slightly better for high earnings quantiles than for quantiles in the lower half of the earnings distribution. Unreported results show that this inference holds for CvM tests with other weight functions and integrating measures we consider.

6 Conclusion

We propose a new test for the correct specification of parametric quantile regression models. The key feature of the test is a projection-based vector-weighted CUSUM process. We show how the projection device enables a straightforward multiplier bootstrap to approximate the critical values and how incorporating the gradient component improves the power of the tests under certain alternatives. An additional benefit of the proposed test is that it facilitates the assessment of the suitability of quantile regression models across a continuum of quantiles. Simulation evidence and an empirical example concerning the specification of the earnings-experience relationship for MLB players illustrate the feasibility of our approach for datasets of moderate sizes.

Funding

Song's research is supported by the National Natural Science Foundation of China [Grant Numbers 72373007 and 72333001]. Song also gratefully acknowledges the research support from the Center for Statistical Science of Peking University, and the Key Laboratory of Mathematical Economics and Quantitative Finance (Peking University) of the Ministry of Education, China. Yang's research is supported by the Fundamental Research Funds for the Central Universities.

References

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

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20(1):105–134.

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

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

Bierens, H. J. and Wang, L. (2012). Integrated conditional moment tests for parametric conditional distributions. *Econometric Theory*, 28(2):328–362.

Chang, N. (1990). Weak convergence of a self-consistent estimator of a survival function with doubly censored data. *Annals of Statistics*, 18:391–404.

Conde-Amboage, M., Cesar, S.-S., and Gonzalez-Manteiga, W. (2015). A lack-of-fit test for quantile regression models with high-dimensional covariates. *Computational Statistics and Data Analysis*, 88:128–138.

Dong, C., Li, G., and Feng, X. (2019). Lack-of-fit tests for quantile regression models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(3):629–648.

Engle, R. F. and Manganelli, S. (2004). Caviar. *Journal of Business & Economic Statistics*, 22(4):367–381.

Escanciano, J. C. (2006a). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22(6):1030–1051.

Escanciano, J. C. (2006b). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, 101(474):531–541.

Escanciano, J. C. (2009). On the lack of power of omnibus specification tests. *Econometric Theory*, 25(01):162–194.

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

Escanciano, J. C. and Goh, S. C. (2019). Quantile-regression inference with adaptive control of size. *Journal of the American Statistical Association*, 114(527):1382–1393.

Escanciano, J. C., Jacho-Chavez, D. T., and Lewbel, A. (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics*, 178(3):426–443.

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

Eubank, R. L. and Hart, J. D. (1993). Commonality of Cusum, von Neumann and smoothing-based goodness-of-fit tests. *Biometrika*, 80(1):89–98.

Feng, X., Liu, Q., and Wang, C. (2023). A lack-of-fit test for quantile regression process models. *Statistics & Probability Letters*, 192:109680.

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

Hoaglin, D. C. and Velleman, P. F. (1995). A critical look at some analyses of major league baseball salaries. *The American Statistician*, 49(3):277–285.

Horowitz, J. L. and Lee, S. (2009). Testing a parametric quantile-regression model with an endogenous explanatory variable against a nonparametric alternative. *Journal of Econometrics*, 152:141–152.

Horowitz, J. L. and Spokoiny, V. (2002). An adaptive rate-optimal test of linearity for median regression models. *Journal of the American Statistical Association*, 97:822–835.

Horvath, P., Li, J., Liao, Z., and Patton, A. J. (2022). A consistent specification test for dynamic quantile models. *Quantitative Economics*, 13(1):125–151.

Kim, T.-H. and White, H. (2003). Estimation, inference, and specification testing for possibly misspecified quantile regression. In *Maximum likelihood estimation of misspecified models: twenty years later*, pages 107–132. Emerald Group Publishing Limited.

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

Koenker, R. and d'Orey, V. (1987). Computing regression quantiles. *Applied Statistics*, 36:383–393.

Koenker, R. and Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71:265–283.

Koul, H. L. and Stute, W. (1999). Nonparametric model checks for time series. *Annals of Statistics*, 27:204–236.

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

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285.

Mukherjee, K. (1999). Asymptotics of quantiles and rank scores in nonlinear time series. *Journal of Time Series Analysis*, 20(2):173–192.

Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. *Probability and Statistics*, The Harald Cramer Volume. Almqvist and Wiksell, Stockholm:213–234.

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, 146(1):170–184.

Rosenblatt, M. (1969). Conditional probability density and regression estimate. In Krishnaiah, P. R., editor, *Multivariate Analysis–II*, pages 25–31. Academic Press.

Sant'Anna, P. H. and Song, X. (2019). Specification tests for the propensity score. *Journal of Econometrics*, 210(2):379–404.

Strasser, H. (1990). Global extrapolation of local efficiency. *Statistics and Decisions*, 8:11–26.

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

Van Der Vaart, A. W. and Wellner, J. A. (2023). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Whang, Y. J. (2006). Consistent specification testing for quantile regression models. *Corbae, D., Durlauf, S.N., Hansen, B.E. (Eds.), Econometric Theory and Practice: Frontiers of Analysis and Applied Research, Cambridge University Press, New York*, pages 288–308.

Zhang, F. and Li, Q. (2017). A continuous threshold expectile model. *Computational Statistics & Data Analysis*, 116:49–66.

Zhang, L., Wang, H. J., and Zhu, Z. (2014). Testing for change points due to a covariate threshold in quantile regression. *Statistica Sinica*, 24(4):1859–1877.

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

Appendices

A Preliminary lemmas

We begin with an important result of **?** that allows for the bounding of entropy numbers and the verification of stochastic equicontinuity for processes indexed by both Euclidean and function-valued parameters. The following notations are used. For generic Banach space Θ with associated norm $\|\cdot\|_\Theta$, the covering number $N(\epsilon, \Theta, \|\cdot\|_\Theta)$ of Θ is the minimal number N for which there exist ϵ-neighborhoods $\left\{\left\{\theta : \|\theta - \theta_j\|_\Theta \le \epsilon\right\}, \|\theta_j\|_\Theta < \infty,\ j = 1, \ldots, N\right\}$ covering Θ. A bracket $[l_j, u_j]$ is the set of elements $\theta \in \Theta$ such that $l_j \le \theta \le u_j$. The covering number with bracketing $N_{[\cdot]}(\epsilon, \Theta, \|\cdot\|_\Theta)$ is the minimal N for which there exist ϵ-brackets $\left\{[l_j, u_j] : \|l_j - u_j\|_\Theta \le \epsilon, \|l_j\|_\Theta, \|u_j\|_\Theta < \infty, j = 1, \ldots, N\right\}$ covering Θ. An envelope function H for a class $\mathcal{H}$ is defined as a measurable function such that $H(x) \ge \sup_{h\in\mathcal{H}} |h(x)|$. Define the L_2 norm $\|g\|_2^2 := \mathbb{E}\left[g^2(X, Y)\right]$ and the entropy number

$$J_{[\cdot]}(\delta, \mathcal{H}, \|\cdot\|_2) := \int_0^\delta \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{H}, \|\cdot\|_2)}d\varepsilon.$$

Other definitions of concepts from empirical process theory may be found in Van Der Vaart and Wellner (2023).

Lemma A.1 *Denote $\mathcal{F} := \{(x, y) \to f(x, y, \theta, h) : \theta \in \Theta,\ h \in \mathcal{H}\}$ as a function class, where Θ and $\mathcal{H}$ are Banach spaces with associated norms $\|\cdot\|_\Theta$ and $\|\cdot\|_\mathcal{H}$, respectively. Assume that*

$$\mathbb{E}\left[\sup_{\|\theta_1-\theta_2\|_\Theta\le\delta}\ \sup_{\|h_1-h_2\|_\mathcal{H}\le\delta}\left|f(Y, X, \theta_1, h_1) - f(Y, X, \theta_2, h_2)\right|^2\right] \le C\delta^s \tag{1}$$

for some constant $s \in (0, 2]$. Then,

1. *$N_{[\cdot]}(\epsilon, \mathcal{F}, \|\cdot\|_2) \le N\left([\frac{\epsilon}{2C}]^{\frac{2}{s}}, \Theta, \|\cdot\|_\Theta\right) \times N\left([\frac{\epsilon}{2C}]^{\frac{2}{s}}, \mathcal{H}, \|\cdot\|_\mathcal{H}\right)$ for any $\epsilon > 0$;*
2. *If, in addition to (1), Θ is a compact subset of $\mathbb{R}^p$ and $\int_0^\infty \sqrt{\log N(\epsilon^{2/s}, \mathcal{H}, \|\cdot\|_\mathcal{H})}d\epsilon < \infty$, then for any sequence of positive constants $\delta_n = o(1)$,*

$$\sup_{\|\theta_1-\theta_2\|_\Theta\le\delta_n}\ \sup_{\|h_1-h_2\|_\mathcal{H}\le\delta_n} \|\mathbb{G}_n[f(Y, X, \theta_1, h_1)] - \mathbb{G}_n[f(Y, X, \theta_2, h_2)]\| = o_p(1),$$

where

$$\mathbb{G}_n[f(Y, X, \theta, h)] := \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{f(Y_i, X_i, \theta, h) - \mathbb{E}[f(Y_i, X_i, \theta, h)]\right\}.$$

Proof of Lemma A.1. The results of Lemma A.1 follow from Theorem 3 of **?**. $\mathcal{Q.E.D.}$

Define the following CUSUM process

$$S_n(\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\psi_\tau(Y_i, X_i, b) - \mathbb{E}[\psi_\tau(Y_i, X_i, b) | X_i]\} \xi_n(X_i, \gamma)$$

which is indexed by $\gamma = (b, \tau, x) \in \Omega := \mathcal{B} \times \mathcal{T} \times \Pi$, where $\mathcal{B}$ is a class of bounded, Lipschitz $\mathbb{R}^p$-valued functions on $\mathcal{T}$ and Π the parameter space of x.[4] Define the class $\mathcal{W}_n := \{\xi_n(X_i, \gamma) : \gamma \in \Omega\}$ with a envelope function $\Xi_n(\cdot)$ satisfying that $\mathbb{E}[(|N(X_i)|^2 + |M(X_i)|^4)|\Xi_n(X_i)|^2] = O(1)$ and $\mathbb{E}[|M(X_i)\Xi_n(X_i)|^2 \mathbb{1}(|M(X_i)|\Xi_n(X_i) > \epsilon\sqrt{n})] = o(1)$ for each $\epsilon > 0$.

Consider the following pseudo metric $\rho(\gamma_1, \gamma_2) = \|b_1 - b_2\|_{\mathcal{T}} + |\tau_1 - \tau_2| + d(x_1, x_2)$, where $\gamma_j = (b_j, \tau_j, x_j)$, $j = 1, 2$, and $d(\cdot, \cdot)$ is a semimetric that makes Π into a totally bounded space. Also, assume that $\sup_{\rho(\gamma_1,\gamma_2)<\delta} \|M(\cdot)[\xi_n(\cdot, \gamma_1) - \xi_n(\cdot, \gamma_2)]\|_2 \le C\delta$ for a sufficiently small $\delta > 0$.

Lemma A.2 *(Lemma A.2 in Escanciano and Goh (2014)) Let $\mathcal{H}$ and $\mathcal{G}$ be classes of functions that have square-integrable envelopes H and G, respectively. Then,*

$$N(2e\|HG\|_2, \mathcal{H} \cdot \mathcal{G}, \|\cdot\|_2) \le N(e\|H\|_2, \mathcal{H}, \|\cdot\|_2) \times N(e\|G\|_2, \mathcal{G}, \|\cdot\|_2),$$

and

$$N_{[\cdot]}(Ce, \mathcal{H} \cdot \mathcal{G}, \|\cdot\|_2) \le N_{[\cdot]}(e, \mathcal{H}, \|\cdot\|_2) \times N_{[\cdot]}(e, \mathcal{G}, \|\cdot\|_2).$$

Lemma A.3 *Assume Assumptions 3.1–3.7 hold. In addition, the class $\mathcal{W}_n$ satisfies the previous conditions and $J_{[\cdot]}(\delta_n, \mathcal{W}_n, \|\cdot\|_2) \to 0$ for every $\delta_n \downarrow 0$. Then, the process $S_n(\gamma)$ is ρ-stochastically equicontinuous.*

Proof of Lemma A.3. Define $W_i = (Y_i, X_i^\top)^\top$ and the class $\Psi := \{w \to \bar{\psi}_\tau(w, b) : (b, \tau) \in \mathcal{B} \times \mathcal{T}\}$, where for any $(b, \tau) \in \mathcal{B} \times \mathcal{T}$, $\bar{\psi}_\tau(W, b) := \psi_\tau(W, b) - \mathbb{E}[\psi_\tau(W, b) \,|\, X]$. For any fixed (b_1, τ_1) and $\delta \in (0, 1)$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{b:\|b_1-b\|_{\mathcal{T}}\le\delta^2} \sup_{\tau:|\tau_1-\tau|\le\delta^2} \left\|\bar{\psi}(W_i, b_1, \tau_1) - \bar{\psi}(W_i, b, \tau)\right\|^2\right] \\
&\le C\mathbb{E}\left[\sup_{\tau:|\tau_1-\tau|\le\delta^2} \sup_{b:\|b_1-b\|_{\mathcal{T}}\le\delta^2} \|\psi_{\tau_1}(W_i, b_1) - \psi_\tau(W_i, b)\|^2\right] \\
&\quad + C\mathbb{E}\left[\sup_{\tau:|\tau_1-\tau|\le\delta^2} \sup_{b:\|b_1-b\|_{\mathcal{T}}\le\delta^2} \|\mathbb{E}[\psi_{\tau_1}(W_i, b_1)|X_i] - \mathbb{E}[\psi_\tau(W_i, b)|X_i]\|^2\right] \\
&\le C\mathbb{E}\|\dot{m}(X, b_1(\tau_1))\|^2 \Bigg[\sup_{\tau:|\tau_1-\tau|\le\delta^2} |\tau_1 - \tau|^2 \\
&\quad + \mathbb{E}\left[F\big(m(X, b_1(\tau_1)) + C|M(X)|\delta^2 \,|\, X\big) - F\big(m(X, b_1(\tau_1)) - C|M(X)|\delta^2 \,|\, X\big)\right]\Bigg]
\end{aligned}$$

[4] For the indicator weighting function $w(X, x) = \mathbb{1}(\beta^\top X \le u)$, we have $x = (\beta^\top, u)^\top$ and $\Pi = \mathbb{R}^{d+1}$. For the exponential weight function $w(X, x) = \exp(\mathrm{i}x^\top X)$, we have Π as a generic compact subset of $\mathbb{R}^d$ containing the origin.

$$
\begin{aligned}
&+ C\mathbb{E}\left[\|\dot{m}(X, b_1(\tau_1))\|^2 \|M(X)\|^2 \sup_{\tau:|\tau_1-\tau|\le\delta^2} \sup_{b:\|b_1-b\|_{\mathcal{T}}\le\delta^2} \|b_1(\tau_1) - b(\tau)\|^2\right] \\
&+ C\mathbb{E}\left[\|N(X)\|^2 \sup_{\tau:|\tau_1-\tau|\le\delta^2} \sup_{b:\|b_1-b\|_{\mathcal{T}}\le\delta^2} \|b_1(\tau_1) - b(\tau)\|^2\right] \\
\le{}& C\delta^2.
\end{aligned}
$$

Applying Lemmas A.1-A.2 and Theorem 2.7.17 in Van Der Vaart and Wellner (2023), we have $J_{[\cdot]}(\delta_n, \Psi \times \mathcal{W}_n, \|\cdot\|_2) \to 0$ for every $\delta_n \downarrow 0$. The desired result follows from the proof of Theorem 19.28 in Van Der Vaart (1998). $\mathcal{Q.E.D}$.

The next result concerns the uniform convergence rate for the kernel estimator $\hat{f}_{i\tau} := \hat{f}_h(m(X_i, \hat{\theta}_n(\tau)) \mid X_i)$ given in (14). Similar to Escanciano and Goh (2014), we view $\hat{f}_{i\tau}$ as a function of $\hat{\theta}_n$ and write

$$
\hat{f}_h(x, \theta) := \frac{|\mathcal{T}|}{mh_m} \sum_{j=1}^{m} K\left(\frac{m(x, \theta(\tau)) - m(x, \theta(\tau_j))}{h_m}\right), \tag{2}
$$

where $\{\tau_j\}$ is a random sample from the uniform distribution in $\mathcal{T}$ with $m \to \infty$ as $n \to \infty$, and h_m denotes a possibly data-dependent bandwidth parameter.

Lemma A.4 *Suppose that Assumptions 3.1–3.6 hold. Let $B(\theta^*, \delta) = \{\theta \in \mathcal{B} : \|\theta_0 - \theta^*\|_{\mathcal{T}} < \delta\}$. Then, under H_0, for any sequence of positive constants $\delta_n = o(1)$,*

$$
\begin{aligned}
&\sup_{a_m \le h \le b_m} \sup_{\tau \in \mathcal{T}, \theta \in B(\theta^*, \delta_n)} \max_{1 \le i \le n} \left|\hat{f}_h(X_i, \theta) - f\left(m(X_i, \theta^*(\tau))|X_i\right)\right| \\
&= \mathcal{O}_p\left(\sqrt{\frac{\log a_m^{-1} \vee \log\log m}{m a_m}} + b_m^2 + \delta_n\right).
\end{aligned}
$$

Under H_1, the above conclusion still holds except that $f(m(X_i, \theta^*(\tau))|X_i)$ is replaced by $f^*(m(X_i, \theta^*(\tau))|X_i)$ as defined above Theorem 3.2.

Proof of Lemma A.4. This result follows from Lemma B.5 of Escanciano et al. (2014). We omit the details here for brevity. $\mathcal{Q.E.D}$.

Lemma A.5 *Suppose that Assumptions 3.1–3.4 hold. Then, under the null hypothesis,*

$$
\begin{aligned}
\sup_{(x,\tau) \in \Pi \times \mathcal{T}} \Bigg|&\breve{R}_n(x, \tau) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(Y_i, X_i, \theta_0) w(X_i, x) \\
&+ G^\top(x, \theta_0(\tau)) \sqrt{n}(\hat{\theta}_n(\tau) - \theta_0(\tau))\Bigg| = o_p(1).
\end{aligned}
$$

Proof of Lemma A.5. Apply Lemma A.3 with $\mathcal{W}_n = \{X \to w(X, x) : x \in \Pi\}$, which, for indicator weighting case, satisfies the conditions of the lemma by problem 14 on Van Der Vaart and Wellner (2023, p.152), and for exponential weighting case, satisfies the conditions of the lemma by the continuity of $x \to \exp(\mathrm{i}x^\top X)$, Theorem 2.7.11 of Van Der Vaart and Wellner (2023) and Assumption 3.1(i). It is also easy to show that under Assumptions 3.1-3.4, $\hat{\theta}_n \in \mathcal{B}$ with probability tending to one and $\theta_0 \in \mathcal{B}$. Hence,

$$\sup_{(x,\tau)\in\Pi\times\mathcal{T}} \left|S_n(\hat{\theta}_n,\tau,x) - S_n(\theta_0,\tau,x)\right| = o_p(1),$$

which is equivalent under the null to

$$\begin{aligned}\sup_{(x,\tau)\in\Pi\times\mathcal{T}} \Bigg|\breve{R}_n(x,\tau) - \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\tau(Y_i,X_i,\theta_0)w(X_i,x) \\ + \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\mathbb{E}[\psi_\tau(Y_i,X_i,\theta_0)|X_i] - \mathbb{E}[\psi_\tau(Y_i,X_i,\hat{\theta}_n)|X_i]\Big)\, w(X_i,x)\Bigg| = o_p(1). \end{aligned} \quad (3)$$

Applying a mean value argument, we obtain

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\mathbb{E}[\psi_\tau(Y_i,X_i,\theta_0)|X_i] - \mathbb{E}[\psi_\tau(Y_i,X_i,\hat{\theta}_n)|X_i]\Big)\, w(X_i,x) \\ &= \frac{1}{n}\sum_{i=1}^n \Bigg(f(m(X_i,\tilde{\theta}_n(\tau))|X_i)\dot{m}(X_i,\tilde{\theta}_n(\tau))\dot{m}^\top(X_i,\tilde{\theta}_n(\tau)) \\ &\quad + \ddot{m}(X_i,\tilde{\theta}_n(\tau))\big(F(m(X_i,\tilde{\theta}_n(\tau))|X_i) - \tau\big)\Bigg) w(X_i,x)\sqrt{n}(\hat{\theta}_n(\tau) - \theta_0(\tau)) + o_p(1)\end{aligned}$$

uniformly in $(x,\tau) \in \Pi\times\mathcal{T}$, where $\ddot{m}(X,\theta) = \partial^2 m(X,\theta)/\partial\theta\partial\theta^\top$, and $\tilde{\theta}_n(\tau)$ lies between $\theta_0(\tau)$ and $\hat{\theta}_n(\tau)$ for each $\tau \in \mathcal{T}$ almost surely. By our assumptions, under H_0 it holds that $n^{-1}\sum_{i=1}^n \ddot{m}(X_i,\tilde{\theta}_n(\tau))(F(m(X_i,\tilde{\theta}_n(\tau))|X_i) - \tau)w(X_i,x) = o_p(1)$ and

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n f(m(X_i,\tilde{\theta}_n(\tau))|X_i)\dot{m}(X_i,\tilde{\theta}_n(\tau))\dot{m}^\top(X_i,\tilde{\theta}_n(\tau))w(X_i,x) \\ &= \frac{1}{n}\sum_{i=1}^n f(m(X_i,\theta_0(\tau))|X_i)\dot{m}(X_i,\theta_0(\tau))\dot{m}^\top(X_i,\theta_0(\tau))w(X_i,x) + o_p(1) \\ &= G(x,\theta_0(\tau)) + o_p(1)\end{aligned}$$

uniformly in $(x,\tau) \in \Pi\times\mathcal{T}$, where the last equality follows from the Glivenko–Cantelli Theorem, since $\{X \to f(m(X,\theta_0(\tau))|X)\dot{m}(X,\theta_0(\tau))\dot{m}^\top(X,\theta_0(\tau))w(X,x) : (x,\tau) \in \Pi\times\mathcal{T}\}$ is Glivenko–Cantelli by Lemma A.2 in Escanciano and Goh (2014). Hence, we obtain the expansion

$$\sup_{(x,\tau)\in\Pi\times\mathcal{T}} \Bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\mathbb{E}[\psi_\tau(Y_i,X_i,\theta_0)|X_i] - \mathbb{E}[\psi_\tau(Y_i,X_i,\hat{\theta}_n)|X_i]\Big)\, w(X_i,x)$$

$$
- G(x,\theta_0(\tau))\sqrt{n}(\hat{\theta}_n(\tau)-\theta_0(\tau))\Big| = o_p(1),
$$

which together with (3) yields the lemma. $\mathcal{Q.E.D}$.

Lemma A.6 *Define* $\breve{S}_n(\tau) = n^{-1/2}\sum_{i=1}^n \hat{g}_n(X_i,\hat{\theta}_n(\tau))\psi_\tau(Y_i,X_i,\hat{\theta}_n)$*. Then, under Assumptions 3.1–3.6,*

$$
\sup_{(x,\tau)\in\Pi\times\mathcal{T}}\left|\breve{S}_n(\tau) - \frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0(\tau))\psi_\tau(Y_i,X_i,\theta_0) + \Delta(\theta_0(\tau))\sqrt{n}(\hat{\theta}_n(\tau)-\theta_0(\tau))\right| = o_p(1)
$$

Proof of Lemma A.6. We apply Lemma A.3 with $\mathcal{W}_n = \{\bar{x} \to \hat{f}_h(\bar{x},b)\dot{m}_j(\bar{x},b)\dot{m}_l(\bar{x},b) : b\in\mathcal{B}\}$, where $j,l\in\{1,\ldots,p\}$, and $\dot{m}_j(\bar{x},b)$ denotes the jth element of $\dot{m}(\bar{x},b)$. By Lemma A.4, we can take as envelope for the class the function $\Xi_n(X_i) = C|M(X_i)|^2$, which trivially satisfies the Lindeberg conditions of the theorem. Similarly, we have uniformly in $n\geq 1$ that $\mathbb{E}\left[\sup_{\theta:\|\theta_1-\theta\|_{\mathcal{T}}<\delta}\left|\hat{f}_h(X,\theta_1)\dot{m}_j(X,\theta_1)\dot{m}_l(X,\theta_1) - \hat{f}_h(X,\theta)\dot{m}_j(X,\theta)\dot{m}_l(X,\theta)\right|^2\right] \leq C\delta^2$. Hence, it follows from Lemma A.1 and Van Der Vaart and Wellner (2023, Theorem 2.7.11) that the assumptions in Lemma A.3 hold for this class. The stochastic equicontinuity implies

$$
\begin{aligned}
\breve{S}_n(\tau) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0(\tau))\psi_\tau(Y_i,X_i,\theta_0(\tau)) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \big(\hat{g}_n(X_i,\theta_0(\tau)) - g(X_i,\theta_0(\tau))\big)\psi_\tau(Y_i,X_i,\theta_0(\tau)) \\
&- \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_n(X_i,\hat{\theta}_n(\tau))\Big(\mathbb{E}\left[\psi_\tau(Y_i,X_i,\theta_0(\tau)) \mid X_i\right] - \mathbb{E}\big[\psi_\tau(Y_i,X_i,\hat{\theta}_n(\tau)) \mid X_i\big]\Big) + o_p(1) \\
:=& \breve{S}_{1n}(\tau) + \breve{S}_{2n}(\tau) - \breve{S}_{3n}(\tau) + o_p(1),
\end{aligned}
$$

where $\hat{g}_n(X_i,\theta(\tau)) = \hat{f}_h(m(X_i,\theta(\tau))|X_i)\dot{m}(X_i,\theta(\tau))\dot{m}^\top(X_i,\theta(\tau))$ for $\theta\in\mathcal{B}$.

Applying Theorem 19.28 of Van Der Vaart (1998) to the class $\{(\mathbf{x},\mathbf{y}) \to \{\hat{f}_h(\mathbf{x},\theta_0) - f(m(\mathbf{x}^\top\theta_0(\tau))|\mathbf{x})\}\dot{m}_j(\mathbf{x},\theta_0(\tau))\dot{m}_l(\mathbf{x},\theta_0(\tau))\psi_\tau(\mathbf{y},\mathbf{x},\theta_0) : \tau\in\mathcal{T}\}$ yields that $\breve{S}_{2n}$ is stochastically equicontinuous. Additionally, for a fixed τ, the lth element of $\breve{S}_{2n}$ satisfies

$$
\begin{aligned}
&\text{Var}\big[\breve{S}_{2n,l}(\tau) \mid \{\tau_j\}_{j=1}^m\big] \\
&\leq \mathbb{E}\Big[|\dot{m}(X_i,\theta_0(\tau))|^3\big(\hat{f}_h(X_i,\theta_0) - f(m(X_i,\theta_0(\tau))|X_i)\big)^2 \mid \{\tau_j\}_{j=1}^m\Big] = o_p(1)
\end{aligned}
$$

by Lemma A.4 and Assumption 3.3(ii), which implies that $\sup_{\tau\in\mathcal{T}}\|\breve{S}_{2n}(\tau)\| = o_p(1)$. Finally, applying the mean-value theorem yields

$$
\begin{aligned}
\breve{S}_{3n}(\tau) =& \frac{1}{n}\sum_{i=1}^n f(m(X_i,\tilde{\theta}_n(\tau))|X_i)\hat{f}_h(m(X_i,\hat{\theta}_n(\tau))|X_i)\dot{m}(X_i,\hat{\theta}_n(\tau))\dot{m}^\top(X_i,\hat{\theta}_n(\tau)) \\
&\times \dot{m}(X_i,\tilde{\theta}_n(\tau))\dot{m}^\top(X_i,\tilde{\theta}_n(\tau))\sqrt{n}(\hat{\theta}_n(\tau)-\theta_0(\tau)) + o_p(1)
\end{aligned}
$$

$$:= \widetilde{\Delta}_{n,h}(\tau)\sqrt{n}(\hat{\theta}_n(\tau) - \theta_0(\tau)) + o_p(1),$$

where $\tilde{\theta}_n(\tau)$ lies between $\theta_0(\tau)$ and $\hat{\theta}_n(\tau)$ for each $\tau \in \mathcal{T}$ a.s.. By the consistency of $\hat{\theta}_n$, definition (8), Lemma A.4, the Glivenko–Cantelli theorem, we can easily see that $\widetilde{\Delta}_{n,h}(\tau) = n^{-1}\sum_{i=1}^n g(X_i, \tilde{\theta}_n(\tau))g^\top(X_i, \hat{\theta}_n(\tau)) + o_p(1) = n^{-1}\sum_{i=1}^n g(X_i, \theta_0(\tau))g^\top(X_i, \theta_0(\tau)) + o_p(1) = \Delta(\theta_0(\tau)) + o_p(1)$, uniformly in $\tau \in \mathcal{T}$. The lemma follows by combining the above results. $\mathcal{Q.E.D}$.

Lemma A.7 *Under Assumptions 3.1–3.4,*

$$\sup_{(x,\tau)\in\Pi\times\mathcal{T}} \left|\hat{G}_n(x, \hat{\theta}_n(\tau)) - G(x, \theta_0(\tau))\right| = o_p(1).$$

Proof of Lemma A.7. The proof follows from Lemma A.4, and the Glivenko–Cantelli Theorem. $\mathcal{Q.E.D}$.

Lemma A.8 *Under Assumptions 3.1–3.4,*

$$\sup_{\tau\in\mathcal{T}} \left|\hat{\Delta}_n^{-1}(\hat{\theta}_n(\tau)) - \Delta^{-1}(\theta_0(\tau))\right| = o_p(1).$$

Proof of Lemma A.8. The proof follows from Lemma A.4, the Glivenko–Cantelli Theorem, and the continuous mapping theorem. $\mathcal{Q.E.D}$.

Lemma A.9 *For the weight function of $w(X, x) = \mathbb{1}\left(\beta^\top X \le u\right)$, the closed-form expressions of the CvM test statistics with the spectral and Frobenius norms are given by (17).*

Proof of Lemma A.9. Recalling the definition of the projection operator $\mathcal{P}_{n,\tau}\mathbb{1}\left(\beta^\top X_i \le u\right)$, we have

$$\begin{aligned}
\mathcal{B}_n &:= \frac{1}{m}\sum_{r=1}^m \int_\Pi \hat{R}_n(x, \tau_\tau)\hat{R}_n^\top(x, \tau_\tau)\Psi^{\text{lin}}(dx) \\
&= \frac{1}{mn}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n \int_\Pi \mathcal{P}_{n,\tau_r}\mathbb{1}\left(\beta^\top X_i \le u\right)\psi_{\tau_r}(Y_i, X_i, \hat{\theta}_n) \\
&\quad \times \psi_{\tau_r}^\top(Y_j, X_j, \hat{\theta}_n)\mathcal{P}_{n,\tau_r}^\top\mathbb{1}\left(\beta^\top X_j \le u\right)F_{n,\beta}(du)d\beta \\
&= \frac{1}{mn^2}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n\sum_{v=1}^n \int_{\mathbb{S}^d} \mathcal{P}_{n,\tau_r}\mathbb{1}\left(\beta^\top X_i \le \beta^\top X_v\right)\psi_{\tau_r}(Y_i, X_i, \hat{\theta}_n) \\
&\quad \times \psi_{\tau_r}^\top(Y_j, X_j, \hat{\theta}_n)\mathcal{P}_{n,\tau_r}^\top\mathbb{1}\left(\beta^\top X_j \le \beta^\top X_v\right)d\beta.
\end{aligned} \tag{4}$$

Denote $\psi_{\tau,k}(Y, X, \theta) = \dot{m}_k(X, \theta(\tau))\left(\tau - \mathbb{1}\left(Y \le m(X, \theta(\tau))\right)\right)$ the k-th element of $\psi_\tau(Y, X, \theta)$, and $\hat{g}_{n,k}(X, \theta(\tau))$ the kth column of $\hat{g}_n(X, \theta(\tau))$, $k = 1, \ldots, p$. Let e_k be a $p \times 1$ vector with kth element being one and others being zero. Then, the (k, l)th element of $\mathcal{B}_n$ can be written as

$$\frac{1}{mn^2}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n\sum_{v=1}^n \int_{\mathbb{S}^d} e_k^\top\mathcal{P}_{n,\tau_r}\mathbb{1}\left(\beta^\top X_i \le \beta^\top X_v\right)\psi_{\tau_r}(Y_i, X_i, \hat{\theta}_n)$$

$$
\begin{aligned}
&\times \psi_{\tau_r}^\top(Y_i, X_i, \hat\theta_n)\mathcal{P}_{n,\tau_r}^\top \mathbb{1}\left(\beta^\top X_j \le \beta^\top X_v\right) e_l \, d\beta \\
&= \frac{1}{mn^2}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n\sum_{v=1}^n \psi_{\tau_r,k}(Y_i, X_i, \hat\theta_n)\psi_{\tau_r,l}(Y_j, X_j, \hat\theta_n) B_{ijv} \\
&\quad - \frac{1}{mn^3}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n\sum_{v=1}^n\sum_{w=1}^n \psi_{\tau_r,k}(Y_i, X_i, \hat\theta_n)\psi_{\tau_r}^\top(Y_j, X_j, \hat\theta_n) \\
&\quad \times \hat g_n^\top(X_j, \hat\theta_n(\tau_r))\hat\Delta_n^{-1}(\hat\theta_n(\tau_r))\hat g_{n,l}(X_w, \hat\theta_n(\tau_r)) B_{iwv} \\
&\quad - \frac{1}{mn^3}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n\sum_{v=1}^n\sum_{u=1}^n \psi_{\tau_r,l}(Y_j, X_j, \hat\theta_n)\psi_{\tau_r}^\top(Y_i, X_i, \hat\theta_n) \\
&\quad \times \hat g_n^\top(X_i, \hat\theta_n(\tau_r))\hat\Delta_n^{-1}(\hat\theta_n(\tau_r))\hat g_{n,k}(X_u, \hat\theta_n(\tau_r)) B_{juv} \\
&\quad + \frac{1}{mn^4}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n\sum_{v=1}^n\sum_{u=1}^n\sum_{w=1}^n \hat g_{n,k}^\top(X_u, \hat\theta_n(\tau_r))\hat\Delta_n^{-1}(\hat\theta_n(\tau_r))\hat g_n(X_i, \hat\theta_n(\tau_r))\psi_{\tau_r}(Y_i, X_i, \hat\theta_n) \\
&\quad \times \hat g_{n,l}^\top(X_w, \hat\theta_n(\tau_r))\hat\Delta_n^{-1}(\hat\theta_n(\tau_r))\hat g_n(X_j, \hat\theta_n(\tau_r))\psi_{\tau_r}(Y_j, X_j, \hat\theta_n) B_{uwv} \\
&:= \frac{1}{mn^2}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n \xi_k(Y_i, X_i, \tau_r) B_{ij} \xi_l(Y_j, X_j, \tau_r),
\end{aligned}
$$

where B_{ij} is the summation of the proportion of spherical wedge volume between vectors $X_i - X_v$ and $X_j - X_v$, that is,

$$
\begin{aligned}
B_{ij} =& \sum_{v=1}^n \int_{\mathbb{S}^d} \mathbb{1}\left(\beta^\top X_i \le \beta^\top X_v\right)\mathbb{1}\left(\beta^\top X_j \le \beta^\top X_v\right) d\beta \\
=& \sum_{v=1}^n \left(\pi - \arccos\left(\frac{(X_i - X_v)^\top (X_j - X_v)}{\|X_i - X_v\|\|X_j - X_v\|}\right)\right)\frac{\pi^{\frac{d}{2}-1}}{\Gamma(\frac{d}{2}+1)},
\end{aligned}
$$

and

$$
\xi_k(Y_i, X_i, \tau) = \psi_{\tau,k}(Y_i, X_i, \hat\theta_n) - \hat g_{n,k}^\top(X_i, \hat\theta_n(\tau))\hat\Delta_n^{-1}(\hat\theta_n(\tau))\frac{1}{n}\sum_{s=1}^n \hat g_n(X_s, \hat\theta_n(\tau))\psi_\tau(Y_s, X_s, \hat\theta_n).
$$

Then, we can readily obtain the matrix expression of $\mathcal{B}_n$ as:

$$
\mathcal{B}_n = \frac{1}{mn^2}\sum_{r=1}^m\sum_{i=1}^n\sum_{j=1}^n \xi_i(\tau_r) B_{ij} \xi_j^\top(\tau_r),
$$

where

$$
\xi_i(\tau) = \psi_\tau(Y_i, X_i, \hat\theta_n) - \hat g_n^\top(X_i, \hat\theta_n(\tau))\hat\Delta_n^{-1}(\hat\theta_n(\tau))\frac{1}{n}\sum_{s=1}^n \hat g_n(X_s, \hat\theta_n(\tau))\psi_\tau(Y_s, X_s, \hat\theta_n).
$$

Finally, for the test statistics based on the spectral and Frobenius norms, we take the largest eigenvalue and the trace, respectively. $\mathcal{Q.E.D}$.

Lemma A.10 *For the weight function $w(X,x)=\exp(\mathrm{i}x^\top X)$, the closed-form expressions of test statistics with Gaussian and Laplace integrating measures are given by* (20) *and* (21)*, respectively.*

Proof of Lemma A.10. The difference between Lemma A.9 and Lemma A.10 is the weight functions. For the exponential weight function equipped with the Gaussian integrating measure, we have

$$U_{ij}=\int_{\mathbb{R}}\exp\{\mathrm{i}x^\top(X_i-X_j)-0.5\cdot\|x\|^2\}dx=\sqrt{\frac{\pi}{2}}\exp\left\{-\frac{\|X_i-X_j\|^2}{2}\right\}.$$

Likewise, for the exponential weight function equipped with the Laplace integrating measure, we have

$$L_{ij}=\int_{\mathbb{R}}\exp\{\mathrm{i}x^\top(X_i-X_j)-\|x\|_1\}dx=\frac{2}{1+\|X_i-X_j\|^2}.$$

We omit other derivations here for brevity. $\mathcal{Q.E.D}$.

B Proof of theorems

Proof of Theorem 3.1. We first show that the process $\hat{R}_n$ is asymptotically equivalent to R_{n0} in (12) under the null, that is,

$$\begin{aligned}
&\hat{R}_n(x,\tau)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n w(X_i,x)\psi_\tau(Y_i,X_i,\hat{\theta}_n)-\hat{G}_n^\top(x,\hat{\theta}_n(\tau))\hat{\Delta}_n^{-1}(\hat{\theta}_n(\tau))\frac{1}{\sqrt{n}}\sum_{i=1}^n\hat{g}_n(X_i,\hat{\theta}_n(\tau))\psi_\tau(Y_i,X_i,\hat{\theta}_n)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n w(X_i,x)\psi_\tau(Y_i,X_i,\theta_0)-\mathbb{E}[f(m(X,\theta_0(\tau))|X)\dot{m}(X,\theta_0(\tau))\dot{m}^\top(X,\theta_0(\tau))w(X,x)]\\
&\times\sqrt{n}(\hat{\theta}_n(\tau)-\theta_0(\tau))-G^\top(x,\theta_0(\tau))\Delta^{-1}(\theta_0(\tau))\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0(\tau))\psi_\tau(Y_i,X_i,\theta_0)\\
&+G^\top(x,\theta_0(\tau))\sqrt{n}(\hat{\theta}_n(\tau)-\theta_0(\tau))+o_p(1)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n w(X_i,x)I_p\psi_\tau(Y_i,X_i,\theta_0)-G^\top(x,\theta_0(\tau))\Delta^{-1}(\theta_0(\tau))\\
&\times\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0(\tau))\psi_\tau(Y_i,X_i,\theta_0)+o_p(1)\\
=&R_{n0}(x,\tau)+o_p(1),
\end{aligned}$$

where the second equality follows from Lemmas A.5-A.8.

Then, the weak convergence of $R_{n0}(x,\tau)$ follows from the joint weak convergence of $n^{-1/2}\sum_{i=1}^n w(X_i,x)I_p\psi_\tau(Y_i,X_i,\theta_0)$ and $n^{-1/2}\sum_{i=1}^n g(X_i,\theta_0(\tau))\psi_\tau(Y_i,X_i,\theta_0)$ as $G(x,\theta_0(\tau))$ and

$\Delta^{-1}(\tau)$ are uniform continuous in $\mathcal{T}$. A standard multivariate CLT implies the convergence of finite-dimensional distributions. The joint asymptotic equicontinuity follows from that of the marginals. Specifically, the asymptotic equicontinuity of $n^{-1/2}\sum_{i=1}^n w(X_i,x)I_p\psi_\tau(Y_i,X_i,\theta_0)$ follows from Lemma A.3 applied to $\mathcal{B}=\{\theta_0\}$ and $\mathcal{W}=\{X\to w(X,x):x\in\Pi\}$, while the asymptotic equicontinuity of $n^{-1/2}\sum_{i=1}^n g(X_i,\theta_0(\tau))\psi_\tau(Y_i,X_i,\theta_0)$ follows from Lemma A.3 applied to $\mathcal{B}=\{\theta_0\}$ and $\mathcal{W}=\{X\to g_{kl}(X,\theta_0(\tau)):\tau\in\mathcal{T}\}$, where $g_{kl}(X,\theta_0(\tau))$ denotes the (k,l)th element of $g(X,\theta_0(\tau))$. Note that for all $\epsilon\in(0,1)$, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\sup_{\tau_2:|\tau_1-\tau_2|<\epsilon}\big|g_{kl}(X,\theta_0(\tau_1))-g_{kl}(X,\theta_0(\tau_2))\big|^2\Big]\\
&\le C\mathbb{E}\Big[\sup_{\tau_2:|\tau_1-\tau_2|<\epsilon}\big|f(m(X,\theta_0(\tau_1))|X)-f(m(X,\theta_0(\tau_2))|X)\big|^2\big|\dot m_k(X,\theta_0(\tau_1))\dot m_l(X,\theta_0(\tau_1))\big|^2\Big]\\
&\quad+C\mathbb{E}\Big[\sup_{\tau_2:|\tau_1-\tau_2|<\epsilon}\big|\dot m_k(X,\theta_0(\tau_1))\dot m_l(X,\theta_0(\tau_1))-\dot m_k(X,\theta_0(\tau_2))\dot m_l(X,\theta_0(\tau_2))\big|^2\Big]\\
&\le C\big(\mathbb{E}|M(X)|^4+\mathbb{E}|M(X)N(X)|^2\big)\epsilon^2\le C\epsilon^2.
\end{aligned}
$$

Hence, by Lemma A.1, the class $\mathcal{W}$ satisfies the conditions of Lemma A.3. $\mathcal{Q.E.D.}$

Proof of Theorem 3.2. Applying similar arguments as those in Lemmas A.5-A.8, we have

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\hat R_n(x,\tau)&=\frac{1}{n}\sum_{i=1}^n w(X_i,x)I_p\psi_\tau(Y_i,X_i,\hat\theta_n)\\
&\quad-\hat G_n^\top(x,\hat\theta_n(\tau))\hat\Delta_n^{-1}(\hat\theta_n(\tau))\frac{1}{n}\sum_{i=1}^n\hat g_n(X_i,\hat\theta_n(\tau))\psi_\tau(Y_i,X_i,\hat\theta_n)\\
&=\frac{1}{n}\sum_{i=1}^n w(X_i,x)I_p\psi_\tau(Y_i,X_i,\theta^*)\\
&\quad-G^{*\top}(x,\theta^*(\tau))\Delta^{*-1}(\theta^*(\tau))\frac{1}{n}\sum_{i=1}^n g^*(X_i,\theta^*(\tau))\psi_\tau(Y_i,X_i,\theta^*)+o_p(1)
\end{aligned}
$$

under the alternative hypothesis and Assumption 3.8. Using similar arguments as Lemma A.3, we can show that the classes of functions $\{(X,Y)\to w(X,x)\psi_\tau(Y,X,\theta^*):(x,\tau)\in\Pi\times\mathcal{T}\}$ and $\{(X,Y)\to g_{kl}(X,\theta^*(\tau))\psi_\tau(Y,X,\theta^*):\tau\in\mathcal{T}\}$ are Glivenko–Cantelli. It follows from the Glivenko–Cantelli theorem that $\sup_{(x,\tau)\in\Pi\times\mathcal{T}}\big|(1/n)\sum_{i=1}^n w(X_i,x)\psi_\tau(Y_i,X_i,\theta^*)-\mathbb{E}[w(X,x)\psi_\tau(Y,X,\theta^*)]\big|=o_p(1)$, and $\sup_{(x,\tau)\in\Pi\times\mathcal{T}}\big|(1/n)\sum_{i=1}^n g^*(X_i,\theta^*(\tau))\psi_\tau(Y_i,X_i,\theta^*)-\mathbb{E}[g^*(X,\theta^*(\tau))\psi_\tau(Y,X,\theta^*)]\big|=o_p(1)$ under H_1. Theorem 3.2 immediately follows. $\mathcal{Q.E.D.}$

Proof of Theorem 3.3. Note that Lemmas A.5-A.8 still hold under H_{1n}. Following along the lines of Theorem 3.1, we have $\hat R_n(x,\tau)=R_{n0}(x,\tau)+o_p(1)$ uniformly in $(x,\tau)\in\Pi\times\mathcal{T}$, where

$$
R_{n0}(x,\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[w(X_i,x)I_p\psi_\tau(Y_i,X_i,\theta_0)-G^\top(x,\theta_0(\tau))\Delta^{-1}(\theta_0(\tau))g(X_i,\theta_0(\tau))\psi_\tau(W_i,\theta_0)\big]
$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{P}_\tau w(X_i, x) \left(\psi_\tau(Y_i, X_i, \theta_0) - \frac{1}{\sqrt{n}} \dot{m}(X_i, \theta_0(\tau)) d(X_i, \tau) \right)$$
$$+ \frac{1}{n} \sum_{i=1}^{n} \mathcal{P}_\tau w(X_i, x) \dot{m}(X_i, \theta_0(\tau)) d(X_i, \tau).$$

Under H_{1n}, the first term on the right-hand side converges weakly to the zero-mean Gaussian process R_∞ as defined in Theorem 3.1. The second term converges in probability to $D(x, \tau)$ uniformly in $(x, \tau) \in \Pi \times \mathcal{T}$, since $\{X \to d(X, \tau) w(X, x) \dot{m}(X, \theta_0(\tau)) : (x, \tau) \in \Pi \times \mathcal{T}\}$ and $\{X \to d(X, \tau) g(X, \theta_0(\tau)) \dot{m}(X, \theta_0(\tau)) : \tau \in \mathcal{T}\}$ are Glivenko–Cantelli as can be shown in a similar way as Lemma A.3. $\mathcal{Q.E.D}$.

Proof of Theorem 4.1. Denote $\breve{R}_n^*(x, \tau) := (1/\sqrt{n}) \sum_{i=1}^{n} \psi_\tau(Y_i, X_i, \hat{\theta}_n) w(X_i, x) V_i$ and $\breve{S}_n^*(\tau) := (1/\sqrt{n}) \sum_{i=1}^{n} \hat{g}_n(X_i, \hat{\theta}_n(\tau)) \psi_\tau(Y_i, X_i, \hat{\theta}_n) V_i$. As in Theorem 3.1, we have

$$\hat{R}_n^*(x, \tau) = \breve{R}_n^*(x, \tau) - \hat{G}_n^\top(x, \theta_n(\tau)) \hat{\Delta}_n^{-1}(\theta_n(\tau)) \breve{S}_n^*(\tau).$$

Define $\mathcal{M} := \{(Y, X, V) \to \psi_\tau(Y, X, \theta) w(X, x) V : (x, \tau) \in \Pi \times \mathcal{T} \text{ and } \theta \in \mathcal{B}\}$ and $\mathcal{N} := \{(Y, X, V) \to \psi_\tau(Y, X, \theta) \hat{f}_h(X, \theta) \dot{m}_k(X, \theta(\tau)) \dot{m}_l(X, \theta(\tau)) V : \tau \in \mathcal{T} \text{ and } \theta \in \mathcal{B}\}$. Then, using similar arguments as in Lemma A.3, we have that $\mathcal{M}$ and $\mathcal{N}$ are both Donsker classes. From the consistency of $\hat{\theta}_n$ and the stochastic equicontinuity, we have

$$\sup_{(x,\tau) \in \Pi \times \mathcal{T}} \left| \breve{R}_n^*(x, \tau) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(Y_i, X_i, \theta_0) w(X_i, x) V_i \right| = o_p(1),$$

and

$$\sup_{\tau \in \mathcal{T}} \left| \breve{S}_n^*(\tau) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(X_i, \theta_0(\tau)) \psi_\tau(Y_i, X_i, \theta_0) V_i \right| = o_p(1).$$

Theorem 4.1(i) then follows from the multiplier CLT (see Theorem 2.9.6 of Van Der Vaart and Wellner (2023)) and the continuous mapping theorem. Theorem 4.1(ii) can be shown using similar arguments as those in the first part. The only difference is that we replace θ_0 with θ^*, and $g(X_i, \theta(\tau))$ with $g^*(X_i, \theta(\tau))$ defined above Theorem 3.2. $\mathcal{Q.E.D}$.

C Additional simulation results

C.1 Additional simulation results with $d = 2$

In this section, we present the simulation results for the CvM tests equipped with the indicator weight function, as well as for CvM tests with the exponential weight function and the Laplace-type integrating measure. The settings of the Monte Carlo experiments are the same as in Section 5.2. Tables C.1 and C.3 below present the rejection probabilities of the

CvM tests equipped with the indicator weight function under DGPs 1–10. Tables C.2 and C.4 present the same results for the tests equipped with the exponential weight function and the Laplace-type integrating measure. The null hypothesis corresponding to Tables C.1 and C.2 is (1) with $m(X,\theta(\tau)) = \theta^0(\tau) + X^\top\theta^1(\tau)$, while the null hypothesis corresponding to Tables C.3 and C.4 is (1) with $m(X,\theta(\tau)) = \theta^0(\tau) + \exp(X^\top\theta^1(\tau))$, where $\theta(\tau) = (\theta^0(\tau), \theta^1(\tau)^\top)^\top$.

Our findings from Tables C.1–C.4 are as follows. First, the empirical levels of the proposed test statistics are close to the nominal levels for moderately large sample sizes. This holds for both types of weight functions and null hypotheses we consider. Notably, the non-projection-based tests are significantly oversized when the sample size is small, suggesting the need to introduce the projection device. Second, the empirical power of the proposed test increases with sample size, as expected. For the indicator weight function case, the power of the proposed test against the linearity null hypothesis is higher than the no-gradient test when $n = 50$. For moderately large sample sizes, the two tests have comparable empirical power. Finally, the CvM functional with nuclear norm typically exhibits higher power than the functional with spectral norm. This pattern is especially obvious for gradient-based CvM tests with exponential weight functions.

C.2 Simulation results with $d = 6$

This subsection presents simulation results to test the null hypothesis (1) with a relatively high-dimensional covariate X. We generate data from the following models:

- DGP 1′: $Y_i = 1 + X_i^\top\theta_0 + \epsilon_i$,
- DGP 2′: $Y_i = 1 + X_i^\top\theta_0 + e_i$,
- DGP 3′: $Y_i = 1 + X_i^\top\theta_0 + 0.25(\sum_{j=1}^6 X_{ji}^2 + \sum_{j=1}^6\sum_{k\neq j}^6 X_{ji}X_{ki}) + \epsilon_i$,
- DGP 4′: $Y_i = 1 + X_i^\top\theta_0 + \exp(-0.5(1 + X_i^\top\theta_0)) + \epsilon_i$,
- DGP 5′: $Y_i = 1 + X_i^\top\theta_0 + \sqrt{|X_i^\top\theta_0|}\,\mathrm{sgn}(X_i^\top\theta_0) + \epsilon_i$,
- DGP 6′: $Y_i = 1 + X_i^\top\theta_0 + \sin(X_i^\top\theta_0) + \epsilon_i$,

where $X_i = (X_{1i}, \ldots, X_{6i})^\top$ is a 6-dimensional random vector with i.i.d. $N(0,1)$ margins, and θ_0 is a 6-dimensional vector of ones. The ϵ_i and e_i are i.i.d. $N(0,1)$ and $\exp(1)$ random variables, respectively, and X_i, ϵ_i and e_i are mutually independent. The null hypothesis is given by (1), with $m(X,\theta_0(\tau)) = \theta_0^0(\tau) + X^\top\theta_0^1(\tau)$, $\theta_0(\tau) = (\theta_0^0(\tau), \theta_0^1(\tau)^\top)^\top$ being the unknown parameters, and $\tau \in \mathcal{T} = [0.10, 0.90]$. Clearly, this hypothesis holds under DGPs 1′ and 2′. The rest of the Monte Carlo parameters are the same as Section 5.2.

Table C.5 gives the rejection frequencies of the CvM tests under DGPs 1′–6′ with projection-based indicator weight function. From the results of DGPs 1′ and 2′, we see that the proposed test exhibits a good level of accuracy when the sample size $n \geq 100$, whereas

Table C.1: Rejection frequencies of the CvM tests with the linear projection-based weights under DGPs 1–6.

	$n=50$			$n=100$			$n=200$		
DGP 1	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{lin,F}}$	0.092	0.047	0.007	0.105	0.058	0.012	0.110	0.056	0.019
$CvM_n^{\text{lin,E}}$	0.085	0.040	0.012	0.104	0.053	0.009	0.111	0.056	0.018
$CvM_n^{\text{lin,F,np}}$	0.199	0.086	0.018	0.151	0.073	0.019	0.128	0.066	0.020
$CvM_n^{\text{lin,E,np}}$	0.149	0.071	0.018	0.130	0.065	0.013	0.123	0.065	0.017
$CvM_n^{\text{lin,ng}}$	0.088	0.038	0.006	0.096	0.048	0.013	0.105	0.061	0.021
DGP 2									
$CvM_n^{\text{lin,F}}$	0.119	0.056	0.014	0.116	0.055	0.014	0.094	0.049	0.014
$CvM_n^{\text{lin,E}}$	0.118	0.055	0.014	0.116	0.052	0.013	0.099	0.049	0.017
$CvM_n^{\text{lin,F,np}}$	0.239	0.124	0.040	0.156	0.085	0.027	0.125	0.061	0.016
$CvM_n^{\text{lin,E,np}}$	0.173	0.093	0.029	0.145	0.077	0.020	0.115	0.054	0.021
$CvM_n^{\text{lin,ng}}$	0.101	0.046	0.011	0.104	0.054	0.013	0.097	0.048	0.016
DGP 3									
$CvM_n^{\text{lin,F}}$	0.653	0.513	0.283	0.938	0.875	0.723	1.000	1.000	0.984
$CvM_n^{\text{lin,E}}$	0.653	0.529	0.286	0.942	0.876	0.714	1.000	1.000	0.981
$CvM_n^{\text{lin,F,np}}$	0.794	0.664	0.393	0.961	0.922	0.787	1.000	1.000	0.988
$CvM_n^{\text{lin,E,np}}$	0.754	0.642	0.380	0.958	0.921	0.760	1.000	0.999	0.985
$CvM_n^{\text{lin,ng}}$	0.645	0.501	0.258	0.939	0.880	0.718	1.000	0.999	0.985
DGP 4									
$CvM_n^{\text{lin,F}}$	0.234	0.132	0.032	0.384	0.257	0.107	0.666	0.540	0.344
$CvM_n^{\text{lin,E}}$	0.220	0.120	0.027	0.357	0.243	0.095	0.621	0.505	0.278
$CvM_n^{\text{lin,F,np}}$	0.396	0.229	0.071	0.472	0.334	0.147	0.714	0.589	0.387
$CvM_n^{\text{lin,E,np}}$	0.321	0.180	0.051	0.409	0.281	0.115	0.666	0.539	0.312
$CvM_n^{\text{lin,ng}}$	0.209	0.119	0.030	0.373	0.240	0.100	0.642	0.527	0.306
DGP 5									
$CvM_n^{\text{lin,F}}$	0.190	0.103	0.026	0.335	0.206	0.091	0.643	0.504	0.260
$CvM_n^{\text{lin,E}}$	0.203	0.112	0.032	0.327	0.210	0.079	0.630	0.495	0.260
$CvM_n^{\text{lin,F,np}}$	0.336	0.188	0.055	0.430	0.278	0.113	0.687	0.543	0.284
$CvM_n^{\text{lin,E,np}}$	0.269	0.168	0.048	0.380	0.251	0.099	0.656	0.518	0.264
$CvM_n^{\text{lin,ng}}$	0.187	0.089	0.021	0.347	0.216	0.088	0.671	0.533	0.289
DGP 6									
$CvM_n^{\text{lin,F}}$	0.332	0.213	0.073	0.638	0.477	0.245	0.932	0.866	0.680
$CvM_n^{\text{lin,E}}$	0.336	0.220	0.067	0.615	0.479	0.252	0.929	0.859	0.688
$CvM_n^{\text{lin,F,np}}$	0.504	0.352	0.119	0.709	0.558	0.293	0.943	0.897	0.731
$CvM_n^{\text{lin,E,np}}$	0.449	0.291	0.100	0.678	0.532	0.268	0.933	0.883	0.708
$CvM_n^{\text{lin,ng}}$	0.325	0.201	0.052	0.640	0.492	0.251	0.954	0.894	0.722

Table C.2: Rejection frequencies of the CvM tests with the Laplace weights under DGPs 1–6.

	$n = 50$			$n = 100$			$n = 200$		
DGP 1	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{lap,F}}$	0.089	0.040	0.008	0.103	0.057	0.014	0.113	0.064	0.022
$CvM_n^{\text{lap,E}}$	0.078	0.040	0.011	0.109	0.053	0.009	0.112	0.067	0.022
$CvM_n^{\text{lap,F,np}}$	0.115	0.054	0.013	0.113	0.061	0.017	0.113	0.065	0.022
$CvM_n^{\text{lap,E,np}}$	0.096	0.043	0.014	0.108	0.053	0.009	0.111	0.070	0.020
$CvM_n^{\text{lap,ng}}$	0.085	0.041	0.008	0.099	0.059	0.012	0.111	0.063	0.024
DGP 2									
$CvM_n^{\text{lap,F}}$	0.122	0.055	0.012	0.114	0.051	0.011	0.104	0.052	0.014
$CvM_n^{\text{lap,E}}$	0.111	0.058	0.019	0.116	0.059	0.011	0.108	0.054	0.017
$CvM_n^{\text{lap,F,np}}$	0.151	0.075	0.014	0.121	0.058	0.014	0.110	0.056	0.014
$CvM_n^{\text{lap,E,np}}$	0.126	0.068	0.020	0.124	0.059	0.009	0.106	0.057	0.016
$CvM_n^{\text{lap,ng}}$	0.125	0.056	0.014	0.113	0.052	0.014	0.094	0.051	0.015
DGP 3									
$CvM_n^{\text{lap,F}}$	0.580	0.436	0.196	0.892	0.830	0.626	0.999	0.993	0.967
$CvM_n^{\text{lap,E}}$	0.600	0.471	0.242	0.923	0.837	0.647	0.999	0.998	0.969
$CvM_n^{\text{lap,F,np}}$	0.645	0.507	0.253	0.914	0.853	0.653	0.999	0.995	0.972
$CvM_n^{\text{lap,E,np}}$	0.651	0.505	0.274	0.931	0.864	0.664	0.999	0.997	0.972
$CvM_n^{\text{lap,ng}}$	0.639	0.499	0.259	0.925	0.870	0.712	1.000	0.999	0.981
DGP 4									
$CvM_n^{\text{lap,F}}$	0.194	0.106	0.029	0.328	0.218	0.082	0.577	0.448	0.251
$CvM_n^{\text{lap,E}}$	0.194	0.104	0.031	0.312	0.195	0.072	0.546	0.410	0.219
$CvM_n^{\text{lap,F,np}}$	0.241	0.133	0.033	0.362	0.240	0.092	0.602	0.464	0.264
$CvM_n^{\text{lap,E,np}}$	0.216	0.114	0.033	0.335	0.207	0.082	0.564	0.417	0.229
$CvM_n^{\text{lap,ng}}$	0.206	0.116	0.029	0.358	0.233	0.091	0.623	0.512	0.283
DGP 5									
$CvM_n^{\text{lap,F}}$	0.229	0.119	0.033	0.407	0.293	0.114	0.735	0.602	0.369
$CvM_n^{\text{lap,E}}$	0.196	0.107	0.029	0.349	0.227	0.090	0.657	0.508	0.246
$CvM_n^{\text{lap,F,np}}$	0.278	0.137	0.043	0.430	0.303	0.117	0.749	0.614	0.369
$CvM_n^{\text{lap,E,np}}$	0.214	0.123	0.028	0.352	0.231	0.085	0.660	0.512	0.249
$CvM_n^{\text{lap,ng}}$	0.219	0.106	0.032	0.384	0.252	0.100	0.706	0.561	0.318
DGP 6									
$CvM_n^{\text{lap,F}}$	0.388	0.262	0.091	0.709	0.565	0.341	0.956	0.924	0.812
$CvM_n^{\text{lap,E}}$	0.338	0.200	0.069	0.617	0.460	0.230	0.927	0.859	0.663
$CvM_n^{\text{lap,F,np}}$	0.453	0.304	0.104	0.727	0.582	0.352	0.956	0.928	0.819
$CvM_n^{\text{lap,E,np}}$	0.365	0.222	0.078	0.626	0.474	0.230	0.926	0.867	0.670
$CvM_n^{\text{lap,ng}}$	0.380	0.245	0.077	0.698	0.545	0.305	0.956	0.920	0.793

Table C.3: Rejection frequencies of the CvM tests with the linear projection-based weights.

	$n = 50$			$n = 100$			$n = 200$		
DGP 7	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{lin,F}}$	0.116	0.053	0.007	0.116	0.058	0.015	0.097	0.054	0.014
$CvM_n^{\text{lin,E}}$	0.112	0.060	0.010	0.113	0.052	0.018	0.096	0.052	0.012
$CvM_n^{\text{lin,F,np}}$	0.161	0.080	0.011	0.122	0.065	0.018	0.106	0.059	0.015
$CvM_n^{\text{lin,E,np}}$	0.139	0.062	0.013	0.118	0.063	0.020	0.101	0.060	0.015
$CvM_n^{\text{lin,ng}}$	0.111	0.056	0.008	0.108	0.052	0.010	0.106	0.055	0.011
DGP 8									
$CvM_n^{\text{lin,F}}$	0.096	0.050	0.009	0.122	0.063	0.012	0.105	0.051	0.010
$CvM_n^{\text{lin,E}}$	0.103	0.044	0.008	0.122	0.061	0.013	0.105	0.054	0.014
$CvM_n^{\text{lin,F,np}}$	0.138	0.065	0.014	0.132	0.076	0.016	0.109	0.058	0.011
$CvM_n^{\text{lin,E,np}}$	0.113	0.057	0.014	0.131	0.068	0.016	0.115	0.059	0.014
$CvM_n^{\text{lin,ng}}$	0.098	0.055	0.013	0.127	0.068	0.013	0.094	0.044	0.010
DGP 9									
$CvM_n^{\text{lin,F}}$	0.267	0.167	0.054	0.491	0.381	0.196	0.759	0.688	0.471
$CvM_n^{\text{lin,E}}$	0.272	0.172	0.060	0.479	0.357	0.178	0.734	0.651	0.431
$CvM_n^{\text{lin,F,np}}$	0.322	0.219	0.086	0.528	0.415	0.222	0.775	0.697	0.499
$CvM_n^{\text{lin,E,np}}$	0.286	0.202	0.070	0.496	0.377	0.200	0.748	0.671	0.457
$CvM_n^{\text{lin,ng}}$	0.277	0.183	0.073	0.539	0.419	0.228	0.831	0.739	0.540
DGP 10									
$CvM_n^{\text{lin,F}}$	0.220	0.134	0.046	0.337	0.242	0.100	0.569	0.448	0.264
$CvM_n^{\text{lin,E}}$	0.227	0.144	0.053	0.374	0.257	0.113	0.598	0.496	0.297
$CvM_n^{\text{lin,F,np}}$	0.286	0.186	0.077	0.410	0.297	0.132	0.627	0.524	0.331
$CvM_n^{\text{lin,E,np}}$	0.260	0.180	0.078	0.421	0.308	0.145	0.648	0.563	0.349
$CvM_n^{\text{lin,ng}}$	0.203	0.126	0.042	0.312	0.199	0.084	0.502	0.370	0.216

the no-gradient test is somewhat undersized. Moreover, the test based on the non-projection process appears to be severely oversized, although it tends to exhibit the correct size as the sample size increases. In terms of the power performance, both CvM_n^{lin} and $CvM_n^{\text{lin,ng}}$ have high empirical power under DGPs 3′ and 4′, which is similar to the pattern found in Table C.1. However, for DGPs 5′ and 6′, the proposed gradient-weighted test appears to be more powerful in detecting the deviation from the null. These illustrate the advantage of our test when the dimension of the covariate X is moderately large. Finally, since the $CvM_n^{\text{lin,np}}$ test fails to control the Type I error reasonably, we do not compare its power performance with that of the other tests.

Tables C.6 and C.7 display the rejection frequencies of the CvM tests with exponential weight function, Gaussian-type and Laplace-type integrating measure, respectively. It is seen that all the CvM tests are undersized in this case, even when the sample size is relatively large. Regarding power performance, the no-gradient test $CvM_n^{\text{gau,ng}}$ achieves the best empirical

Table C.4: Rejection frequencies of the CvM tests with the Laplace weights.

	$n = 50$			$n = 100$			$n = 200$		
DGP 7	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{lap,F}}$	0.099	0.041	0.006	0.112	0.065	0.014	0.107	0.053	0.013
$CvM_n^{\text{lap,E}}$	0.083	0.035	0.008	0.111	0.055	0.013	0.103	0.052	0.012
$CvM_n^{\text{lap,F,np}}$	0.132	0.061	0.008	0.121	0.068	0.020	0.113	0.059	0.016
$CvM_n^{\text{lap,E,np}}$	0.101	0.043	0.007	0.118	0.060	0.017	0.108	0.056	0.014
$CvM_n^{\text{lap,ng}}$	0.096	0.042	0.007	0.121	0.057	0.013	0.105	0.054	0.009
DGP 8									
$CvM_n^{\text{lap,F}}$	0.097	0.046	0.008	0.113	0.052	0.012	0.104	0.044	0.010
$CvM_n^{\text{lap,E}}$	0.089	0.044	0.007	0.122	0.054	0.013	0.109	0.049	0.014
$CvM_n^{\text{lap,F,np}}$	0.135	0.070	0.008	0.126	0.062	0.014	0.109	0.045	0.010
$CvM_n^{\text{lap,E,np}}$	0.095	0.049	0.008	0.130	0.059	0.011	0.110	0.050	0.013
$CvM_n^{\text{lap,ng}}$	0.106	0.042	0.008	0.108	0.062	0.014	0.095	0.047	0.009
DGP 9									
$CvM_n^{\text{lap,F}}$	0.223	0.129	0.039	0.425	0.298	0.131	0.699	0.603	0.393
$CvM_n^{\text{lap,E}}$	0.188	0.108	0.030	0.388	0.243	0.106	0.650	0.541	0.311
$CvM_n^{\text{lap,F,np}}$	0.271	0.162	0.048	0.464	0.324	0.153	0.713	0.619	0.408
$CvM_n^{\text{lap,E,np}}$	0.213	0.120	0.036	0.400	0.263	0.113	0.658	0.551	0.329
$CvM_n^{\text{lap,ng}}$	0.237	0.139	0.044	0.444	0.320	0.140	0.738	0.614	0.412
DGP 10									
$CvM_n^{\text{lap,F}}$	0.181	0.113	0.037	0.291	0.180	0.072	0.494	0.373	0.201
$CvM_n^{\text{lap,E}}$	0.149	0.090	0.023	0.234	0.142	0.050	0.412	0.290	0.153
$CvM_n^{\text{lap,F,np}}$	0.230	0.138	0.047	0.341	0.228	0.086	0.552	0.449	0.262
$CvM_n^{\text{lap,E,np}}$	0.178	0.108	0.037	0.290	0.183	0.072	0.493	0.387	0.221
$CvM_n^{\text{lap,ng}}$	0.173	0.113	0.029	0.277	0.159	0.065	0.465	0.337	0.176

power across all the DGPs we consider. It also outperforms our CvM_n^{lin} test under DGPs 5′ and 6′. Comparing the simulation results across different weight functions, we see that none of the weight functions (or integration measures) is superior to the others for all models when $d = 6$.

Table C.5: Rejection frequencies of the CvM tests with indicator weights under DGPs $1'$–$6'$.

	$n=50$			$n=100$			$n=200$		
DGP $1'$	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{lin,F}}$	0.114	0.039	0.000	0.103	0.053	0.004	0.095	0.039	0.004
$CvM_n^{\text{lin,E}}$	0.071	0.028	0.000	0.092	0.045	0.007	0.094	0.046	0.015
$CvM_n^{\text{lin,F,np}}$	0.887	0.696	0.228	0.401	0.223	0.060	0.178	0.096	0.023
$CvM_n^{\text{lin,E,np}}$	0.268	0.113	0.013	0.158	0.077	0.018	0.115	0.055	0.020
$CvM_n^{\text{lin,ng}}$	0.088	0.020	0.000	0.071	0.025	0.002	0.066	0.032	0.004
DGP $2'$									
$CvM_n^{\text{lin,F}}$	0.127	0.045	0.004	0.110	0.049	0.003	0.105	0.035	0.003
$CvM_n^{\text{lin,E}}$	0.075	0.028	0.005	0.078	0.037	0.006	0.088	0.034	0.008
$CvM_n^{\text{lin,F,np}}$	0.890	0.693	0.245	0.437	0.238	0.070	0.212	0.113	0.023
$CvM_n^{\text{lin,E,np}}$	0.252	0.119	0.021	0.150	0.075	0.020	0.126	0.059	0.011
$CvM_n^{\text{lin,ng}}$	0.086	0.022	0.001	0.073	0.024	0.000	0.079	0.025	0.003
DGP $3'$									
$CvM_n^{\text{lin,F}}$	0.758	0.539	0.192	0.991	0.977	0.850	1	1	1
$CvM_n^{\text{lin,E}}$	0.598	0.422	0.156	0.973	0.934	0.742	1	1	1
$CvM_n^{\text{lin,F,np}}$	0.999	0.990	0.822	0.999	0.997	0.982	1	1	1
$CvM_n^{\text{lin,E,np}}$	0.890	0.744	0.416	0.993	0.981	0.915	1	1	1
$CvM_n^{\text{lin,ng}}$	0.821	0.602	0.229	0.996	0.989	0.929	1	1	1
DGP $4'$									
$CvM_n^{\text{lin,F}}$	0.412	0.257	0.051	0.781	0.634	0.333	0.989	0.965	0.883
$CvM_n^{\text{lin,E}}$	0.286	0.174	0.038	0.687	0.539	0.267	0.974	0.951	0.839
$CvM_n^{\text{lin,F,np}}$	0.981	0.911	0.544	0.962	0.898	0.659	0.997	0.988	0.958
$CvM_n^{\text{lin,E,np}}$	0.571	0.382	0.131	0.820	0.711	0.430	0.991	0.982	0.930
$CvM_n^{\text{lin,ng}}$	0.408	0.221	0.039	0.788	0.614	0.306	0.985	0.970	0.878
DGP $5'$									
$CvM_n^{\text{lin,F}}$	0.137	0.047	0.002	0.168	0.083	0.011	0.254	0.156	0.050
$CvM_n^{\text{lin,E}}$	0.093	0.035	0.002	0.161	0.080	0.016	0.298	0.184	0.070
$CvM_n^{\text{lin,F,np}}$	0.910	0.697	0.263	0.544	0.320	0.091	0.435	0.299	0.098
$CvM_n^{\text{lin,E,np}}$	0.306	0.145	0.032	0.261	0.147	0.049	0.415	0.289	0.111
$CvM_n^{\text{lin,ng}}$	0.105	0.031	0.001	0.129	0.039	0.003	0.205	0.106	0.026
DGP $6'$									
$CvM_n^{\text{lin,F}}$	0.161	0.044	0.003	0.240	0.112	0.028	0.466	0.301	0.118
$CvM_n^{\text{lin,E}}$	0.112	0.043	0.006	0.226	0.122	0.029	0.500	0.369	0.167
$CvM_n^{\text{lin,F,np}}$	0.933	0.760	0.305	0.636	0.440	0.144	0.684	0.503	0.241
$CvM_n^{\text{lin,E,np}}$	0.360	0.197	0.045	0.391	0.232	0.077	0.667	0.518	0.254
$CvM_n^{\text{lin,ng}}$	0.134	0.037	0.004	0.187	0.083	0.005	0.400	0.225	0.051

Table C.6: Rejection frequencies of the CvM tests with exponential weights under DGPs 1′–6′.

	$n = 50$			$n = 100$			$n = 200$		
DGP 1′	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{gau,F}}$	0.066	0.006	0.000	0.056	0.008	0.000	0.046	0.013	0.001
$CvM_n^{\text{gau,E}}$	0.010	0.004	0.000	0.023	0.007	0.000	0.035	0.014	0.002
$CvM_n^{\text{gau,F,np}}$	0.153	0.025	0.000	0.076	0.011	0.000	0.053	0.015	0.001
$CvM_n^{\text{gau,E,np}}$	0.011	0.004	0.000	0.020	0.009	0.000	0.034	0.016	0.001
$CvM_n^{\text{gau,ng}}$	0.082	0.009	0.000	0.077	0.019	0.001	0.066	0.019	0.002
DGP 2′									
$CvM_n^{\text{gau,F}}$	0.076	0.010	0.000	0.066	0.016	0.001	0.062	0.019	0.001
$CvM_n^{\text{gau,E}}$	0.019	0.003	0.001	0.026	0.011	0.000	0.045	0.020	0.000
$CvM_n^{\text{gau,F,np}}$	0.168	0.027	0.000	0.081	0.020	0.001	0.068	0.018	0.001
$CvM_n^{\text{gau,E,np}}$	0.018	0.005	0.000	0.024	0.010	0.001	0.039	0.020	0.001
$CvM_n^{\text{gau,ng}}$	0.101	0.020	0.001	0.079	0.024	0.002	0.074	0.032	0.006
DGP 3′									
$CvM_n^{\text{gau,F}}$	0.530	0.215	0.008	0.944	0.816	0.425	1	1	0.997
$CvM_n^{\text{gau,E}}$	0.245	0.105	0.012	0.829	0.692	0.375	0.999	0.998	0.986
$CvM_n^{\text{gau,F,np}}$	0.710	0.341	0.023	0.963	0.866	0.490	1	1	0.998
$CvM_n^{\text{gau,E,np}}$	0.294	0.148	0.019	0.842	0.721	0.415	0.999	0.998	0.990
$CvM_n^{\text{gau,ng}}$	0.757	0.469	0.093	0.995	0.959	0.788	1	1	1
DGP 4′									
$CvM_n^{\text{gau,F}}$	0.193	0.030	0.000	0.450	0.230	0.044	0.873	0.727	0.410
$CvM_n^{\text{gau,E}}$	0.054	0.017	0.000	0.246	0.135	0.027	0.720	0.586	0.302
$CvM_n^{\text{gau,F,np}}$	0.364	0.090	0.001	0.558	0.290	0.068	0.909	0.789	0.498
$CvM_n^{\text{gau,E,np}}$	0.055	0.015	0.000	0.241	0.129	0.032	0.710	0.577	0.289
$CvM_n^{\text{gau,ng}}$	0.311	0.074	0.002	0.590	0.363	0.114	0.922	0.839	0.600
DGP 5′									
$CvM_n^{\text{gau,F}}$	0.096	0.008	0.001	0.115	0.025	0.001	0.265	0.112	0.018
$CvM_n^{\text{gau,E}}$	0.016	0.003	0.000	0.047	0.016	0.002	0.159	0.087	0.020
$CvM_n^{\text{gau,F,np}}$	0.200	0.035	0.001	0.149	0.038	0.004	0.267	0.117	0.018
$CvM_n^{\text{gau,E,np}}$	0.018	0.004	0.000	0.050	0.017	0.003	0.165	0.089	0.022
$CvM_n^{\text{gau,ng}}$	0.128	0.018	0.000	0.151	0.045	0.007	0.292	0.142	0.032
DGP 6′									
$CvM_n^{\text{gau,F}}$	0.147	0.020	0.001	0.287	0.106	0.007	0.727	0.515	0.200
$CvM_n^{\text{gau,E}}$	0.027	0.009	0.000	0.108	0.047	0.010	0.442	0.283	0.096
$CvM_n^{\text{gau,F,np}}$	0.295	0.068	0.003	0.335	0.123	0.017	0.738	0.522	0.200
$CvM_n^{\text{gau,E,np}}$	0.030	0.011	0.000	0.098	0.049	0.009	0.439	0.286	0.094
$CvM_n^{\text{gau,ng}}$	0.179	0.036	0.002	0.323	0.141	0.020	0.734	0.542	0.217

Table C.7: Rejection frequencies of the CvM tests with the Laplace weights under DGPs 1′–6′.

	$n = 50$			$n = 100$			$n = 200$		
DGP 1′	10%	5%	1%	10%	5%	1%	10%	5%	1%
$CvM_n^{\text{lap,F}}$	0.047	0.002	0.000	0.034	0.002	0.000	0.027	0.007	0.001
$CvM_n^{\text{lap,E}}$	0.005	0.003	0.000	0.013	0.002	0.000	0.026	0.011	0.001
$CvM_n^{\text{lap,F,np}}$	0.160	0.012	0.000	0.058	0.004	0.000	0.032	0.007	0.000
$CvM_n^{\text{lap,E,np}}$	0.006	0.003	0.000	0.012	0.004	0.000	0.026	0.011	0.002
$CvM_n^{\text{lap,ng}}$	0.051	0.000	0.000	0.039	0.005	0.001	0.034	0.011	0.001
DGP 2′									
$CvM_n^{\text{lap,F}}$	0.050	0.004	0.000	0.044	0.003	0.000	0.035	0.006	0.000
$CvM_n^{\text{lap,E}}$	0.009	0.001	0.000	0.014	0.004	0.000	0.026	0.008	0.000
$CvM_n^{\text{lap,F,np}}$	0.174	0.013	0.000	0.065	0.014	0.000	0.048	0.009	0.001
$CvM_n^{\text{lap,E,np}}$	0.009	0.004	0.000	0.018	0.005	0.000	0.027	0.010	0.000
$CvM_n^{\text{lap,ng}}$	0.057	0.006	0.000	0.044	0.008	0.000	0.042	0.010	0.000
DGP 3′									
$CvM_n^{\text{lap,F}}$	0.479	0.142	0.002	0.941	0.787	0.359	1	1	0.994
$CvM_n^{\text{lap,E}}$	0.205	0.082	0.005	0.847	0.705	0.377	1	1	0.995
$CvM_n^{\text{lap,F,np}}$	0.747	0.336	0.008	0.965	0.850	0.442	1	1	0.996
$CvM_n^{\text{lap,E,np}}$	0.263	0.114	0.015	0.860	0.737	0.414	1	1	0.994
$CvM_n^{\text{lap,ng}}$	0.726	0.365	0.045	0.989	0.957	0.737	1	1	1
DGP 4′									
$CvM_n^{\text{lap,F}}$	0.151	0.015	0.000	0.439	0.194	0.025	0.904	0.770	0.427
$CvM_n^{\text{lap,E}}$	0.043	0.007	0.000	0.253	0.127	0.031	0.792	0.666	0.373
$CvM_n^{\text{lap,F,np}}$	0.415	0.081	0.000	0.579	0.283	0.050	0.938	0.828	0.495
$CvM_n^{\text{lap,E,np}}$	0.042	0.009	0.000	0.241	0.132	0.036	0.799	0.682	0.376
$CvM_n^{\text{lap,ng}}$	0.219	0.035	0.000	0.539	0.292	0.061	0.938	0.844	0.581
DGP 5′									
$CvM_n^{\text{lap,F}}$	0.059	0.004	0.000	0.072	0.015	0.000	0.191	0.072	0.008
$CvM_n^{\text{lap,E}}$	0.006	0.000	0.000	0.032	0.014	0.001	0.125	0.068	0.020
$CvM_n^{\text{lap,F,np}}$	0.212	0.023	0.001	0.119	0.020	0.001	0.212	0.079	0.009
$CvM_n^{\text{lap,E,np}}$	0.007	0.000	0.000	0.034	0.014	0.000	0.125	0.068	0.019
$CvM_n^{\text{lap,ng}}$	0.059	0.005	0.000	0.075	0.021	0.000	0.187	0.071	0.007
DGP 6′									
$CvM_n^{\text{lap,F}}$	0.082	0.008	0.001	0.189	0.047	0.002	0.562	0.341	0.084
$CvM_n^{\text{lap,E}}$	0.016	0.003	0.000	0.071	0.026	0.007	0.333	0.228	0.070
$CvM_n^{\text{lap,F,np}}$	0.289	0.042	0.002	0.263	0.074	0.004	0.600	0.361	0.092
$CvM_n^{\text{lap,E,np}}$	0.018	0.002	0.000	0.073	0.029	0.006	0.348	0.211	0.075
$CvM_n^{\text{lap,ng}}$	0.084	0.008	0.001	0.174	0.050	0.002	0.539	0.306	0.067