A consistent specification test for expectile models

Xiaojun Song*   Zixin Yang†
Peking University   Peking University

November 6, 2024

Abstract

In this article, we propose a nonparametric test for the correct specification of parametric expectile models over a continuum of expectile levels. The test is based on continuous functions of a residual-marked empirical process. We show that the test is consistent and has nontrivial power against a sequence of local alternatives approaching the null at a parametric rate. Since the limiting distribution of the test statistic is non-Gaussian, we propose a simple multiplier bootstrap procedure to approximate the critical values. A Monte Carlo study shows that the asymptotic results provide good approximations for small sample sizes.

Keywords: *Expectile regression, Specification tests, Marked empirical process, Multiplier bootstrap.*
JEL Classification: C12, C31, C52

*Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. Email: `sxj@gsm.pku.edu.cn`.

†Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing, 100871, China. Email: `yangzixin@stu.pku.edu.cn`.

1 Introduction

Regression beyond the mean of a conditional distribution has been the focus of a wide range of econometric analysis. One of the most popular tools is quantile regression (Koenker and Bassett, 1978), which allows researchers to fit flexible models to the entire conditional distribution function. Another frequently used approach is expectile regression introduced by Aigner et al. (1976) and Newey and Powell (1987). It is based on a novel concept called "expectiles", which are measures of location similar to quantiles but determined by tail expectations rather than tail probabilities. Precisely, given a continuous response variable Y and a regressor vector X, the αth conditional expectile of Y given $X = x$ is defined as

$$\mu_\alpha(x) = \arg\min_{\xi \in \mathbb{R}} \mathbb{E}\big[\rho_\alpha(Y - \xi) \mid X = x\big],$$

where $\rho_\alpha(\epsilon) = |\alpha - 1(\epsilon \leq 0)|\epsilon^2$ with $1(\cdot)$ being the indicator function. Clearly, the expectile regression can be interpreted as a least squares analogue of regression quantile estimation. Quantile regression is naturally more dominant in the literature due to its intuitive interpretation as the inverse of the distribution function. However, expectile regression also have some desired features. In particular, the expectile regression coefficient is easier to compute and reasonably efficient under normality conditions (Efron, 1991). The asymptotic covariance matrix can be estimated without the need of estimating the conditional density function in a nonparametric way. Therefore, expectile regression offers a convenient and relatively efficient method of summarizing the conditional response distribution.

Expectile regression has found wide applications in various fields, including risk assessment (Kuan et al., 2009; Kim and Lee, 2016; Xu et al., 2022), labor economics (Dawber et al., 2022; Bonaccolto-Töpfer and Bonaccolto, 2023) as well as forecasting performance evaluation (Guler et al., 2017). In the finance applications, the expectiles, also known as the expectile-Value at Risk, play a central role in the class of coherent risk measures as they provide a lower bound for elicitable risk measures. These features make the expectile a important tool for financial risk management and decision-making.

All the above mentioned work concentrated on parametric models. It is well known that inference procedures within parametric quantile models depend crucially on the validity of the specified parametric functional forms for the range of quantiles under consideration. Much the same can be expected for expectile regression. For instance, the conditional autoregressive expectile model proposed by Kuan et al. (2009), which has been widely used in assessing the tail risk of asset returns (see e.g. Xu et al., 2018), depends on the linear expectile specification. Once the linear form of the conditional expectile function is misspecified, the estimators would lead to misleading inferences and inaccurate predictions. Therefore, it is important to develop powerful tests for the correct specification of parametric conditional expectiles over a possibly continuous range of expectiles of interest. This is the main purpose of the present paper.

In this article, we propose omnibus specification tests of parametric expectile models over a continuum of expectile levels. The null hypothesis is $H_0 : \mu_\alpha(X) = m(X, \theta_0(\alpha))$ a.s. for some $\theta_0(\cdot)$ in the corresponding parameter space and for all $\alpha \in \mathcal{A}$, where $m(\cdot, \theta(\alpha))$ is a known function up to a finite dimensional vector $\theta(\alpha)$ and $\mathcal{A}$ is a compact subset of $(0, 1)$ which comprises the range of expectile levels of interest. In the context of quantile regression, such testing problem has been studied by Escanciano and Velasco (2010) and Escanciano and Goh (2014). However, to the best of our knowledge, no consistent test has been proposed where the targeted quantity is the conditional expectile function. Our test is based on the fact that H_0 can be characterized by the infinite set of conditional moment restrictions

$$\mathbb{E}\left[\left|1(Y \le m(X, \theta_0(\alpha))) - \alpha\right|(Y - m(X, \theta_0(\alpha))) \mid X\right] = 0, \text{ a.s.} \tag{1.1}$$

for some $\theta_0(\cdot) : \mathcal{A} \to \Theta \subset \mathbb{R}^p$ and for all $\alpha \in \mathcal{A}$.[1] The proposed tests are functionals of a expectile-marked empirical process that characterize condition (1.1). The asymptotic theory is largely complicated by the fact that (1.1) involves an infinite number of conditional moment restrictions, indexed by $\alpha \in \mathcal{A}$. We solve this technical difficulty using delicate weak convergence results for empirical processes similar to Escanciano and Velasco (2010). It turns out that the asymptotic null distributions of the test statistics depend on the specification under the null and the data generating process (DGP). Therefore, we propose to implement the test with the assistance of a multiplier bootstrap. As a byproduct, the sampling properties of expectile regression process under misspecification are also investigated. Such results may be of independent interest due to its usefulness in the construction of simultaneous (joint) confidence regions for population coefficient vector.

The rest of the paper is organized as follows. In Section 2, we introduce the residual-marked empirical process, which serves as the basic ingredient of the test statistic. The asymptotic distributions of the new tests under the null, fixed and local alternatives are established in Section 3. A multiplier bootstrap procedure for approximating the asymptotic null distribution of tests is also considered. In Section 4, we present a simulation exercise assessing the finite-sample performance of tests. Section 5 present the results of an application to modelling murder rate in the USA. The last section concludes.

[1] Consider the general expectile regression model $Y = \mu_\alpha(X) + \epsilon_\alpha$, where ϵ_α is the "expectile error" whose αth conditional expectile is 0. By Equation (2.7) in Newey and Powell (1987), $\mathbb{E}[|\alpha - 1(\epsilon_\alpha < 0)|\epsilon_\alpha|X] = 0$. Further, under H_0, $\mu_\alpha(X) = m(X, \theta_0(\alpha))$ for some $\theta_0(\alpha) \in \Theta$. Combining these results leads to (1.1).

2 Testing framework

Let $W = (Y, X')' \in \mathbb{R}^{1+d}$, where Y is the response variable and X is a d-dimensional vector of explanatory variables. Denote $\psi(W, \theta(\alpha), \alpha) = |1\{Y \leq m(X, \theta(\alpha))\} - \alpha|(Y - m(X, \theta(\alpha)))$. As mentioned in the introduction, we aim to test the null hypothesis

$$H_0 : \mathbb{P}\left(\mathbb{E}\left[\psi(W, \theta_0(\alpha), \alpha)\,|X\right] = 0\right) = 1 \;\text{ for some } \theta_0 \in \mathcal{B} \;\text{ and for all }\; \alpha \in \mathcal{A}, \tag{2.1}$$

(for some $\theta_0 \in \mathcal{B}$, or for some $\theta_0(\tau) \in \Theta$?)

against the nonparametric alternatives

$$H_1 : \mathbb{P}\left(\mathbb{E}\left[\psi(W, \theta(\alpha), \alpha)\,|X\right] \neq 0\right) > 0 \;\text{ for some } \alpha \in \mathcal{A} \;\text{ and for all }\; \theta(\alpha) \in \Theta \subset \mathbb{R}^p, \tag{2.2}$$

where $\mathcal{A}$ is a compact subinterval of $(0, 1)$, and $\mathcal{B}$ is a family of uniformly bounded functions from $\mathcal{A}$ to $\Theta \subset \mathbb{R}^p$. It is clear that when $\mathcal{A} = \{0.5\}$, testing H_0 reduces to the classical problem of testing a parametric conditional mean model.

Testing (2.1) is new and challenging since it involves an infinite number of conditional moments indexed by $\alpha \in \mathcal{A}$. To obtain a consistent test, in the spirits of Bierens (1982), Stute (1997) and Escanciano (2006*b*), we transform H_0 to an infinite number of unconditional moment restrictions over a parametric family of functions as follows:

$$\mathbb{E}\left[\psi(W, \theta_0(\alpha), \alpha) w(X, x)\right] = 0 \text{ for all } x \in \Pi, \tag{2.3}$$

where $\Pi \subseteq \mathbb{R}^s$ for some $s \in \mathbb{N}$ is a properly chosen space and the parametric family $\{w(\cdot, x) : x \in \Pi\}$ is a class of weight functions such that the equivalence between (2.1) and (2.3) holds. Commonly used weight functions include (i) the indicator weight function $w(X, x) = 1\{X \leq x\}$ and $s = d$, see, e.g., Stute (1997) and Koul and Stute (1999); (ii) the exponential weight function $w(X, x) = \exp\{\mathrm{i}x'X\}$ and $s = d$, where $\mathrm{i} = \sqrt{-1}$ denotes the imaginary unit, see, e.g., Bierens (1982) and Bierens (1990); and (iii) the projected indicator weight function $w(X, x) = 1\{\gamma'X \leq \delta\}$ and $s = d + 1$, where $x = (\gamma', \delta)' \in \Pi_{\text{pro}} = \mathbb{S}^d \times [-\infty, \infty]$, with $\mathbb{S}^d$ the unit ball in $\mathbb{R}^d$, i.e., $\mathbb{S}^d = \{\gamma \in \mathbb{R}^d : \|\gamma\| = 1\}$, see Escanciano (2006*a*).

Henceforth, for the sake of illustration, we focus on the indicator weight function $w(X, x) = 1\{X \leq x\}$ and $\Pi = \mathbb{R}^d$. Other weights can be analyzed similarly. Given a random sample $\{W_i = (Y_i, X_i')'\}_{i=1}^n$ and a parameter value $\theta \in \mathcal{B}$, we consider the expectile-marked empirical process indexed by $x \in \mathbb{R}^d$, $\alpha \in \mathcal{A}$ and $\theta \in \mathcal{B}$,

$$S_n(x, \alpha, \theta) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i, \theta(\alpha), \alpha) 1\{X_i \leq x\}.$$

Associated to S_n are the expectile-marked error and residual processes, respectively, defined by

$$R_n(x,\alpha) \equiv S_n(x,\alpha,\theta_0) \text{ and } \hat{R}_n(x,\alpha) \equiv S_n(x,\alpha,\hat{\theta}_n), \tag{2.4}$$

where $\hat{\theta}_n(\alpha)$ denotes a $\sqrt{n}$-consistent estimator for $\theta_0(\alpha)$. The null is likely to hold when the process $\hat{R}_n(x,\alpha)$ is close to zero for almost all $(x^\top,\alpha)^\top \in \mathbb{R}^d \times \mathcal{A}$.

In the following, we focus on the residual process $\hat{R}_n$ with $\hat{\theta}_n$ given by the expectile regression estimator for θ_0. This estimator is first proposed by Newey and Powell (1987) for linear expectile regression models. In our context, it is defined as

$$\hat{\theta}_n(\alpha) = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \rho_\alpha(Y_i - m(X_i,\theta)). \tag{2.5}$$

The estimator $\hat{\theta}_n(\alpha)$ is easy to compute due to the differentiability and convexity of the asymmetric square loss function $\rho_\alpha(\cdot)$. In the special case where $m(x,\theta) = x'\theta$, the expectile regression estimator of $\theta_0(\alpha)$ can be obtained by iterated weighted least squares method.

The test statistics are based on a distance from the standardized sample analogue of $\mathbb{E}[\psi(W,\theta_0(\alpha),\alpha)1(X \le x)]$ to zero, i.e., on a norm of $\hat{R}_n$. Following Escanciano and Velasco (2010), we consider the following types of norms: the Cramér-von Mises (CvM) functional

$$CvM_n := \int_{\mathcal{T}} \left|\hat{R}_n(x,\alpha)\right|^2 dF_n(x)dW(\alpha), \tag{2.6}$$

and the Kolmogorov-Smirnov (KS) functional

$$KS_n := \sup_{\alpha\in\mathcal{A}} \int_{\mathbb{R}^d} \left|\hat{R}_n(x,\alpha)\right|^2 dF_n(x), \tag{2.7}$$

where $\mathcal{T} := \mathbb{R}^d \times \mathcal{A}$, $F_n(x) = n^{-1}\sum_{i=1}^n 1(X_i \le x)$ is the empirical distribution function of X, and W is some integrating measure on $\mathcal{A}$. We reject the null hypothesis H_0 for "large" values of CvM_n and KS_n. Practical issues about the computation of the test statistic CvM_n and KS_n are deferred to Section 4.

3 Asymptotic theory

In what follows we establish the limiting distributions of the empirical process $\hat{R}_n(\cdot,\cdot)$ under the null hypothesis, fixed alternatives, and a sequence of local alternatives approaching H_0 at a parametric rate. We adopt the following notation. For a generic function $g : \mathcal{Z} \to \mathbb{R}$, define $\|g\|_{\mathcal{Z}} \equiv \sup_{z\in\mathcal{Z}} |g(z)|$. We study the weak convergence of $\hat{R}_n(\cdot,\cdot)$ as an element of $l^\infty(\mathcal{T})$, the space of real-valued functions that are uniformly bounded on $\mathcal{T}$, where, with slight abuse of notation, $\mathcal{T} \equiv \overline{\mathbb{R}}^d \times \mathcal{A}$. The space $l^\infty(\mathcal{T})$ is equipped with the supremum norm $\|\cdot\|_{\mathcal{T}}$. Let "$\Longrightarrow$" denote weak convergence on $(l^\infty(\mathcal{T}), \mathcal{B}_{d_\infty})$ 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 (1996). Define the family of conditional distributions $F(y|X) = \mathbb{P}(Y \leq y|X)$. Let $\mathcal{X}_X \subset \mathbb{R}^d$ be the support of X and $\mathcal{Y}_{\mathcal{A}} \equiv \{x'\theta_0(\alpha) : x \in \mathcal{X}_X, \alpha \in \mathcal{A}\}$.

3.1 Asymptotic null distribution

To study the asymptotic distribution of $\hat{R}_n$, we assume the following basic conditions.

Assumption 1. *(i) $\{W_i = (Y_i, X_i^\top)^\top\}_{i=1}^n$ is a sequence of i.i.d. random $(d+1)$-variates with $\mathbb{E}[\|W\|^{4+\epsilon}] < \infty$ for some $\epsilon > 0$; (ii) The conditional density $\{f(y|x) : x \in \mathbb{R}^d\}$ is continuous with respect to y and is uniformly bounded on $y \in \mathcal{Y}_{\mathcal{A}}$ and $x \in \mathcal{X}_X$.*

Assumption 2. *The parameter space Θ is compact in $\mathbb{R}^p$, and the true parameter $\theta_0(\alpha)$ belongs to the interior of Θ for each $\alpha \in \mathcal{A}$, and $\theta_0 \in \mathcal{B}$. The class $\mathcal{B}$ is such that*

$$\int_0^\infty (\log(N_{[\cdot]}(\delta, \mathcal{B}, \|\cdot\|_{\mathcal{B}})))^{1/2} d\delta < \infty. \tag{3.1}$$

Assumption 3. *(i) The function $m(\cdot, b)$ is twice continuously differentiable at each $b \in \Theta_0$ with Θ_0 the interior subset of Θ; the derivative $\dot{m}(X, b) \equiv (\partial/\partial b)m(X, b)$ satisfies that there exist F_X-measurable functions $M(\cdot)$ and $\bar{M}(\cdot)$, such that $\sup_{b\in\Theta} \|\dot{m}(x, b)\| \leq M(x)$ and for any x, b_1, b_2, $\|\dot{m}(x, b_1) - \dot{m}(x, b_2)\| \leq \bar{M}(x)\|b_1 - b_2\|$ with $\mathbb{E}|M(X)|^2 < \infty$ and $\mathbb{E}|\bar{M}(X)|^2 < \infty$; (ii) $\mathbb{E}\sup_{b\in\Theta} |m(X, b)|^2 < \infty$; (iii) For each $\theta \in \mathcal{B}$, $\mathbb{E}\big[\sup_{|\alpha_1-\alpha_2|\leq\delta} |m(X, \theta(\alpha_1)) - m(X, \theta(\alpha_2))|^2\big] < C\delta^2$; (iv) The matrix $\Delta(\theta(\alpha)) \equiv \mathbb{E}\left[|\alpha - 1(Y \leq m(X, \theta(\alpha)))|\dot{m}(X, \theta(\alpha))\dot{m}(X, \theta(\alpha))^\top\right]$ is nonsingular in a neighborhood of $\theta(\alpha) = \theta_0(\alpha)$.*

Assumption 1 is standard in the model checks literature. The i.i.d. assumption is introduced for analytical simplicity, and could be relaxed to a weaker martingale or mixing condition using similar techniques in Escanciano and Velasco (2010). Assumption 2 restricts the complexity of $\mathcal{B}$ in terms of bracketing number. If $m(\cdot, \theta(\alpha))$ is linear, we may take $\mathcal{B}$ as the class of all uniformly bounded monotone functions on $\mathcal{A}$, which trivially satisfies the condition 3.1, see Theorem 2.7.5 in Van Der Vaart (1998). Assumption 3 is standard in inference about nonlinear models. In particular, it guarantees that a asymptotic linear representation of the process $\sqrt{n}(\hat{\theta}_n(\alpha) - \theta_0(\alpha))$ can be obtained. For testing the correct specification of linear expectile models, Assumption 3(i)-(ii) will be vacuous and Assumption 3(iv) reduce to the nonsingularity of $\mathbb{E}[XX^\top]$.

We now examine the limit distribution of the process R_n. Under Assumption 1 and H_0, R_n is a zero-mean square-integrable process for $(x^\top, \alpha)^\top \in \mathcal{T}$. A standard multivariate central limit theorem implies that the finite-dimensional distributions of R_n converge to those of a multivariate normal distribution with a zero mean vector and covariance matrix given by

$$\mathbb{K}\left(v_1, v_2\right) = \mathbb{E}\left[\psi(W, \theta_0(\alpha_1), \alpha_1)\psi(W, \theta_0(\alpha_2), \alpha_2)1\{X \leq x_1 \wedge x_2\}\right], \tag{3.2}$$

where $v_1 = (x_1^\top, \alpha_1)^\top$ and $v_2 = (x_2^\top, \alpha_2)^\top$ represent generic elements of $\mathcal{T}$ and $\wedge$ denotes the minimum. The following theorem is an extension of the finite-dimensional convergence of R_n to weak convergence in the space $\mathcal{T}$.

Theorem 3.1. *Under H_0 and Assumptions 1-3, $R_n \Longrightarrow R_\infty$, where R_∞ is a Gaussian process with zero mean and covariance function* (3.2)*.*

In practice, θ_0 is unknown and has to be estimated from a sample $\{W_i\}_{i=1}^n$ by an estimator θ_n. However, replacing θ_0 with θ_n has a non-negligible effect on the asymptotic null distribution of the empirical process S_n. To deal with such parametric estimation effect, we need to examine the asymptotic linear representation of the estimator θ_n being used. For the expectile regression estimator given in (2.5), we show in the Appendix that under H_0 and some mild conditions, uniformly in $\alpha \in \mathcal{A}$,

$$Q_n(\alpha) \equiv \sqrt{n}(\hat{\theta}_n(\alpha) - \theta_0(\alpha)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n l(W_i, \theta_0(\alpha), \alpha) + o_p(1),$$

where $l(\cdot)$ is a p-dimensional vector defined as

$$l(W_i, \theta_0(\alpha), \alpha) := J^{-1}(\alpha)\psi(W_i, \theta_0(\alpha), \alpha)\dot{m}(X_i, \theta_0(\alpha)) \tag{3.3}$$

and $J(\alpha) \equiv \mathbb{E}[|\alpha - 1(Y \leq m(X, \theta_0(\alpha)))|\dot{m}(X, \theta_0(\alpha))\dot{m}^\top(X, \theta_0(\alpha))]$. The above result extends the finite-dimensional result in Newey and Powell (1987, Theorem 3) to uniform consistency and weak convergence of the expectile regression process $\hat{\theta}_n(\cdot)$. It directly follows that under H_0, the process $Q_n(\cdot)$ converges weakly to a Gaussian process $Q(\cdot)$ with zero mean and covariance function $\mathbb{K}_Q(\alpha_1, \alpha_2) = J^{-1}(\alpha_1)\Sigma(\alpha_1, \alpha_2)J^{-1}(\alpha_2)$, where

$$\Sigma(\alpha_1, \alpha_2) = \mathbb{E}\big[\psi(W, \theta_0(\alpha_1), \alpha_1)\psi(W, \theta_0(\alpha_2), \alpha_2)\dot{m}(X, \theta_0(\alpha_1))\dot{m}^\top(X, \theta_0(\alpha_2))\big]. \tag{3.4}$$

The next result shows the parametric estimation effect on the asymptotic null distribution of $\hat{R}_n$.

Theorem 3.2. *Under H_0 and Assumptions 1-3,*

$$\sup_{(x,\alpha)\in\mathcal{T}} \left| \hat{R}_n(x, \alpha) - R_n(x, \alpha) + G^\top(x, \theta_0(\alpha), \alpha) n^{-1/2} \sum_{i=1}^n l(W_i, \theta_0(\alpha), \alpha) \right| = o_p(1),$$

where $G(x,\theta_0(\alpha),\alpha) \equiv \mathbb{E}\left[(\alpha + (1-2\alpha)F(m(X,\theta_0(\alpha))|X))\,\dot{m}(X,\theta_0(\alpha))1(X \le x)\right]$.

As a consequence of Theorem 3.2, we obtain the following corollary.

Corollary 3.1. *Under H_0 and Assumptions 1-3,*

$$\hat{R}_n(\cdot) \Longrightarrow R^1_\infty(\cdot) \equiv R_\infty(\cdot) - G^\top(\cdot,\theta_0(\cdot),\cdot)Q(\cdot).$$

Corollary 3.1 and the continuous mapping theorem (CMT), see e.g. Theorem 1.3.6 in Van Der Vaart and Wellner (1996), yield the asymptotic null distributions of continuous functionals of $\hat{R}_n(x,\alpha)$, including the test statistics CvM_n and KS_n given in (2.6) and (2.7). Specifically, under the assumptions of Corollary 3.1 and H_0, for any continuous and even functional $\Gamma(\cdot)$, we have that $\Gamma(\hat{R}_n) \xrightarrow{d} \Gamma(R^1_\infty)$. Note that the asymptotic null distribution of continuous functionals of $\hat{R}_n$ depends on the DGP and the specification under the null in a complex way. We propose a simple multiplier bootstrap procedure to approximate the null distribution, see Section 3.3 below.

3.2 Asymptotic power

In this section, we study the consistency properties of tests based on $\hat{R}_n(\cdot,\cdot)$. First, we show that these tests are consistent against all fixed alternatives provided that a mild regularity condition is satisfied.

Assumption 4. *Under H_1, there exists a $\theta^* \in \mathcal{B}$ such that* $\sup_{\alpha\in\mathcal{A}} \|\hat{\theta}_n(\alpha) - \theta^*(\alpha)\| = o_p(1)$.

The $\theta^*(\alpha)$, if exists, solves the population minimization problem $\min_{\theta\in\Theta} \mathbb{E}[\rho_\alpha(Y - m(X,\theta))]$. Maybe Assumption 4 is a direct consequence of some preliminary conditions.

Theorem 3.3. *Under H_1 and Assumptions 1-4,*

$$\sup_{(x,\alpha)\in\mathcal{T}} \left| \frac{1}{\sqrt{n}}\hat{R}_n(x,\alpha) - \mathbb{E}\big[\psi(W,\theta^*(\alpha),\alpha)1(X \le x)\big] \right| = o_p(1).$$

Theorem 3.3 and the CMT yield that

$$\int_{\mathcal{T}} \left|n^{-1/2}\hat{R}_n(x,\alpha)\right|^2 dF_n(x)dW(\alpha) \xrightarrow{p} \int_{\mathcal{T}} \left|\mathbb{E}[\psi(W,\theta^*(\alpha),\alpha)1(X \le x)]\right|^2 dF(x)dW(\alpha) > 0$$

provided that W is absolute continuous with respect to the Lebesgue measure on $\mathcal{A}$, and $\mathbb{E}[\psi(W,\theta^*(\cdot),\cdot)1(X \le \cdot)]$ is different from zero in a subset with positive Lebesgue measure on $\mathcal{T}$. In such a situation, the test statistic CvM_n will diverge to ∞ under any fixed alternative, and the test will be consistent against all directions in the alternative hypothesis.

Finally we analyze the asymptotic distribution of $\hat{R}_n(\cdot,\cdot)$ under a sequence of local alternatives converging to null at a parametric rate $n^{-1/2}$:

$$H_{1n} : \mathbb{E}\left[\psi(W, \theta_0(\alpha), \alpha)\,|X\right] = \frac{b(X,\alpha)}{\sqrt{n}} + r_n(X,\alpha) \text{ a.s. for some } \theta_0 \in \mathcal{B} \text{ and for all } \alpha \in \mathcal{A}. \tag{3.5}$$

We require the function $b(\cdot,\cdot)$ and $r_n(\cdot,\cdot)$ to satisfy the following assumption.

Assumption 5. *(i)* $\mathbb{E}[\sup_{\alpha\in\mathcal{A}} |b(X,\alpha)|] < \infty$ *and* $\mathbb{E}[\sqrt{n}\sup_{\alpha\in\mathcal{A}} |r_n(X,\alpha)|] = o(1)$*; (ii) The function* $b(X,\cdot)$ *is continuous in* $\mathcal{A}$*, a.s..*

As shown in Lemma A.1 in the Appendix, under the local alternatives, the process $Q_n(\cdot)$ does not converge to a centered Gaussian process but has a local shift. Precisely, under H_{1n}, we have uniformly in $\alpha \in \mathcal{A}$,

$$\sqrt{n}(\hat{\theta}_n(\alpha) - \theta_0(\alpha)) = \xi_b(\alpha) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l(W_i, \theta_0(\alpha), \alpha) + o_p(1),$$

where $\xi_b(\alpha) \equiv J^{-1}(\alpha)\mathbb{E}[b(X,\alpha)\dot{m}(X,\alpha)]$, and the function $l(\cdot)$ is defined in (3.3).

Theorem 3.4. *Suppose that Assumptions 1-3 and 5 hold. Under* H_{1n}*, we have*

$$\hat{R}_n \Longrightarrow R^1_\infty + D_b,$$

where R^1_∞ *is the process defined in Corollary 3.1, and* D_b *is a deterministic shift function given by*

$$D_b(x, \theta_0(\alpha), \alpha) \equiv \mathbb{E}[b(X,\alpha)1(X \le x)] - \xi_b^\top(\alpha)G(x, \theta_0(\alpha), \alpha).$$

Note that $D_b \equiv 0 \Leftrightarrow b(X,\alpha) = [\alpha + (1-2\alpha)F(m(X,\theta_0(\alpha))|X)]\dot{m}^\top(X,\theta_0(\alpha))\xi_b(\alpha)$ a.s.. Thus a test based on $\hat{R}_n$ is able to detect asymptotically local alternatives H_{1n} with $b(X,\alpha)$ not parallel to the vector $[\alpha + (1-2\alpha)F(m(X,\theta_0(\alpha))|X)]\dot{m}(X,\theta_0(\alpha))$.

3.3 Multiplier bootstrap approximation

We see in Section 3.1 that the asymptotic null distribution of $\hat{R}_n(\cdot,\cdot)$ depends in a complex way on the DGP, thus critical values for continuous functionals of $\hat{R}_n(\cdot,\cdot)$ cannot be tabulated in general. To overcome this problem, we propose a multiplier bootstrap procedure in the same spirit as He and Zhu (2003). The proposed procedure is easy to implement and does not require computing new parameter estimates at each bootstrap replication. It also does not involve any need to select tuning parameters apart from the number of bootstrap replications, as opposed to the subsampling schemes where inferences are in general sensitive to a researcher's subjective choice of subsample size.

The method advocated in this section is to approximate the asymptotic distribution of a continuous functional $\Gamma(\hat{R}_n)$ with that of $\Gamma(\hat{R}_n^*)$, where $\hat{R}_n^*(\cdot,\cdot)$ is a simple multiplier-bootstrap approximation of $\hat{R}_n(\cdot,\cdot)$ given by

$$\hat{R}_n^*(x,\alpha) = \frac{1}{\sqrt{n}}\sum_{i=1}^n V_i\left[\psi(W_i,\hat{\theta}_n(\alpha),\alpha)1(X_i\le x) - \hat{G}_n'(x,\hat{\theta}_n(\alpha),\alpha)\hat{l}(W_i,\hat{\theta}_n(\alpha),\alpha)\right], \quad (3.6)$$

where $\{V_i\}_{i=1}^n$ is the sequence of i.i.d. random variables with zero mean, unit variance, bounded support and also independent of the sample $\{W_i\}_{i=1}^n$, the vector $\hat{l}(\cdot)$ is defined as in (3.3) with $J(\alpha)$ substituted by

$$\hat{J}_n(\alpha) = \frac{1}{n}\sum_{i=1}^n \left|\alpha - 1(Y_i \le m(X_i,\hat{\theta}_n(\alpha)))\right|\dot{m}(X_i,\hat{\theta}_n(\alpha))\dot{m}^\top(X_i,\hat{\theta}_n(\alpha)),$$

and

$$\hat{G}_n(x,\hat{\theta}_n(\alpha),\alpha) = \frac{1}{n}\sum_{i=1}^n \left|1\big(Y_i\le m(X_i,\hat{\theta}_n(\alpha))\big)-\alpha\right|\dot{m}(X_i,\hat{\theta}_n(\alpha))1(X_i\le x)$$

is a consistent estimate of the function $G(x,\theta_0(\alpha),\alpha)$ defined in Theorem 3.2. Notably, no nonparametric estimation is required in obtaining $\hat{J}_n(\alpha)$ and $\hat{G}_n(x,\hat{\theta}_n(\alpha),\alpha)$.

It is shown in the Appendix that uniformly in $(x^\top,\alpha)^\top\in\mathcal{T}$,

$$\hat{R}_n^*(x,\alpha) = \frac{1}{\sqrt{n}}\sum_{i=1}^n V_i\left[\psi(W_i,\theta_0(\alpha),\alpha)1(X_i\le x) - G'(x,\theta_0(\alpha),\alpha)l(W_i,\theta_0(\alpha),\alpha)\right] + o_p(1).$$

Note that the first part is simply the infeasible multiplier bootstrap version of $\hat{R}_n(x,\alpha)$.

The bootstrap empirical distribution of $\Gamma(\hat{R}_n^*)$, i.e., $\hat{F}_n^*(t|\{W_i\}_{i=1}^n) = P(\Gamma(\hat{R}_n^*)\le t|\{W_i\}_{i=1}^n)$ is shown to be a consistent estimate of the asymptotic null distribution function of $\Gamma(\hat{R}_n)$, i.e. $F_\infty(t) = P(\Gamma(R_\infty^1)\le t)$. The null hypothesis will be rejected at the τ-level of significance when $\Gamma(\hat{R}_n)\ge c_{n,\tau}^*$, where $c_{n,\tau}^*$ is such that $\hat{F}_n^*(c_{n,\tau}^*|\{W_i\}_{i=1}^n) = 1-\tau$. We can also use bootstrap p-values in this context. In this case, the null could be rejected if $p_n^*<\tau$, where $p_n^* = P\big(\Gamma(\hat{R}_n^*)\ge\Gamma(\hat{R}_n)|\{W_i\}_{i=1}^n\big)$.

Denote by "$\overset{*}{\Longrightarrow}$ in probability" the weak convergence in probability under the bootstrap law, i.e., conditional on the original sample $\{W_i\}_{i=1}^n$, see Van Der Vaart and Wellner (1996).

Theorem 3.5. *Suppose that Assumptions 1-3 hold.*

(i) Under H_0 and H_{1n}, $\hat{R}_n^*\overset{*}{\Longrightarrow}R_\infty^1$ in probability, where R_∞^1 is the Gaussian process defined in Corollary 3.1. For any continuous functional $\Gamma(\cdot)$ from $l^\infty(\mathcal{T})$ to $\mathbb{R}$, we have

$\Gamma(\hat{R}_n^*) \xrightarrow{d} \Gamma(R_\infty^1)$ *in probability.*

(ii) Under H_1, $\hat{R}_n^* \overset{*}{\Longrightarrow} R_\infty^{*1}$ *in probability, where* R_∞^{*1} *is defined in the same way as* R_∞^1 *expect that* θ_0 *is replaced by* θ^*. *For any continuous functional* $\Gamma(\cdot)$ *from* $l^\infty(\mathcal{T})$ *to* $\mathbb{R}$, *we have* $\Gamma(\hat{R}_n^*) \xrightarrow{d} \Gamma(R_\infty^{*1})$ *in probability.*

Theorem 3.5 implies the consistency of our multiplier-bootstrap test against all alternatives, provided that Γ is such that $\Gamma(f) = 0 \Leftrightarrow f = 0$ a.s. Moreover, it can be shown that our bootstrap test preserves the asymptotic local power properties. Details are omitted to save space.

4 Simulation

In this section, we investigate the finite sample properties of test statistics CvM_n and KS_n given in (2.6) and (2.7) by simulation studies. Samples are generated from the model

$$Y_i = X_{1i} + X_{2i} + c\sigma_i + u_i, \quad i = 1, \ldots, n, \tag{4.1}$$

where $\sigma_i = X_{1i}^2 + X_{2i}^2 + X_{1i} * X_{2i}$. The random variables X_{1i} and X_{2i} are taken to be i.i.d. $N(0,1)$ and mutually independent. The error u_i is generated in four ways: (i) $u_i \overset{i.i.d.}{\sim} N(0,1)$, (ii) $u_i = \nu_i/\sqrt{2}$ with $\nu_i \overset{i.i.d.}{\sim} t(4)$, (iii) $u_i \sim \exp(1)$, and (iv) $u_i = m_i/\sqrt{0.86}$ with $m_i \overset{i.i.d.}{\sim} 0.9N(0,1) + 0.1N(1,5)$, and is independent of $\{X_{1i}, X_{2i}\}_{i=1}^n$. We aim to test the correct specification of a linear conditional expectile function. The null hypothesis corresponds to the model given above with $c = 0$. It follows that the model for the conditional expectile function under the null is given by $\mu_\alpha(\mathbf{X}_i) = \mathbf{X}_i^\top \theta_0(\alpha)$, where $\mathbf{X}_i = (1, X_{1i}, X_{2i})^\top$ and $\theta_0(\alpha) = (\psi(\alpha), 1, 1)^\top$ with $\psi(\alpha)$ being the αth expectile of the distribution of u_i.

We use a sample size of $n = 100$ and consider a subinterval of expectiles given by $\mathcal{A} = [0.1, 0.9]$. To calculate the test statistics, we consider W in (2.6) as a uniform discrete distribution over a grid of $\mathcal{A}$ in $m = 30$ equidistributed points from 0.1 to 0.9. The number of Monte Carlo replications is set to 1,000, and the number of sequences of bootstrap multipliers generated for each replication is set to 200.

Table 1 presents the rejection frequencies of the CvM_n and KS_n test based on the bootstrapped p-values. The nominal levels are given by 10%, 5% and 1%. First, under the null ($c = 0$), the empirical rejection probabilities of both the CvM_n and KS_n tests are close to the nominal levels. Second, the empirical power are reasonably high for both tests. As one might expect, larger values of $|c|$ imply higher power for the tests considered. The CvM test appears to dominates the KS test for all the DGPs considered when $c \neq 0$. The above pattern appears to be robust to the distribution of the error u_i.

Table 1: Rejection frequencies under DGP (4.1)

	$u_i \sim N(0,1)$						$u_i \sim t(4)$					
	CvM_n			KS_n			CvM_n			KS_n		
c	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
-0.3	0.935	0.871	0.615	0.925	0.851	0.585	0.900	0.809	0.558	0.878	0.795	0.556
-0.2	0.682	0.556	0.309	0.674	0.528	0.295	0.676	0.522	0.268	0.659	0.516	0.268
-0.1	0.283	0.171	0.067	0.263	0.159	0.059	0.269	0.156	0.039	0.234	0.140	0.039
0	0.105	0.062	0.015	0.102	0.057	0.015	0.098	0.033	0.005	0.085	0.030	0.005
0.1	0.298	0.183	0.058	0.284	0.176	0.053	0.254	0.158	0.049	0.231	0.145	0.047
0.2	0.704	0.565	0.302	0.688	0.553	0.286	0.671	0.539	0.290	0.658	0.525	0.285
0.3	0.924	0.866	0.634	0.919	0.854	0.611	0.907	0.806	0.599	0.891	0.796	0.585
	$u_i \sim 0.9N(0,1) + 0.1N(1,5)$						$u_i \sim exp(1)$					
	CvM_n			KS_n			CvM_n			KS_n		
c	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
-0.3	0.914	0.833	0.630	0.906	0.820	0.617	0.917	0.842	0.630	0.896	0.824	0.608
-0.2	0.702	0.579	0.318	0.688	0.562	0.291	0.726	0.570	0.336	0.641	0.508	0.290
-0.1	0.310	0.191	0.073	0.304	0.173	0.064	0.290	0.166	0.053	0.209	0.119	0.036
0	0.114	0.060	0.012	0.111	0.055	0.011	0.087	0.038	0.015	0.078	0.036	0.011
0.1	0.290	0.191	0.065	0.275	0.171	0.059	0.255	0.145	0.039	0.196	0.106	0.030
0.2	0.677	0.537	0.293	0.662	0.514	0.275	0.673	0.536	0.275	0.584	0.468	0.216
0.3	0.934	0.855	0.648	0.926	0.832	0.628	0.884	0.779	0.546	0.828	0.717	0.472

It is also of interest to check the performance of the test in presence of conditional heteroskedasticity. To this end, we consider the following DGPs:

$$Y_i = X_{1i} + X_{2i} + c\sigma_i + f(X_i)u_i, \tag{4.2}$$

where $f(X_i) = 1+(X_{1i}+X_{2i})/4$, $\sigma_i = X_{1i}^2+X_{2i}^2+X_{1i}*X_{2i}$, and $\{(X_{1i}, X_{2i})\}_{i=1}^n$ are generated in the same way as above. We aim to test the validity of a linear conditional expectile function. The null hypothesis corresponds to (4.2) with $c = 0$. The conditional expectile function is given by $\mu_\alpha(\mathbf{X}_i) = \mathbf{X}_i^\top \theta_0(\alpha)$, where $\theta_0(\alpha) = \theta_{0,1} + \psi(\alpha)\theta_{0,2}$, $\theta_{0,1} = (0,1,1)^\top$ and $\theta_{0,2} = (1, 0.25, 0.25)^\top$.

Table 2 presents the rejection frequencies of CvM and KS statistics under DGP (4.2). Similar as above, the performance of both statistics is satisfactory in terms of size and power. Comparing with Table 1, we find that the empirical power of the test statistics becomes slightly higher in the presence of conditional heteroskedasticity.

Table 2: Rejection frequencies under DGP (4.2)

	$u_i \sim N(0,1)$						$u_i \sim t(4)$					
	CvM_n			KS_n			CvM_n			KS_n		
c	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
-0.3	0.923	0.866	0.674	0.912	0.856	0.647	0.897	0.828	0.632	0.890	0.815	0.624
-0.2	0.737	0.607	0.359	0.712	0.589	0.333	0.700	0.580	0.340	0.687	0.561	0.332
-0.1	0.332	0.202	0.082	0.324	0.193	0.076	0.277	0.154	0.047	0.255	0.141	0.044
0	0.107	0.047	0.011	0.097	0.045	0.012	0.089	0.040	0.009	0.077	0.032	0.008
0.1	0.333	0.207	0.081	0.305	0.203	0.079	0.270	0.159	0.052	0.248	0.144	0.051
0.2	0.722	0.605	0.354	0.714	0.590	0.333	0.727	0.599	0.359	0.716	0.571	0.355
0.3	0.934	0.876	0.661	0.929	0.868	0.641	0.901	0.836	0.651	0.891	0.826	0.643

	$u_i \sim 0.9N(0,1) + 0.1N(1,5)$						$u_i \sim exp(1)$					
	CvM_n			KS_n			CvM_n			KS_n		
c	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
-0.3	0.940	0.870	0.653	0.929	0.857	0.636	0.932	0.863	0.659	0.914	0.852	0.650
-0.2	0.742	0.622	0.367	0.728	0.611	0.345	0.762	0.611	0.362	0.685	0.537	0.316
-0.1	0.324	0.195	0.072	0.308	0.191	0.066	0.321	0.217	0.059	0.240	0.134	0.037
0	0.119	0.057	0.014	0.109	0.051	0.015	0.106	0.047	0.012	0.108	0.038	0.009
0.1	0.319	0.208	0.076	0.303	0.189	0.074	0.276	0.166	0.052	0.221	0.122	0.032
0.2	0.730	0.593	0.336	0.712	0.576	0.318	0.701	0.557	0.339	0.619	0.488	0.278
0.3	0.942	0.873	0.676	0.935	0.868	0.652	0.901	0.835	0.617	0.860	0.762	0.539

5 Empirical application

Predicting excess stock returns has been the subject of an extensive literature in financial risk analysis (Welch and Goyal, 2007; Campbell and Thompson, 2008). In many situations, knowledge is required for the entire return distribution or specific parts (e.g., left tail) of the return distribution, which suggests expectile regression as a desirable modelling strategy. For example, Chen et al. (2020) examined the dependence structure between stock returns and 11 predictive variates based on linear expectile regression. Bai et al. (2022) proposed model averaging for expectile regression and applied the method to predict the expectiles of excess stock returns. In what follows, we examine whether the linear specification adopted in above works is supported across a broad range of expectiles. The dataset is taken from Welch and Goyal (2007) and contains 367 quarterly economic observations from January, 1927 to September, 2018. The response is the equity premium, which is the return on the S&P 500 Index minus the return on treasury bill. We used 11 quarterly predictors following Chen et al. (2020), which include dividend yield, earnings-price ratio, book-to-market ratio, net equity expansion, stock variance, treasury bill rate, term spread, long-term rate of return for government bonds, default yield spread, default return spread and inflation.

We apply the proposed omnibus testing approach to assess the fit of the linear expectile regression model

$$\mu_\alpha(Y_t|X_{t-1}) = \theta_0(\alpha) + X_{t-1}^\top \theta_1(\alpha), \quad \alpha \in \mathcal{A}, \tag{5.1}$$

where Y_t denotes the equity premium at time t, and X_{t-1} is a 11-dimensional predictive vector at time $t-1$, and $\theta_0(\alpha)$ and $\theta_1(\alpha)$ are unknown coefficients. We take $\mathcal{A}$ as a grid of $m = 30$ equidistributed points $\{\alpha_j\}_{j=1}^m$ from α_1 to α_2, where $[\alpha_1, \alpha_2]$ denotes one of $[0.01, 0.10]$, $[0.10, 0.50]$ and $[0.50, 0.90]$.

Table 3: The bootstrapped p-values over various ranges of expectiles

Tests	[0.01,0.10]	[0.10,0.50]	[0.50,0.90]
KS_n	0.082	0.387	0.395
CvM_n	0.092	0.255	0.327

Table 3 reports the bootstrapped p-values of KS and CvM tests of (5.1). The distributions of the KS and CvM statistic are approximated using 400 sequences of bootstrap multipliers. It is clear from Table 3 that the linear expectile regression model (5.1) is rejected at 10% level over the range of levels from 0.01 to 0.1, implying that the impact of the predictors could be nonlinear in the left tail of the excess stock return distribution. In contrast, when the focus is on expectile range [0.10,0.50] and [0.50,0.90], the linear expectile regression model (5.1) cannot be rejected at 10% level.

6 Conclusion

We propose a simple test for the correct specification of linear conditional expectile models against semiparametric alternatives over a range of expectile levels. The approach has the appealing property of delivering a consistent test. We propose a easy-to-implement multiplier bootstrap procudure to facilitate the simulation of critical values. Simulations illustrate the advantages of our approach in finite samples. Our methods provide a powerful tool that can be used by practitioners to assess the plausibility of standard market risk models.

Appendix

A Some useful lemmas

Throughout the appendix, C denotes a generic constant that does not depend on n and may vary from line to line. For a random variable X, let $\|X\|_2 \equiv [\mathbb{E}(X^2)]^{1/2}$. For a generic

Banach space $\mathcal{G}$, denote the covering number of with bracketing of $\mathcal{G}$ by $N_{[\cdot]}(\epsilon, \mathcal{G}, \|\cdot\|_2)$ and the entropy number by

$$J_{[\cdot]}(\delta, \mathcal{G}, \|\cdot\|_2) \equiv \int_0^\delta \sqrt{\log N_{[\cdot]}(\epsilon, \mathcal{G}, \|\cdot\|_2)} d\epsilon.$$

The detailed definition of the above notation can be found in Van Der Vaart and Wellner (1996).

Similar to Escanciano and Goh (2014), we first state a weak convergence theorem. Let

$$V_n(\gamma) = n^{-1/2} \sum_{i=1}^n \big(\psi(W_i, \theta(\alpha), \alpha) - \mathbb{E}\big[\psi(W_i, \theta(\alpha), \alpha) \mid X_i\big]\big) 1(X_i \le x),$$

which is indexed by $\gamma = (\theta, \alpha, x) \in \Gamma := \mathcal{B} \times \mathcal{A} \times \overline{\mathbb{R}}^d$. Consider the following pseudo-metric

$$\rho(\gamma_1, \gamma_2) \equiv \|\theta_1 - \theta_2\|_{\mathcal{A}} + |\alpha_1 - \alpha_2| + |F_X(x_1) - F_X(x_2)|, \tag{A.1}$$

where $\gamma_j = (\theta_j, \alpha_j, x_j) \in \Gamma$, $j = 1, 2$.

Theorem A.1. *Under Assumptions 1-3(i), the process $V_n(\gamma)$ is ρ-stochastically equicontinuous.*

Proof of Theorem A.1. For $(\theta, \alpha) \in \mathcal{B} \times \mathcal{A}$, define $\zeta(W, \theta(\alpha), \alpha) = 1(Y \le m(X, \theta(\alpha))) - \alpha$, $h(W, \theta, \alpha) = \psi(W, \theta(\alpha), \alpha) - \mathbb{E}\big[\psi(W, \theta(\alpha), \alpha) \mid X\big]$, and $\mathcal{H} \equiv \{w \mapsto h(w, \theta, \alpha) : (\theta, \alpha) \in \mathcal{B} \times \mathcal{A}\}$. Fix $(\theta_1, \alpha_1) \in \mathcal{B} \times \mathcal{A}$. By the triangle inequality, we have

$$\begin{aligned}
&\mathbb{E}\Bigg[\sup_{\theta:\|\theta-\theta_1\|_{\mathcal{A}} \le \delta} \sup_{\alpha:|\alpha-\alpha_1| \le \delta} \big|h(W, \theta_1, \alpha_1) - h(W, \theta, \alpha)\big|^2\Bigg] \\
&\le C_1 \mathbb{E}\Bigg[\sup_{\theta:\|\theta-\theta_1\|_{\mathcal{A}} \le \delta} \sup_{\alpha:|\alpha-\alpha_1| \le \delta} \big|\psi(W, \theta_1(\alpha_1), \alpha_1) - \psi(W, \theta(\alpha), \alpha)\big|^2\Bigg] \\
&\quad + C_1 \mathbb{E}\Bigg[\sup_{\theta:\|\theta-\theta_1\|_{\mathcal{A}} \le \delta} \sup_{\alpha:|\alpha_1-\alpha| \le \delta} \big|\mathbb{E}[\psi(W, \theta_1(\alpha_1), \alpha_1) \mid X] - \mathbb{E}[\psi(W, \theta(\alpha), \alpha) \mid X]\big|^2\Bigg] \\
&\le C_2 \mathbb{E}\Bigg[\sup_{\theta:\|\theta-\theta_1\|_{\mathcal{A}} \le \delta} \sup_{\alpha:|\alpha-\alpha_1| \le \delta} \big|\zeta(W, \theta_1(\alpha_1), \alpha_1) - \zeta(W, \theta(\alpha), \alpha)\big|^2 (Y - m(X, \theta_1(\alpha_1)))^2\Bigg] \\
&\quad + C_2 \mathbb{E}\Bigg[\sup_{\theta:\|\theta-\theta_1\|_{\mathcal{A}} \le \delta} \sup_{\alpha:|\alpha-\alpha_1| \le \delta} \big|m(X, \theta(\alpha)) - m(X, \theta_1(\alpha_1))\big|^2\Bigg] \\
&\le C_3 \mathbb{E}\Bigg[\sup_{\theta \in \mathcal{B}} \sup_{\alpha \in \mathcal{A}} \delta^2 (Y - m(X, \theta(\alpha)))^2 + \sup_{\theta:\|\theta-\theta_1\|_{\mathcal{A}} \le \delta} \sup_{\alpha:|\alpha-\alpha_1| \le \delta} \big|m(X, \theta(\alpha)) - m(X, \theta_1(\alpha_1))\big|^2\Bigg] \\
&\le C_4 \delta^2,
\end{aligned}$$

where the second inequality follows from conditional Jansen inequality, and the third inequality follows from the fact that $|(\zeta(W,\theta_1(\alpha_1),\alpha_1)-\zeta(W,\theta(\alpha),\alpha))\cdot(Y-m(X,\theta_1(\alpha_1)))| \leq \big[|1(Y\leq m(X,\theta(\alpha)))-1(Y\leq m(X,\theta_1(\alpha_1)))|+|\alpha_1-\alpha|\big]\cdot|Y-m(X,\theta_1(\alpha_1))| \leq |m(X,\theta_1(\alpha_1))-m(X,\theta(\alpha))|+|\alpha_1-\alpha|\cdot|Y-m(X,\theta_1(\alpha_1))|$.

Let $\mathscr{W}=\{\mathbf{x}\mapsto 1\{\mathbf{x}\leq x\}: x\in[-\infty,+\infty]^d\}$. Chen et al. (2003, Theorem 3), Lemma A.2 in Escanciano and Goh (2014) and Assumption 2 yield that $J_{[\cdot]}(\delta_n,\mathcal{H}\times\mathscr{W},\|\cdot\|_2)\to 0$ for every $\delta_n\downarrow 0$. It follows from Theorem 19.28 of Van Der Vaart (1998) that the process $V_n(\gamma)$ is ρ-stochastically equicontinuous. □

The next lemma gives a uniform linear representation of the ER process $\hat{\theta}_n(\cdot)$ in (2.5). This is essential in establishing the asymptotic null distribution of $\hat{R}_n$. For $(\alpha,b)\in\mathcal{A}\times\Theta$, let $\mathcal{Q}_n(\alpha,b):=\mathbb{E}_n[\rho_\alpha(Y-m(X,b))]$ and $\mathcal{Q}_\infty(\alpha,b):=\mathbb{E}[\rho_\alpha(Y-m(X,b))]$.

Assumption 6. *For all $\delta>0$, $\inf_{\theta:\|\theta-\theta_0\|_{\mathcal{A}}\geq\delta}\sup_{\alpha\in\mathcal{A}}|\mathcal{Q}_\infty(\alpha,\theta(\alpha))-\mathcal{Q}_\infty(\alpha,\theta_0(\alpha))|>0$.*

Assumption 6 is a well-separated condition used for proving the consistency of the process $\widehat{\theta}_n(\cdot)$. It is trivially satisfied when $m(\cdot,\theta(\alpha))$ is a linear function and Assumption 3 in the main text holds.

Lemma A.1. *Suppose that Assumptions 1-3, 5 and 6 hold. Then under H_{1n} in (3.5), we have the sample ER process $\hat{\theta}_n(\cdot)$ defined in (2.5) is uniformly consistent, and uniformly in $\alpha\in\mathcal{A}$,*

$$\sqrt{n}(\hat{\theta}_n(\alpha)-\theta_0(\alpha))=\xi_b(\alpha)+\frac{1}{\sqrt{n}}\sum_{i=1}^n l(W_i,\theta_0(\alpha),\alpha)+o_p(1), \tag{A.2}$$

where $l(\cdot)$ is a p-dimensional vector defined as

$$l(W_i,\theta_0(\alpha),\alpha):=J^{-1}(\alpha)\psi(W_i,\theta_0(\alpha),\alpha)\dot{m}(X_i,\theta_0(\alpha)) \tag{A.3}$$

with $J(\alpha):=\mathbb{E}[|\alpha-1(Y\leq m(X,\theta_0(\alpha)))|\dot{m}(X,\theta_0(\alpha))\dot{m}^\top(X,\theta_0(\alpha))]$, and $\xi_b(\alpha)\equiv J^{-1}(\alpha)\mathbb{E}[b(X,\alpha)\dot{m}(X,\theta_0(\alpha))]$.

Proof of Lemma A.1. We apply the arguments in Angrist et al. (2006, Theorem 3). For $W=(Y,X')'$, denote $\mathbb{E}_n[f(W)]\equiv n^{-1}\sum_{i=1}^n f(W_i)$ and $\mathbb{G}_n[f(W)]\equiv n^{-1/2}\sum_{i=1}^n(f(W_i)-\mathbb{E}[f(W_i)])$. For a matrix A, let $\lambda_{\min}(A)$ denote its minimum eigenvalue.

We first establish the uniform consistency of the sample ER process $\hat{\theta}_n(\cdot)$. By Theorem 5.7 of Van Der Vaart (1998), it suffices to show that

$$\sup_{\theta\in\mathcal{B},\alpha\in\mathcal{A}}\left|\mathcal{Q}_n(\alpha,\theta(\alpha))-\mathcal{Q}_\infty(\alpha,\theta(\alpha))\right|=o_p(1).$$

The pointwise convergence follows from Khinchine law of large numbers. The empirical process $(\alpha,\theta)\mapsto\mathcal{Q}_n(\alpha,\theta(\alpha))$ is stochastically equicontinuous. Indeed, let $\mathcal{F}=\{(x',y)'\to$

$\rho_\alpha(y - m(x, \theta(\alpha))) : \theta \in \mathcal{B}, \alpha \in \mathcal{A}\}$. Fix $\theta_1 \in \mathcal{B}$ and $\alpha_1 \in \mathcal{A}$. It follows from Assumptions 1 and 3 that for $\delta > 0$, $\mathbb{E} \sup_{\alpha:|\alpha-\alpha_1|\leq\delta} \sup_{\theta:\|\theta-\theta_1\|_{\mathcal{A}}\leq\delta} |\rho_{\alpha_1}(Y - m(X, \theta_1(\alpha_1))) - \rho_\alpha(Y - m(X, \theta(\alpha)))| \leq C\delta$, where C is a finite constant. By the proof of Theorem 3 in Chen et al. (2003), we have for any $\epsilon > 0$, $N_{[\cdot]}(\epsilon, \mathcal{F}, \|\cdot\|_1) \leq N(\epsilon/(2C), \mathcal{A}, \|\cdot\|_{\mathcal{A}}) \times N(\epsilon/(2C), \mathcal{B}, \|\cdot\|_{\mathcal{B}})$, which is finite due to Assumption 2. By Theorem 2.4.1 of Van Der Vaart and Wellner (1996), the convergence also holds uniformly.

By the property of $\hat{\theta}_n(\alpha)$ and the above uniform convergence, we have $\mathcal{Q}_n(\alpha, \hat{\theta}_n(\alpha)) \leq \mathcal{Q}_n(\alpha, \theta_0(\alpha)) = \mathcal{Q}_\infty(\alpha, \theta_0(\alpha)) + o_p(1)$. Therefore,

$$\begin{aligned}\mathcal{Q}_\infty(\alpha, \hat{\theta}_n(\alpha)) - \mathcal{Q}_\infty(\alpha, \theta_0(\alpha)) &\leq \mathcal{Q}_\infty(\alpha, \hat{\theta}_n(\alpha)) - \mathcal{Q}_n(\alpha, \hat{\theta}_n(\alpha)) + o_p(1) \\ &\leq \sup_{\theta\in\mathcal{B},\alpha\in\mathcal{A}} |\mathcal{Q}_n(\alpha, \theta(\alpha)) - \mathcal{Q}_\infty(\alpha, \theta(\alpha))| + o_p(1) = o_p(1).\end{aligned}$$

By Assumption 6, there exists for every $\epsilon > 0$ a number $\eta > 0$ such that $\sup_{\alpha\in\mathcal{A}} |\mathcal{Q}_\infty(\alpha, \theta(\alpha)) - \mathcal{Q}_\infty(\alpha, \theta_0(\alpha))| > \eta$ for every θ such that $\|\theta - \theta_0\|_{\mathcal{A}} \geq \epsilon$. Thus, the event $\{\|\hat{\theta}_n - \theta_0\|_{\mathcal{A}} \geq \epsilon\}$ is contained in the event $\{\sup_{\alpha\in\mathcal{A}} |\mathcal{Q}_\infty(\alpha, \hat{\theta}_n(\alpha)) - \mathcal{Q}_\infty(\alpha, \theta_0(\alpha))| > \eta\}$. The probability of the latter event converges to 0 in view of the preceding display. This proves that $\sup_{\alpha\in\mathcal{A}} \|\hat{\theta}_n(\alpha) - \theta_0(\alpha)\| = o_p(1)$.

Next, we establish asymptotic Gaussianity of the sample ER process. With abuse of notation, let $\psi_\alpha(\lambda) \equiv |\alpha - 1(\lambda < 0)| \cdot \lambda$. Then we have $\partial\rho_\alpha(Y_i - m(X_i, b))/\partial b = -2\dot{m}(X_i, b)\psi_\alpha(Y_i - m(X_i, b))$. Recall that $\hat{\theta}_n(\alpha)$ minimizes $\mathcal{Q}_n(\alpha, \theta) = \mathbb{E}_n(\rho_\alpha(Y - m(X, \theta)))$. By the first order condition, we have for all $\alpha \in \mathcal{A}$,

$$\sqrt{n}\mathbb{E}_n \left[\psi_\alpha(Y - m(X, \hat{\theta}_n(\alpha)))\dot{m}(X, \hat{\theta}_n(\alpha))\right] = 0. \tag{A.4}$$

Second, $(\alpha, b) \mapsto \mathbb{G}_n[\psi_\alpha(Y - m(X, b))\dot{m}(X, b)]$ is stochastically equicontinuous over $\mathcal{A} \times \Theta$, with respect to the $L_2(P)$ pseudometric $\rho((\alpha', b'), (\alpha'', b''))^2 := \max_{j=1,\ldots,p} \mathbb{E}[(\psi_{\alpha'}(Y - m(X, b'))\dot{m}_j(X, b') - \psi_{\alpha''}(Y - m(X, b''))\dot{m}_j(X, b''))^2]$ for $j = 1, \ldots, p$. By Assumption 3, we can show that the function class $\{\mathbf{w} \to \psi_\alpha(\mathbf{y} - m(\mathbf{x}, b))\dot{m}(\mathbf{x}, b) : \alpha \in \mathcal{A}, b \in \Theta\}$ satisfies that $|\psi_{\alpha'}(\mathbf{y} - m(\mathbf{x}, b'))\dot{m}(\mathbf{x}, b') - \psi_{\alpha''}(\mathbf{y} - m(\mathbf{x}, b''))\dot{m}(\mathbf{x}, b'')| \leq |\alpha' - \alpha''|F_1(\mathbf{w}) + \|b' - b''\|F_2(\mathbf{w})$ for some square-integrable functions F_1 and F_2. By Theorem 2.7.1 of Van Der Vaart and Wellner (1996), this class is Donsker with a square-integrable envelope. Stochastic equicontinuity then is part of being Donsker.

Third, by stochastic equicontinuity of $(\alpha, b) \mapsto \mathbb{G}_n[\psi_\alpha(Y - m(X, b))\dot{m}(X, b)]$ we have that

$$\begin{aligned}&\mathbb{G}_n[\psi_\alpha(Y - m(X, \hat{\theta}_n(\alpha)))\dot{m}(X, \hat{\theta}_n(\alpha))] \\ &= \mathbb{G}_n[\psi_\alpha(Y - m(X, \theta_0(\alpha)))\dot{m}(X, \theta_0(\alpha))] + o_p(1) \quad \text{in } l^\infty(\mathcal{A}),\end{aligned} \tag{A.5}$$

which follows from $\sup_{\alpha\in\mathcal{A}} \|\hat{\theta}_n(\alpha) - \theta_0(\alpha)\| = o_p(1)$ and resulting convergence with respect to the pseudometric $\sup_{\alpha\in\mathcal{A}} \rho[(\alpha, \hat{\theta}_n(\alpha)), (\alpha, \theta_0(\alpha))]^2 = o_p(1)$. The latter is immediate from $\sup_{\alpha\in\mathcal{A}} \rho[(\alpha, \theta(\alpha)), (\alpha, b(\alpha))]^2 \leq C \sup_{\alpha\in\mathcal{A}} \|\theta(\alpha) - b(\alpha)\|^2$ for some constant $C < \infty$.

Furthermore, the following expansion is valid uniformly in α:

$$\begin{aligned}\mathbb{E}[\psi_\alpha(Y - m(X,\theta))\dot{m}(X,\theta)]|_{\theta=\hat{\theta}_n(\alpha)} = &-[J(\alpha) + o_p(1)]\big(\hat{\theta}_n(\alpha) - \theta_0(\alpha)\big) \\ &+ n^{-1/2}\mathbb{E}[b(X,\alpha)\dot{m}(X,\theta_0)] + o_p(1).\end{aligned} \tag{A.6}$$

By Taylor expansion, $\mathbb{E}[\psi_\alpha(Y - m(X,\theta))\dot{m}(X,\theta)]|_{\theta=\hat{\theta}_n(\alpha)} = \mathbb{E}[\psi_\alpha(Y - m(X,\theta_0))\dot{m}(X,\theta_0)]| - \mathbb{E}[|\alpha - 1(Y \leq m(X,b(\alpha)))|\dot{m}(X,b(\alpha))\dot{m}(X,b(\alpha))^\top + \psi_\alpha(Y - m(X,b(\alpha)))\ddot{m}(X,b(\alpha))]|_{b(\alpha)=\theta^*(\alpha)} \times (\hat{\theta}_n(\alpha) - \theta_0(\alpha))$, where $\theta^*(\alpha)$ is on the line connecting $\widehat{\theta}(\alpha)$ and $\theta_0(\alpha)$ for each α. Then (A.6) follows by Assumption 3, the uniform consistency of $\hat{\theta}_n(\alpha)$, (3.5), and the assumed uniform continuity the mapping $y \mapsto F(y|x)$, uniformly in x over the support of X.

Fourth, we have that

$$\begin{aligned}o_p(1) = &\mathbb{E}[b(X,\alpha)\dot{m}(X,\theta_0)] - [J(\cdot) + o_p(1)]\sqrt{n}\big(\hat{\theta}_n(\cdot) - \theta_0(\cdot)\big) \\ &+ \mathbb{G}_n[\psi_\cdot(Y - m(X,\theta_0(\cdot)))\dot{m}(X,\theta_0(\cdot))]\end{aligned} \tag{A.7}$$

because the left-hand side of (A.4) is equal to the left-hand side of $n^{1/2}$ times (A.6) plus the left-hand side of (A.5). Since $\lambda_{\min}(J(\alpha)) \geq \lambda > 0$, we have $\sup_{\alpha\in\mathcal{A}} \big\|\mathbb{E}[b(X,\alpha)\dot{m}(X,\theta_0)] + \mathbb{G}_n[\psi_\cdot(Y - m(X,\theta_0(\cdot)))\dot{m}(X,\theta_0(\cdot))] + o_p(1)\big\| \geq (\sqrt{\lambda} + o_p(1)) \cdot \sup_{\alpha\in\mathcal{A}} \sqrt{n}\|\widehat{\theta}_n(\alpha) - \theta_0(\alpha)\|$.

Fifth, under H_0 and Assumption 1, the conditional expectile function $m(X,\theta_0(\alpha))$ is continuously differentiable with respect to α, see Theorem 1 in Newey and Powell (1987). Hence $\alpha \mapsto \mathbb{G}_n[\psi_\alpha(Y - m(X,\theta_0(\alpha)))\dot{m}(X,\theta_0(\alpha))]$ is stochastically equicontinuous over $\mathcal{A}$ for the pseudometric given by $\rho(\alpha', \alpha'') := \rho((\alpha', \theta_0(\alpha')), (\alpha^{\top\top}, \theta_0(\alpha'')))$. Stochastic equicontinuity of $\alpha \mapsto \mathbb{G}_n[\psi_\alpha(Y - m(X,\theta_0(\alpha)))\dot{m}(X,\theta_0(\alpha))]$ and a multivariate central limit theorem imply that

$$\mathbb{G}_n[\psi_\cdot(Y - m(X,\theta_0(\cdot)))\dot{m}(X,\theta_0(\cdot))] \Rightarrow z(\cdot) \quad \text{in } l^\infty(\mathcal{A}), \tag{A.8}$$

where $z(\cdot)$ is a Gaussian process with covariance function $\Sigma(\cdot,\cdot)$ given in (3.4). Therefore, $\sup_{\alpha\in\mathcal{A}} \|\sqrt{n}(\hat{\theta}_n(\alpha) - \theta_0(\alpha))\| = O_p(1)$. Finally, the latter fact and (A.7) imply that in $l^\infty(\mathcal{A})$,

$$\sqrt{n}\big(\widehat{\theta}_n(\cdot) - \theta_0(\cdot)\big) = \xi_b(\cdot) + J^{-1}(\alpha)\mathbb{G}_n[\psi_\cdot(Y - m(X,\theta_0(\cdot)))\dot{m}(X,\theta_0(\cdot))] + o_p(1).$$

This completes the proof. □

B Proof of main theorems

Proof of Theorem 3.1. The finite-dimensional convergence follows from a standard multivariate central limit theorem. The asymptotic equicontinuity of $(1/\sqrt{n})\sum_{i=1}^{n}\psi(W_i,\theta_0(\alpha),\alpha)1(X_i \le x)$ follows from an application of Theorem A.1 with $\mathcal{B} = \{\theta_0\}$. $\square$

Proof of Theorem 3.2. From Lemma A.1, we obtain that $\theta_n \in \mathcal{B}$ with probability tending to one. Hence, by Theorem A.1 we have that

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

Under the null, this is equivalent to

$$\begin{aligned}\sup_{(x,\alpha)\in\mathcal{T}}\Bigg|&\hat{R}_n(x,\alpha) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(W_i,\theta_0(\alpha),\alpha)1(X_i\le x) + \frac{1}{\sqrt{n}}\\ &\times\sum_{i=1}^{n}\big(\mathbb{E}[\psi(W_i,\theta_0(\alpha),\alpha)|X_i] - \mathbb{E}[\psi(W_i,\hat{\theta}_n(\alpha),\alpha)|X_i]\big)1(X_i\le x)\Bigg| = o_p(1).\end{aligned}$$

Note that $\int_{-\infty}^{\alpha}(y-\alpha)f(y|x)dy$ is continuously differential in α, and the derivative is $-\int_{-\infty}^{\alpha}f(y|x)dy$. Applying the mean-value theorem we have

$$\begin{aligned}o_p(1) = &\sup_{(x,\alpha)\in\mathcal{T}}\Bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\mathbb{E}[\psi(W_i,\theta_0(\alpha),\alpha)|X_i] - \mathbb{E}[\psi(W_i,\hat{\theta}_n(\alpha),\alpha)|X_i]\big)\\ &\times 1(X_i\le x) - \sqrt{n}(\hat{\theta}_n(\alpha)-\theta_0(\alpha))^\top\\ &\times\frac{1}{n}\sum_{i=1}^{n}\big[(\alpha-(1-2\alpha)F(m(X_i,\tilde{\theta}_n(\alpha))|X_i))\dot{m}(X_i,\tilde{\theta}_n(\alpha))1(X_i\le x)\big]\Bigg|\\ = &\sup_{(x,\alpha)\in\mathcal{T}}\Bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\mathbb{E}[\psi(W_i,\theta_0(\alpha),\alpha)|X_i] - \mathbb{E}[\psi(W_i,\hat{\theta}_n(\alpha),\alpha)|X_i]\big)\\ &\times 1(X_i\le x) - \sqrt{n}(\hat{\theta}_n(\alpha)-\theta_0(\alpha))^\top\\ &\times\frac{1}{n}\sum_{i=1}^{n}\big[\alpha-(1-2\alpha)F(m(X_i,\theta_0(\alpha))|X_i)\big]\dot{m}(X_i,\theta_0(\alpha))1(X_i\le x)\Bigg| + o_p(1)\\ = &\sup_{(x,\alpha)\in\mathcal{T}}\Bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big(\mathbb{E}[\psi(W_i,\theta_0(\alpha),\alpha)|X_i] - \mathbb{E}[\psi(W_i,\hat{\theta}_n(\alpha),\alpha)|X_i]\big)\\ &\times 1(X_i\le x) - \sqrt{n}(\hat{\theta}_n(\alpha)-\theta_0(\alpha))^\top G(x,\theta_0(\alpha),\alpha) + o_p(1),\end{aligned}$$

where $\tilde{\theta}_n(\cdot)$ is such that $\left|\tilde{\theta}_n(\alpha)-\theta_0(\alpha)\right| \le \left|\hat{\theta}_n(\alpha)-\theta_0(\alpha)\right|$ *a.s.* for each $\alpha\in\mathcal{A}$. The second equality is from the uniform convergence of $\hat{\theta}_n$ and Assumption 3, and the last equality follows from the Glivenko–Cantelli Theorem, since $(x,\alpha) \to [\alpha - (1 - 2\alpha)F(m(X,\theta_0(\alpha))|X)]\dot{m}(X,\theta_0(\alpha))1(X \leq x)$ is Glivenko–Cantelli. The last two displays and Lemma A.1 lead to Theorem 3.2.

Proof of Theorem 3.3 Recall that $\Gamma = \mathcal{B} \times \mathcal{A} \times \overline{\mathbb{R}}^d$ and we endow it with the metric $\rho(\cdot,\cdot)$ as in (A.1). Given $\gamma = (\theta,\alpha,x) \in \Gamma$, let $B(\gamma,\delta)$ be the open ball of radius δ around γ, i.e., $B(\gamma,\delta) = \{\gamma_1 \in \Gamma : \rho(\gamma,\gamma_1) < \delta\}$. Using similar arguments as Theorem A.1, we obtain that for each $\gamma = (\theta,\alpha,x) \in \Gamma$,

$$\lim_{\delta\to 0} \mathbb{E}\left[\sup_{\gamma_1 \in B(\gamma,\delta)} |\psi(W,\theta(\alpha),\alpha)1(X \leq x) - \psi(W,\theta_1(\alpha_1),\alpha_1)1(X \leq x_1)|^2\right] = 0.$$

By uniform law of large numbers, we have

$$\sup_{\theta\in\mathcal{B}} \sup_{(x,\alpha)\in\mathcal{T}} \left|\frac{1}{n}\sum_{i=1}^{n} \left[\psi(W_i,\theta(\alpha),\alpha)1(X_i \leq x)\right] - \mathbb{E}[\psi(W,\theta(\alpha),\alpha)1(X \leq x)]\right| = o_p(1).$$

Further,

$$\begin{aligned}\mathbb{E}[\psi(W,\theta(\alpha),\alpha)1(X \leq x)]|_{\theta=\hat{\theta}_n} = {} & \mathbb{E}[\psi(W,\theta^*(\alpha),\alpha)1(X \leq x)]|_{\theta=\theta^*} \\ & + (\hat{\theta}_n(\alpha) - \theta^*(\alpha))^\intercal G(x,\theta^*(\alpha),\alpha) + o_p(1).\end{aligned}$$

Hence, from the last display and Assumption 4, we conclude that

$$\sup_{(x,\alpha)\in\mathcal{T}} \left|\frac{1}{n}\sum_{i=1}^{n} \left[\psi(W_i,\hat{\theta}_n(\alpha),\alpha)1(X_i \leq x)\right] - \mathbb{E}[\psi(W,\theta^*(\alpha),\alpha)1(X \leq x)]\right| = o_p(1).$$

□

Proof of Theorem 3.4 The proof parallels to the proof of Theorem 3.2. Notice that Theorem A.1 still holds under H_{1n}. A uniform law of large numbers holds for $n^{-1}\sum_{i=1}^{n} b(X_i,\alpha)1(X_i \leq x)$ under the conditions of Assumption 5. Then Theorem 3.4 follows from this and Lemma A.1. □

Proof of Theorem 3.5 As in Theorem 3.2, we have the following decomposition

$$\hat{R}_n^*(x,\alpha) = \hat{S}_n^*(x,\alpha) - \hat{G}_n'(x,\hat{\theta}_n(\alpha),\alpha)\hat{J}(\alpha)^{-1}\hat{S}_{1n}^*(\alpha),$$

where

$$\hat{S}_n^*(x,\alpha) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i, \hat{\theta}_n(\alpha), \alpha) 1(X_i \leq x) V_i,$$

$$\hat{S}_{1n}^*(\alpha) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i, \hat{\theta}_n(\alpha), \alpha) \dot{m}(X_i, \hat{\theta}_n(\alpha)) V_i,$$

The class of functions $\{(W,V) \to \psi(W, \theta(\alpha), \alpha) 1(X \leq x) V : \alpha \in \mathcal{A}, \theta \in \mathcal{B}, x \in \overline{\mathbb{R}}^d\}$ and $\{(W,V) \to \psi(W, \theta(\alpha), \alpha) \dot{m}(X, \theta(\alpha)) V : \theta \in \mathcal{B}, \alpha \in \mathcal{A}\}$ are Donsker, as can be shown by applying Lemma A.1 of Escanciano and Goh (2014). We obtain from a stochastic equicontinuity argument and the uniform consistency of $\hat{\theta}_n$ that uniformly in $(x,\alpha) \in \mathcal{T}$,

$$\hat{S}_n^*(x,\alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i, \theta_0(\alpha), \alpha) 1(X_i \leq x) V_i + o_p(1),$$

and

$$\hat{S}_{1n}^*(\alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(W_i, \theta_0(\alpha), \alpha) \dot{m}(X_i, \theta_0(\alpha)) V_i + o_p(1).$$

Also, we have $\sup_{(x,\alpha)\in\mathcal{T}} |\hat{G}_n(x, \hat{\theta}_n(\alpha), \alpha) - G(x, \theta_0(\alpha), \alpha)| = o_p(1)$ and $\sup_{\alpha\in\mathcal{A}} |\hat{J}_n^{-1}(\alpha) - J^{-1}(\alpha)| = o_p(1)$. Thus

$$\hat{R}_n^*(x,\alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \left[\psi(W_i, \theta_0(\alpha), \alpha) 1(X_i \leq x) - G'(x, \theta_0(\alpha), \alpha) l(W_i, \theta_0(\alpha), \alpha)\right] + o_p(1).$$

The theorem follows from multiplier central limit theorem, see Theorem 2.9.2 of Van Der Vaart and Wellner (1996), and the continuous mapping theorem. □

References

Aigner, D. J., Amemiya, T. and Poirier, D. J. (1976), 'On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function', *International economic review (Philadelphia)* **17**(2), 377–396.

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.

Bai, Y., Jiang, R. and Zhang, M. (2022), 'Optimal model averaging estimator for expectile regressions', *Journal of statistical planning and inference* **217**, 204–223.

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: Journal of the Econometric Society* pp. 1443–1458.

Bonaccolto-Töpfer, M. and Bonaccolto, G. (2023), 'Gender wage inequality: new evidence from penalized expectile regression', *Journal of economic inequality* **21**(3), 511–535.

Campbell, J. Y. and Thompson, S. B. (2008), 'Predicting excess stock returns out of sample: Can anything beat the historical average?', *The Review of financial studies* **21**(4), 1509–1531.

Chen, T., Su, Z., Yang, Y. and Ding, S. (2020), 'Efficient estimation in expectile regression using envelope models', *Electronic Journal of Statistics* **14**(1), 143 – 173.

Chen, X., Linton, O. and Van Keilegom, I. (2003), 'Estimation of semiparametric models when the criterion function is not smooth', *Econometrica* **71**(5), 1591–1608.

Dawber, J., Salvati, N., Fabrizi, E. and Tzavidis, N. (2022), 'Expectile regression for multi-category outcomes with application to small area estimation of labour force participation', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **185**(S2), S590–S619.

Efron, B. (1991), 'Regression percentiles using asymmetric squared error loss', *Statistica Sinica* **1**(1), 93–125.

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

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

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

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

Guler, K., Ng, P. T. and Xiao, Z. (2017), 'Mincer–Zarnowitz quantile and expectile regressions for forecast evaluations under aysmmetric loss functions', *Journal of Forecasting* **36**(6), 651–679.

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

Kim, M. and Lee, S. (2016), 'Nonlinear expectile regression with application to value-at-risk and expected shortfall estimation', *Computational Statistics & Data Analysis* **94**, 1–19.

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

Koul, H. L. and Stute, W. (1999), 'Nonparametric model checks for time series', *The Annals of Statistics* **27**(1), 204–236.

Kuan, C.-M., Yeh, J.-H. and Hsu, Y.-C. (2009), 'Assessing value at risk with CARE, the Conditional Autoregressive Expectile models', *Journal of Econometrics* **150**(2), 261–270. Recent Development in Financial Econometrics.

Newey, W. K. and Powell, J. L. (1987), 'Asymmetric least-squares estimation and testing', *Econometrica* **55**(4), 819–847.

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

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. (1996), *Weak convergence and empirical processes*, Springer.

Welch, I. and Goyal, A. (2007), 'A Comprehensive Look at The Empirical Performance of Equity Premium Prediction', *The Review of Financial Studies* **21**(4), 1455–1508.

Xu, W., Hou, Y. and Li, D. (2022), 'Prediction of extremal expectile based on regression models with heteroscedastic extremes', *Journal of Business & Economic Statistics* **40**(2), 522–536.

Xu, X., Mihoci, A. and Härdle, W. K. (2018), 'lCARE – localizing conditional autoregressive expectiles', *Journal of Empirical Finance* **48**, 198–220.