Expectile Regression under Misspecification

Zixin Yang
Peking University

Xiaojun Song
Peking University

December 1, 2024

Abstract

Allowing for possible misspecification of a linear conditional expectile function, this paper provides an interpretation and asymptotic distribution theory for the expectile regression (ER) parameter in Newey and Powell (1987). The first result on interpretation shows that ER minimizes a weighted mean-squared error loss function. The weighting function is related to an average of the conditional distribution function of the dependent variable near the true conditional expectile. The weighted least squares interpretation of ER is further used to derive an omitted variables bias formula. We also establish the asymptotic distribution of the ER process under the misspecification of the conditional expectile function. The result provides a foundation for simultaneous confidence intervals and a basis for global tests of hypotheses about distributions.

Keywords: *Conditional expectile function, misspecfication, asymmetric least squares, best linear approximation.*
JEL Classification: C14, C21

1 Introduction

Ordinary least squares (OLS) has become a baseline approach for fitting parsimonious models to a conditional expectation function. Its focus is primarily on evaluating the effect of a set of covariates on the conditional mean of the response variable. In many economic applications, however, more aspects than the mean of the conditional distribution of the response given the covariates are of interest, and the covariate effects may not be homogeneous and/or the noise variables exhibit heavy and asymmetric tails. To capture heterogeneity in the set of covariates at different locations of the response distribution, methods such as quantile regression (Koenker and Bassett, 1978) and asymmetric least squares regression (expectile regression, ER) (Newey and Powell, 1987) have been widely used. We refer the readers to Koenker (2005) for a comprehensive overview of quantile regression, and Newey and Powell (1987) and Man et al. (2024) for expectile regressions.

Both quantiles and expectiles provide complete characterization of an entire distribution. In fact, expectiles are exactly quantiles of a transformed version of the original distribution (Jones, 1994). Expectiles are related to means of location in the same way as quantiles are related to the median. In the context of regression, ER can be interpreted as a least squares analog of regression quantile estimation. Quantile regression is more dominant in the literature due to its intuitive interpretation as the inverse of the distribution function. However, ER also has some desired features. First, the ER coefficient is easier to compute and reasonably efficient under normality conditions (Efron, 1991). The asymptotic covariance matrix can be estimated without the need to estimate the conditional density function in a nonparametric way. Therefore, ER 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 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 an important tool for financial risk management and decision-making.

All the above-mentioned papers, however, assume that the conditional ER model is correctly specified. When the model is misspecified, confidence intervals and hypothesis tests based on the conventional covariance matrix are invalid. Although the properties of OLS under misspecification are well-known, to date, no theory has been proposed for inference when a conditional expectile model may be misspecified. In this paper, we study robust inferences of expectile regression. Our contribution is twofold. First, we show that expectile regression is the best linear approximation to the conditional expectile function using a weighted mean squared error loss function. We also illustrate how this approximation property can be used to interpret multivariate ER coefficients as partial regression coefficients and to develop an omitted variables bias formula for ER. A second contribution is to investigate the limiting distribution of the ER process that accounts for possible misspecification of a linear conditional expectile function. The results can be used to test the global hypothesis of the conditional distribution as well as construct simultaneous confidence intervals.

The rest of the article is organized as follows. Section 2 presents two interpretations of the ER vector under misspecification. The approximation properties are used to interpret multivariate ER coefficients as partial regression coefficients and to develop an omitted variables bias formula in Section 3. We present in Section 4 the asymptotic theory of the ER process under possible misspecification of the conditional expectile function. Section 5 illustrates the usefulness of the theory by a set of Monte Carlo experiments. The last section concludes. The Appendix provides the proof of the theorems.

2 Interpreting ER under misspecification

Given a continuous response variable Y and a $d \times 1$ regressor vector X, we are interested in the conditional expectile function (CEF) of Y given X. Assuming integrability, the CEF is defined as

$$\mu_\tau(Y|X) \in \arg\min_{q(X)} \mathbb{E}[\rho_\tau(Y - q(X))],$$

where $\rho_\tau(\lambda) = |\tau - 1(\lambda < 0)| \cdot \lambda^2$ for $\tau \in (0,1)$, and the minimum is over the set of measurable functions of X. Clearly, the CEF can be viewed as an asymmetric generalization of the conditional mean function, which is $\mu_{1/2}(Y|X)$. It offers insights into the entire conditional distribution of Y given X in much the same way that the conditional quantile function does while enjoying both computational convenience and statistical efficiency under normality conditions (Newey and Powell, 1987).

It may be possible to capture important features of the CEF using a linear model. This motivates linear ER. The linear ER vector solves the population minimization problem

$$\beta(\tau) \equiv \arg\min_{\beta \in \mathbb{R}^d} \mathbb{E}[\rho_\tau(Y - X^\mathsf{T}\beta)]. \tag{1}$$

If $\mu_\tau(Y|X)$ is linear, the ER minimand will find it. Otherwise, the ER vector provides the best linear predictor for Y under the asymmetric square loss function ρ_τ. The first-order conditions for this minimization problem can be shown to imply that for each τ, $\beta(\tau)$ is a solution of the equation

$$\beta(\tau) = \left\{\mathbb{E}\left[|\tau - 1(Y \le X^\mathsf{T}\beta(\tau))|XX^\top\right]\right\}^{-1}\mathbb{E}\left[|\tau - 1(Y \le X^\mathsf{T}\beta(\tau))|XY\right].$$

Misspecification of the regression function may affect the asymmetric least squares coefficients $\beta(\tau)$, since in general $X^\top\beta(\tau)$ is an approximation to the actual conditional expectile function $\mu_\tau(Y|X)$. Our first aim is to establish the nature of the approximation to conditional expectiles that ER provides. Similar to Angrist et al. (2006), we show that the population ER vector minimizes a weighted sum of squared specification errors. For any expectile index $\tau \in (0,1)$, define the ER specification error as

$$D_\tau(X,\beta) \equiv X^\intercal\beta - \mu_\tau(Y|X).$$

Also, let ϵ_τ be an expectile-specific residual, defined as the deviation of the response variable from the conditional expectile of interest,

$$\epsilon_\tau \equiv Y - \mu_\tau(Y|X),$$

Let $F_{\epsilon_\tau}(e|X)$ and $f_{\epsilon_\tau}(e|X)$ be the conditional distribution function and density of ϵ_τ, evaluated at $\epsilon_\tau = e$, respectively. The following theorem shows that ER is a weighted approximation to the unknown CEF.

Theorem 1. *(Approximation Property) Suppose that (i) the conditional density $f_Y(y|X)$ exists a.s., and (ii) $\mathbb{E}|Y|^2 < \infty$ and $\mathbb{E}\|X\|^2 < \infty$. Then, for each $\tau \in (0,1)$, a unique solution $\beta(\tau)$ to (1) exists, and*

$$\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \mathbb{E}[w_\tau(X,\beta)\cdot D_\tau^2(X,\beta)] \tag{2}$$

where

$$\begin{aligned} w_\tau(X,\beta) &= \tau + 2(1-2\tau)\int_0^1 (1-u)\cdot F_{\epsilon_\tau}(uD_\tau(X,\beta)|X)du \\ &= \tau + 2(1-2\tau)\int_0^1 (1-u)\cdot F_Y(u\cdot X^\top\beta + (1-u)\cdot\mu_\tau(Y|X)|X)du \geq 0. \end{aligned}$$

Theorem 1 shows that the population ER coefficient vector $\beta(\tau)$ minimizes the expected weighted mean-squared approximation error with weighting function $w_\tau(X,\beta)$. The weights depend on the index τ as well as the average distribution function of the response variable over a line from the point of approximation, $X^\top\beta$, to the true conditional expectile, $\mu_\tau(Y|X)$. When $\tau = 0.5$, the weights are equal to the constant 0.5, and (2) reduces to the well-known fact that OLS estimates provide the minimum mean-squared error linear approximation to a conditional expectation function $\mathbb{E}(Y|X)$.

As in Angrist et al. (2006), we refer to the function $w_\tau(X,\beta)$ as *importance weights* since this function determines the importance the ER minimand gives to points in the support of

X for a given distribution of X. To understand what determines the shape of the importance weights, we consider the following approximation. Suppose that Y has a conditional density, then for β in the neighborhood of $\beta(\tau)$,

$$\begin{aligned} w_\tau(X,\beta) &= \tau + (1-2\tau)F_Y(\mu_\tau(Y|X)|X) + r_\tau(X), \\ \big|r_\tau(X)\big| &\leq 1/3 \cdot |D_\tau(X,\beta)| \cdot \bar{f}(X). \end{aligned}$$

Here $r_\tau(X)$ is a remainder term, and the density $f_Y(y|X)$ is assumed to be bounded in y in absolute value by $\bar{f}(X)$ a.s.. Clearly, the weights $\tau + (1-2\tau)F_Y(\mu_\tau(Y|X)|X)$ are the primary determinants of the importance weights.

Similar to Angrist et al. (2006), we also derive a second approximation property for ER. This property is particularly well suited to the development of a partial regression decomposition and the derivation of an omitted variables bias formula for ER.

Theorem 2. *(Iterative Approximation Property) Suppose that (i) the conditional density $f_Y(y|X)$ exists and is bounded a.s., and (ii) $\mathbb{E}|Y|^2$, $\mathbb{E}[\mu_\tau(Y|X)^2]$, and $\mathbb{E}\|X\|^2$ is finite. Then $\bar{\beta}(\tau) = \beta(\tau)$ uniquely solve the equation*

$$\bar{\beta}(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \mathbb{E}\big[\bar{w}_\tau(X,\bar{\beta}(\tau)) \cdot D_\tau^2(X,\beta)\big],$$

where

$$\begin{aligned} \bar{w}_\tau(X,\bar{\beta}(\tau)) &= \tau + (1-2\tau)\int_0^1 F_{\epsilon_\tau}\big(u \cdot D_\tau(X,\bar{\beta}(\tau))|X\big)du \\ &= \tau + (1-2\tau)\int_0^1 F_Y\big(u \cdot X^\top\bar{\beta}(\tau) + (1-u)\cdot\mu_\tau(Y|X)|X\big)du. \end{aligned}$$

The point of the above result is that ER solves a weighted least squares approximation problem $\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d}\mathbb{E}[\bar{w}_\tau(X)\cdot D_\tau^2(X,\beta)]$, where the weight is a function of X only. Theorem 2 also shows that the ER coefficient is the unique fixed point to an iterated minimum distance approximation. As in Theorem 1, when $\tau \neq 1/2$, the weighting function $\bar{w}_\tau(X)$ is related to the conditional distribution function of the dependent variable. In particular, when Y has a smooth conditional density around the relevant expectile, we have, by a Taylor approximation,

$$\begin{aligned} \bar{w}_\tau(X,\beta(\tau)) &= \tau + (1-2\tau)F_Y\big(\mu_\tau(Y|X)|X\big) + \bar{r}_\tau(X), \\ \big|\bar{r}_\tau(X)\big| &\leq 1/2 \cdot |D_\tau(X,\beta(\tau))| \cdot \bar{f}(X). \end{aligned}$$

Here, $\bar{r}_\tau(X)$ is a remainder term, and the density $f_Y(y|X)$ is assumed to be bounded in y in absolute value by $\bar{f}(X)$ a.s.. When either $D_\tau(X, \beta(\tau))$ or $\bar{f}(X)$ is small, we have

$$\bar{w}_\tau(X, \beta(\tau)) \approx w_\tau(X, \beta(\tau)) \approx \tau + (1-2\tau)F_Y(\mu_\tau(Y|X)|X), \tag{3}$$

i.e., the approximate weighting function is the same as derived in Theorem 1.

3 Partial ER and omitted variable bias

The approximation property established in Section 2 can be used to derive an omitted variables bias formula and a partial ER concept, similar to the relationship between partial regression and OLS. The partial ER is defined with regard to a partition of the regressor vector X into a variable X_1 and the remaining variables X_2 along with the corresponding partition of ER coefficients $\beta(\tau)$ into $\beta_1(\tau)$ and $\beta_2(\tau)$. Decompose

$$\mu_\tau(Y|X) = \pi_\mu^\top X_2 + q_\tau(Y|X), \text{ where } \mathbb{E}\big[\bar{w}_\tau(X)\cdot X_2 \cdot q_\tau(Y|X)\big] = 0,$$

and

$$X_1 = \pi_1^\top X_2 + V_1, \text{ where } \mathbb{E}\big[\bar{w}_\tau(X)\cdot X_2 \cdot V_1^\top\big] = 0.$$

Here, $q_\tau(Y|X)$ and V_1 are residuals created by a weighted linear projection of $\mu_\tau(Y|X)$ and X_1 on X_2, respectively, using $\bar{w}_\tau(X) = \bar{w}(X, \beta(\tau))$ defined in Theorem 2 as the weight.[1] Using the above decomposition and Theorem 2 and standard least squares algebra, we have

$$\beta_1(\tau) = \arg\min_{\beta_1} \mathbb{E}\big[\bar{w}_\tau(X)\big(q_\tau(Y|X) - V_1^\top\beta_1\big)^2\big].$$

and also $\beta_1(\tau) = \arg\min_{\beta_1} \mathbb{E}\big[\bar{w}_\tau(X)\big(\mu_\tau(Y|X) - V_1^\top\beta_1\big)^2\big]$. This shows that $\beta_1(\tau)$ is a partial ER coefficient in the sense that it can be obtained from a weighted least squares regression of $\mu_\tau(Y|X)$ on X_1, once we have partialled out the effect of X_2. Both the first-step and second-step regressions are weighted by $\bar{w}_\tau(X)$.

An omitted variable bias formula for ER can be similarly derived. Suppose we are interested in an expectile regression with explanatory variables $X = (X_1^\top, X_2^\top)^\top$, but X_2 is not available. We run ER on X_1 only, obtaining the coefficient vector $\gamma_1(\tau) = \arg\min_{\gamma_1} \mathbb{E}[\rho_\tau(Y - X_1^\top\gamma_1)]$. The long regression coefficient vectors are given by $(\beta_1^\top(\tau), \beta_2^\top(\tau))^\top = \arg\min_{\beta_1,\beta_2} \mathbb{E}[\rho_\tau(Y - X_1^\top\beta_1 - X_2^\top\beta_2)]$. Following along the lines of Theorem 2, we can easily find that

$$\gamma_1(\tau) = \beta_1(\tau) + \big(\mathbb{E}[\tilde{w}_\tau(X)\cdot X_1X_1^\top]\big)^{-1}\mathbb{E}[\tilde{w}_\tau(X)\cdot X_1R_\tau(X)], \tag{4}$$

[1]Thus, $\pi_\mu = \mathbb{E}[\bar{w}_\tau(X)X_2X_2^\top]^{-1}\mathbb{E}[\bar{w}_\tau(X)X_2\mu_\tau(Y|X)]$ and $\pi_1 = \mathbb{E}[\bar{w}_\tau(X)X_2X_2^\top]^{-1}\mathbb{E}[\bar{w}_\tau(X)X_2X_1^\top]$.

where $R_\tau(X) = \mu_\tau(Y|X) - X_1^\top \beta_1(\tau)$, $\tilde{w}_\tau(X) = \tau + (1-2\tau)\int_0^1 F_{\epsilon_\tau}(u \cdot D_\tau(X, \gamma_1(\tau)) \mid X)du$, and $D_\tau(X, \gamma_1) = X_1^\top \gamma_1 - \mu_\tau(Y|X)$. Here $R_\tau(X)$ is the part of the CEF not explained by the linear function of X_1 in the long ER. If the CEF is linear, then $R_\tau(X) = X_2^\top \beta_2(\tau)$.

For the mean regression ($\tau = 0.5$), we have $\tilde{w}_\tau(X) = 1/2$ and (4) reduces to $\gamma_1(1/2) = \beta_1(1/2) + (\mathbb{E}[X_1 X_1^\top])^{-1}\mathbb{E}[X_1(Y - X_1^\top \beta_1(1/2))]$. However, when $\tau \neq 0.5$, the regression of omitted variables on included variables is weighted by $\tilde{w}_\tau(X)$, which is a function of X. The effect of omitted variables appears through the remainder term $R_\tau(X)$. In practice, it seems reasonable to think of this as being approximated by the omitted linear part $X_2^\top \beta_2(\tau)$.

4 ER inference under misspecification

In this section, we focus on robust inference methods for the ER process. Specifically, we study the distribution theory for expectile regressions in large samples without imposing the restriction that the CEF is linear. Although not consistent for the true nonlinear CEF, expectile regression consistently estimates the approximations to the CEF given in Theorems 1 and 2. We quantify the sampling uncertainty in estimates of these approximations by deriving the asymptotic distribution of the sample ER process, which is defined as

$$\widehat{\beta}_n(\tau) = \arg\min_{\beta \in \mathbb{R}^d} n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - X_i^\top \beta), \tag{5}$$

where $\tau \in \mathcal{T}$ is a compact subset in $(0,1)$. We target the entire ER process because this enables us to test global hypotheses about conditional distributions and construct simultaneous (joint) confidence regions. Newey and Powell (1987) investigated pointwise inference for multiple expectile coefficients assuming the correct specification of a linear CEF. To our knowledge, the impact of specification error on the ER coefficient has not been discussed in the existing literature. Theorem 3 below fills this gap by establishing joint asymptotic normality for the ER process.

Assumption 1. *(a) $\{W_i = (Y_i, X_i^\top)\}_{i=1}^n$ is a sequence of independent and identically distributed (i.i.d.) random variates. (b) The conditional density $f_Y(y|X = x)$ exists, and is uniformly continuous in y for almost all x over the support of X. (c) $\mathbb{E}\|W\|^{4+\eta} < \infty$ for some $\eta > 0$. (d) $\mathbb{E}[XX^\top]$ is nonsingular.*

Theorem 3. *Suppose that Assumption 1 holds. The expectile regression process is uniformly consistent, i.e., $\sup_{\tau \in \mathcal{T}} \|\widehat{\beta}_n(\tau) - \beta(\tau)\| = o_p(1)$, and $J(\cdot)\sqrt{n}(\widehat{\beta}_n(\cdot) - \beta(\cdot))$ converges in distribution to a zero mean Gaussian process $z(\cdot)$, where $J(\tau) \equiv \mathbb{E}[|\tau - 1(Y \leq X^\top \beta(\tau))|XX^\top]$,* *and $z(\cdot)$ is defined by its covariance function $\Sigma(\tau;\tau') \equiv \mathbb{E}[z(\tau)z(\tau')']$ with*

$$\Sigma(\tau;\tau') = \mathbb{E}\big[\big|\big(\tau - 1(Y \leq X^\intercal\beta(\tau))\big)\big(\tau' - 1(Y \leq X^\intercal\beta(\tau'))\big)\big|\big(Y - X^\intercal\beta(\tau)\big)\big(Y - X^\intercal\beta(\tau')\big)XX^\intercal\big].$$

Alternatively, it can be expressed as

$$\begin{aligned}\Sigma(\tau;\tau') =& \mathbb{E}\big[\big|\big(\tau - 1(\epsilon_\tau \leq D_\tau(X,\beta(\tau)))\big)\big(\tau' - 1(\epsilon_{\tau'} \leq D_{\tau'}(X,\beta(\tau')))\big)\big| \\ & \times \big(\epsilon_\tau - D_\tau(X,\beta(\tau))\big)\big(\epsilon_{\tau'} - D_{\tau'}(X,\beta(\tau'))\big)XX^\intercal\big].\end{aligned}$$

If the model is correctly specified, i.e. $\mu_\tau(Y|X) = X^\top\beta(\tau)$ a.s., then $\Sigma(\tau;\tau')$ simplifies to

$$\Sigma(\tau;\tau') = \mathbb{E}\big[\big|\big(\tau - 1(\epsilon_\tau \leq 0)\big)\big(\tau' - 1(\epsilon_{\tau'} \leq 0)\big)\big|\epsilon_\tau\epsilon_{\tau'}XX^\intercal\big].$$

Theorem 3 establishes joint asymptotic normality for the entire ER process under possible misspecification of the conditional expectile function. The simplest corollary of Theorem 3 is that any finite collection of $\sqrt{n}(\widehat{\beta}(\tau_k) - \beta(\tau))$, $k = 1, 2, \ldots$, is asymptotically jointly normal, with asymptotic covariance between the kth and lth subsets equal to $J(\tau_k)^{-1}\Sigma(\tau_k, \tau_l)J(\tau_l)^{-1}$. Under correct specification, this corresponds to the result in Newey and Powell (1987, Theorem 3). The implication for the standard OLS distribution theory when $\tau = 1/2$ is also obvious.

Inference on the ER process is useful for testing basic hypotheses of the form

$$R(\tau)\beta(\tau) = r(\tau), \quad \text{for all } \tau \in \mathcal{T}, \tag{6}$$

where $R(\tau)$ denotes an $l \times d$ matrix with $l \leq d$, and $r(\tau)$ is an $l \times 1$ vector. We assume that both $R(\tau)$ and $r(\tau)$ are continuous in τ over $\mathcal{T}$. The components $R(\tau)$ and $r(\tau)$ could be functions of the conditional distribution and thus need to be estimated. The hypothesis above embeds several interesting hypotheses about the parameters of the conditional expectile function. For example, we may be interested in testing whether a subset of variables $j \in \{k+1, \ldots, d\}$ enters models for all conditional expectiles with zero coefficients, i.e., whether $\beta_j(\tau) = 0$ for all $\tau \in \mathcal{T}$ and $j = k+1, \ldots, d$. This corresponds to $R(\tau) = [\mathbf{0}_{(d-k)\times k} \quad I_{d-k}]$ and $r(\tau) = \mathbf{0}_{d-k}$.

Denote the estimators of $R(\tau)$ and $r(\tau)$ as $\widehat{R}_n(\tau)$ and $\widehat{r}_n(\tau)$, respectively. We next consider the empirical process

$$v_n(\tau) \equiv \widehat{R}_n(\tau)\widehat{\beta}_n(\tau) - \widehat{r}_n(\tau),$$

and derive from it the test statistics $\Gamma(v_n)$ for some continuous function $\Gamma(\cdot)$ from $\ell^\infty(\mathcal{T})$ to

$\mathbb{R}$. Two commonly used functionals are Kolmogorov-type functional given by

$$K_n = \sqrt{n} \sup_{\tau \in \mathcal{T}} \left\| v_n(\tau) \right\|_{\widehat{V}(\tau)}, \tag{7}$$

and the Cramér–von Mises (CvM)-type functional given by

$$C_n = n \int_{\mathcal{T}} \left\| v_n(\tau) \right\|^2_{\widehat{V}(\tau)} d\Phi(\tau). \tag{8}$$

Here, $\|a\|_V$ denotes $\sqrt{a^\top V a}$, and $\widehat{V}(\tau)$ is a symmetric weight matrix such that $\widehat{V}(\tau) = V(\tau) + o_p(1)$, uniformly in τ, where $V(\tau)$ is some positive definite and continuous symmetric matrix, and Φ is some integrating measure on $\mathcal{T}$.

Remark 1. *Tests of homoskedasticity and conditional symmetry can also be based on the hypothesis given in (6). Assume that data are generated from the linear model* $Y_i = X_i^\top \beta + (X_i^\top \gamma) u_i$, *where* $\{X_i\}_{i=1}^n$ *is a sequence of i.i.d. regression vector of dimension* d *with first element* $X_{i1} = 1$, *and* $\{u_i\}$ *is a sequence of i.i.d. error terms that is independent with* $\{X_i\}_{i=1}^n$. *Partition* $\beta = (\beta_0, \beta_1^\top)^\top$, *and* $\beta(\tau) = (\beta_0(\tau), \beta_1(\tau)^\top)^\top$, *where* β_0 *and* $\beta_0(\tau)$ *are scalars, and* β_1 *and* $\beta_1(\tau)$ *are* $(d-1) \times 1$ *slope vectors. Then, testing of homoskedasticity is equivalent to testing whether the slope coefficient does not vary across expectiles, i.e.,* $\beta_1(\tau) = \beta_1$ *for some* β_1 *and for all* τ. *This corresponds to* $R(\tau) = [\mathbf{0}_{d-1}, I_{d-1}]$ *and* $r(\tau) = \beta_1$. *We can estimate* $r(\tau)$ *in this case by* $\widehat{\beta}_{1n} = R(\tau)(\sum_{i=1}^n X_i X_i^\top)^{-1} \sum_{i=1}^n X_i Y_i$, *and base the test on the empirical process* $v_n(\tau) = \widehat{\beta}_{1n}(\tau) - \widehat{\beta}_{1n}$.

Nonsymmetry can be detected by checking whether the symmetrically placed ER estimator averages up to the least squares estimator. As is shown by Newey and Powell (1987, Theorem 2), if the distribution of Y_i *conditional on* X_i *is symmetric around* $X_i^\top \beta$ *with probability one, then* $[\beta(\tau) + \beta(1-\tau)]/2 = \beta(1/2)$. *This implies that tests for symmetry can be based on the empirical process* $v_n(\tau) = [\widehat{\beta}_n(\tau) + \widehat{\beta}_n(1-\tau)]/2 - \widehat{\beta}_n$, *where* $\widehat{\beta}_n = (\sum_{i=1}^n X_i X_i^\top)^{-1} \sum_{i=1}^n X_i Y_i$ *is the OLS estimator of* $\beta(1/2)$.

In the following, we study the weak convergence of $v_n(\cdot)$ as elements of $l^\infty(\mathcal{T})$, the space of real-valued functions that are uniformly bounded on $\mathcal{T}$. Denote $\Rightarrow$ as the weak convergence on $l^\infty(\mathcal{T})$, see Van Der Vaart and Wellner (1996, Definition 1.3.3) and Dudley (2014, p.136). We make the following assumptions.

Assumption 2. *For each sample size* n, $\{(Y_i, X_i^\top)\}_{i=1}^n$ *is a sequence of i.i.d. variables satisfying that*

$$R(\tau)\beta_n(\tau) - r(\tau) = g(\tau),$$

where for a fixed continuous function $p(\tau) : \mathcal{T} \to \mathbb{R}^l$, *either (a)* $g(\tau) = p(\tau)/\sqrt{n}$, *or (b)* $g(\tau) = p(\tau) \neq 0$.

Assumption 3. *(a) Under the local alternative in Assumption 2(a), the expectile estimates and nuisance parameter estimates satisfy that* $\sqrt{n}J(\cdot)(\widehat{\beta}_n(\cdot) - \beta_n(\cdot)) \Rightarrow z(\cdot)$, $\sqrt{n}(\widehat{R}_n(\cdot) - R(\cdot)) \Rightarrow \rho(\cdot)$, *and* $\sqrt{n}(\widehat{r}_n(\cdot) - r(\cdot)) \Rightarrow \zeta(\cdot)$ *jointly in* $l^\infty(\mathcal{T})$, *where* $J(\cdot)$ *and* $z(\cdot)$ *are defined in Theorem 3, and* (z, ρ, ζ) *is a zero mean continuous Gaussian process with a non-degenerate covariance kernel. (b) Under the global alternative in Assumption 2(b), the same holds, except that the limit* $(z, \tilde{\rho}, \tilde{\zeta})$ *does not need to have the same distribution as in Assumption 3(a), and may depend on the alternative.*

Corollary 1 describes the limiting distribution of the process $v_n(\cdot)$ and the test statistics.

Corollary 1. *(i) Suppose that Assumptions 1, 2(a) and 3 hold. Then, in* $l^\infty(\mathcal{T})$

$$\sqrt{n}v_n(\cdot) \Rightarrow v(\cdot) = v_0(\cdot) + p(\cdot), \quad v_0(\cdot) \equiv R(\cdot)J(\cdot)^{-1}z(\cdot) + \rho(\cdot)\beta(\cdot) - \zeta(\cdot).$$

where $\beta(\tau) = \lim_{n\to\infty} \beta_n(\tau)$. *Under the null hypothesis (6),* $p = 0$, *the test statistic* $K_n \xrightarrow{d} \mathcal{K} \equiv \sup_{\tau\in\mathcal{T}} \|v_0(\tau)\|_{V(\tau)}$, *and* $C_n \xrightarrow{d} \mathcal{C} \equiv \int_{\mathcal{T}} \|v_0(\tau)\|^2_{V(\tau)} d\Psi(\tau)$.

(ii) Suppose that Assumptions 1, 2(b) and 3 hold. Then, $\sqrt{n}(v_n(\cdot) - p(\cdot)) \Rightarrow \tilde{v}(\cdot)$, *where* $\tilde{v}(\tau) = R(\tau)J(\tau)^{-1}\tilde{z}(\tau) + \tilde{\rho}(\cdot)\beta(\cdot) - \tilde{\zeta}(\cdot)$. *Moreover,* K_n *and* C_n *converges in probability to* $+\infty$.

Corollary 1 shows that under the null hypothesis, the Kolmogorov-type and CvM-type test statistics have nonstandard limit distributions. A consistent estimate of the critical value can be simulated by a multiplier bootstrap procedure. The following assumption facilitates the simulation of the asymptotic distributions of $\widehat{R}_n(\cdot)$ and $\widehat{r}_n(\cdot)$.

Assumption 4. *Under the hypothesis (6), uniformly in* $\tau \in \mathcal{T}$,

$$\sqrt{n}(\widehat{R}_n(\tau) - R(\tau)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} L_\tau(Y_i, X_i, R(\tau)) + o_p(1),$$

where $L_\tau(Y_i, X_i, R(\tau))$ *is an* $l \times d$ *matrix taking values in* $\Xi_R \subset \mathbb{R}^{l\times d}$ *such that* $\mathbb{E}[L_\tau(Y_i, X_i, R(\tau))] = 0$ *and* $\mathbb{E}[vec(L_\tau(Y_i, X_i, R(\tau)))vec'(L_\tau(Y_i, X_i, R(\tau)))]$ *exists and is positive definite for all* τ*; and*

$$\sqrt{n}(\widehat{r}_n(\tau) - r(\tau)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_\tau(Y_i, X_i, r(\tau)) + o_p(1),$$

where $l_\tau(Y_i, X_i, r(\tau))$ *is an* $l \times 1$ *vector taking values in* $\Xi_r \subset \mathbb{R}^l$ *such that* $\mathbb{E}[l_\tau(Y_i, X_i, r(\tau))] = 0$ *and* $\mathbb{E}[l_\tau(Y_i, X_i, r(\tau))l'_\tau(Y_i, X_i, r(\tau))]$ *exists and is positive definite for all* τ. *The function classes* $\{W \to L_\tau(Y, X, R) : \tau \in \mathcal{T}, R \in \Xi_R\}$ *and* $\{W \to l_\tau(Y, X, R) : \tau \in \mathcal{T}, R \in \Xi_r\}$ *are Donsker.*

Let $\widehat{J}_n(\tau)$ be a consistent estimate of $J(\tau)$ and $\psi_\tau(u) \equiv |1(u < 0) - \tau| \cdot u$. Define the bootstrap version of the test statistic K_n and C_n as

$$K_n^* \equiv \sqrt{n} \sup_{\tau \in \mathcal{T}} \|v_n^*(\tau)\|_{\widehat{V}(\tau)}, \quad \text{and} \quad C_n^* \equiv n \int_{\mathcal{T}} \left\|v_n^*(\tau)\right\|^2_{\widehat{V}(\tau)} d\tau,$$

where

$$\sqrt{n} v_n^*(\tau) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big\{ L_\tau(Y_i, X_i, \widehat{R}_n(\tau)) \widehat{\beta}_n(\tau) \\ + \widehat{R}_n(\tau) \widehat{J}_n(\tau)^{-1} \psi_\tau(Y_i - X_i^\top \widehat{\beta}_n(\tau)) X_i - l_\tau(Y_i, X_i, \widehat{r}_n(\tau)) \Big\} V_i,$$

and $\{V_i\}_{i=1}^n$ is a sequence of i.i.d. random variables with zero mean, unit variance, bounded support, and independent with $\{W_i\}_{i=1}^n$. A popular example is i.i.d. Bernoulli variables $\{V_i\}$ with $P(V = 1 - \kappa) = \kappa/\sqrt{5}$ and $P(V = \kappa) = 1 - \kappa/\sqrt{5}$, where $\kappa = (\sqrt{5} + 1)/2$ (Mammen, 1993). The bootstrap empirical distribution of K_n^*, i.e., $\hat{F}_n^*(t|\{W_i\}_{i=1}^n) = \mathbb{P}[K_n^* \leq t|\{W_i\}_{i=1}^n]$ is shown to be a consistent estimate of the asymptotic null distribution function of $\mathcal{K}$, i.e., $F_\infty(t) = \mathbb{P}[\mathcal{K} \leq t]$. Hence, the null hypothesis (6) will be rejected at the α-level of significance whenever $K_n \geq c_{n,\alpha}^*$, where $c_{n,\alpha}^*$ is such that $\hat{F}_n^*(c_{n,\alpha}^*|\{W_i\}_{i=1}^n) = 1 - \alpha$. The critical values for C_n can be found similarly.

The next theorem establishes the asymptotic validity of the multiplier bootstrap procedure proposed above. Let $\underset{*}{\Rightarrow}$ denote the weak convergence under the bootstrap law, i.e., conditional on the original samples $\{W_i\}_{i=1}^n$, see Van Der Vaart and Wellner (1996).

Theorem 4. *Suppose that Assumptions 1-4 hold.*

(i) Under the null hypothesis (6), $\sqrt{n} v_n^*(\cdot) \underset{*}{\Rightarrow} v_0(\cdot)$ a.s. in $l^\infty(\mathcal{T})$, where $v_0(\cdot)$ is the Gaussian process defined in Corollary 1(i). Additionally, $K_n^* \underset{*}{\xrightarrow{d}} \mathcal{K}$ a.s. and $C_n^* \underset{*}{\xrightarrow{d}} \mathcal{C}$ a.s..

(ii) Under the alternative in Assumption 2(b), $\sqrt{n} v_n^*(\cdot) \underset{*}{\Rightarrow} \tilde{v}(\cdot)$ a.s. in $l^\infty(\mathcal{T})$, where $\tilde{v}(\cdot)$ is the Gaussian process defined in Corollary 1(ii). Additionally, $K_n^* \underset{*}{\xrightarrow{d}} \tilde{\mathcal{K}} \equiv \sup_{\tau \in \mathcal{T}} \|\tilde{v}(\tau)\|_{V(\tau)}$ a.s. and $C_n^* \underset{*}{\xrightarrow{d}} \tilde{\mathcal{C}} \equiv \int_{\mathcal{T}} \|\tilde{v}(\tau)\|^2_{V(\tau)} d\Phi(\tau)$ a.s..

On the one hand, under the null hypothesis (6), both K_n and K_n^* converge in distribution to $\mathcal{K}$. Hence, a test based on the bootstrapped critical value would yield an asymptotically correct level for K_n. On the other hand, under the alternative in Assumption 3(b), K_n diverges to infinity with probability approaching one, whereas K_n^* is stochastically bounded conditional on the original sample. This implies that the test based on the bootstrapped critical value is consistent against all fixed alternatives. Similar conclusions can be drawn for the CvM test statistic. Therefore, the multiplier bootstrap procedure is asymptotically valid.

For statistical inferences based on Theorem 3 and Corollary 1, it is necessary to consistently estimate $J(\tau)$. It can be easily estimated by the following sample analogue

$$\widehat{J}_n(\tau) = \frac{1}{n}\sum_{i=1}^{n} |\tau - 1(Y_i \leq X_i^\top \widehat{\beta}_n(\tau))| X_i X_i^\top.$$

We show in the Appendix that this estimate is consistent uniformly in τ under the conditions of Theorem 4. Compared with the tests based on the quantile regression process, the covariance matrix of the ER estimator does not involve the conditional density function of the dependent variable, so no nonparametric techniques are needed to obtain its consistent estimate. It also circumvents completely the course of dimensionality.

5 Simulation

In this section, we investigate the finite sample properties of test statistics K_n and C_n by simulation studies. We assume that the linear expectile function is correctly specified. Specifically, we generate data from the following location-scale model:

$$Y_i = X_i^\top \beta + \sigma_i \epsilon_i, \quad i = 1, \ldots, n, \tag{9}$$

where $X_i = (X_{1i}, X_{2i})^\top$, and $\sigma_i = 1 + X_i^\top \eta_1 + 1(\epsilon_i > 0) X_i^\top \eta_2$. The random variables X_{1i} and X_{2i} are taken to be i.i.d. $N(0,1)$ and mutually independent. The errors $\epsilon_i \overset{i.i.d.}{\sim} N(0,1)$ and are independent of $\{X_i\}_{i=1}^n$. We aim to test the homoskedasticity and conditional symmetry, as illustrated in Remark 1. For the former test, the null hypothesis corresponds to the model (9) with $\eta_1 = \eta_2 = 0$. For the latter test, the null hypothesis corresponds to the model (9) with $\eta_2 = 0$.

We consider sample sizes $n \in \{100, 200\}$ and a subinterval of expectiles given by $\mathcal{T} = [0.1, 0.9]$. To calculate the test statistics, we consider K_n and C_n in (7) and (8), respectively, with $\widehat{V}(\tau) = I_l$ and Φ being a uniform discrete distribution over a grid of 60 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 5 displays the rejection frequencies of the K_n and C_n tests for homoskedasticity. The nominal levels are given by 10%, 5%, and 1%. The parameters η_1 and η_2 in model (9) are set to be $\eta_1 = (c_1, c_1)^\top$ and $\eta_2 = (c_2, c_2)^\top$, where c_1 and c_2 take values in $\{0, 0.25, 0.50\}$. First, under the null ($c_1 = c_2 = 0$), both the K_n and C_n tests are moderately oversized when $n = 100$, while the size inflation is reduced as the sample size increases. The empirical power is reasonably high for both tests. The CvM test appears to dominate the K_n test for all the DGPs considered when $c_1 \neq 0$ or $c_1 \neq 0$.

Table 1: Rejection frequencies for the test of homoskedasticity

		$n = 100$						$n = 200$					
		K_n			C_n			K_n			C_n		
c_1	c_2	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
0	0	0.127	0.083	0.031	0.135	0.083	0.027	0.114	0.058	0.018	0.125	0.063	0.023
0.25	0	0.966	0.933	0.837	0.989	0.977	0.924	0.999	0.997	0.992	0.999	0.999	0.998
0.50	0	0.997	0.992	0.967	0.998	0.997	0.993	0.999	0.998	0.997	1	1	0.999
0	0.25	0.729	0.626	0.458	0.685	0.589	0.374	0.966	0.926	0.801	0.931	0.877	0.711
0.25	0.25	0.997	0.992	0.968	0.998	0.996	0.990	1	1	0.999	1	1	1
0.50	0.25	0.975	0.957	0.905	0.986	0.975	0.955	0.996	0.993	0.989	1	1	0.993
0	0.50	0.986	0.963	0.898	0.975	0.949	0.874	1	1	0.999	1	0.999	0.997
0.25	0.50	0.992	0.985	0.958	0.997	0.993	0.982	0.998	0.997	0.997	1	0.999	0.999
0.50	0.50	0.947	0.915	0.840	0.959	0.946	0.894	0.996	0.990	0.968	0.995	0.995	0.990

Table 2 displays the rejection frequencies of the K_n and C_n tests for symmetry. The parameters η_1 and η_2 take the same values as in the test of homoskedasticity above. Under the null ($c_2 = 0$), the empirical rejection probabilities of both K_n and C_n tests are close to the nominal levels. When $c_2 \neq 0$, our tests exhibit nontrivial power. As the sample size increases, the empirical power of both tests significantly improves. Note that larger values of $|c_2|$ imply higher power for the tests considered, whereas larger values of $|c_1|$ imply lower power of the tests. This is expected because in the presence of the heteroskedasticity due to the term $X_i^\top \eta_1$ in σ_i, the signal/noise ratio is smaller than the case with pure nonsymmetry.

Table 2: Rejection frequencies for the test of symmetry

		$n = 100$						$n = 200$					
		K_n			C_n			K_n			C_n		
c_1	c_2	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
0	0	0.109	0.062	0.017	0.096	0.049	0.013	0.094	0.044	0.015	0.096	0.042	0.010
0.25	0	0.107	0.050	0.011	0.093	0.050	0.014	0.118	0.061	0.022	0.113	0.059	0.019
0.50	0	0.097	0.056	0.017	0.092	0.044	0.012	0.100	0.049	0.011	0.091	0.050	0.014
0	0.25	0.252	0.168	0.068	0.256	0.165	0.068	0.499	0.393	0.196	0.540	0.411	0.206
0.25	0.25	0.172	0.105	0.041	0.169	0.108	0.035	0.208	0.130	0.060	0.229	0.142	0.060
0.50	0.25	0.156	0.087	0.031	0.129	0.073	0.023	0.166	0.095	0.036	0.168	0.101	0.037
0	0.50	0.484	0.336	0.188	0.485	0.370	0.194	0.844	0.736	0.535	0.863	0.763	0.562
0.25	0.50	0.272	0.178	0.078	0.247	0.167	0.079	0.447	0.342	0.187	0.469	0.350	0.191
0.50	0.50	0.187	0.109	0.042	0.169	0.099	0.041	0.286	0.190	0.086	0.296	0.200	0.090

6 Conclusion

This paper shows how linear ER provides a weighted least squares approximation to an unknown and possibly nonlinear conditional expectile function. The ER approximation property is further used to examine a partial expectile regression relationship and omitted variables bias. All the results can be seen as an asymmetric generalization of OLS. We also present a misspecification-robust distribution theory for the ER process. This provides a foundation for simultaneous confidence intervals and a basis for global tests of hypotheses about distributions.

Appendix

Proof of Theorem 1.

The proof proceeds by proving the equivalence of the two objective functions. We have

$$\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \left\{\mathbb{E}[\rho_\tau(\epsilon_\tau - D_\tau(X,\beta))] - \mathbb{E}[\rho_\tau(\epsilon_\tau)]\right\}.$$

By definition of ρ_τ, it follows further that

$$\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \left\{\mathbb{E}[\mathcal{A}(X,\beta)] - 2\mathbb{E}[\mathcal{B}(X,\beta)] + \mathbb{E}[\mathcal{C}(X,\beta)]\right\}, \tag{1}$$

where

$$\begin{aligned}
\mathcal{A}(X,\beta) &\equiv \mathbb{E}[|\tau - 1\{\epsilon_\tau \le D_\tau(X,\beta)\}|D_\tau^2(X,\beta) \mid X],\\
\mathcal{B}(X,\beta) &\equiv \mathbb{E}[|\tau - 1\{\epsilon_\tau \le D_\tau(X,\beta)\}|\epsilon_\tau D_\tau(X,\beta) \mid X],\\
\mathcal{C}(X,\beta) &\equiv (1-2\tau)\mathbb{E}[\epsilon_\tau^2[1\{\epsilon_\tau < D_\tau(X,\beta)\} - 1\{\epsilon_\tau < 0\}] \mid X].
\end{aligned}$$

It can be shown that

$$\mathcal{A}(X,\beta) = \left[\tau + (1-2\tau)F_{\epsilon_\tau}(D_\tau(X,\beta) \mid X)\right]\cdot D_\tau^2(X,\beta),$$

For $\mathcal{B}(X,\beta)$, suppose first that $D_\tau(X,\beta) > 0$. By the first order condition, we have $\mathbb{E}[|\tau - 1\{\epsilon_\tau \le 0\}|\epsilon_\tau \mid X] = 0$. Then

$$\begin{aligned}
\mathcal{B}(X,\beta) &= \mathbb{E}\{[|\tau - 1\{\epsilon_\tau \le D_\tau(X,\beta)\}| - |\tau - 1\{\epsilon_\tau \le 0\}|]\epsilon_\tau D_\tau(X,\beta) \mid X\}\\
&= (1-2\tau)\int_0^{D_\tau(X,\beta)} y f_{\epsilon_\tau}(y|X)dy \cdot D_\tau(X,\beta)\\
&= (1-2\tau)\int_0^1 u f_{\epsilon_\tau}(uD_\tau(X,\beta)|X)du \cdot D_\tau^3(X,\beta).
\end{aligned} \tag{2}$$

Finally,

$$
\begin{aligned}
\mathcal{C}(X,\beta) &= (1-2\tau)\int_0^{D_\tau(X,\beta)} y^2 f_{\epsilon_\tau}(y|X)dy \\
&= (1-2\tau)\int_0^1 u^2 f_{\epsilon_\tau}(uD_\tau(X,\beta)|X)du \cdot D_\tau^3(X,\beta).
\end{aligned} \tag{3}
$$

Combining the above results yields $\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \mathbb{E}[w_\tau(X,\beta)\cdot D_\tau^2(X,\beta)]$, where

$$
\begin{aligned}
w_\tau(X,\beta) &= \tau + (1-2\tau)\Big(F_{\epsilon_\tau}(D_\tau(X,\beta)\,|\,X) - \int_0^1 D_\tau(X,\beta)u(2-u)f_{\epsilon_\tau}(uD_\tau(X,\beta)\,|\,X)du\Big) \\
&= \tau + (1-2\tau)\Big(F_{\epsilon_\tau}(D_\tau(X,\beta)\,|\,X) - \int_0^1 u(2-u)dF_{\epsilon_\tau}(uD_\tau(X,\beta)\,|\,X)\Big) \\
&= \tau + 2(1-2\tau)\int_0^1 (1-u)F_{\epsilon_\tau}(uD_\tau(X,\beta)\mid X)du.
\end{aligned}
$$

Here, the third equation follows from integration by parts. A similar argument shows that (2)-3 also hold when $D_\tau(X,\beta) < 0$. □

Proof of Theorem 2.

The proof proceeds by showing that

$$
\beta(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \mathbb{E}[\rho_\tau(Y - X^\top\beta)], \tag{4}
$$

is equivalent to the fixed point $\bar{\beta}(\tau)$ that uniquely solves

$$
\bar{\beta}(\tau) = \arg\min_{\beta\in\mathbb{R}^d} \mathbb{E}\big[\bar{w}_\tau(X,\bar{\beta}(\tau))\cdot D_\tau^2(X,\beta)\big]. \tag{5}
$$

Both objective functions are finite by the stated conditions.

By convexity of (4) in β, any fixed point $\beta = \bar{\beta}(\tau)$ solves the first-order condition $\mathcal{F}(\beta) = 2\mathbb{E}\big[\bar{w}_\tau(X,\beta)D_\tau(X,\beta)X\big] = 0$. Since (4) is convex on β, the ER vector solves the first-order condition $\mathcal{H}(\beta) = \mathbb{E}[\mathcal{H}(X,\beta)] = 0$, where

$$
\mathcal{H}(X,\beta) = -2\mathbb{E}\big[\big|\tau - 1(\epsilon_\tau < D_\tau(X,\beta))\big|(\epsilon_\tau - D_\tau(X,\beta))X \mid X\big].
$$

An argument similar to that used to establish equation (2) yields

$$
\begin{aligned}
\mathcal{H}(X,\beta) = \;& 2\mathbb{E}\big[\big|\tau - 1(\epsilon_\tau < D_\tau(X,\beta))\big|\big|X\big]D_\tau(X,\beta)X \\
& - 2\mathbb{E}\big[\big(\big|\tau - 1(\epsilon_\tau < D_\tau(X,\beta))\big| - \big|\tau - 1(\epsilon_\tau < 0)\big|\big)\epsilon_\tau \mid X\big]X \\
=\;& 2\big[\tau + (1-2\tau)F_{\epsilon_\tau}(D_\tau(X,\beta)\mid X)\big]\cdot D_\tau(X,\beta)X
\end{aligned}
$$

$$
\begin{aligned}
&- 2(1-2\tau)\int_0^1 u f_{\epsilon_\tau}\big(uD_\tau(X,\beta)|X\big)du \cdot D_\tau^2(X,\beta)X \\
&= 2\bigg(\tau + (1-2\tau)\int_0^1 F_{\epsilon_\tau}\big(u\cdot D_\tau(X,\beta)|X\big)du\bigg)\cdot D_\tau(X,\beta)X \\
&= 2\bar{w}_\tau(X,\beta)D_\tau(X,\beta)X.
\end{aligned}
$$

The functions $\mathcal{F}(\beta)$ and $\mathcal{H}(\beta)$ are therefore identical. Since $\beta = \beta(\tau)$ uniquely satisfies $\mathcal{H}(\beta) = 0$, it also uniquely satisfies $\mathcal{F}(\beta) = 0$. Therefore, $\beta = \beta(\tau) = \bar{\beta}(\tau)$ is the unique solution to both (4) and (5). □

Proof of Theorem 3.

We follow the approach of Angrist et al. (2006); see the proof of Theorem 3 therein. For $W = (Y, X^\top)^\top$, denote $\mathbb{E}_n[f(W)] = n^{-1}\sum_{i=1}^n f(W_i)$ and $\mathbb{G}_n[f(W)] = n^{-1/2}\sum_{i=1}^n (f(W_i) - \mathbb{E}[f(W_i)])$. For a square matrix A, let $\lambda_{\min}(A)$ denote its minimum eigenvalue. For a normed space of real functions $(\mathcal{G}, \rho)$, let $N_{[\cdot]}(\epsilon, \mathcal{G}, \rho)$ denote the covering number with bracketing, and $J_{[\cdot]}(\delta, \mathcal{G}, \rho) = \int_0^\delta \sqrt{\log N_{[\cdot]}(\epsilon, \mathcal{G}, \rho)}d\epsilon$, see Van Der Vaart (1998, p.270) for details.

For each $\tau \in \mathcal{T}$, $\widehat{\beta}_n(\tau)$ minimizes $Q_n(\beta,\tau) = \mathbb{E}_n[\rho_\tau(Y - X^\top\beta) - \rho_\tau(Y - X^\top\beta(\tau))]$. Define $Q_\infty(\beta,\tau) = \mathbb{E}[\rho_\tau(Y - X^\top\beta) - \rho_\tau(Y - X^\top\beta(\tau))]$. It can be shown that $\mathbb{E}\|W\|^2 < \infty$ implies that $Q_\infty(\beta,\tau) < \infty$. Therefore, $Q_\infty(\beta,\tau)$ is finite and, from the lines of Theorem 3 in Newey and Powell (1987), it is uniquely minimized at $\beta(\tau)$ for each τ in $\mathcal{T}$.

We first show the uniform convergence, namely for any compact set $\mathcal{B}$, $Q_n(\tau,\beta) = Q_\infty(\tau,\beta) + o_p(1)$ uniformly in $(\tau,\beta) \in \mathcal{T}\times\mathcal{B}$. Pointwise convergence follows the Khinchine law of large numbers. The empirical process $(\tau,\beta) \mapsto Q_n(\tau,\beta)$ is stochastically equicontinuous because $|Q_n(\tau',\beta') - Q_n(\tau'',\beta'')| \le C_{1n}\cdot|\tau'-\tau''| + C_{2n}\cdot\|\beta'-\beta''\|$, where $C_{1n} = 2\mathbb{E}_n\|W\|^2(1 + \sup_{\beta\in\mathcal{B}}\|\beta\|^2) = O_p(1)$, and $C_{2n} = C\mathbb{E}_n\|W\|^2(1+\sup_{\beta\in\mathcal{B}}\|\beta\|) = O_p(1)$. Hence, the convergence also holds uniformly. The uniform consistency of $\widehat{\beta}_n$ can be shown in a similar way as Angrist et al. (2006, Theorem 3), and the details are omitted.

Next, we establish the asymptotic Gaussianity of the sample ER process. Denote $\psi_\tau(\lambda) = |\tau - 1(\lambda < 0)|\cdot\lambda$. Then we have $\partial\rho_\tau(Y_i - X_i^\top\beta)/\partial\beta = -2X_i\psi_\tau(Y_i - X_i^\top\beta)$. Recall that $\widehat{\beta}_n(\tau)$ minimizes $Q_n(\tau,\beta) = \mathbb{E}_n[\rho_\tau(Y - X^\top\beta)]$. By the first order condition, we have for all $\tau \in \mathcal{T}$,

$$
\mathbb{E}_n[\psi_\tau(Y - X^\top\widehat{\beta}_n(\tau))X] = 0. \tag{6}
$$

Second, $(\tau,\beta) \mapsto \mathbb{G}_n[\psi_\tau(Y - X^\top\beta)X]$ is stochastically equicontinuous over $\mathcal{T}\times\mathcal{B}$, where $\mathcal{B}$ is any compact set, with respect to the $L_2(P)$ pseudometric $\rho((\tau',\beta'),(\tau'',\beta''))^2 \equiv \max_{j=1,\ldots,d}\mathbb{E}[(\psi_{\tau'}(Y - X^\top\beta')X_j - \psi_{\tau''}(Y - X^\top\beta'')X_j)^2]$. To see this, note that $|\psi_{\tau'}(Y - X^\top\beta')X - \psi_{\tau''}(Y - X^\top\beta'')X| \le |\tau'-\tau''|F_1(W) + \|\beta'-\beta''\|F_2(W)$, where $F_1(W) = \|W\|^2(1+\sup_{\beta\in\mathcal{B}}\|\beta\|)$ and $F_2(W) = 2\|X\|^2$ are two square integrable functions. Applying Theorem 3 of Chen et al. (2003), we have $J_{[\cdot]}(\delta_n, H, \rho) \to 0$ for every $\delta_n \downarrow 0$, where H

denotes the function class $\{W \to \psi_\tau(Y - X^\top\beta)X : \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$. Besides, it is easy to see that the class H has an envelope function with finite $(2+\delta)$th moment for some $\delta > 0$ under Assumption 1(c). Stochastic equicontinuity follows from the proof of Theorem 19.28 in Van Der Vaart (1998).

Third, by stochastic equicontinuity of $(\tau, \beta) \mapsto \mathbb{G}_n[\psi_\tau(Y - X^\top\beta)X]$ we have that

$$\mathbb{G}_n[\psi_\tau(Y - X^\top\widehat{\beta}_n(\tau))X] = \mathbb{G}_n[\psi_\tau(Y - X^\top\beta(\tau))X] + o_p(1) \quad \text{in } l^\infty(\mathcal{T}), \tag{7}$$

which follows from $\sup_{\tau\in\mathcal{T}} \|\widehat{\beta}_n(\tau) - \beta(\tau)\| = o_p(1)$ and resulting convergence with respect to the pseudometric $\sup_{\tau\in\mathcal{T}} \rho[(\tau, \widehat{\beta}_n(\tau)), (\tau, \beta(\tau))]^2 = o_p(1)$. The latter is immediate from $\sup_{\tau\in\mathcal{T}} \rho[(\tau, \beta(\tau)), (\tau, b(\tau))]^2 \leq C \sup_{\tau\in\mathcal{T}} \|\beta(\tau) - b(\tau)\|^2$, where $C = 2\mathbb{E}\|W\|^4 < \infty$.

Furthermore, the following expansion is valid uniformly in τ:

$$\mathbb{E}[\psi_\tau(Y - X^\top\beta)X]|_{\beta=\widehat{\beta}_n(\tau)} = -[J(\tau) + o_p(1)]\big(\widehat{\beta}_n(\tau) - \beta(\tau)\big). \tag{8}$$

By Taylor expansion, $\mathbb{E}[\psi_\tau(Y - X^\top\beta)X]|_{\beta=\widehat{\beta}_n(\tau)} = -\mathbb{E}[|\tau - 1(Y \leq X^\top b(\tau))|XX^\intercal]|_{b(\tau)=\beta_n^*(\tau)} \times (\widehat{\beta}_n(\tau) - \beta(\tau))$, where $\beta_n^*(\tau)$ is on the line connecting $\widehat{\beta}_n(\tau)$ and $\beta(\tau)$ for each τ. Then (8) follows by the uniform consistency of $\widehat{\beta}_n(\tau)$, the uniform continuity of the mapping $y \mapsto f_Y(y|x)$, uniformly in x over the support of X and Assumption 1(c).

Fourth, we have that

$$o_p(1) = -[J(\cdot) + o_p(1)]\sqrt{n}\big(\widehat{\beta}_n(\cdot) - \beta(\cdot)\big) + \mathbb{G}_n[\psi_\cdot(Y - X^\top\beta(\cdot))X], \tag{9}$$

because $n^{1/2}$ times the left-hand side of (6) is equal to the left-hand side of $n^{1/2}$ times (8) plus the left-hand side of (7). Since $\lambda_{\min}(J(\tau)) \geq \lambda > 0$, uniformly in $\tau \in \mathcal{T}$,

$$\sup_{\tau\in\mathcal{T}} \big\|\mathbb{G}_n[\psi_\cdot(Y - X^\top\beta(\cdot))X] + o_p(1)\big\| \geq (\sqrt{\lambda} + o_p(1)) \cdot \sup_{\tau\in\mathcal{T}} \sqrt{n}\|\widehat{\beta}_n(\tau) - \beta(\tau)\|. \tag{10}$$

Fifth, the mapping $\tau \mapsto \beta(\tau)$ is continuous by the implicit function theorem and stated assumptions. Because $\beta(\tau)$ solves $\mathbb{E}[|\tau - 1(Y \leq X^\intercal\beta)|(Y - X^\intercal\beta)X] = 0$, $d\beta(\tau)/d\tau = -J(\tau)^{-1}\mathbb{E}[|Y - X^\top\beta(\tau)|X]$. Hence $\tau \mapsto \mathbb{G}_n[\psi_\tau(Y - X^\top\beta(\tau))X]$ is stochastically equicontinuous over $\mathcal{T}$ for the pseudometric given by $\rho(\tau', \tau'') \equiv \rho((\tau', \beta(\tau')), (\tau'', \beta(\tau'')))$. Stochastic equicontinuity of $\tau \mapsto \mathbb{G}_n[\psi_\tau(Y - X^\top\beta(\tau))X]$ and a multivariate central limit theorem imply that

$$\mathbb{G}_n[\psi_\cdot(Y - X^\top\beta(\cdot))X] \Rightarrow z(\cdot) \quad \text{in } l^\infty(\mathcal{T}), \tag{11}$$

where $z(\cdot)$ is a Gaussian process with covariance function $\Sigma(\cdot, \cdot)$. Therefore, the left-hand side of (10) is $O_p(n^{-1/2})$, implying $\sup_{\tau\in\mathcal{T}} \|\sqrt{n}(\widehat{\beta}_n(\tau) - \beta(\tau))\| = O_p(1)$. Finally, the latter fact and (9)-(11) imply that in $l^\infty(\mathcal{T})$,

$$J(\cdot)\sqrt{n}\big(\widehat{\beta}_n(\cdot)-\beta(\cdot)\big)=\mathbb{G}_n[\psi_\cdot(Y-X^\intercal\beta(\cdot))X]+o_p(1)\Rightarrow z(\cdot).$$

This completes the proof of the theorem. $\square$

Proof of Corollary 1.

Under the local alternative in Assumption 2(a), $R(\tau)\beta_n(\tau)-r(\tau)-p(\tau)/\sqrt{n}=0$; thus

$$\begin{aligned}\sqrt{n}v_n(\tau)&=\sqrt{n}\big(\widehat{R}_n(\tau)\widehat{\beta}_n(\tau)-\widehat{r}_n(\tau)\big)\\&=R(\tau)\sqrt{n}\left[\widehat{\beta}_n(\tau)-\beta_n(\tau)\right]+\sqrt{n}\left[\widehat{R}_n(\tau)-R(\tau)\right]\widehat{\beta}_n(\tau)-\sqrt{n}[\widehat{r}_n(\tau)-r(\tau)]+p(\tau).\end{aligned}$$

Corollary 1(i) follows from Assumption 3(a), the consistency of $\widehat{\beta}_n(\cdot)$ and the continuous mapping theorem.

Under any fixed alternative, $R(\tau)\beta(\tau)-r(\tau)-p(\tau)=0$ for some continuous function $p(\tau)\neq 0$; thus

$$\begin{aligned}\sqrt{n}(v_n(\tau)-p(\tau))&=\sqrt{n}\big(\widehat{R}_n(\tau)\widehat{\beta}_n(\tau)-\widehat{r}_n(\tau)-p(\tau)\big)\\&=R(\tau)\sqrt{n}\left[\widehat{\beta}_n(\tau)-\beta_n(\tau)\right]+\sqrt{n}\left[\widehat{R}_n(\tau)-R(\tau)\right]\widehat{\beta}_n(\tau)-\sqrt{n}[\widehat{r}_n(\tau)-r(\tau)].\end{aligned}$$

Corollary 1(ii) follows immediately from Assumption 3(b) and the consistency of $\widehat{\beta}_n(\cdot)$. $\square$

Proof of Theorem 4.

(i) In a similar manner as Theorem 3, we can show that the class of functions $\{(W,V)\to\psi_\tau(Y-X^\top\beta)XV:\tau\in\mathcal{T},\beta\in\mathcal{B}\}$ is Donsker. It follows from a stochastic equicontinuity argument and the consistency of $\widehat{\beta}_n(\tau)$ that, uniformly in $\tau\in\mathcal{T}$ we have the convergences

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_\tau(Y_i-X_i^\top\widehat{\beta}_n(\tau))X_iV_i=\frac{1}{\sqrt{n}}\sum_{i=1}^n\psi_\tau(Y_i-X_i^\top\beta(\tau))X_iV_i+o_p(1).\tag{12}$$

Similarly, using Assumption 4, we have uniformly in $\tau\in\mathcal{T}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n L_\tau(Y_i,X_i,\widehat{R}_n(\tau))V_i=\frac{1}{\sqrt{n}}\sum_{i=1}^n L_\tau(Y_i,X_i,R(\tau))V_i+o_p(1).\tag{13}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n l_\tau(Y_i,X_i,\widehat{r}_n(\tau))V_i=\frac{1}{\sqrt{n}}\sum_{i=1}^n l_\tau(Y_i,X_i,r(\tau))V_i+o_p(1).\tag{14}$$

By Assumptions 3(a), 4 and the multiplier central limit theorem, see Van Der Vaart and Wellner (1996, Theorem 2.9.7), we have that $n^{-1/2}J(\cdot)\sum_{i=1}^n\psi_\cdot(Y_i-X_i^\top\beta(\cdot))X_iV_i$,

$n^{-1/2}\sum_{i=1}^n L.(Y_i, X_i, R(\cdot))V_i$, and $n^{-1/2}\sum_{i=1}^n l.(Y_i, X_i, r(\cdot))V_i$ converge in distribution to $z(\cdot)$, $\rho(\cdot)$ and $\zeta(\cdot)$, respectively, under the bootstrap law. The above results and the consistency of $\widehat{R}_n$ and $\widehat{\beta}_n$ yields that $\sqrt{n}v_n^*(\cdot) \underset{*}{\Rightarrow} v_0(\cdot)$ a.s.. The rest of the part of Theorem 4(i) follows from the continuous mapping theorem.

(ii) Under the fixed alternative, Equations (12)-(14) remain valid. Thus, Theorem 4(ii) can be shown in the same way as above. We omit the details. $\square$

Proof of uniform consistency of $\widehat{J}_n(\tau)$

Recall that $\widehat{J}_n(\tau) = n^{-1}\sum_{i=1}^n |\tau - 1(Y_i \le X_i^\top \widehat{\beta}_n(\tau))| X_i X_i^\top$. We will show that

$$\widehat{J}_n(\tau) - J(\tau) = o_p(1) \quad \text{uniformly in } \tau \in \mathcal{T}. \tag{15}$$

It can be easily verified that $\{W \to |\tau - 1(Y \le X^\top\beta)|XX^\top : \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$ is Donsker, and hence a Glivenko–Cantelli class, for any compact set $\mathcal{B}$, e.g., using Theorem 2.10.6 in Van Der Vaart and Wellner (1996). This implies that $\mathbb{E}_n[|\tau - 1(Y \le X^\top\beta)|XX^\top] - \mathbb{E}[|\tau - 1(Y \le X^\top\beta)|XX^\top] = o_p(1)$ uniformly in $(\tau, \beta) \in \mathcal{T} \times \mathcal{B}$. The latter, the continuity of $\mathbb{E}[|\tau - 1(Y \le X^\top\beta)|XX^\top]$ in (τ, β), and the consistency of $\widehat{\beta}_n$ imply (15).

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.

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

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 multicategory 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.

Dudley, R. M. (2014), *Definition of Donsker Classes*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, p. 133–174.

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

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.

Jones, M. (1994), ‘Expectiles and M-quantiles are quantiles’, *Statistics & Probability Letters* **20**(2), 149–153.

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. (2005), *Quantile regression*, Vol. no. 38, Cambridge University Press, Cambridge;New York;.

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

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.

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

Man, R., Tan, K. M., Wang, Z. and Zhou, W.-X. (2024), ‘Retire: Robust expectile regression in high dimensions’, *Journal of Econometrics* **239**(2), 105459.

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

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.

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.