Interpretation and Semiparametric Efficiency in Quantile Regression under Misspecification

Ying-Ying Lee

Department of Economics, University of Oxford, Manor Road Building, Manor Road,
Oxford OX1 3UQ, UK; E-Mail: ying-ying.lee@economics.ox.ac.uk

Abstract: Allowing for misspecification in the linear conditional quantile function, this paper provides a new interpretation and the semiparametric efficiency bound for the quantile regression parameter $\beta(\tau)$ in Koenker and Bassett (1978). The first result on interpretation shows that under a mean-squared loss function, the probability limit of the Koenker–Bassett estimator minimizes a weighted distribution approximation error, defined as $F_Y(X'\beta(\tau)|X) - \tau$, *i.e.*, the deviation of the conditional distribution function, evaluated at the linear quantile approximation, from the quantile level. The second result implies that the Koenker–Bassett estimator semiparametrically efficiently estimates the quantile regression parameter that produces parsimonious descriptive statistics for the conditional distribution. Therefore, quantile regression shares the attractive features of ordinary least squares: interpretability and semiparametric efficiency under misspecification.

Keywords: semiparametric efficiency bounds; misspecification; conditional quantile function; conditional distribution function; best linear approximation

JEL classification: C14; C21

1. Introduction

This paper revisits the approximation properties of the linear quantile regression under misspecification ([1–3]). The quantile regression estimator, introduced by the seminal paper of Koenker and Bassett [4], offers parsimonious summary statistics for the conditional quantile function and is computationally tractable. Since the development of the estimator, researchers have frequently used quantile regression, in conjunction with ordinary least squares regression, to analyse how the outcome variable responds to the explanatory variables. For example, to model wage structure in labour economics, Angrist, Chernozhukov, and Fernández-Val [1] study returns to education at different points in the wage distribution and changes in inequality over time. A thorough review of recent developments in quantile regression can be found in [5]. The object of interest of this paper is the quantile regression (QR) parameter that is the probability limit of the Koenker–Bassett estimator without assuming the true conditional quantile function to be linear. Two results are presented: a new interpretation and the semiparametric efficiency bound for the QR parameter.

The topic of interest is the conditional distribution function (CDF) of a continuous response variable Y given the regressor vector X, denoted as $F_Y(y|X)$. An alternative for the CDF is the conditional quantile function (CQF) of Y given X, defined as $Q_\tau(Y|X) := \inf\{y : F_Y(y|X) \geq \tau\}$ for any quantile index $\tau \in (0, 1)$. Assuming integrability, the CQF minimizes the check loss function

$$Q_\tau(Y|X) \in \arg\min_{q \in \mathcal{Q}} E\left[\rho_\tau(Y - q(X))\right]$$

where $\mathcal{Q}$ is the set of measurable functions of X, $\rho_\tau(u) = u(\tau - \mathbf{1}_{\{u \leq 0\}})$ is known as the check function and $\mathbf{1}_{\{\cdot\}}$ is the indicator function. A linear approximation to the CQF is provided by the QR parameter $\beta(\tau)$, which solves the population minimization problem

$$\beta(\tau) := \arg\min_{b \in R^d} E\left[\rho_\tau(Y - X'b)\right] \tag{1}$$

assuming the integrability and uniqueness of the solution, and d is the dimension of X. The QR parameter $\beta(\tau)$ provides a simple summary statistic for the CQF. The QR estimator introduced in [4] is the sample analogue

$$\hat{\beta}(\tau) \in \arg\min_{b \in R^d} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'b) \tag{2}$$

for the random sample $(Y_i, X_i', i \leq n)$ on the random variables (Y, X'). By the equivalent first-order condition, this estimator $\hat{\beta}(\tau)$ is also the generalized method of moments (GMM) estimator based on the unconditional moment restriction ([6,7])

$$E[(\tau - \mathbf{1}_{\{Y \leq X'\beta(\tau)\}})X] = 0 \tag{3}$$

This paper focuses on the population QR parameter defined by (1) or equivalently (3).

If the CQF is modelled to be linear in the covariates $Q_\tau(Y|X) = X'\beta(\tau)$ or $F_Y(X'\beta(\tau)|X) = \tau$, the coefficient $\beta(\tau)$ satisfies the conditional moment restriction

$$E[\tau - \mathbf{1}_{\{Y \leq X'\beta(\tau)\}}|X] = 0 \tag{4}$$

almost surely. In the theoretical and applied econometrics literature, this linear QR model is often assumed to be correctly specified. Nevertheless, a well-known crossing problem arises: the CQF for different quantiles may cross at some values of X, except when $\beta(\tau)$ is the same for all τ. A logical monotone requirement is violated for $Q_\tau(Y|X)$ or its estimator to be weakly increasing in the probability index τ given X. The crossing problem for estimation could be treated by rearranging the estimator (for example, see [8] and the references therein.[1]). However, the crossing problem remains for the population CQF, suggesting that the linear QR model (4) is inherently misspecified. That is, there is no $\beta(\tau) \in R^d$ satisfying the conditional moment (4) almost surely. Therefore, the parameter of interest in this paper is the QR parameter $\beta(\tau)$ defined by (1) or (3) without the linear CQF assumption in (4). We can view $\beta(\tau)$ as the pseudo-true value of the linear QR model under misspecification. As the Koenker–Bassett QR estimator is widely used, it is important to understand the approximation nature of the estimand.

For the mean regression counterpart, ordinary least squares (OLS) consistently estimates the linear conditional expectation and minimizes mean-squared error loss for fitting the conditional expectation under misspecification. Chamberlain [9] proves the semiparametric efficiency of the OLS estimator, which provides additional justification for the widespread use of OLS. The attractive features of OLS, interpretability and semiparametric efficiency, under misspecification, motivate my investigation of parallel properties in QR. I study how this QR parameter approximates the CQF and the CDF and calculate its semiparametric efficiency bound.

The first contribution of this paper is on how $\beta(\tau)$ minimizes the distribution approximation error, defined by $F_Y(X'\beta(\tau)|X) - \tau$, under a mean-squared loss function. The first-order condition (3)

---

[1]   Chernozhukov, Fernández-Val, and Galichon [8] rearrange an estimator $\hat{Q}_\tau(Y|X)$ to be monotonic. The original estimator can be computationally tractable. The rearranged monotonic estimated CDF is $\hat{F}_Y(y|X) = \int_0^1 \mathbf{1}_{\{\hat{Q}_\tau(Y|X) \leq y\}} d\tau$. The rearranged quantile estimation is $\hat{Q}_\tau^*(Y|X) = \inf\{y : \hat{F}_Y(y|X) \geq \tau\}$.

can be understood as the orthogonality condition of the covariates X and the distribution approximation error in the projection model. I show that the QR parameter $\beta(\tau)$ minimizes the mean-squared distribution approximation error, inversely weighted by the conditional density function $f_Y(X'\beta(\tau)|X)$. Angrist, Chernozhukov, and Fernández-Val [1] (henceforth ACF) show that $\beta(\tau)$ minimizes the mean-squared quantile specification error, defined by $Q_\tau(Y|X) - X'\beta(\tau)$, using a weight primarily determined by the conditional density. ACF's study, as well as my own results, suggests that QR approximates the CQF more accurately at points with more observations, but the corresponding CDF evaluated at the approximated point $F_Y(X'\beta(\tau)|X)$ is more distant from the targeted quantile level τ. This trade-off is controlled by the conditional density, which is distinct from OLS approximating the conditional mean, because the distribution and quantile functions are generally nonlinear operators. This observation is novel and increases the understanding of how the QR summarizes the outcome distribution. A numerical example in Figure 1 in Section 4 illustrates this finding.

The second result is the semiparametric efficiency bound of the $\beta(\tau)$. Chamberlain's results in [9] on the mean regression based on differentiable moment restrictions cannot be applied to semiparametric efficiency for QR, due to the lack of moment function differentiability in (3). Although Ai and Chen [10] provide general results for sequential moment restrictions containing unknown functions, which could cover the quantile regression setting, I calculate the efficiency bound accommodating regularity conditions specifically for the QR parameter $\beta(\tau)$ using the method of Severini and Tripathi [11]. It follows that the misspecification-robust asymptotic variance of the QR estimator $\hat{\beta}(\tau)$ in (2) attains this bound, which means no regular[2] estimator for (3) has smaller asymptotic variance than $\hat{\beta}(\tau)$. This result might be expected for an M-estimator, but, to my knowledge, the QR application has not been demonstrated and discussed rigorously in any publication. Furthermore, I calculate the efficiency bounds for jointly estimating QR parameters at a finite number of quantiles for both linear projection (3) and linear QR (4) models. Employing the widely-used method of Newey [12], Newey and Powell [13] find the semiparametric efficiency bound for $\beta(\tau)$ of the correctly-specified linear CQF in (4). Note that the efficiency bounds for (3) do not imply the bounds for (4); nor does the converse hold.

In Section 2, I discuss the interpretation of the misspecified QR model in terms of approximating the CDF and the CQF. The theorems for the semiparametric efficiency bounds are in Section 3. In Section 4, I discuss the parallel properties of QR and OLS. The paper is concluded by a review of some existing efficient estimators for linear projection model (3) and linear QR model (4).

2. Interpreting QR under Misspecification

Let Y be a continuous response variable and X be a $d \times 1$ regressor vector. The quantile-specific residual is defined as the distance between the response variable and the CQF, $\varepsilon_\tau := Y - Q_\tau(Y|X)$ with the conditional density $f_{\varepsilon_\tau}(e|X)$ at $\varepsilon_\tau = e$ or $f_Y(y|X)$ at $Y = y = e + Q_\tau(Y|X)$ for any $\tau \in (0,1)$. This is a semiparametric problem in the sense that the distribution functions of ε_τ and X, as well as the CQF, are unspecified and unrestricted other than by the following assumptions, which are standard in QR models. I assume the following regularity conditions, based on the conditions of Theorem 3 in ACF.

(R1) $(Y_i, X_i, i \leq n)$ are independent and identically distributed on the probability space $(\Omega, \mathcal{F}, P)$ for each n;

(R2) the conditional density $f_Y(y|X = x)$ exists and is bounded and uniformly continuous in y, uniformly in x over the support of X;

(R3) $J(\tau) := E[f_Y(X'\beta(\tau)|X)XX']$ is positive definite for all $\tau \in (0,1)$, where $\beta(\tau)$ is uniquely defined in (1);

---

[2]   See [12] for the definition of regular estimators.

(R4) $E\|X\|^{2+\epsilon} < \infty$ for some $\epsilon > 0$;

(R5) $f_Y(X'\beta(\tau)|X)$ to be bounded away from zero.

The identification of the pseudo-true parameter $\beta(\tau)$ is assumed in (R3). The bounded conditional density function of the continuous response variable Y given X in (R2) is needed for the existence of the CQF for any $\tau \in (0,1)$. The uniform continuity guarantees the existence and differentiability of the distribution function, *i.e.*, $dF_Y(y|X)/dy = f_Y(y|X)$ and $F_Y(y|X) = \int_{-\infty}^{y} f_Y(u|X)du$ with probability one. (R4) is used for the asymptotic normality of $\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$. The covariates X are allowed to contain discrete components. (R5) guarantees that the objective function defined below in Equation (6) is finite $\forall \beta \in R^d$, where $\beta(\tau)$ is the parameter of interest uniquely defined by Equation (1).

The parameter of interest $\beta(\tau)$ is equivalent to solving

$$E\left[X\left(F_Y(X'\beta(\tau)|X) - \tau\right)\right] = 0 \tag{5}$$

by applying the law of iterated expectations on Equation (3). Equation (5) states that X is orthogonal to the distribution approximation error $F_Y(X'\beta(\tau)|X) - \tau$. The following theorem interprets QR via a weighed mean-squared error loss function on the distribution approximation error.

Theorem 1. *Assume (R1)–(R5). Then, $\bar{\beta}(\tau) = \beta(\tau)$ solves the equation*

$$\bar{\beta}(\tau) = \arg\min_{b \in R^d} E\left[\left(f_Y(X'\bar{\beta}(\tau)|X)\right)^{-1}\left(F_Y(X'b|X) - \tau\right)^2\right]. \tag{6}$$

Furthermore, if $E\left[\left(f_Y(X'b|X) + (F_Y(X'b|X) - \tau)f_Y'(X'b|X)/f_Y(X'b|X)\right)XX'\right]$ is positive definite at $b = \beta(\tau)$, then $\bar{\beta}(\tau) = \beta(\tau)$ is the unique solution to this problem (6).

Proof of Theorem 1. *The objective function in (6) is finite by the assumptions. Any fixed point $b = \bar{\beta}(\tau)$ would solve the first-order condition, $E\left[X(F_Y(X'b|X) - \tau)\right] = 0$. By the law of iterated expectations, (3) implies the above first-order condition. Therefore, $\beta(\tau)$ solves (6). When the second-order condition holds, i.e., $E\left[\left(f_Y(X'b|X) + (F_Y(X'b|X) - \tau)f_Y'(X'b|X)/f_Y(X'b|X)\right)XX'\right]$ is positive definite at $b = \beta(\tau)$, $\beta(\tau)$ solves (6) uniquely.* □

Theorem 1 states that $\beta(\tau)$ is the unique fixed point to an iterated minimum distance approximation, with a weight of a function of X only. The mean-squared loss makes it clear how the linear function matches the CDF to the targeted probability of interest. The loss function puts more weight on points where the conditional density $f_Y(X'\beta(\tau)|X)$ is small. As a result, the distribution approximation error is smaller at points with smaller conditional density.

Now, I discuss the approximation nature of QR based on the distributional approximation error and quantile specification error. ACF interpret QR as the minimizer of the weighted mean-squared error loss function for quantile specification error, defined as the deviation between the approximation point $X'\beta(\tau)$ and the true CQF $Q_\tau(Y|X)$,[3]

$$\beta(\tau) = \arg\min_{\beta \in R^d} E\left[\bar{w}_\tau(X, \beta(\tau))\left(X'\beta - Q_\tau(Y|X)\right)^2\right], \text{ where} \tag{7}$$

$$\bar{w}_\tau(X, \beta(\tau)) = \frac{1}{2}\int_0^1 f_{\varepsilon_\tau}\left(u\left(X'\beta(\tau) - Q_\tau(Y|X)\right)\Big|X\right)du. \tag{8}$$

ACF define $\bar{w}_\tau(X, \beta(\tau))$ in (8) to be the importance weights that are the averages of the response variable over a line connecting the approximation point $X'\beta(\tau)$ and the true CQF. ACF note that

---

[3]   For estimation, [14] studies different approaches based on the distribution regression and quantile regression.

the regressors contribute disproportionately to the QR estimate and the primary determinant of the importance weight is the conditional density.

Moreover, the first-order condition implied by (7) $E[\bar{w}_\tau(X, \beta(\tau))X(X'\beta(\tau) - Q_\tau(Y|X))] = 0$ is a weighted orthogonal condition of the quantile specification error. A Taylor expansion provides intuition to connect the distribution approximation error and the quantile specification error: $(f_Y(X'\beta|X))^{-1}(F_Y(X'\beta|X) - \tau)^2 \approx f_Y(X'\beta|X)(Q_\tau(Y|X) - X'\beta)^2$ by $f_Y(X'\beta|X) = f_{\varepsilon_\tau}(X'\beta - Q_\tau(Y|X)|X)$. This observation implies the quantile specification error is smaller at points where the conditional density $f_Y(X'\beta|X)$ is larger. On the other hand, the distribution approximation error is larger at points with larger $f_Y(X'\beta|X)$. Comparing with the OLS, where the mean operator is linear, the CDF and its inverse operator, the CQF, are generally nonlinear. The distribution approximation error can be interpreted as the distance after a nonlinear transformation by the CDF, $F_Y(X'\beta(\tau)|X) - F_Y(Q_\tau(Y|X)|X)$. A Taylor expansion linearizes the distribution function to the quantile specification error multiplied by the conditional density function. The conditional density plays a crucial role on weighting the distribution approximation error and the quantile specification error. The above discussion provides additional insights to how the QR parameter approximates the CQF and fits the CDF to the targeted quantile level.

Remark 1 (Mean-squared loss under misspecification). *The linear function $X'\beta(\tau)$ is the best linear approximation under the check loss function in (1). While $\beta(.5)$ is the least absolute derivations estimation, the QR parameter $\beta(\tau)$ for $\tau \neq 0.5$ is the best linear predictor for a response variable under the asymmetric loss function $\rho_\tau(\cdot)$ in (1). ACF note that the prediction under the asymmetric check loss function is often not the object of interest in empirical work, with the exception of the forecasting literature, for example [15]. For the mean regression counterpart, OLS consistently estimates the linear conditional expectation and minimizes mean-squared error loss for fitting the conditional expectation under misspecification. The robust nature of OLS also motivates research on misspecification in panel data models. For example, Galvao and Kato [16] investigate linear panel data models under misspecification. The pseudo-true value of the fixed effect estimator provides the best partial linear approximation to the conditional mean given the explanatory variables and the unobservable individual effect.*[4]

3. The Semiparametric Efficiency Bounds

Section 3.1 presents the semiparametric efficiency bound for the unconditional moment restriction (3). Section 3.2 discusses the existing results on the semiparametric efficiency bound for the conditional moment restriction (4).

3.1. QR under Misspecification

I calculate the semiparametric efficiency bound for the unconditional moment restriction (3) by the approach of Severini and Tripathi [11].

Theorem 2. *Assume (R1)–(R4). The semiparametric efficiency bound for estimating the population QR parameter $\beta(\tau)$, defined in (1) or equivalently (3), is $J(\tau)^{-1}\Gamma(\tau, \tau)J(\tau)^{-1}$, where $J(\tau)$ is defined in (R3) and*

$$\Gamma(\tau_i, \tau_j) := E\left[\left(\tau_i - \mathbf{1}_{\{Y < X'\beta(\tau_i)\}}\right)\left(\tau_j - \mathbf{1}_{\{Y < X'\beta(\tau_j)\}}\right)XX'\right]$$

for any $\tau_i, \tau_j \in \mathcal{T} :=$ a closed subset of $[\epsilon, 1 - \epsilon]$ for $\epsilon > 0$.

In general, the semiparametrically-efficient joint asymptotic covariance of the estimators for $(\beta'(\tau_1), \beta'(\tau_2), ..., \beta'(\tau_m))'$ is $J(\tau_i)^{-1}\Gamma(\tau_i, \tau_j)J(\tau_j)^{-1}$, for any $\tau_i, \tau_j \in \mathcal{T}$, $i, j = 1, 2, ..., m$, for a finite integer $m \geq 1$.

---

[4]   Galvao and Kato [16] show that misspecification affects the bias correction and convergence rate of the estimator and provide a misspecification-robust inference procedure. In panel models under time series misspecification, Lee [17] proposes bias reduction methods for the incidental parameter.

Proof of Theorem 2. *See the Appendix.*   □

My proof accommodates the regularity assumptions for quantile regression and modifies Section 9 of [11]. For example, the covariate X can contain discrete components, by constructing two tangent spaces for the conditional density of Y given X and the marginal density of X, respectively. In the efficiency bound, $J(\tau) := E\left[f_Y(X'\beta(\tau)|X)XX'\right]$ is obtained by assuming the interchangeability of integration and differentiation for the nonsmooth check function.[5]

The method in [11] has been used in the monotone binary model in [18], Lewbel [19] latent variable model in [20] and the partial linear single index model in [21], for example. I work in the Hilbert space of tangent vectors of the square-root density functions and using the Riesz–Fréchet representation theorem. Another equivalent approach in [12] works in a Hilbert space of random variables and uses the projection on the linear space spanned by the scores from the one-dimensional subproblems to find the efficient influence function. The efficiency bound is then the second moment of the efficient influence function, $J(\tau)^{-1}X(\tau - \mathbf{1}_{\{Y \le X'\beta\}})$. Newey's efficient influence function is the score function evaluated at the unique representers by the Riesz–Fréchet theorem used in [11]; a more detailed comparison of these two approaches is discussed in [11].

ACF show that the QR process $\hat{\beta}(\cdot)$ is asymptotically mean-zero Gaussian with the covariance function $J(\tau_1)^{-1}\Gamma(\tau_1, \tau_2)J(\tau_2)^{-1}$ for any $\tau_1, \tau_2 \in \mathcal{T}$, which is the semiparametric efficiency bound in Theorem 2. This asymptotic covariance under misspecification for a single quantile, $J(\tau)^{-1}\Gamma(\tau, \tau)J(\tau)^{-1}$, has been presented in [2] and [3]. Hahn [3] further shows the QR estimator is well approximated by the bootstrap distribution, even when the linear quantile restriction is misspecified. An alternative estimator for the misspecification-robust asymptotic covariance matrix of $\hat{\beta}(\tau)$ is the nonparametric kernel method in ACF.

3.2. QR for Linear Specification

Assuming the linear QR model in (4) is correctly specified, *i.e.*, $Q_\tau(Y|X) = X'\beta(\tau)$ almost surely, the asymptotic covariance for the QR process $\hat{\beta}(\cdot)$ derived by ACF is simplified to $J(\tau_1)^{-1}\Gamma_0(\tau_1, \tau_2)J(\tau_2)^{-1}$, where $\Gamma_0(\tau_1, \tau_2) := (\min\{\tau_1, \tau_2\} - \tau_1\tau_2)E[XX']$ for any $\tau_1, \tau_2 \in (0, 1)$. The asymptotic covariance $J(\tau)^{-1}\Gamma_0(\tau, \tau)J(\tau)^{-1}$ for a single quantile τ, first derived by Powell [7], is widely used for inference in most empirical studies, which implicitly assume correct specification.

The semiparametric efficiency bound for the correctly-specified quantile regression (4) is $\tau(1-\tau)\left(E\left[XX'f_{\varepsilon_\tau}^2(0|X)\right]\right)^{-1}$, where $f_Y(X'\beta(\tau)|X) = f_{\varepsilon_\tau}(0|X)$ a.s. and $E\left[XX'f_{\varepsilon_\tau}^2(0|X)\right]$ is assumed to be finite and nonsingular. This is first calculated in [13] using the method developed in [12]. If, in addition, the conditional density function of ε_τ given X is independent of X, *i.e.*, $f_Y(Q_\tau(Y|X)|X) = f_{\varepsilon_\tau}(0|X) = f_{\varepsilon_\tau}(0)$, and $f_{\varepsilon_\tau}(0) > 0$, the semiparametric efficiency bound becomes $\tau(1-\tau)\left(E[XX']\right)^{-1}/f_{\varepsilon_\tau}^2(0)$. This asymptotic covariance is attained by $\hat{\beta}(\tau)$, first shown in [4]. This has an interesting resemblance to the fact that the OLS estimator is semiparametrically efficient in a homoskedastic regression model, *i.e.*, $e = Y - X'\beta$, $E[e|X] = 0$, and $E[e^2|X] = E[e^2]$.

I further show, in general, that the semiparametrically-efficient joint asymptotic covariance of the estimators for $(\beta'(\tau_1), ..., \beta'(\tau_m))'$ is

$$\left(\min\{\tau_i, \tau_j\} - \tau_i\tau_j\right)\left(E\left[XX'f_{\varepsilon_{\tau_i}}(0|X)f_{\varepsilon_{\tau_j}}(0|X)\right]\right)^{-1} \tag{9}$$

for any $\tau_i, \tau_j \in \mathcal{T}$, $i, j = 1, 2, ..., m$, for any finite integer $m \ge 1$. The regularity conditions imposed, (R1), (R2) and (R4), are weaker than the assumptions in [13]; for example, they assume $f(\varepsilon, X)$ is

---

[5]   Severini and Tripathi construct the tangent space for the continuous and bounded joint density $f(X, Y)$ in Section 9 of [11]. Additionally, they define J on the derivative of the moment restriction.

absolutely continuous in ε, which implies uniform continuity in (R2). See the Appendix for the detailed proof for (9).

4. Discussion and Conclusions

Misspecification is a generic phenomenon; especially in quantile regression (QR), the true conditional quantile function (CQF) might be nonlinear or different functions of the covariates at different quantiles. Table 1 summarizes the parallel properties of QR and OLS. Under misspecification, the pseudo-true OLS coefficient can be interpreted as the best linear predictor of the conditional mean function, $E[Y|X]$, in the sense that the coefficient minimizes the mean-squared error of the linear approximation to the conditional mean. The approximation properties of OLS have been well studied (see, for example, [22]). With respect to the QR counterpart, I present the inverse density-weighted mean-squared error loss function based on the distribution approximation error $F_Y(X'\beta|X) - \tau$. This result complements the interpretation based on the quantile specification error in [1]. My results imply that the Koenker–Bassett estimator is semiparametrically efficient for misspecified linear projection models and correctly specified linear quantile regression models when $f_Y(Q_\tau(Y|X)|X) = f_{\varepsilon_\tau}(0|X)$ does not depend on X. Alternatively, the smoothed empirical likelihood estimator using the unconditional moment restriction in [23] has the same asymptotic distribution as the Koenker-Bassett estimator and, hence, attains the efficiency bound.

Table 1. Summary properties of OLS and quantile regression (QR).

	OLS	QR
Linear Projection Model		
objective minimized *(interpretation)*	$E[(Y - X'\beta)^2]$ $E[(E[Y|X] - X'\beta)^2]$	$E[\rho_\tau(Y - X'\beta(\tau))]$ $E[\bar{w}_\tau(Q_\tau(Y|X) - X'\beta(\tau))^2]$ $E[f_Y(X'\beta(\tau)|X)^{-1}(F_Y(X'\beta(\tau)|X) - \tau)^2]$
unconditional moment *(interpretation)*	$E[X(Y - X'\beta)] = 0$ $E[X(E[Y|X] - X'\beta)] = 0$	$E[X(\mathbf{1}_{\{Y \le X'\beta(\tau)\}} - \tau)] = 0$ $E[X(F_Y(X'\beta(\tau)|X) - \tau)] = 0$ $E[\bar{w}_\tau X(X'\beta(\tau) - Q_\tau(Y|X))] = 0$
efficient estimators	$\arg\min_{\beta \in R^d} \frac{1}{n} \sum_{i=1}^{n}(Y_i - X_i'\beta)^2$ $= (\sum_{i=1}^{n} X_i X_i')^{-1}(\sum_{i=1}^{n} X_i Y_i)$ (OLS)	$\arg\min_{\beta \in R^d} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta)$ (Koenker–Bassett)
asymptotic covariance	$Q^{-1}\Omega Q^{-1}$ *	$J^{-1}\Gamma J^{-1}$
efficiency bounds	Chamberlain (1987) [9]	Theorem 2
Linear Regression Model		
conditional moment	$E[Y|X] = X'\beta$	$Q_\tau(Y|X) = X'\beta(\tau)$ or $F_Y(X'\beta(\tau)|X) = \tau$
efficiency bounds	Chamberlain (1987) [9] [†]	Newey and Powell (1990) [13]
homoscedasticity-type condition	$var[Y|X] = \sigma^2$	$f_{\varepsilon_\tau}(0|X) = f_{\varepsilon_\tau}(0)$
efficient estimators	OLS	Koenker–Bassett

* $Q = E[XX']$ and $\Omega = E[XX'e^2]$ where $e = Y - X'\beta$; [†] The feasible generalized least squares estimator is semiparametrically efficient, for example.

Under the linear quantile regression model, the Koenker–Bassett estimator consistently estimates the true $\beta(\tau)$, although it is not semiparametrically efficient given heteroskedasticity. Researchers have proposed many efficient estimators for the correctly-specified linear quantile regression parameter, for example the one-step score estimator in [13], the smoothed conditional empirical likelihood estimator in [24] and the sieve minimum distance (SMD) estimator in [25,26]. However, for all of these estimators, the pseudo-true values under misspecification are different, and their interpretations have not been thoroughly studied. Therefore, the semiparametric efficiency bounds

of these pseudo-true values are also different. For example, an unweighted SMD estimator converges to a pseudo-true value β_{SMD} that minimizes $E\left[(F_Y(X'\beta|X) - \tau)^2\right]$.[6] The first-order condition is $E\left[X\left(F_Y(X'\beta_{SMD}|X) - \tau\right)f_Y(X'\beta_{SMD}|X)\right] = 0$, which is the unconditional moment used in [13] for the semiparametrically efficient GMM estimator under correct specification. The conditional density weight is similar to the generalized least squares in the mean regression in that it uses a weight function of the conditional variance to construct an efficient estimator.

It is interesting to note that the pseudo-true value of the SMD estimator minimizes $E\left[(F_Y(X'\beta|X) - \tau)^2\right] \approx E\left[f_Y^2(Q_\tau(Y|X)|X)(X'\beta - Q_\tau(Y|X))^2\right]$. The distribution approximation error is weighted evenly over the support of X for β_{SMD}, in contrast to the QR parameter, which is weighted more at points with smaller conditional density in Theorem 1. Therefore, the SMD estimator might have more desirable and reasonable approximation properties than QR. Nevertheless, the SMD estimator is computationally more demanding than the Koenker–Bassett estimator. A numerical example in Figure 1 illustrates how the Koenker and Bassett (KB) and SMD estimators approximate the CQF and the CDF.
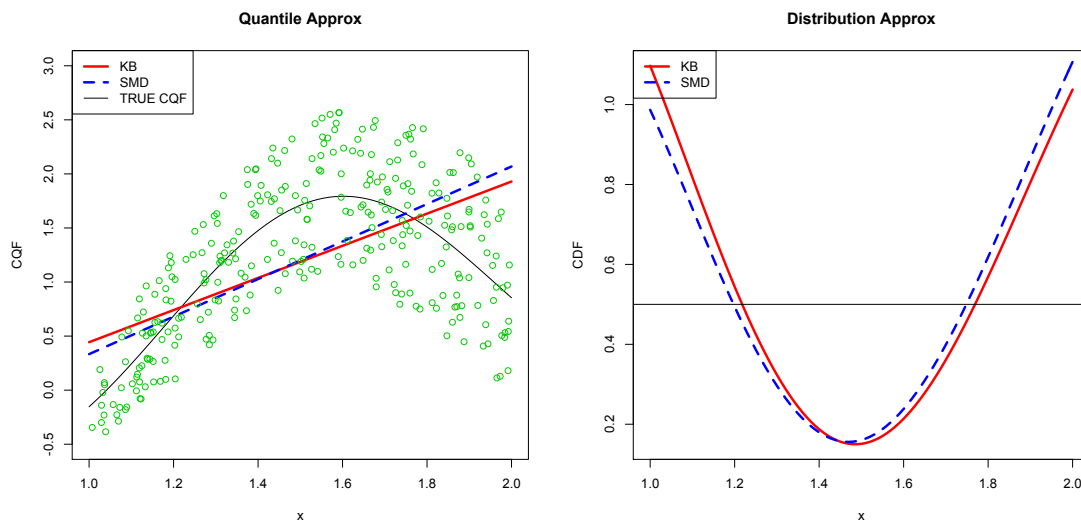

Figure 1. This numerical example is constructed by $X \sim Uniform[1,2]$, $e|X = x \sim Uniform[0,x]$ and $Y = \cos(2X) + e$. Therefore, $f_Y(y|X) = 1/X$, $F_Y(y|X) = (y - \cos(2X))/X$ and $Q_\tau(Y|X) = \tau X + \cos(4X)$. Set $\tau = 0.5$ for the median. The red solid line is for the QR parameter β_{KB} defined in (3) and estimated by the Koenker-Bassett (KB) estimator. The blue dashed line is the approximation by the SMD estimator β_{SMD} minimizing $E[(F_Y(X'\beta|X) - \tau)^2]$. The approximations are $X'\beta_{KB} = -0.324 + 0.161X$ and $X'\beta_{SMD} = -0.204 + 0.078X$. The left panel shows the linear approximations $X'\beta_{KB}$, $X'\beta_{SMD}$ and the true CQF. The green circles are 300 random draws from the DGP. The right panel shows the corresponding CDFs $F_Y(X'\beta_{KB}|X)$ and $F_Y(X'\beta_{SMD}|X)$. For smaller x where the conditional density is larger, the quantile specification error of SMD is smaller than that of KB in the left panel. For the distribution approximation error in the right panel, SMD weights more evenly over the support of X, while KB has smaller distribution approximation error at larger x with smaller density.

This discussion leads to open-ended questions: What is an appropriate linear approximation or a meaningful summary statistic for the nonlinear CQF? How should economists measure the

---

6　The conditional moment restriction in (4) can be expressed as $m(X, \beta) = \tau - F_Y(X'\beta|X) = 0$. In [26], an unweighted penalized sieve minimum distance estimator minimizes a possibly penalized consistent estimate of the minimum distance criterion, $E[m(X, \beta)^2]$.

marginal effect of the covariates on the CQF? An approach that circumvents this problem is measuring the average marginal response of the covariates on the CQF directly. The average quantile derivative, defined as $E[W(X)\nabla Q_\tau(Y|X)]$ where $W(X)$ is a weight function, offers such a succinct summary statistic ([27]). Sasaki [28] investigates the question that quantile regressions may misspecify true structural functions. He provides a causal interpretation of the derivative of the CQF, which identifies a weighted average of heterogeneous structural partial effects among the subpopulation of individuals at the conditional quantile of interest. Sasaki's work adds economic content to this misspecified question. This paper complements the prior literature on understanding how the QR statistically summarizes the outcome distribution.

Conflicts of Interest: The author declares no conflict of interest.

Appendix

Proof of Theorem 3. *I implement the main results in [11]. I start with the definitions and construct the Hilbert space. The unknown probability density or mass function of the random vector $(Y, X')' \in \Omega = S_Y \times S_X$ with respect to the measure P, the products of the Lebesgue measure μ_Y and μ_X,[7] is written as $f(Y, X) = f(Y|X)f(X) := \psi_0^2(Y|X)\phi_0^2(X)$. The functionals ψ_0 and ϕ_0 belong to the following sets defined by the regularity conditions*

$$\Psi_Y := \left\{ \psi \in S_Y \times S_X \to R, \psi^2(Y|X) > 0, bounded\ and\ uniformly\ continuous\ in\ y, \right.$$
$$\left. uniformly\ in\ x\ over\ the\ support\ of\ X, \int_{S_Y} \psi^2(y|X)dy = 1 \right\}\ and$$

$$\Phi := \left\{ \phi \in L^2(S_X; \mu_X), \phi^2(X) > 0, \int_{S_X} \phi^2(x)\mu_X(dx) = 1, \right.$$
$$\left. \int_{S_X} \|x\|^{2+\epsilon}\phi^2(x)\mu_X(dx) < \infty, for\ some\ \epsilon > 0 \right\}.$$

Let $\mathcal{A} := \Psi_Y \times \Phi$.

Definition 1. *A vector $\dot{\xi} = (\psi, \phi)$ is said to be tangent to $\mathcal{A}$ at ξ_0 if it is the slope of $\xi_t := (\psi_t, \phi_t)$ at $t = 0$, i.e., $lim_{t\to 0}\|t^{-1}(\xi_t - \xi_0) - \dot{\xi}\| = 0$.*

Definition 2. *The tangent space to $\mathcal{A}$ at the true value ξ_0, denoted as $\overline{lin\ T(\mathcal{A}, \xi_0)}$ is the smallest linear space, which is closed under the L^2-norm and contains all $\dot{\xi} \in L^2(\Omega; \mu_Y \times \mu_X)$ tangent to $\mathcal{A}$ at ξ_0.*

Severini and Tripahi [11] show that the tangent space $\overline{lin\ T(\mathcal{A}, \xi_0)}$ is the product of $\overline{lin\ T(\Psi_Y, \psi_0)}$ and $\overline{lin\ T(\Phi, \phi_0)}$, where

$$\overline{lin\ T(\Psi_Y, \psi_0)} := \left\{ \dot{\psi} \in L^2(\Omega; \mu_Y \times X), \int_{S_Y} \dot{\psi}(y|X)\psi_0(y|X)\mu_Y(dy) = 0\ with\ probability\ 1(w.p.1) \right\}$$

$$\overline{lin\ T(\Phi, \phi_0)} := \left\{ \dot{\phi} \in L^2(S_X; \mu_X), \int_{S_X} \dot{\phi}(x)\phi_0(x)\mu_X(dx) = 0 \right\}.$$

The pseudo-true model is the unconditional moment restriction $E_0\big[(\tau - \mathbf{1}_{\{Y \leq X'\beta_0\}})X\big] = 0$ in (3). Here, E_0 is the expectation with respect to the true density functions $\xi_0 = (\psi_0, \phi_0)$, and β_0 denotes the pseudo-true $\beta(\tau)$ for notational simplicity. The objective is to estimate the efficiency bound for estimating β_0. Equivalently, I can instead look at the efficiency bound for estimating the functional $\eta(\psi_0, \phi_0) := c'\beta_0$ for any arbitrary vector $c \in R^d$. Severini and Tripahi [11] parameterize

---

7   μ_X may not be a Lebesgue measure, since I allow discrete components in the covariates X.

$\xi_0 = (\psi_0, \phi_0) \in \mathcal{A}$ and β_0 as a one-dimensional subproblem. For some $t_0 > 0$, let $t \mapsto (\xi_t, \beta_t)$ be a curve from $[0, t_0]$ into $\mathcal{A} \times R^d$, which passes through (ξ_0, β_0) at $t = 0$. That is, estimating $\eta(\xi_t) = c'\beta_t = t$ at the true parameter $t = 0$ is equivalent to estimating $t = 0$. The likelihood of estimating t using a single observation $(Y, X')'$ is given by $\psi_t^2(Y|X)\phi_t^2(X)$. Therefore, the score function for estimating $t = 0$ is

$$S_0(Y, X) := \frac{d}{dt} \log \left[\psi_t^2(Y|X)\phi_t^2(X) \right]\big|_{t=0} = 2\frac{\dot{\psi}(Y|X)}{\psi_0(Y|X)} + 2\frac{\dot{\phi}(X)}{\phi_0(X)}.$$

Then, the Fisher information at $t = 0$ can be written as

$$i_F = E[S_0(Y, X)S_0'(Y, X)] = \int_{S_X} \int_{S_Y} S_0(y, x)S_0'(y, x)\psi_0^2(y|x)\phi_0^2(x)\mu_Y(dy)\mu_X(dx)$$

$$= 4E_X \left[\int_{S_Y} \dot{\psi}(y|X)\dot{\psi}'(y|X)\mu_Y(dy) \right] + 4\int_{S_X} \dot{\phi}(x)\dot{\phi}'(x)\mu_X(dx)$$

$$:= < (\dot{\psi}, \dot{\phi}), (\dot{\psi}, \dot{\phi}) >_F,$$

where the third equality is because $\dot{\xi}_0 = (\dot{\psi}_0, \dot{\phi}_0) \in \overline{lin\ T(\mathcal{A}, \xi_0)}$, and E_X denotes integrals with respect to the distribution of X. Therefore, the Fisher information inner product $< \cdot, \cdot >_F$ and the corresponding norm $\| \cdot \|_F$ are defined as

$$< \dot{\xi}_1, \dot{\xi}_2 >_F := 4E_X \left[\int_{S_Y} \dot{\psi}_1(y|X)\dot{\psi}_2'(y|X)\mu_Y(dy) \right] + 4\int_{S_X} \dot{\phi}_1(x)\dot{\phi}_2'(x)\mu_X(dx) \text{ and}$$

$$\|\dot{\xi}_1\|_F^2 = \|(\dot{\psi}_1, \dot{\phi}_1)\|_F^2 := < (\dot{\psi}_1, \dot{\phi}_1), (\dot{\psi}_1, \dot{\phi}_1) >_F$$

for any $\dot{\xi}_1, \dot{\xi}_2 \in \overline{lin\ T(\mathcal{A}, \xi_0)}$, which is a closed subset of $L^2(\Omega; P)$. Hence, I have constructed the Hilbert space $(\overline{lin\ T(\mathcal{A}, \xi_0)}, < \cdot, \cdot >_F)$.

Now, I am ready to derive the efficiency bounds. It is known that the information inequality holds for all regular estimators, *i.e.*, the asymptotic covariance of the estimator $\geq 1/i_F = \|\dot{\xi}_0\|_F^{-2}$. The semiparametric bound can be interpreted as the supremum of the asymptotic covariance over the parametric submodels. By [11], the lower bound is

$$l.b. = sup_{\{\dot{\xi} \in (\overline{lin\ T(\mathcal{A}, \xi_0)}): \dot{\xi} \neq 0, \nabla\eta(\dot{\xi})=1\}} \|\dot{\xi}\|_F^{-2} = sup_{\{\dot{\xi} \in (\overline{lin\ T(\mathcal{A}, \xi_0)}): \|\dot{\xi}\|_F=1\}} |\nabla\eta(\dot{\xi})|^2$$

$$= \|\nabla\eta\|_*^2 = \|\xi^*\|_F^2. \tag{10}$$

The third equality is the norm of the linear functional $\nabla\eta$, the path-wise derivative of η (p. 105 in [29]). The forth equality is from the Riesz–Fréchet theorem: there exists a unique $\xi^* \in \overline{lin\ T(\mathcal{A}, \xi_0)}$ for the continuous linear functional $\nabla\eta$ on the Hilbert space $(\overline{lin\ T(\mathcal{A}, \xi_0)}, < \cdot, \cdot >_F)$, such that $\nabla\eta(\dot{\xi}) \leq \xi^*, \dot{\xi} >_F$ for all $\dot{\xi} \in \overline{lin\ T(\mathcal{A}, \xi_0)}$, *i.e.*,

$$\nabla\eta(\dot{\psi}, \dot{\phi}) = c'\dot{\beta} = < (\psi^*, \phi^*), (\dot{\psi}, \dot{\phi}) >_F$$

$$= 4E_X \left[\int_{S_Y} \psi^*\dot{\psi}'\mu_Y(dy) \right] + 4\int_{S_X} \phi^*\dot{\phi}'\mu_X(dx). \tag{11}$$

Therefore, to find the lower bound by (10), I need to find ξ^*, which is known as the representer of the continuous linear functionals $\nabla\eta$.

The submodel $(\psi_t, \phi_t, \beta_t)$ is also required to satisfy the unconditional moment restriction, $E_t[(\mathbf{1}_{\{Y \leq X'\beta_t\}} - \tau)X] = 0$. For any $\tau_1, \tau_2 \in (0, 1)$, $\beta_t := (\beta_t'(\tau_1), \beta_t'(\tau_2))' := (\beta_{1t}', \beta_{2t}')'$. I simultaneously estimate $\beta_0 = (\beta_{10}', \beta_{20}')'$, a $2d$-dimensional vector, so the unconditional moment restriction is

$$\int_{S_X} \int_{S_Y} \left(\begin{array}{c} (\mathbf{1}_{\{y \leq x'\beta_{1t}\}} - \tau_1)x \\ (\mathbf{1}_{\{y \leq x'\beta_{2t}\}} - \tau_2)x \end{array} \right) \psi_t^2(y|x)\phi_t^2(x)\mu_Y(dy)\mu_X(dx) = 0.$$

Taking the derivative with respect to t evaluated at $t = 0$,[8]

$$0 = \int_{S_X} \begin{pmatrix} xx' f_Y(x'\beta_{10}|x)\dot{\beta}_1 \\ xx' f_Y(x'\beta_{20}|x)\dot{\beta}_2 \end{pmatrix} \phi_0^2(x)\mu_X(dx)$$

$$+ 2\int_{S_X}\int_{S_Y} \begin{pmatrix} x\mathbf{1}_{\{y \le x'\beta_{10}\}} \\ x\mathbf{1}_{\{y \le x'\beta_{20}\}} \end{pmatrix} \psi_0(y|x)\dot{\psi}'(y|x)\mu_Y(dy)\phi_0^2(x)\mu_X(dx)$$

$$+ 2\int_{S_X} \begin{pmatrix} x(F_Y(x'\beta_{10}|x) - \tau_1) \\ x(F_Y(x'\beta_{20}|x) - \tau_2) \end{pmatrix} \phi_0(x)\dot{\phi}'(x)\mu_X(dx),$$

where the second term uses the fact that if $\dot{\psi} \in \overline{lin\, T(\Psi_Y, \psi_0)}$, then $\int_{S_Y} \psi_0\dot{\psi}\mu_Y(dy) = 0$. Note that $\int_{S_X} xx' f_Y(x'\beta_0|x)\phi_0^2(x)\mu_X(dx) = E_0[XX' f_Y(X'\beta_0|X)] = J(\tau)$, which is assumed to be positive definite by (R3), so $J(\tau)^{-1}$ exists. Define

$$D := \begin{pmatrix} J(\tau_1) & 0 \\ 0 & J(\tau_2) \end{pmatrix},$$

so D^{-1} exists. Then

$$\begin{pmatrix} \dot{\beta}_1 \\ \dot{\beta}_2 \end{pmatrix} = -2D^{-1}\Big[\int_{S_X}\int_{S_Y} \begin{pmatrix} x\mathbf{1}_{\{y \le x'\beta_{10}\}} \\ x\mathbf{1}_{\{y \le x'\beta_{20}\}} \end{pmatrix} \psi_0(y|x)\dot{\psi}'(y|x)\mu_Y(dy)\phi_0^2(x)\mu_X(dx)$$

$$+ \int_{S_X} \begin{pmatrix} x(F_Y(x'\beta_{10}|x) - \tau_1) \\ x(F_Y(x'\beta_{20}|x) - \tau_2) \end{pmatrix} \phi_0(x)\dot{\phi}'(x)\mu_X(dx)\Big]. \tag{12}$$

I confirm that $\nabla\eta(\dot{\xi}) = c'\dot{\beta}$ is a continuous linear functional on $\overline{lin\, T(\mathcal{A}, \xi_0)}$, so η is indeed path-wise differentiable. Using (12) to replace $c'\dot{\beta}$ in the left-hand side of (11), I can find the representer for $\nabla\eta$ as

$$\phi^*(x) = -\frac{1}{2}c'D^{-1}\begin{pmatrix} (F_Y(x'\beta_{10}|x) - \tau_1)x \\ (F_Y(x'\beta_{20}|x) - \tau_2)x \end{pmatrix}\phi_0(x) \text{ and}$$

$$\psi^*(y|x) = -\frac{1}{2}c'D^{-1}\begin{pmatrix} (\mathbf{1}_{\{y \le x'\beta_{10}\}} - F_Y(x'\beta_{10}|x))x \\ (\mathbf{1}_{\{y \le x'\beta_{20}\}} - F_Y(x'\beta_{20}|x))x \end{pmatrix}\psi_0(y|x)$$

because $\dot{\psi} \in \overline{lin\, T(\Psi_Y, \psi_0)}$. It can be easily checked that $(\psi^*, \phi^*) \in \overline{lin\, T(\mathcal{A}, \xi_0)}$. For notational ease, denote $\mathbf{1}_i := \mathbf{1}_{\{y \le x'\beta_{i0}\}}$ and $F_i := F_Y(x'\beta_{i0}|x)$, $i = 1, 2$. Then, the lower bound for regular $\sqrt{n}$-consistent estimators of $c'\beta_0$ is

$$\|(\psi^*, \phi^*)\|_F^2$$

$$= c'D^{-1}\Big\{\int_{S_X} \begin{pmatrix} (F_1 - \tau_1)^2 xx' & (F_1 - \tau_1)(F_2 - \tau_2)xx' \\ (F_1 - \tau_1)(F_2 - \tau_2)xx' & (F_2 - \tau_2)^2 xx' \end{pmatrix}\phi_0^2(x)\mathbf{1}_{\{\phi_0(x) > 0\}}\mu_X(dx)$$

$$+ E\begin{pmatrix} E\left[(\mathbf{1}_1 - F_1)^2|X\right]XX' & E\left[(\mathbf{1}_1 - F_1)(\mathbf{1}_2 - F_2)|X\right]XX' \\ E\left[(\mathbf{1}_1 - F_1)(\mathbf{1}_2 - F_2)|X\right]XX' & E\left[(\mathbf{1}_2 - F_2)^2|X\right]XX' \end{pmatrix}\Big\}D^{-1}c$$

$$= cD^{-1}\begin{pmatrix} \Gamma(\tau_1, \tau_1) & \Gamma(\tau_1, \tau_2) \\ \Gamma(\tau_1, \tau_2) & \Gamma(\tau_2, \tau_2) \end{pmatrix}D^{-1}c,$$

---

[8] The interchange of differentiation and integration is allowed, assumed throughout [11], by the smoothness of $\xi_t(Y, X)$ in $t \in [0, t_0]$ by the construction of regular parametric submodels; see [12] for details.

where

$$\Gamma(\tau_1, \tau_2) := E\Big[E\big[(F_1 - \tau_1)(F_2 - \tau_2) + (\mathbf{1}_1 - F_1)(\mathbf{1}_2 - F_2)|X\big]XX'\Big]$$

$$= E\Big[E\big[\tau_1\tau_2 - \tau_1 F_2 - \tau_2 F_1 + \mathbf{1}_1\mathbf{1}_2|X\big]XX'\Big]$$

$$= E\Big[E\big[(\tau_1 - \mathbf{1}_{\{y \leq X'\beta_{10}\}})(\tau_2 - \mathbf{1}_{\{y \leq X'\beta_{20}\}})|X\big]XX'\Big]$$

$$= E\Big[(\tau_1 - \mathbf{1}_{\{y \leq X'\beta_{10}\}})(\tau_2 - \mathbf{1}_{\{y \leq X'\beta_{20}\}})XX'\Big]$$

by the law of iterated expectations, and so, $\Gamma(\tau, \tau) = E\Big[(\tau - \mathbf{1}_{\{y \leq X'\beta_0\}})^2 XX'\Big]$.

Therefore, the lower bound for estimating $\beta(\tau)$ is $J(\tau)^{-1}\Gamma(\tau, \tau)J(\tau)^{-1}$. The asymptotic covariance of the estimators for $\beta(\tau_1)$ and $\beta(\tau_2)$ cannot be smaller than $J(\tau_1)^{-1}\Gamma(\tau_1, \tau_2)J(\tau_2)^{-1}$.

$\square$

Remark 2. *Consider the efficiency bound for estimating one single quantile $\beta(\tau)$ using Newey's approach in [12]. Severini and Triphathi [11] claim that the efficient influence function for $c'\beta(\tau)$ is $2\psi^*/\psi_0 + 2\phi^*/\phi_0 = c'J(\tau)^{-1}X(\tau - F_Y(X'\beta|X)) + c'J(\tau)^{-1}X(F_Y(X'\beta|X) - \mathbf{1}_{\{Y \leq X'\beta\}}) = c'J(\tau)^{-1}X(\tau - \mathbf{1}_{\{Y \leq X'\beta\}})$. Then, the efficient influence function for β_0 is $(E[SS'])^{-1}S$, where the efficient score $S = J(\tau)\Gamma(\tau, \tau)^{-1}X(\tau - \mathbf{1}_{\{Y \leq X'\beta\}})$. Newey shows that the semiparametric bound is $(E[SS'])^{-1}$.*

Proof of the Semiparametric Efficiency Bound for the Linear QR Model (4). Under the linear specification, $F_Y(X'\beta_0|X) = \tau$. The random vectors (Y, X) satisfy the conditional moment restriction $E\big[\mathbf{1}_{\{Y \leq X'\beta_0\}} - \tau|X\big] = 0$, i.e., $\int_{S_Y}(\mathbf{1}_{\{y \leq X'\beta_0\}} - \tau)\psi_0^2(y|X)\mu_Y(dy) = 0$, where the joint distribution of (Y, X) is $\psi_0^2(Y|X)\phi_0^2(X)$. The Hilbert space $(\overline{lin\ T(\mathcal{A}, \zeta_0)}, <\cdot, \cdot>_F)$ and $\mathcal{A} = (\Psi_Y, \Phi)$ are defined in the proof of Theorem 2. Consider any $\tau_1.\tau_2 \in (0,1)$, $\beta_t := (\beta_t'(\tau_1), \beta_t'(\tau_2))' := (\beta_{1t}', \beta_{2t}')'$. The parameterized submodel $(\psi_t, \phi_t, \beta_t)$ also has to satisfy the moment condition

$$\int_{S_Y}\begin{pmatrix}(\mathbf{1}_{\{y \leq X'\beta_{1t}\}} - \tau_1) \\ (\mathbf{1}_{\{y \leq X'\beta_{2t}\}} - \tau_2)\end{pmatrix}\psi_t^2(y|X)\mu_Y(dy) = 0.$$

Taking the derivative with respect to t evaluated at $t = 0$, I have

$$\begin{pmatrix}f_Y(X'\beta_1|X)X'\dot{\beta}_1 \\ f_Y(X'\beta_2|X)X'\dot{\beta}_2\end{pmatrix} + 2\int_{S_Y}\begin{pmatrix}(\mathbf{1}_{\{y \leq X'\beta_{10}\}} - \tau_1) \\ (\mathbf{1}_{\{y \leq X'\beta_{20}\}} - \tau_2)\end{pmatrix}\psi_0(y|X)\dot{\psi}'(y|X)dy = 0, \tag{13}$$

where $f_Y(y|X) = \psi_0^2(y|X)$. Define

$$D(X) := \begin{pmatrix}f_Y(X'\beta_1|X)X' & 0 \\ 0 & f_Y(X'\beta_2|X)X'\end{pmatrix}.$$

Note that (13) has an over-identifying moment restriction that cannot uniquely solve $\dot{\beta}$. To locally identify $\dot{\beta}$, [11] give the sufficient condition by $W(X)$, which is some nonsingular (w.p.1) 2×2 matrix, such that $E[D(X)'W(X)D(X)]$ is nonsingular. By assumption, $E[XX'f_Y^2(X'\beta(\tau)|X)] = E[XX'f_{\varepsilon_\tau}^2(0|X)]$ exists and is nonsingular, so the same holds for $E[D'(X)D(X)]$. Hence, I can choose $W(X) = \mathbf{1}$, the identity matrix. Multiplying (13) by $D'(X)$,

$$D'(X)D(X)\begin{pmatrix}\dot{\beta}_1 \\ \dot{\beta}_2\end{pmatrix} + D'(X)2\int_{S_Y}\begin{pmatrix}(\mathbf{1}_{\{y \leq X'\beta_{10}\}} - \tau_1) \\ (\mathbf{1}_{\{y \leq X'\beta_{20}\}} - \tau_2)\end{pmatrix}\psi_0(y|X)\dot{\psi}'(y|X)dy = 0.$$

Taking expectations on both sides with respect to X and solve for $\dot{\beta}$,

$$\begin{pmatrix} \dot{\beta}_1 \\ \dot{\beta}_2 \end{pmatrix} = -2\left(E[D'(X)D(X)]\right)^{-1}E\left[D'(X)2\int_{S_Y}\begin{pmatrix} (\mathbf{1}_{\{y\leq X'\beta_{10}\}} - \tau_1) \\ (\mathbf{1}_{\{y\leq X'\beta_{20}\}} - \tau_2) \end{pmatrix}\psi_0(y|X)\psi'(y|X)dy\right].$$

Then, for any arbitrary $c \in R^{2d}$, the representer for $\nabla\eta((\dot{\psi},\dot{\phi})) = c'\dot{\beta}$ is

$$\psi^*(y|X) = -\frac{1}{2}c'\left(E[D'(X)D(X)]\right)^{-1}D'(X)\begin{pmatrix} (\mathbf{1}_{\{y\leq X'\beta_{10}\}} - \tau_1) \\ (\mathbf{1}_{\{y\leq X'\beta_{20}\}} - \tau_2) \end{pmatrix}\psi_0(y|X)$$
$$\in \overline{lin\ T(\Psi_Y,\psi_0)}$$

by the conditional moment restriction. Additionally, $\phi^* = 0$, since ϕ_0 is just ancillary in this conditional moment case. Define $A := \left(E[D'(X)D(X)]\right)^{-1}$ and $\mathbf{1}_i := \mathbf{1}_{\{Y\leq X'\beta_{i0}\}}$ for $i = 1,2$ for notational ease. Without loss of generality, assume $\tau_1 < \tau_2$. Therefore, the lower bound is

$$\|(\psi^*,\phi^*)\|_F^2$$
$$= c'E\left[AD'(X)\begin{pmatrix} E\left[(\mathbf{1}_1 - \tau_1)^2\big|X\right] & E\left[(\mathbf{1}_1 - \tau_1)(\mathbf{1}_2 - \tau_2)\big|X\right] \\ E\left[(\mathbf{1}_1 - \tau_1)(\mathbf{1}_2 - \tau_2)\big|X\right] & E\left[(\mathbf{1}_2 - \tau_2)^2\big|X\right] \end{pmatrix}D(X)A\right]c$$
$$= c'AE\left[D'(X)\begin{pmatrix} \tau_1(1-\tau_1) & \tau_1(1-\tau_2) \\ \tau_1(1-\tau_2) & \tau_2(1-\tau_2) \end{pmatrix}D(X)\right]Ac$$
$$= c'\begin{pmatrix} \tau_1(1-\tau_1)\left\{E\left[XX'f_{\varepsilon_{\tau_1}}^2(0|X)\right]\right\}^{-1} & \tau_1(1-\tau_2)\left\{E\left[XX'f_{\varepsilon_{\tau_1}}(0|X)f_{\varepsilon_{\tau_2}}(0|X)\right]\right\}^{-1} \\ \tau_1(1-\tau_2)\left\{E\left[XX'f_{\varepsilon_{\tau_1}}(0|X)f_{\varepsilon_{\tau_2}}(0|X)\right]\right\}^{-1} & \tau_2(1-\tau_2)\left\{E\left[XX'f_{\varepsilon_{\tau_2}}^2(0|X)\right]\right\}^{-1} \end{pmatrix}c,$$

since $f_Y(X'\beta|X) = f_{\varepsilon_\tau}(0|X)$ for correct specification.   □

Remark 3. *Consider the efficiency bound for estimating one single quantile $\beta(\tau)$ by Newey's approach in [12]. Severini and Triphathi [11] claim that Newey's efficient influence function for $c'\beta(\tau)$ is $2\psi^*/\psi_0 = c'\left(E[f_{\varepsilon_\tau}^2(0|X)XX']\right)^{-1}f_{\varepsilon_\tau}(0|X)X(\tau - \mathbf{1}_{\{Y\leq X'\beta\}})$. Then, the efficient influence function for β_0 is $(E[SS'])^{-1}S$, where the efficient score $S = (\tau - \tau^2)^{-1}f_{\varepsilon_\tau}(0|X)X(\tau - \mathbf{1}_{\{Y\leq X'\beta\}})$. Newey shows the semiparametric bound is $(E[SS'])^{-1}$.*

References

1. Angrist, J.; Chernozhukov, V.; Fernández-Val, I. Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* **2006**, *74*, 539–563.
2. Kim, T.-H.; White, H. Estimation, inference, and specification testing for possibly misspecified quantile regression. *Adv. Econom.* **2003**, *17*, 107–132.
3. Hahn, J. Bayesian bootstrap of the quantile regression estimator: A large sample study. *Int. Econ. Rev.* **1997**, *38*, 795–808.
4. Koenker, R.; Bassett, G. Regression quantile. *Econometrica* **1978**, *46*, 33–50.
5. Koenker, R. *Quantile Regression. Econometric Society Monographs*; Cambridge University Press: New York, NY, USA, 2005.
6. Powell, J. Least absolute deviations estimation for the censored regression model. *J. Econom.* **1984**, *25*, 303–325.
7. Powell, J. Censored regression quantiles. *J. Econom.* **1986**, *32*, 143–155.
8. Chernozhukov, V.; Fernández-Val, I.; Galichon, A. Quantile and probability curves without crossing. *Econometrica* **2010**, *78*, 1093–1125.
9. Chamberlain, G. Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econom.* **1987**, *34*, 305–334.
10. Ai, C.; Chen, X. The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *J. Econom.* **2012**, *170*, 442–457.

11. Severini, T.A.; Tripathi, G. A simplified approach to computing efficiency bounds in semiparametric models. *J. Econom.* **2001**, *102*, 23–66.
12. Newey, W.K. Semiparametric efficiency bounds. *J. Appl. Econom.* **1990**, *5*, 99–135.
13. Newey, W.K.; Powell , J.L. Efficient estimation of linear and type 1 censored regression models under conditional quantile restrictions. *Econom. Theory* **1990**, *6*, 295–317.
14. Koenker, R.; Leorato, S.; Peracchi, F. *Distributional vs. Quantile Regression*; Ceis Working Paper No. 300. 2013. Available online:http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2368737 (accessed on 20 November 2015).
15. Komunjer, I. Quasi-maximum likelihood estimation for conditional quantiles. *J. Econom.* **2005**, *128*, 137–164.
16. Galvao, A.F.; Kato, K. Estimation and inference for linear panel data models under misspecification when both n and t are large. *J. Bus. Econ. Stat.* **2014**, *32*, 285–309.
17. Lee, Y. Bias in dynamic panel models under time series misspecification. *J. Econom.* **2012**, *169*, 54–60.
18. Magnac, T.; Maurin, E. Identification and information in monotone binary models. *J. Econom.* **2007**, *139*, 76–104.
19. Lewbel, A. Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica* **1998**, *66*, 105–122.
20. Jacho-Chávez, D.T. Efficiency bounds for semiparametric estimation of inverse conditional-density-weighted functions. *Econom. Theory* **2009**, *25*, 847–855.
21. Chen, T.; Parker, T. Semiparametric efficiency for partially linear single-index regression models. *J. Multivar. Anal.* **2014**, *130*, 376–386.
22. White, H. Using least squares to approximate unknown regression functions. *Int. Econ. Rev.* **1980**, *21*, 149–170.
23. Whang, Y.-J. Smoothed empirical likelihood methods for quantile regression models. *Econom. Theory* **2006**, *22*, 173–205.
24. Otsu, T. Conditional empirical likelihood estimation and inference for quantile regression models. *J. Econom.* **2008**, *142*, 508–538.
25. Chen, X.; Pouzo, D. Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *J. Econom.* **2009**, *152*, 46–60.
26. Chen, X.; Pouzo, D. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica* **2012**, *80*, 277–321.
27. Chaudhuri, P.; Doksum, K.; Samarov, A. On average derivative quantile regression. *Ann. Stat.* **1997**, *25*, 715–744.
28. Sasaki, Y. What do quantile regressions identify for general structural functions? *Econom. Theory* **2015**, *31*, 1102–1116.
29. Luenberger, D.G., *Optimization by Vector Space Methods.* Wiley: New York, NY, USA, 2005.