Proofs of Theorems and Corollaries for "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure"

Joshua Angrist, Victor Chernozhukov, and Iván Fernández-Val

This supplement to the paper "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure" provides added technical details related to the proofs of the theorems and corollaries not included in the main text. In particular, it contains the proofs for Theorem 3 on the uniform consistency and asymptotic Gaussianity of the sample QR process and the proofs for the Corollaries of this Theorem, along with the proofs of uniform consistency for the estimators of the components of the covariance kernel.

1 Proof of Theorems 3 and its Corollaries

The proof has two steps.[1] The first step establishes uniform consistency of the sample QR process. The second step establishes asymptotic Gaussianity of the sample QR process.[2] For $W = (Y, X)$, let $\mathbb{E}_n[f(W)]$ denote $n^{-1}\sum_{i=1}^n f(W_i)$ and $\mathbb{G}_n[f(W)]$ denote $n^{-1/2}\sum_{i=1}^n (f(W_i) - E[f(W_i)])$. If $\widehat{f}$ is an estimated function, $\mathbb{G}_n[\widehat{f}(W)]$ denotes $n^{-1/2}\sum_{i=1}^n (f(W_i) - E[f(W_i)])_{f=\widehat{f}}$. For a matrix A, mineig $[A]$ denotes the minimum eigenvalue of A.

1.1 Uniform consistency of $\hat{\beta}(\cdot)$

For each τ in $\mathcal{T}$, $\hat{\beta}(\tau)$ minimizes $Q_n(\tau, \beta) := \mathbb{E}_n[\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))]$. Define $Q_\infty(\tau, \beta) := E[\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))]$. It is easy to show that $E\|X\| < \infty$ implies that $E|\rho_\tau(Y - X'\beta) - \rho_\tau(Y - X'\beta(\tau))| < \infty$. Therefore, $Q_\infty(\tau, \beta)$ is finite, and by the stated assumptions, 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 $(\tau, \beta) \in \mathcal{T} \times \mathcal{B}$. This statement holds pointwise by the Khinchine law of large numbers. The empirical process $(\tau, \beta) \mapsto Q_n(\tau, \beta)$ is stochastically equicontinuous since $|Q_n(\tau', \beta') - Q_n(\tau'', \beta'')| \leq C_{1n} \cdot |\tau' - \tau''| + C_{2n} \cdot \|\beta' - \beta''\|$, where $C_{1n} = 2 \cdot \mathbb{E}_n\|X\| \cdot \sup_{\beta \in \mathcal{B}} \|\beta\| = O_p(1)$ and $C_{2n} = 2 \cdot \mathbb{E}_n\|X\| = O_p(1)$. Hence, the convergence also holds uniformly.

Next, we show uniform consistency. Consider a collection of closed balls $B_M(\beta(\tau))$ of radius M and center $\beta(\tau)$, and let $\beta_M(\tau) = \beta(\tau) + \delta_M(\tau) \cdot v(\tau)$, where $v(\tau)$ is a direction vector with unity norm $\|v(\tau)\| = 1$ and $\delta_M(\tau)$ is a positive scalar such that $\delta_M(\tau) \geq M$. Then uniformly in $\tau \in \mathcal{T}$, $(M/\delta_M(\tau)) \cdot (Q_n(\tau, \beta_M(\tau)) - Q_n(\tau, \beta(\tau))) \overset{(a)}{\geq} Q_n(\tau, \beta_M^*(\tau)) - Q_n(\tau, \beta(\tau)) \overset{(b)}{\geq} Q_\infty(\tau, \beta_M^*(\tau)) - Q_\infty(\tau, \beta(\tau)) + o_{p^*}(1) \overset{(c)}{>} \epsilon_M + o_{p^*}(1)$, for some $\epsilon_M > 0$; where (a) follows by convexity in β, for $\beta_M^*(\tau)$ the point of the boundary of $B_M(\beta(\tau))$ on the line connecting $\beta_M(\tau)$ and $\beta(\tau)$; (b) follows by the uniform convergence established above; and (c) follows since $\beta(\tau)$ is the unique minimizer of $Q_\infty(\beta, \tau)$ uniformly in $\tau \in \mathcal{T}$, by

[1] Basic concepts used in the proof, including weak convergence in the space of bounded functions, stochastic equicontinuity, Donsker and Vapnik-Červonenkis (VC) classes, are defined as in Van der Vaart and Wellner (1996).

[2] The step does not rely on Pollard's (1991) convexity argument, as this argument does not apply to the process case.

convexity and assumption (iii). Hence for any $M > 0$, the minimizer $\hat{\beta}(\tau)$ must be in the radius-M ball centered at $\beta(\tau)$ uniformly for all $\tau \in \mathcal{T}$, with probability approaching one.

1.2 Asymptotic Gaussianity of $\sqrt{n}(\hat{\beta}(\cdot) - \beta(\cdot))$

First, by the computational properties of $\hat{\beta}(\tau)$, for all $\tau \in \mathcal{T}$ (cf. Theorem 3.3 in Koenker and Bassett, 1978) we have that $\|\mathbb{E}_n[\varphi_\tau(Y - X'\hat{\beta}(\tau))X]\| \leq \text{const} \cdot \sup_{i \leq n} \|X_i\|/n$, where $\varphi_\tau(u) = \tau - 1\{u \leq 0\}$. Note that $E\|X_i\|^{2+\varepsilon} < \infty$ implies $\sup_{i \leq n} \|X_i\| = o_{p^*}(n^{1/2})$, since $P\left(\sup_{i \leq n} \|X_i\| > n^{1/2}\right) \leq nP(\|X_i\| > n^{1/2}) \leq nE\|X_i\|^{2+\varepsilon}/n^{\frac{2+\varepsilon}{2}} = o(1)$. Hence uniformly in $\tau \in \mathcal{T}$,

$$\sqrt{n}\mathbb{E}_n\left[\varphi_\tau(Y - X'\hat{\beta}(\tau))X\right] = o_p(1). \tag{1}$$

Second, $(\tau, \beta) \mapsto \mathbb{G}_n\left[\varphi_\tau\left(Y - X'\beta\right)X\right]$ is stochastically equicontinuous over $\mathcal{B} \times \mathcal{T}$, where $\mathcal{B}$ is any compact set, with respect to the $L_2(P)$ pseudometric

$$\rho((\tau', \beta'), (\tau'', \beta''))^2 := \max_{j \in 1,\ldots,d} E\left[\left(\varphi_{\tau'}\left(Y - X'\beta'\right)X_j - \varphi_{\tau''}\left(Y - X'\beta''\right)X_j\right)^2\right],$$

for $j \in 1, ..., d$ indexing the components of X. Note that the functional class $\{\varphi_\tau\left(Y - X'\beta\right)X, \tau \in \mathcal{T}, \beta \in \mathcal{B}\}$ is formed as $(\mathcal{T} - \mathcal{F})X$, where $\mathcal{F} = \{1\{Y \leq X'\beta\}, \beta \in \mathcal{B}\}$ is a VC subgraph class and hence a bounded Donsker class. Hence $\mathcal{T} - \mathcal{F}$ is also bounded Donsker, and $(\mathcal{T} - \mathcal{F})X$ is therefore Donsker with a square integrable envelope $2 \cdot \max_{j \in 1,..d} |X|_j$, by Theorem 2.10.6 in Van der Vaart and Wellner (1996). Stochastic equicontinuity then is part of being Donsker.

Third, by stochastic equicontinuity of $(\tau, \beta) \mapsto \mathbb{G}_n\left[\varphi_\tau\left(Y - X'\beta\right)X\right]$ we have that

$$\mathbb{G}_n\left[\varphi_\tau(Y - X'\hat{\beta}(\tau))X\right] = \mathbb{G}_n\left[\varphi_\tau(Y - X'\beta(\tau))X\right] + o_{p^*}(1), \quad \text{in } \ell^\infty(\mathcal{T}), \tag{2}$$

which follows from $\sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| = o_{p^*}(1)$, and resulting convergence with respect to the pseudometric $\sup_{\tau \in \mathcal{T}} \rho[(\tau, \hat{\beta}(\tau)), (\tau, \beta(\tau))]^2 = o_p(1)$. The latter is immediate from $\sup_{\tau \in \mathcal{T}} \rho\left[(\tau, b(\tau)), (\tau, \beta(\tau))\right]^2 \leq C_3 \cdot \sup_{\tau \in \mathcal{T}} \|b(\tau) - \beta(\tau)\|^{\frac{\varepsilon}{(2+\varepsilon)}}$, where $C_3 = (\bar{f} \cdot (E\|X\|^2)^{1/2})^{\frac{\varepsilon}{2(2+\varepsilon)}} \cdot (E\|X\|^{2+\varepsilon})^{\frac{2}{2+\varepsilon}} < \infty$ and $\bar{f}$ is the a.s. upper bound on $f_Y(Y|X)$. (This follows by the Hölder's inequality and Taylor expansion.)

Further, the following expansion is valid uniformly in $\tau \in \mathcal{T}$

$$E\left[\varphi_\tau(Y - X'\beta)X\right]\Big|_{\beta = \hat{\beta}(\tau)} = \left[J(\tau) + o_p(1)\right]\left(\hat{\beta}(\tau) - \beta(\tau)\right). \tag{3}$$

Indeed, by Taylor expansion $E\left[\varphi_\tau(Y - X'\beta)X\right]|_{\beta = \hat{\beta}(\tau)} = E\left[f_Y(X'b(\tau)|X)XX'\right]|_{b(\tau) = \beta^*(\tau)}(\hat{\beta}(\tau) - \beta(\tau))$, where $\beta^*(\tau)$ is on the line connecting $\hat{\beta}(\tau)$ and $\beta(\tau)$ for each τ, and is different for each row of the Jacobian matrix. Then, (3) follows by the uniform consistency of $\hat{\beta}(\tau)$, and the assumed uniform continuity and boundedness of the mapping $y \mapsto f_Y(y|x)$, uniformly in x over the support of X.

Fourth, since the left hand side (lhs) of (1) = lhs of $n^{1/2}$(3)+ lhs of (2), we have that

$$o_p(1) = [J(\cdot) + o_p(1)](\hat{\beta}(\cdot) - \beta(\cdot)) + \mathbb{G}_n[\varphi_\cdot(Y - X'\beta(\cdot))X]. \tag{4}$$

Therefore, using that mineig $[J(\tau)] \geq \lambda > 0$ uniformly in $\tau \in \mathcal{T}$,

$$\sup_{\tau \in \mathcal{T}} \left\|\mathbb{G}_n[\varphi_\tau(Y - X'\beta(\tau))X] + o_p(1)\right\| \geq (\sqrt{\lambda} + o_p(1)) \cdot \sup_{\tau \in \mathcal{T}} \sqrt{n}\|\hat{\beta}(\tau) - \beta(\tau)\|. \tag{5}$$

Fifth, the mapping $\tau \mapsto \beta(\tau)$ is continuous by the implicit function theorem and stated assumptions. In fact, since $\beta(\tau)$ solves $E\left[(\tau - 1\{Y \leq X'\beta\})X\right] = 0$, $d\beta(\tau)/d\tau = J(\tau)^{-1}E\left[X\right]$. Hence $\tau \mapsto \mathbb{G}_n\left[\varphi_\tau\left(Y - X'\beta(\tau)\right)X\right]$ is stochastically equicontinuous over $\mathcal{T}$ for the pseudo-metric given by $\rho(\tau', \tau'') := \rho((\tau', \beta(\tau')), (\tau'', \beta(\tau'')))$. Stochastic equicontinuity of $\tau \mapsto \mathbb{G}_n\left[\varphi_\tau(Y - X'\beta(\tau))X\right]$ and a multivariate CLT imply that

$$\mathbb{G}_n\left[\varphi_\cdot(Y - X'\beta(\cdot))X\right] \Rightarrow z(\cdot) \text{ in } \ell^\infty(\mathcal{T}), \tag{6}$$

where $z(\cdot)$ is a Gaussian process with covariance function $\Sigma(\cdot,\cdot)$ specified in the statement of Theorem 3. Therefore, the lhs of (5) is $O_p(n^{-1/2})$, implying $\sup_{\tau\in\mathcal{T}} \|\sqrt{n}(\hat\beta(\tau) - \beta(\tau))\| = O_{p^*}(1)$.

Finally, the latter fact and (4)-(6) imply that in $\ell^\infty(\mathcal{T})$

$$J(\cdot)\sqrt{n}(\hat\beta(\cdot) - \beta(\cdot)) = -\mathbb{G}_n\left[\varphi_\cdot(Y - X'\beta(\cdot))\right] + o_{p^*}(1) \Rightarrow z(\cdot). \tag{7}$$

Q.E.D.

1.3 Proof of Corollaries

Proof of Corollary 1. This result follows by the continuous mapping theorem in $\ell^\infty(\mathcal{T})$. Absolute continuity of $\mathcal{K}$ follows from Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998). *Q.E.D.*

Proof of Corollary 2. This result follows by absolute continuity of $\mathcal{K}$. The consistency of subsampling estimator of $\hat\kappa(\alpha)$ follows from Theorem 2.2.1 and Corollary 2.4.1 in Politis, Romano and Wolf (1999), for the case where $V(\tau)$ are known. When $V(\tau)$ is estimated consistently uniformly in $\tau \in \mathcal{T}$, the result follows by an argument similar to the proof of Theorem 2.5.1 in Politis, Romano and Wolf (1999). *Q.E.D.*

1.4 Uniform Consistency of $\widehat\Sigma(\cdot,\cdot)$ and $\widehat J(\cdot)$.

Here it is shown that under the conditions of Theorem 3 and the additional assumption that $E\|X\|^4 < \infty$, the estimates described in the main text are consistent uniformly in $(\tau, \tau') \in \mathcal{T} \times \mathcal{T}'$.[3]

First, recall that $\hat J(\tau) = [1/(2h_n)] \cdot \mathbb{E}_n[1\{|Y_i - X_i'\hat\beta(\tau)| \leq h_n\} \cdot X_i X_i']$. We will show that

$$\hat J(\tau) - J(\tau) = o_{p^*}(1) \text{ uniformly in } \tau \in \mathcal{T}. \tag{8}$$

Note that $2h_n\hat J(\tau) = \mathbb{E}_n[f_i(\hat\beta(\tau), h_n)]$, where $f_i(\beta, h) = 1\{|Y_i - X_i'\beta| \leq h\} \cdot X_i X_i'$. For any compact set B and positive constant H, the functional class $\{f_i(\beta, h), \beta \in B, h \in (0, H]\}$ is a Donsker class with a square-integrable envelope by Theorem 2.10.6 in Van der Vaart and Wellner (1996), since this is a product of a VC subgraph class $\{1\{|Y_i - X_i'\beta| \leq h\},\ \beta \in B,\ h \in (0, H]\}$ and a square integrable random matrix $X_i X_i'$ (recall $E\|X_i\|^4 < \infty$ by assumption). Therefore, $(\beta, h) \mapsto \mathbb{G}_n\left[f_i(\beta, h)\right]$ converges to a Gaussian process in $\ell^\infty(B \times (0, H])$, which implies that $\sup_{\beta\in B, 0<h\leq H} \|\mathbb{E}_n\left[f_i(\beta, h)\right] - E\left[f_i(\beta, h)\right]\| = O_{p^*}(n^{-1/2})$. Letting B be any compact set that covers $\cup_{\tau\in\mathcal{T}}\beta(\tau)$, this implies $\sup_{\tau\in\mathcal{T}} \|\mathbb{E}_n[f_i(\hat\beta(\tau), h_n)] - E[f_i(\beta, h_n)]\big|_{\beta=\hat\beta(\tau)}\| = O_{p^*}(n^{-1/2})$. Hence (8) follows by using $2h_n\hat J(\tau) = \mathbb{E}_n[f_i(\hat\beta(\tau), h_n)]$, $1/(2h_n) \cdot E[f_i(\beta, h_n)]\big|_{\beta=\hat\beta(\tau)} = J(\tau) + o_p(1)$, and the assumption $h_n^2 n \to \infty$.

[3]Note that the result for $\widehat J(\tau)$ is not covered by Powell (1986) because his proof applies only pointwise in τ, whereas we require a uniform result.

Second, we can write $\hat{\Sigma}(\tau,\tau') = \mathbb{E}_n[g_i(\hat{\beta}(\tau),\hat{\beta}(\tau'),\tau,\tau')X_iX_i']$, where $g_i(\beta',\beta'',\tau',\tau'') = (\tau - 1\{Y_i \leq X_i'\beta'\})(\tau' - 1\{Y_i \leq X_i'\beta''\}) \cdot X_iX_i'$. We will show that

$$\hat{\Sigma}(\tau,\tau') - \Sigma(\tau,\tau') = o_{p^*}(1) \text{ uniformly in } (\tau,\tau') \in \mathcal{T} \times \mathcal{T}. \tag{9}$$

It is easy to verify that $\{g_i(\beta',\beta'',\tau',\tau''),(\beta',\beta'',\tau',\tau'') \in B \times B \times \mathcal{T} \times \mathcal{T}\}$ is Donsker and hence a Glivenko-Cantelli class, for any compact set B, e.g., using Theorem 2.10.6 in Van der Vaart and Wellner (1996). This implies that $\mathbb{E}_n\left[g_i(\beta',\beta'',\tau',\tau'')X_iX_i'\right] - E\left[g_i(\beta',\beta'',\tau',\tau'')X_iX_i'\right] = o_{p^*}(1)$ uniformly in $(\beta',\beta'',\tau',\tau'') \in (B \times B \times \mathcal{T} \times \mathcal{T})$. The latter and continuity of $E\left[g_i(\beta',\beta'',\tau',\tau'')X_iX_i'\right]$ in $(\beta',\beta'',\tau',\tau'')$ imply (9).

Q.E.D.

References

[1] Davydov, Y. A., M. A. Lifshits, and N. V. Smorodina (1998): *Local Properties of Distributions of Stochastic Functionals.* Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova. Translations of Mathematical Monographs, 173. American Mathematical Society, Providence, RI.

[2] Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica* 46, pp. 33-50.

[3] Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling.* Springer-Verlag. New York.

[4] Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, no. 2, 186–199.

[5] Van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer Series in Statistics. Springer-Verlag. New York.