Centre for Economic and Financial Research at New Economic School

November 2018

Reconsideration of a simple approach to quantile regression for panel data

Galina Besstremyannaya
Sergei Golovan

Working Paper No 248

CEFIR /NES Working Paper series

Reconsideration of a simple approach to quantile regression for panel data

Galina Besstremyannaya* Sergei Golovan**

November 2018

Abstract

The note discusses a fallacy in the approach proposed by Ivan Canay (2011, *The Econometrics Journal*) for constructing a computationally simple two-step estimator in a quantile regression model with quantile-independent fixed effects. We formally prove that the estimator gives an incorrect inference for the constant term due to violation of the assumption about additive expansion of the first-step estimator, which requires the independence of its terms. Our simulations show that Canay's confidence intervals for the constant term are wrong. Finally, we focus on the fact that finding a $\sqrt{nT}$ consistent within estimator, as required by Canay's procedure, may be problematic. We provide an example of a model, for which we formally prove the non-existence of such an estimator.

JEL Classification Codes: C21, C23

Keywords: quantile regression, panel data, fixed effects, inference

*Centre for Economic and Financial Research at New Economic School, Moscow, Nakhimovsky pr.47, gbesstre@cefir.ru

**New Economic School, Moscow, Skolkovskoe shosse, 45, sgolovan@nes.ru

1 Introduction

Use of panel data quantile regression models dates back to Koenker (2004), who considers the equation

$$Q_{y_{it}}(\tau|x_{ij}) = \alpha_i + x'_{it}\beta(\tau), \quad t = 1, \ldots, T_i, \quad i = 1, \ldots, n,$$

where $Q_{y_{it}}(\tau|x_{ij})$ denotes the value of a given quantile for conditional distribution of the continuous dependent variable y for observation i at period t. The equation specifies the individual effects α_i as additional unknown parameters, but their estimation is difficult since n can be very large in panel datasets.

A solution is apparently offered by a computationally simple estimator by Ivan Canay (2011, *The Econometrics Journal*) for quantile-independent individual effects. Canay (2011) proposes a two-step procedure, which first gives a consistent estimation of individual effects using the within estimator and then applies the pooled version of the panel data quantile regression to the dependent variable, cleared of the estimated individual effects. The Canay estimator is widely used by practitioners and is much cited in the theoretical literature. According to the Wiley online library, there are 115 citations in Web of Science journals (as of November 2018), while Google Scholar gives 377 citations.

However, as we show in this note, the Canay's approach gives an incorrect inference for the constant term because it violates the assumption of additive expansion of the first-step estimator, which requires the independence of its terms. We show that the terms are dependent between different time periods and, as a result, the derivation of the asymptotics of the second-step estimator of the constant term fails. Our simulations find that Canay uses the wrong confidence intervals for the constant term. The bias in the asymptotic standard errors increases with the number of time periods in the panel. Finally, we focus on the fact that Canay's approach depends on finding a $\sqrt{nT}$ consistent within estimator, which may be problematic in a panel data model with individual effects. We provide an example of a model, for which we formally prove the impossibility of constructing such an estimator.

2 Theoretical critique

The approach proposed in Canay's article uses a two-step estimator for the following model

$$Y_{it} = X'_{it}\theta(U_{it}) + \alpha_i, \quad i = 1, \ldots, n, \ t = 1, \ldots, T,$$

where U_{it} does not depend on (X_{it}, α_i).

At the first stage, a $\sqrt{nT}$ consistent estimator $\hat{\theta}_\mu$ of $\theta_\mu = E[\theta(U_{it})]$ is used to compute

$$\hat{\alpha}_i \equiv \frac{1}{T}\sum_{t=1}^{T}[Y_{it} - X'_{it}\hat{\theta}_\mu].$$

The second stage defines $\hat{Y}_{it} \equiv Y_{it} - \hat{\alpha}_i$ and the estimator $\hat{\theta}(\tau)$ as

$$\hat{\theta}(\tau) = \underset{\theta}{\operatorname{argmin}} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\rho_\tau(\hat{Y}_{it} - X'_{it}\theta)v_{it}. \tag{1}$$

The asymptotic properties of the two-step estimator are derived using the key assumption described below, the most important part of which is an additive expansion of $\hat{\theta}_\mu$ with the independence of its terms ψ_{it}.

Assumption 4.2, Canay (2011). *The first-step estimator $\hat{\theta}_\mu$ admits the expansion*

$$\sqrt{nT}(\hat{\theta}_\mu - \theta_\mu) = \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} \sum_{i=1}^{n} \psi_{it} + o_p(1), \tag{2}$$

where ψ_{it} is an i.i.d. sequence of random variables with $E[\psi_{it}] = 0$ and finite $\Omega_{\psi\psi} = E[\psi_{it}\psi_{it}']$.

Assumption 4.2 is then exploited for the derivation of the asymptotic normality of the second-step estimator.[1] Note that the assumption is roughly equivalent to a $\sqrt{nT}$ consistency of the first-step estimator, where $\sqrt{nT}(\hat{\theta}_\mu - \theta_\mu)$ converges to a finite distribution.

Theorem 4.1, Canay (2011). *Let $n/T^s \to 0$ for some $s \in (1, +\infty)$. Under Assumptions 3.2, 4.1 and 4.2*

$$\sup_{\tau \in \mathcal{T}} \|\hat{\theta}(\tau) - \theta(\tau)\| \to_p 0,$$

and

$$\sqrt{nT}(\hat{\theta}(\cdot) - \theta(\cdot)) = [-J_1(\cdot)]^{-1} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \{\phi_\tau(\varepsilon_{it}(\tau))X_{it} + J_2(\cdot)\xi_{it}\} + o_p(1), \tag{3}$$

$$\rightsquigarrow \mathbb{G}(\cdot) \quad \text{in } \ell^\infty(\mathcal{T}), \tag{4}$$

where $\epsilon_{it}(\tau) \equiv Y_{it}^* - X_{it}'\theta(\tau)$, $Y_{it}^* = Y_{it} - \alpha_i$, $\xi_{it} \equiv \mu_X'\psi_{it} - u_{it}$, $u_{it} \equiv Y_{it}^* - X_{it}'\theta_\mu$, $\mu_X = E[X_{it}]$, $J_1(\tau) \equiv J_1(\theta(\tau), \tau, 0)$, $J_2(\tau) \equiv J_2(\theta(\tau), \tau, 0)$, $\mathbb{G}(\cdot)$ is a mean zero Gaussian process with the covariance function $E\mathbb{G}(\tau)\mathbb{G}(\tau')' = J_1(\tau)^{-1}\Psi(\tau, \tau')[J_1(\tau')^{-1}]'$, $\Psi(\tau, \tau')$ is defined in the equation below, and $\ell^\infty(\mathcal{T})$ is the set of uniformly bounded functions on $\mathcal{T}$. The matrix $\Psi(\tau, \tau')$ is given by

$$\Psi(\tau, \tau') = S(\tau, \tau') + J_2(\tau)\Omega_{\xi g}(\tau') + \Omega_{g\xi}(\tau)J_2(\tau')' + J_2(\tau)\Omega_{\xi\xi}J_2(\tau')',$$

where $S(\tau, \tau') \equiv (\min\{\tau, \tau'\} - \tau\tau')E(XX')$, $\Omega_{g\xi}(\tau) \equiv E[g_\tau(W, \theta(\tau))\xi]$, and $\Omega_{\xi\xi} \equiv E[\xi^2]$.

Next, the within estimator is taken to satisfy Assumption 4.2 (see the lemma below) and therefore supposed to be an appropriate first-step estimator. It is then used to construct the asymptotic covariance matrix of the two-step estimator.

Lemma A.4, Canay (2011). *Assume $\Omega_{XX} \equiv E[(X_{it}^s - \mu_X^s)(X_{it}^s - \mu_X^s)']$ is non-singular with finite norm, $n/T^a \to 0$ for some $a \in (0, \infty)$ and let Assumptions 3.2 and 4.1 hold. The within estimator of θ_μ satisfies Assumption 4.2 with the influence function*

$$\psi_{it} = \begin{pmatrix} \psi_{it}^0 \\ \psi_{it}^s \end{pmatrix} \equiv \begin{pmatrix} Y_{it} - \mu_Y - \mu_X^{s\prime}\Omega_{XX}^{-1}(X_{it}^s - \mu_X^s)u_{it} \\ \Omega_{XX}^{-1}(X_{it}^s - \mu_X^s)u_{it} \end{pmatrix},$$

where $X_{it}' = (1, X_{it}^{s\prime})$, $\mu_X^s \equiv E(X_{it}^s)$, $\mu_Y \equiv E(Y_{it})$, u_{it} is i.i.d. with $E[u_{it}|X_i] = 0$ and $E[u_{it}^2|X_i] = X_{it}'\Omega_{UU}X_{it}$, and Ω_{UU} non-singular with finite norm.

However, as we prove in the proposition below, there is a fallacy in Lemma A.4. Namely, the assumption of independence of the first components ψ_{it}^0 is unjustified. So the within estimator does not satisfy Assumption 4.2 and the asymptotic standard errors are incorrect.

[1] Along with Assumption 4.2, which is the focus of this note, Theorem 4.1. in Canay's article uses Assumption 3.2 and Assumption 4.1. The former defines fixed effects as time-independent ("location shifters") and the latter gives the expressions for the terms J_1 and J_2 in the covariance matrix of the first-step estimator.

Proposition 1. *Given the conditions of Lemma A.4 the first components ψ_{it}^0 of the influence vectors ψ_{it} are not independent across time periods if $i = 1, \ldots, n$ is fixed. Therefore, Assumption 4.2 is not satisfied.*

Proof. Consider the model

$$Y_{it} = X_{it}'\theta(U_{it}) + \alpha_i, \quad i = 1, \ldots, n,\ t = 1, \ldots, T.$$

Under the definition of $u_{it} = X_{it}'(\theta(U_{it}) - \theta_\mu)$ from the proof of Lemma A.4, the model can be expressed as

$$Y_{it} = X_{it}'\theta_\mu + \alpha_i + u_{it} = \theta_\mu^0 + X_{it}^{s\prime}\theta_\mu^s + \alpha_i + u_{it}, \quad i = 1, \ldots, n,\ t = 1, \ldots, T,$$

where u_{it} are i.i.d. across i and t (and independent of X_{it}), but α_i are constant for different t when i is fixed. Taking expectations, we obtain

$$\mu_Y = E(Y_{it}) = E(\theta_\mu^0 + X_{it}^{s\prime}\theta_\mu^s + \alpha_i + u_{it}) = \theta_\mu^0 + \mu_X^{s\prime}\theta_\mu^s.$$

(Here we assume that $E(\alpha_i = 0)$, otherwise θ_μ^0 is not identifiable.) This implies

$$Y_{it} - \mu_Y = (X_{it}^s - \mu_X^s)'\theta_\mu^s + \alpha_i + u_{it}$$

and

$$\psi_{it}^0 = Y_{it} - \mu_Y - \mu_X^{s\prime}\Omega_{XX}^{-1}(X_{it}^s - \mu_X^s)u_{it} = \alpha_i + u_{it} + (X_{it}^s - \mu_X^s)'\theta_\mu^s - \mu_X^{s\prime}\Omega_{XX}^{-1}(X_{it}^s - \mu_X^s)u_{it}.$$

The last three terms in the expression for ψ_{it}^0 are i.i.d. across all i and t.

Consider $t \neq t'$. Since ψ_{it}^0 and $\psi_{it'}^0$ contain the same term α_i, they are generally correlated. □

3 Demonstration of incorrect asymptotic distribution in simulations

We replicate the simulations from section 5 of Canay's article in order to show the bias in estimate of the standard error of the constant term. The model is defined as

$$\begin{aligned} Y_{it} &= (\epsilon_{it} - 1) + \epsilon_{it}X_{it} + \alpha_i, \\ \alpha_i &= \gamma(X_{i1} + \cdots + X_{iT} + \eta_{it}) - E(\alpha_i), \end{aligned}$$

where ϵ_{it} is taken to be $N(2, 1)$.

We fix γ, τ (this defines θ), n and T, and generate $B = 1000$ samples. We then compute the confidence intervals for the coefficients using the asymptotic distribution derived by Canay. Next, we compute the coverage probability for these confidence intervals, dividing the number of cases when the actual parameter value falls in the interval by the number of simulations B. Along with examining the 90% confidence intervals, as is standardly done, we also focus on the 80% confidence intervals, since the lower confidence level better illustrates the bias. As shown by the results reported in Table 1, the coverage probabilities for $\theta_1(\tau)$ are close to their true levels, but the coverage probabilities for $\theta_0(\tau)$ are overestimated.

We would also note that the coverage probability goes up with an increase in T. In other words, the intervals become relatively wider, so the ratio of the length of the estimated confidence interval to the length of the true confidence interval grows with T.

Table 1: Coverage probability for the confidence intervals for model 5.1 from Canay (2011)

	$\gamma = 1$, $\tau = 0.5$, $\theta(\tau) = (1, 2)'$							
	80% confidence level				90% confidence level			
	$n = 100$		$n = 1000$		$n = 100$		$n = 1000$	
	$\theta_0(\tau)$	$\theta_1(\tau)$	$\theta_0(\tau)$	$\theta_1(\tau)$	$\theta_0(\tau)$	$\theta_1(\tau)$	$\theta_0(\tau)$	$\theta_1(\tau)$
$T = 5$	0.811	0.740	0.789	0.745	0.910	0.854	0.914	0.853
$T = 10$	0.856	0.793	0.831	0.764	0.925	0.868	0.910	0.865
$T = 50$	0.889	0.783	0.898	0.782	0.962	0.879	0.962	0.879
$T = 100$	0.938	0.792	0.930	0.801	0.986	0.890	0.980	0.900
$T = 200$	0.969	0.826	0.973	0.787	0.997	0.916	0.997	0.878
$T = 500$	0.996	0.778	0.995	0.824	1.000	0.911	1.000	0.912

4 Final remarks on a $\sqrt{nT}$ consistency

It should be noted that finding a $\sqrt{nT}$ consistent within estimator of a constant term is problematic in the model with individual effects α_i. Indeed, a new observation significantly improves the accuracy of the estimator of the constant term only if it contains information about a new individual (hence, about new α_i). Here we provide a simple example of a panel data model with individual effects, for which we strictly prove the non-existence of such an estimator.

Proposition 2. *Let $Y_{it} = \mu + \alpha_i + \varepsilon_{it}$, $i = 1, \dots, n$, $t = 1, \dots, T$, where α_i are i.i.d. $N(0, \sigma_\alpha^2)$, ε_{it} are i.i.d. $N(0, \sigma_\varepsilon^2)$ and α_i are independent of ε_{jt} for all i, j, t ($j = 1, \dots, n$). Suppose σ_α and σ_ε are known. Then, the following inequality holds for any unbiased estimator $\hat{\mu}$ of μ*

$$\mathrm{Var}(\hat{\mu}) \geq \frac{\sigma_\alpha^2 + \sigma_\varepsilon^2/T}{n}.$$

So $\hat{\mu}$ can be only $\sqrt{n}$ consistent, and not $\sqrt{nT}$ consistent.

Proof. The joint distribution of $Y = (Y_{11}, \dots, Y_{1T}, \dots, Y_{n1}, \dots, Y_{nT})'$ is Gaussian with the mean $\boldsymbol{\mu} = (\mu, \dots, \mu)'$ and the covariance matrix $I \otimes \Sigma$, where

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix}.$$

This implies that the Fisher-information for μ is

$$I(\mu) = \iota'(I \otimes \Sigma)^{-1}\iota = \iota'(I \otimes \Sigma^{-1})\iota,$$

where $\iota = (1, \dots, 1)'$ is a unity vector of length nT.

$$\Sigma^{-1} = \frac{1}{\sigma_\varepsilon^2(T\sigma_\alpha^2 + \sigma_\varepsilon^2)} \begin{pmatrix} (T-1)\sigma_\alpha^2 + \sigma_\varepsilon^2 & -\sigma_\alpha^2 & \dots & -\sigma_\alpha^2 \\ -\sigma_\alpha^2 & (T-1)\sigma_\alpha^2 + \sigma_\varepsilon^2 & \dots & -\sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_\alpha^2 & -\sigma_\alpha^2 & \dots & (T-1)\sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix}.$$

Hence, $I(\mu) = \dfrac{nT\sigma_\varepsilon^2}{\sigma_\varepsilon^2(T\sigma_\alpha^2 + \sigma_\varepsilon^2)} = \dfrac{nT}{T\sigma_\alpha^2 + \sigma_\varepsilon^2}$.

An application of the Cramér–Rao bound (see Amemiya (1985), Theorem 1.3.1) finishes the proof. □

References

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Canay, I. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal*, 14(3):368–386.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89.