Прикладная эконометрика, 2021, т. 64, с. 70–82.
Applied Econometrics, 2021, v. 64, pp. 70–82.
DOI: 10.22394/1993-7601-2021-64-70-82

G. Besstremyannaya, S. Golovan[1]

Measuring heterogeneity with fixed effect quantile regression: Long panels and short panels

The desire to capture heterogeneity in the response of the dependent variable to covariates often forces empiricists to employ panel data quantile regression models. Very often practitioners forget the limitations of their datasets in terms of the sample size n and the length of panel T. Yet, quantile regression requires large samples, long panels and small value of the ratio n/T. So the estimator in quantile regression with short panels is biased. The paper reviews the approaches for estimating longitudinal models for quantile regression. We highlight the fact that a method of smoothed quantile regression may be viewed as a remedy for reducing the asymptotic bias of the estimator in short panels, both in case of quantile-dependent and quantile-independent fixed effect specifications.

Keywords: quantile regression; panel data.
JEL classification: C44; C61.

1. Introduction

The popularity of the OLS regression in applied economics may be attributed to the fact that "least-squares methods provide a general approach to estimating conditional mean functions" (Koenker, 2005, p. 1). However, the conditional mean function does not give the full information about the distribution of the dependent variable conditional on covariates. Indeed, in many economic applications the researcher can expect that "the partial effect of an explanatory variable can have very different effects across different segments of a population" (Wooldridge, 2010, p. 449). For instance, in case of the analysis of the loglinear production function, the value of elasticity of output with respect to capital may differ across more productive and less productive firms. Yet, mean regression only enables to obtain the estimate of elasticity for all firms in the sample.

The quantile regression offers an approach to study the impact of the covariates at the "segments" of dependent variable: the analysis is applied to the conditional τ-th quantile of the dependent variable. Instead of extrapolating the results of the mean regression to the tails of the distribution of the dependent variable, quantile regression enables obtaining independent estimates

[1] **Besstremyannaya, Galina** — National Research University Higher School of Economics, Moscow; gbesstremyannya@hse.ru.
Golovan, Sergei — New Economic School, Moscow; sgolovan@nes.ru.

for the impact of covariates in each conditional quantile of the dependent variable. Different values of the estimated coefficients for the explanatory variable obtained in regressions with different values of τ are interpreted as the presence of heterogeneous effect of this explanatory variable. For instance, quantile regression may be used for studying heterogeneous effect of policy reforms and macroeconomic shocks on production, or for evaluating heterogeneous effects of socio-demographic characteristics of consumers on their expenditure.

Another merit of quantile regression is the applicability for efficiency analysis. High values of quantile index (e.g. 0.8, 0.9) may be taken as an approximation of the production possibility frontier, while in case of conditional quantile regression applied to cost function low values of quantile index (e.g. 0.1, 0.2) may serve an approximation for the best cost minimization trajectory[2].

It should be noted that the asymptotic theory for cross-sectional quantile regression is similar to the OLS models and requires large values of sample size n. The differences between the quantile regression and OLS methodology arise in case of longitudinal models. Specifically, the OLS regression with longitudinal data admits short panels, but the asymptotic theory for quantile regression holds only for long panels: the ratio of sample size n to the length of panel T should be small. So the estimator in quantile regression with short panels is biased.

Yet, the low values of n/T are rarely faced by empiricists. In our meta-review of over 80 empirical papers using panel data conditional quantile regression with quantile-independent fixed effects (Besstremyannaya, Golovan, 2019), only 7 percent of papers had $n/T < 1$ and another 21 percent had $1 \le n/T < 10$. All these papers are long macro panels with the annual or quarterly data for countries or regions. Most of the remaining works with short panels are applications in different fields of economics, where the unit of observation is a firm, an individual, a household, an employer etc.

The purpose of the present paper is to highlight the fact that a recently developed method of smoothed quantile regression (Galvao, Kato, 2016) may be viewed as a means for reducing the asymptotic bias of the estimator in short panels, both in case of quantile-dependent and quantile-independent conditional quantile regressions with exogenous covariates. The remainder of the paper is structured as follows. Section 2 outlines the specification and estimation procedure for conditional quantile regression in case of cross-section model, and Section 3 deals with pooled data model. Section 4 describes the Galvao and Kato (2016) smoothing technique for estimating *quantile-dependent* fixed effect model and the possibility to reduce the bias of the smoothed estimator in case of short panels: for instance, through split-panel jackknife estimator of Dhaene and Jochmans (2015). As regards *quantile-independent* fixed effect model, Section 5 outlines the simple estimator suggested by Canay (2011) and mentions the critique of the estimator in view of its asymptotic bias (Besstremyannaya, Golovan, 2019). We note that a new simple estimator of quantile-independent fixed effect model, developed by Chen and Huo (2021), exploits smoothing techniques by Galvao and Kato (2016) and enables to reduce the asymptotic bias in case of short panels. Section 6 contains simulations that show the bias of the estimator in case of short panels with quantile-dependent fixed effects. The conclusion in Section 7 draws attention of practitioners to the need of considering limitations of their datasets in terms of the sample size and the length of panel.

---

[2] See application in banking in (Besstremyannaya, 2017) and in health economics in (Liu et al., 2007).

2. Cross sectional quantile regression model

Denote $Q_\tau\left(Y \mid X=x\right)$ the conditional τ-th quantile of a continuous variable Y under fixed values of covariates x. By definition, for $0<\tau<1$, $\mathrm{P}\left(Y \le Q_\tau\left(Y \middle| X=x\right) \middle| X=x\right)=\tau$. The linear quantile regression considers conditional τ-th quantile of a continuous variable Y as a linear function of covariates X.

The model which was originally formulated in (Koenker, Bassett, 1978) may be specified as follows:

$$Y_i = X_i'\beta\left(U_i\right);\ \text{function } \tau \mapsto X_i'\beta\left(\tau\right) \text{ is increasing in } \tau,$$

where τ is the value of a given quantile for conditional distribution of the dependent variable Y for observation *i*, X is a vector of exogenous variables, and $U_i \perp \left(X_i\right) \sim U\left[0,1\right]$, $i=1,\ldots,n$.

In other words, $Q_\tau\left(Y \mid X=x\right)=X'\beta\left(\tau\right)$ for each $\tau \in \left(0,1\right)$.

A consistent procedure for estimating β involves minimizing the objective function W_n:

$$W_n\left(\tau,\beta\right)=\frac{1}{n}\sum_{i=1}^{n}\rho_\tau\left(Y_i - X_i'\beta\right). \tag{1}$$

Here $\rho_\tau\left(\cdot\right)$ is the Koenker and Bassett (1978) loss function[3]:

$$\rho_\tau\left(u\right)=u\left(\tau-\mathrm{I}\left(u<0\right)\right)=\begin{cases}\tau u, & \text{if } u \ge 0,\\ -\left(1-\tau\right)u, & \text{if } u<0.\end{cases}$$

Koenker (2005) shows that under regularity conditions, $\sqrt{n}\left(\hat{\beta}-\beta\right)$ has a limiting normal distribution and the asymptotic variance matrix Σ may be expresses in the sandwich form (Wooldridge, 2007). Specifically,

$$\Sigma\left(\tau,\tau'\right)=A\left(\tau\right)^{-1}B\left(\tau,\tau'\right)\left[A\left(\tau'\right)^{-1}\right]'$$

of the stochastic process $\hat{\beta}\left(\tau\right)$:

$$\hat{B}\left(\tau,\tau'\right)=\hat{S}\left(\tau,\tau'\right),\ \hat{A}\left(\tau\right)=\frac{1}{2nh_n\left(\tau\right)}\sum_{i=1}^{n}\mathrm{I}\left(\left|\hat{\varepsilon}_i\left(\tau\right)\right| \le h_n\left(\tau\right)\right)X_iX_i',$$

where residuals $\hat{\varepsilon}_i\left(\tau\right)=Y_i - X_i'\hat{\beta}\left(\tau\right)$, and $h_n\left(\tau\right)$ is an appropriately selected bandwidth (for instance, $h_n\left(\tau\right)=\kappa\left[\Phi^{-1}\left(\tau+h_n\right)-\Phi^{-1}\left(\tau-h_n\right)\right]$ with κ equal to median absolute deviation of the τ-th quantile regression residuals and h_n defined in (Koenker, Machado, 1999).

The scores $s_i\left(\tau\right)$ of the objective function in (1) are set as a piecewise derivative

$$s_i\left(\tau\right)=-\frac{\partial\rho_\tau\left(\varepsilon_i\left(\tau\right)\right)}{\partial\beta'}=\psi_\tau\left(\varepsilon_i\left(\tau\right)\right)X_i,$$

[3] As its shape resembles the check mark, the loss function is often referred to as the check function, see (Wooldridge, 2010, Sec. 12.10.1).

where $\psi_\tau(u) = \rho'_\tau(u) = \tau - \mathrm{I}(u < 0)$. The estimator of $S(\tau, \tau')$ in the expression for $\hat{B}(\tau, \tau')$ becomes:

$$\hat{S}(\tau, \tau') = \frac{1}{n} \sum_{i=1}^{n} \hat{s}_i(\tau)' \, \hat{s}_i(\tau') = \frac{1}{n} \sum_{i=1}^{n} \psi_\tau(\hat{\varepsilon}_i(\tau)) \psi_{\tau'}(\hat{\varepsilon}_i(\tau')) X_i X'_i.$$

3. The pooled model of quantile regression

A simple longitudinal version of quantile regression is a pooled model (Wooldridge, 2007):

$$Y_{it} = X'_{it}\beta(U_{it}), \ \tau \mapsto X'_{it}\beta(\tau),$$

where τ denotes the value of a given quantile for conditional distribution of the dependent variable Y for observation i at period t, X is a vector of exogenous variables, and $U_{it} \perp (X_{it}) \sim U[0, 1]$, $i = 1, \ldots, n$, $t = 1, \ldots, T$.

The pooled model of conditional quantile regression is the simplest way to work with the longitudinal data. Note that the pooled model does not contain individual effects α_i. Similarly to the OLS model, pooling the data in conditional quantile regression leads to serial correlation of the error terms for each fixed value of i (i.e. each cluster of observations).

A consistent estimation procedure involves minimizing the quantile regression objective function, where the sums are taken across the values of i and t:

$$W_{nT}(\tau, \beta) = \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \rho_\tau(Y_{it} - X'_{it}\beta).$$

Here ρ_τ is the loss function.

The Wooldridge (2007) correction of the variance matrix in such pooled model accounts for serial correlation within the clusters of observations. Specifically, the scores of the objective function $s_{it}(\tau)$ are computed as a piecewise derivative:

$$s_{it}(\tau) = \frac{\partial \rho_\tau(\varepsilon_{it}(\tau))}{\partial \beta'} = -\psi_\tau(\varepsilon_{it}(\tau)) X_{it},$$

and the asymptotic covariance matrix of the estimates becomes

$$\Sigma(\tau, \tau') = A(\tau)^{-1} B(\tau, \tau') \left[A(\tau')^{-1} \right]',$$

where its components can be estimated as follows:

$$\hat{B}(\tau, \tau') = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{s}_{it}(\tau)' \, \hat{s}_{is}(\tau') = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{t=1}^{T} \psi_\tau(\hat{\varepsilon}_{it}(\tau)) \psi_{\tau'}(\hat{\varepsilon}_{is}(\tau')) X_{it} X'_{is},$$

$$\hat{A}(\tau) = \frac{1}{2nh_n(\tau)} \sum_{i=1}^{n} \sum_{t=1}^{T} \mathrm{I}(|\hat{\varepsilon}_{it}(\tau)| \le h_n(\tau)) X_{it} X'_{it}.$$

Similar approach for accounting for groupwise serial correlation is proposed in (Parente, Santos Silva, 2016): in this case, time may be taken as one of the grouping dimensions.

4. Quantile-dependent fixed effects

4.1. The model

A general form of a panel data quantile regression model is specified in Koenker (2004) as follows:

$$\begin{gathered} Y_{it} = X_{it}'\beta\left(U_{it}\right) + \alpha_i\left(U_{it}\right), \quad U_{it} \sim U\left[0, 1\right], \\ \tau \mapsto X_{it}'\beta\left(\tau\right) + \alpha_i\left(\tau\right) \text{ is monotonically increasing,} \end{gathered} \tag{2}$$

where $\tau \in \left(0, 1\right)$, mapping in (2) is the conditional quantile of the dependent variable Y_{it}, X_{it} is a vector of covariates and $\alpha_i\left(\tau\right)$ are fixed effects, which vary across quantiles.

Note that in case of the OLS model, the conditional mean is a linear function. The linear character of the conditional mean enables conducting the within-transformation in estimating the fixed effect OLS model and thus exclude the individual effects under any length of panel. However, the conditional quantile is a non-linear function, so the within-transformation becomes infeasible. Therefore, the Koenker (2005) model of quantile-dependent fixed effects contains the whole list of individual effects. This leads to incidental parameters problem. So long panels are required to avoid the negative effect of the problem on the properties of the estimator.

The estimator becomes (see (Kato et al., 2012, eq. (2.2))):

$$\left(\left\{\hat{\alpha}_i\right\}, \hat{\beta}\right) = \underset{\{\alpha_i\},\beta}{\operatorname{argmin}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{i=1}^{T} \rho_\tau\left(Y_i - X_i'\beta - \alpha_i\right),$$

and its asymptotic theory requires long panels: n/T must be small. Kato et al. (2012) do not derive an analytical expression for the bias of the estimator making it impossible to apply the estimator to short panels.

4.2. Restrictions on the model in short panels

Approaches to estimate quantile-dependent fixed effect model with short panels impose various restrictions on the model. For example, Machado, Santos Silva (2019) and Li et al. (2003) require additional assumptions about the distribution of dependent variable, while Harding and Lamarche (2016) model fixed effects as functions of covariates.

4.3. Smoothed estimator for short panels

The smoothed quantile regression offers a remedy for keeping the general form of specification in short panels. The approach was proposed by Galvao and Kato (2016) who modify the Koenker (2004) quantile regression objective function through smoothing

$$\min_{\{\alpha_i\},\beta} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}-X_{it}'\beta-\alpha_i\right)\left(\tau-G\left(Y_{it}-X_{it}'\beta-\alpha_i\right)/h\right),$$

where $G(u)=\int_u^{\infty}K(v)\,dv$ is a smoothed analogue of the step function $\mathrm{I}(u<0)$, $K(v)$ is a kernel function, and h is a bandwidth.

Using the smoothing technique, Galvao and Kato (2016) derive the bias of the estimator. To reduce the bias, they recommend two methods: subtracting the asymptotic expression of the bias or using the Dhaene and Jochmans (2015) jackknife split panel correction of the bias.

In case of balanced panels, the Dhaene and Jochmans (2015) procedure requires splitting the panel into two: $i\in\{1,\ldots,n\}$ in each panel, while the time index is $t\in\{1,\ldots,T/2\}$ in the first panel and $t\in\{T/2+1,\ldots,T\}$ in the second panel. The split panel estimator is computed as

$$\hat{\beta}_{1/2}(\tau)=2\hat{\beta}(\tau)-\left(\hat{\beta}_1(\tau)+\hat{\beta}_2(\tau)\right)/2,$$

where $\hat{\beta}(\tau)$, $\hat{\beta}_1(\tau)$, $\hat{\beta}_2(\tau)$ are, respectively, estimators for the full panel, the first part of the panel and the second part of the panel. The split panel estimator under the Dhaene and Jochmans (2015) approach has the same asymptotic variance as the original $\hat{\beta}$ estimator and allows for more reliable inference under short panels.

5. Quantile-independent fixed effects

5.1. The model

The locational shift model assumed that fixed effects do not vary across quantiles. The model is given in (Koenker, 2004) as:

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

where the function $\tau\mapsto X_{it}'\beta(\tau)$ is strictly increasing in τ, U_{it} is uniformly distributed on $[0,1]$ and does not depend on (X_{it},α_i). Here X_{it} do not include the constant term. Individual effects α_i as n additional unknown parameters.

5.2. A simple estimator

It should be noted that estimation of quantile-dependent individual effects in the Koenker (2004) model lead to computational burden owing to piece-wise character of the quantile objective function W_n. This supports the cause for simplifying the computation procedure.

One simple estimator was proposed in (Canay, 2011) for a more restricted the model with quantile-independent fixed effects[4]:

$$Y_{it} = X_{it}'\beta(U_{it}) + \beta_0(U_{it}) + \alpha_i, \quad i = 1,\ldots,n, \ t = 1,\ldots,T, \tag{3}$$

where the function $\tau \mapsto X_{it}'\beta(\tau) + \beta_0(\tau)$ is strictly increasing in τ, U_{it} is uniformly distributed on $[0, 1]$ and does not depend on (X_{it}, α_i). Here the identification condition $\mathrm{E}[\alpha_i] = 0$ is assumed. Individual effects α_i as n additional unknown parameters. At the first step, the approach uses a $\sqrt{nT}$ consistent estimator (e.g. the within estimator) to consistently estimate fixed effects. At the second step, the pooled version of the panel data quantile regression model is applied to the dependent variable cleared of the estimated fixed effects.

Formally, at the first stage, a $\sqrt{nT}$ consistent estimator $\hat{\beta}_\mu$ of $\beta_\mu \equiv \mathrm{E}[\beta(U_{it})]$ is used to compute

$$\hat{\alpha}_i \equiv \frac{1}{T}\sum_{t=1}^{T}\left(Y_{it} - X_{it}'\hat{\beta}_\mu\right).$$

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

$$\hat{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\rho_\tau\left(\hat{Y}_{it} - X_{it}'\beta - \beta_0\right).$$

5.3. Critique of the simple estimator

Canay (2011) claimed that under $n/T^s \to 0$ (where $s > 1$), $\sqrt{nT}\left(\hat{\beta} - \beta\right)$ has a limiting normal distribution with zero mean and an asymptotic variance matrix. However, Besstremyannaya and Golovan (2019) showed that the condition $n/T^s \to 0$ is insufficient for guaranteeing the existence of a limiting distribution and the absence of asymptotic bias of the vector of coefficients. Another error of the Canay (2011) estimator deals with the constant term : Besstremyannaya and Golovan (2019) prove that the asymptotic standard error of the intercept is incorrect even under $n/T \to 0$.

5.4. A new simple estimator based on smoothing

To address the problem of the asymptotic bias of the Canay (2011) estimator, Chen and Huo (2021) develop a new estimator for quantile-independent fixed effects specified by (3). It uses another normalization condition $\mathrm{E}[\beta_0(U_{it})] = 0$, retains the first-step of the Canay (2011) procedure

[4] It should be noted that quantile-dependent and quantile-independent fixed effect conditional quantile regression models, e.g. of Koenker (2005) and of Canay (2011), appear as nested. Yet, to the best of our knowledge there is not formal statistical test to establish the need of using a more general specification with quantile-dependent fixed effects. Arguably, a weak version of such test would be use the Wald statistics for testing the equality of coefficients for individual effects in each pair of conditional quantile regressions for different values of τ.

but modifies the second by introducing the smoothing technique of Galvao and Kato (2016) and it enables to reduce the asymptotic bias of the estimator in case of short panels. Specifically,

$$\hat{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\hat{Y}_{it} - X_{it}'\beta\right)\left(\tau - G\left(\hat{Y}_{it} - X_{it}'\beta\right)/h\right),$$

where $G(u) = \int_u^{\infty} K(v)\,dv$ is a smoothed analogue of the step function $\mathrm{I}(u<0)$, $K(v)$ is a kernel function, and h is a bandwidth.

6. Simulations

This section uses simulations to demonstrate the problems due to the asymptotic bias of the estimates obtained in quantile regression with short panels. Specifically, we examine the asymptotics for the smoothed quantile regression estimator of Galvao and Kato (2016) and the bias-corrected estimator, based on the approach of Dhaene and Jochmans (2015).

The data generating process is taken from Galvao and Kato (2016):

$$Y_{it} = \alpha_i + X_{it} + (1 + 0.5X_{it})U_{it}, \quad X_{it} = 0.3\alpha_i + Z_{it},$$
$$Z_{it} \sim \text{i.i.d. } \chi^2(3), \quad \alpha_i \sim \text{i.i.d. } U[0,1], \quad U_{it} \sim \text{i.i.d. } N(0,1).$$

The model becomes $Q_\tau(Y_{it} \mid X_{it}) = \alpha_i(\tau) + \beta(\tau) X_{it}$, where $\alpha_i(\tau) = \alpha_i + \Phi^{-1}(\tau)$ and $\beta(\tau) = 1 + 0.5\Phi^{-1}(\tau)$. We simulate 250 samples of four different panels: $n = 75$, $T = 48$; $n = 150$, $T = 24$; $n = 300$, $T = 12$; $n = 600$, $T = 6$. The panels differ in their length, but contain the total number of observations equal to 3600. So the effect of the total size of the sample in longitudinal dataset may be neglected when we compare the results across the four panels. The slope $\beta(\tau)$ is estimated for $\tau \in \{0.25, 0.5, 0.75\}$.

We focus on three criteria: the bias of the estimator, the ratio of the standard error of the estimator to the true standard error, the *z*-statistic defined as $z = \dfrac{\hat{\beta}(\tau) - \beta(\tau)}{\mathrm{se}\left(\hat{\beta}(\tau)\right)}$. (The *z*-statistic is employed to show the behavior of the estimator around the its true value.)

Table 1 and Fig. 1 show the performance of the non-modified smoothed quantile regression estimator by Galvao and Kato (2016). For each of the four panels and each of the three quantiles Table 1 gives the values of the bias of the estimator, the expected value of the ratio of the standard error of the estimator to the true standard error, expected value of the *z*-statistic, probability that the 95% confidence interval covers the true value of $\beta(\tau)$.

As may be revealed from Table 1, in each sample and in each quantile, the bias of the estimator is the smallest under the largest length of panel. The bias of the estimator increases with the increase of the ratio n/T, which is particularly observed at quantiles 0.25 and 0.75.

The expected value of the ratio of the standard error of the estimator to the true standard error is in the range of [0.92, 1.03]. So the standard error of the estimator is a good approximation of the true standard error.

Yet, the expected value of the *z*-statistic differs from zero: for instance, the values are in the range of [0.08, 1.7] for quantile 0.25. The probability that the 95% confidence interval covers the true value of $\beta(\tau)$ is close to 0.95 in case of the median regression and in case of long panels and small ratio of n/T . So the difference between the expected value of the z-statistic and zero may be considered negligible. Yet, the difference is statistically significant for panels with large n/T. When the length of panel is 6, for instance, the coverage probability of the 95% confidence interval is only 0.58–0.60 in regressions for quantiles 0.25 and 0.75. So incorrect conclusions are likely to result from the standard inference procedures in case of such short panel.

Table 1. Distribution of $\beta(\tau)$ for different size of panel:
Non-modified smoothed quantile regression estimator by Galvao and Kato (2016)

	$n = 75, T = 48$	$n = 150, T = 24$	$n = 300, T = 12$	$n = 600, T = 6$
		$\tau = 0.25, \beta(\tau) = 0.6628$		
$\text{bias}\left(\hat{\beta}(\tau)\right)$	0.002	0.009	0.021	0.056
$\text{E}\left[\text{se}\left(\hat{\beta}(\tau)\right)/\sigma\left(\hat{\beta}(\tau)\right)\right]$	0.959	0.956	0.970	0.938
$\text{E}\left[z_{\hat{\beta}(\tau)}\right]$	0.080	0.277	0.612	1.671
$\text{P}\left(\left\lvert z_{\hat{\beta}(\tau)}\right\rvert < z_{0.975}\right)$	0.904	0.924	0.892	0.596
		$\tau = 0.5, \beta(\tau) = 1$		
$\text{bias}\left(\hat{\beta}(\tau)\right)$	–0.003	–0.006	–0.003	–0.005
$\text{E}\left[\text{se}\left(\hat{\beta}(\tau)\right)/\sigma\left(\hat{\beta}(\tau)\right)\right]$	0.953	0.915	0.914	1.033
$\text{E}\left[z_{\hat{\beta}(\tau)}\right]$	–0.099	–0.183	–0.097	–0.132
$\text{P}\left(\left\lvert z_{\hat{\beta}(\tau)}\right\rvert < z_{0.975}\right)$	0.940	0.924	0.936	0.944
		$\tau = 0.75, \beta(\tau) = 1.3372$		
$\text{bias}\left(\hat{a}(\hat{o})\right)$	–0.006	–0.010	–0.021	–0.060
$\text{E}\left[\text{se}\left(\hat{\beta}(\tau)\right)/\sigma\left(\hat{\beta}(\tau)\right)\right]$	1.007	1.022	0.976	0.936
$\text{E}\left[z_{\hat{\beta}(\tau)}\right]$	–0.217	–0.299	–0.587	–1.778
$\text{P}\left(\left\lvert z_{\hat{\beta}(\tau)}\right\rvert < z_{0.975}\right)$	0.948	0.920	0.896	0.584

Figure 1 supports the conjecture about the shifted distribution of the z-statistic. Note that the inference procedures in panel data quantile regression assume that the asymptotic distribution of the z-statistic is standard normal, so the actual distribution of the z-statistic should be close to standard normal. Figure 1 plots the standard normal distribution and distribution of the z-statistic for various values of the ratio n/T in our simulations. As may be revealed from Fig. 1, the larger the ratio n/T , the larger is the difference between the distribution of the z-statistic and the standard normal distribution.

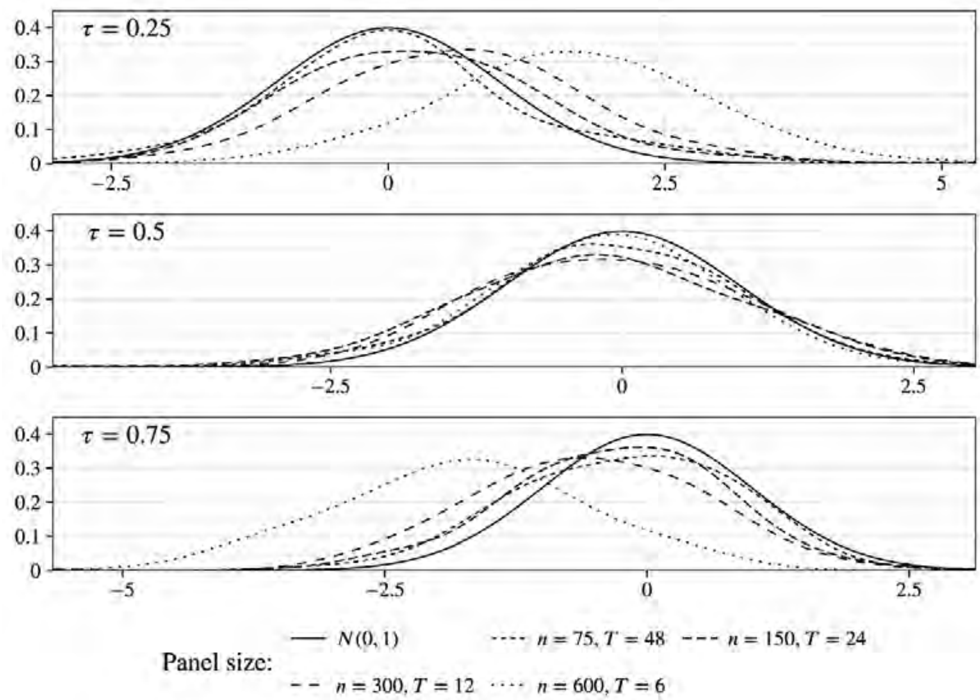

Fig. 1. Kernel density estimates for the z-statistic for $\beta(\tau)$ for different size of panel: Non-modified smoothed quantile regression estimator by Galvao and Kato (2016)

Table 2 and Fig. 2 show the results of similar computations for the bias-corrected smoothed quantile regression estimator according to the approach of Dhaene and Jochmans (2015). The bias becomes negligible: the absolute value of the bias is the range of [0.001, 0.004]. The fact supports the claim that the estimator can be used for shorter panels.

The expected value of the ratio of the standard error of the estimator to the true standard error is in the range of [0.86, 0.93] for panels longer than 6. The value drops to 0.82 at quantiles 0.75 and to 0.85 at quantile 0.25 when the length of panel is only 6. So inference is still not perfect, and more observations are required for better asymptotics.

The expected value of the z-statistic does not differ from zero as much as it did in case of non-modified estimator. The coverage probability of the 95% confidence intervals is rather good. The worst values, which correspond to panel of length 6, are tolerable: 0.89 at quantile 0.25 and 0.87 at quantile 0.75.

As may be revealed from Fig. 2, the z-statistic does not shift as much as in case with the non-modified estimator in short panels.

7. Conclusion

The desire to capture heterogeneity in the response of the dependent variable to covariates often forces empiricists to employ panel data quantile regression models. Very often practitioners forget the limitations of their datasets in terms of the sample size n and the length of panel T. Yet, quantile

Table 2. Distribution of $\beta(\tau)$ for different size of panel: Bias-corrected smoothed quantile regression estimator according to the Dhaene and Jochmans (2015) approach

	$n = 75, T = 48$	$n = 150, T = 24$	$n = 300, T = 12$	$n = 600, T = 6$
		$\tau = 0.25, \beta(\tau) = 0.6628$		
$\text{bias}\left(\hat{\beta}(\tau)\right)$	–0.001	0.003	0.003	0.002
$\text{E}\left[\text{se}\left(\hat{\beta}(\tau)\right)/\sigma\left(\hat{\beta}(\tau)\right)\right]$	0.931	0.895	0.892	0.852
$\text{E}\left[z_{\hat{\beta}(\tau)}\right]$	–0.008	0.107	0.114	0.059
$\text{P}\left(\left\lvert z_{\hat{\beta}(\tau)}\right\rvert < z_{0.975}\right)$	0.884	0.916	0.900	0.888
		$\tau = 0.5, \beta(\tau) = 1$		
$\text{bias}\left(\hat{\beta}(\tau)\right)$	–0.003	–0.006	–0.004	–0.004
$\text{E}\left[\text{se}\left(\hat{\beta}(\tau)\right)/\sigma\left(\hat{\beta}(\tau)\right)\right]$	0.899	0.863	0.864	0.948
$\text{E}\left[z_{\hat{\beta}(\tau)}\right]$	–0.106	–0.195	–0.113	–0.119
$\text{P}\left(\left\lvert z_{\hat{\beta}(\tau)}\right\rvert < z_{0.975}\right)$	0.928	0.904	0.920	0.928
		$\tau = 0.75, \beta(\tau) = 1.3372$		
$\text{bias}\left(\hat{\beta}(\tau)\right)$	–0.004	–0.003	–0.003	–0.004
$\text{E}\left[\text{se}\left(\hat{\beta}(\tau)\right)/\sigma\left(\hat{\beta}(\tau)\right)\right]$	0.938	0.948	0.888	0.821
$\text{E}\left[z_{\hat{\beta}(\tau)}\right]$	–0.137	–0.103	–0.070	–0.136
$\text{P}\left(\left\lvert z_{\hat{\beta}(\tau)}\right\rvert < z_{0.975}\right)$	0.936	0.912	0.920	0.868

regression requires large samples, long panels and small value of the ratio n/T (Kato et al., 2012). At the same time, according to our review of empirical literature which used the Canay (2011) estimator, 45% papers had the value of T less than 10 and only 6 papers out of 81 had a ratio of n/T less than 1 (Besstremyannaya, Golovan, 2019, Table 9).

This paper touched upon the problems of the asymptotic bias of estimators in the fixed effect conditional quantile regressions in absence of long panels. The Galvao and Kato (2016) approach which smooths the quantile loss function and the correction of bias according to the methodology of Dhaene and Jochmans (2015) may be used as helpful tools for reducing the asymptotic bias in short panels.

Acknowledgements. The paper was prepared in the framework of the HSE University's Basic Research Program. We are grateful to Jan Magnus, Anatoly Peresetsky and organizers of the 3rd Applied Econometrics Workshop (HSE University, April 13–14, 2021).

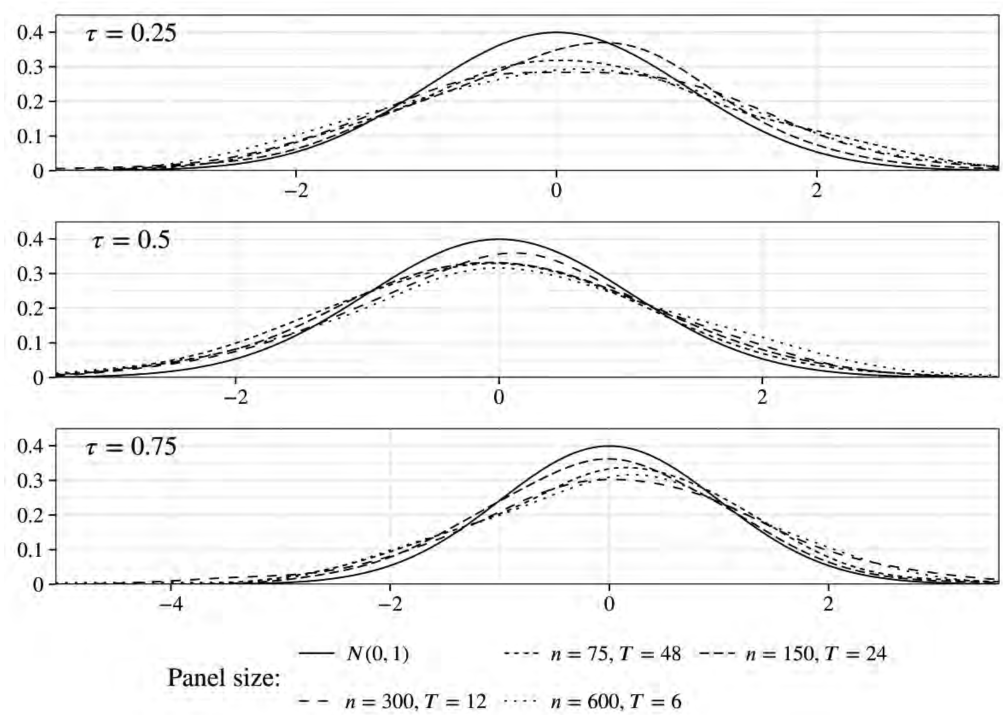

Fig. 2. Kernel density estimates for the z-statistic for $\beta(\tau)$ for different size of panel: Bias-corrected smoothed QR estimator according to the Dhaene and Jochmans (2015) approach

References

Besstremyannaya G. (2017). Heterogeneous effect of the global financial crisis and the Great East Japan Earthquake on costs of Japanese banks. *Journal of Empirical Finance*, 42, 66–89.

Besstremyannaya G., Golovan S. (2019). Reconsideration of a simple approach to quantile regression for panel data. *The Econometrics Journal*, 22 (3), 292–308.

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

Chen L., Huo Y. (2021). A simple estimator for quantile panel data models using smoothed quantile regressions. *The Econometrics Journal*, 24 (2), 247–263.

Dhaene G., Jochmans K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82 (3), 991–1030.

Galvao A. F., Kato K. (2016). Smoothed quantile regression for panel data. *Journal of Econometrics,* 193 (1), 92–112.

Harding M., Lamarche C. (2016). Penalized quantile regression with semiparametric correlated effects: An application with heterogeneous preferences. *Journal of Applied Econometrics*, 32 (2), 342–358.

Kato K., Galvao A. F., Montes-Rojas G. V. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170 (1), 76–91.

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

Koenker R. (2005). *Quantile regression*. Cambrdige University Press.

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

Koenker R., Machado J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94 (448), 1296–1310.

Li H., Lindsay B. G., Waterman R. P. (2003). Efficiency of projected score methods in rectangular array asymptotics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65 (1), 191–208.

Liu C., Laporte A., Ferguson B. (2007). The quantile regression approach to efficiency measurement: Insights from Monte Carlo simulations. *HEDG Working Paper* 07/14.

Machado J. A. F., Parente P. M., Santos Silva J. M. C. (2011). QREG2: Stata module to perform quantile regression with robust and clustered standard errors. *Statistical Software Components*, S457369. Boston College Department of Economics.

Machado J. A. F., Santos Silva J. M. C. (2018). XTQREG: Stata module to compute quantile regression with fixed effects. *Statistical Software Components*, S458523. Boston College Department of Economics.

Machado J. A. F., Santos Silva J. M. C. (2019). Quantiles via moments. *Journal of Econometrics*, 213 (1), 145–73.

Parente P. M., Santos Silva J. M. C. (2016). Quantile regression with clustered data. *Journal of Econometric Methods*, 5 (1), 1–15.

Wooldridge J. M. (2007). Quantile methods. *Lecture Notes 14, NBER Summer Institute*. https://users.nber.org/~confer/2007/si2007/WNE/Slides8-1-07/slides_14_quantilel.pdf.

Wooldridge J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Received 12.05.2021; accepted 14.07.2021.

Appendix. Estimation in Stata

Here we note several availabilities for estimation of quantile regression with longitudinal data in Stata. Firstly, the Machado et al. (2011) package $qreg2$ implements computation of clustered standard errors according to the Parente and Santos Silva (2016) approach. Additional possibilities of this user-written module deal with the analysis of heteroskedasticity.

Secondly, the Machado and Santos Silva (2019) method for estimating quantile-dependent fixed effects under assumption that the expected value and the standard deviation of the dependent variable are linear functions of covariates is available in the package τ (Machado, Santos Silva, 2018).