European Journal of Statistics and Probability
Vol.8, No.1, pp, 74-81, April 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

PANEL QUANTILE REGRESSION WITH PENALIZED FIXED EFFECTS AND CORRELATED RANDOM EFFECTS

Nwakuya Maureen Tobechukwu
Department of Mathematics/Statistics
University of Port Harcourt, Choba, Rivers State.
State. maureen.nwakuya@uniport.edu.ng

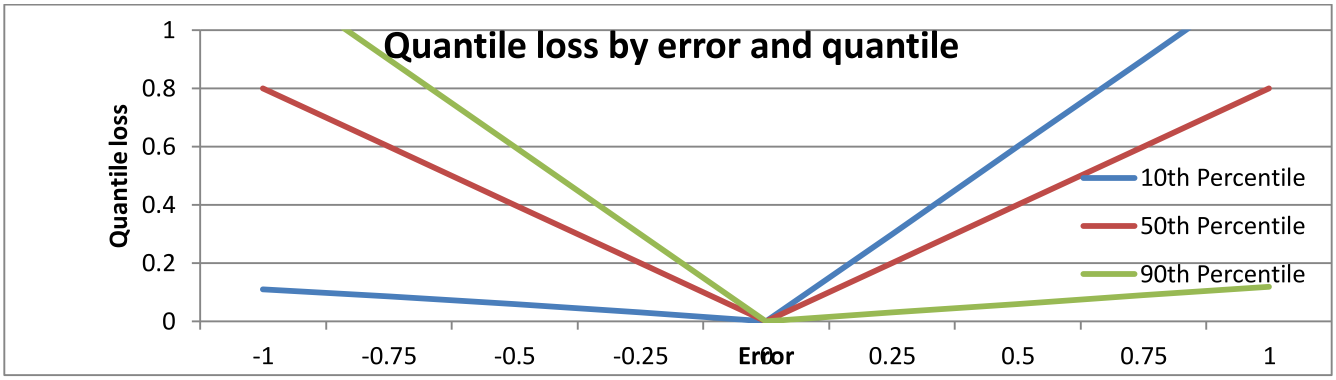
Ijomah Maxwell Azubuike
Department of Mathematics/Statistics
University of Port Harcourt, Choba, Rivers
maxwell.ijomah@uniport.edu.ng

ABSTRACT: *The crucial difficult in estimating covariates effects in panel analysis, is when there is correlation of the unobserved heterogeneity with the covariates and the fact that estimation of conditional mean effects seems potentially limited. Much consideration has not really been given to curb this difficulty especially in the context of quantile regression. In this work Panel Quantile regression was applied in other to investigate the correlated random effects (i.e. effects of the correlation between the covariates and the unobserved heterogeneity) and the penalized fixed effect (i.e. effects after eliminating the unobserved heterogeneity). We employed the use of real data and simulated data sets at different sample sizes. The results showed significant correlated random effect for both covariates in the real data only at the low level (0.25 quantile), but when the unobserved heterogeneity was eliminated both variables were seen to significantly affect the response at the 0.25, 0.5 and 0.75 quantiles of its distribution. The simulation study also confirmed it. We also noticed that as the sample size increases in the simulation study the correlated random effects become insignificant, while the penalized fixed effect and quantile regression effects were evidently significant at all quantiles considered. Comparison of these methods showed that the penalized fixed effect had the least value for both MSE and RMSE. This analysis was done in R environment using the quantreg package.*

KEYWORDS: panel data, unobserved heterogeneity, quantile regression, correlated random effect, penalized fixed effect.

INTRODUCTION

Linear quantile regression analysis is a proven complement to least squares methods (Koenker and Bassett, 1978). Using traditional mean regression will only uncover effects on the mean of the response variable without considering other parts of the distribution of the response variable. Just as linear regressions minimizes the squared-error loss function to predict a single point estimate, quantile regressions minimizes the *quantile loss* in predicting a certain quantile. The quantile loss differs depending on the evaluated quantile, such that more negative errors are penalized more for higher quantiles and more positive errors are penalized more for lower quantiles. The graph below shows how the quantile loss varies with the error, depending on the quantile.



The paper is organized as follows; section 2 presents the linear quantile regression as an extension to linear regression, section 3 provides the panel quantile regression while in section 4, we introduce the models. Section 5 presents analysis with real data while section 6 involves a Monte Carlo simulation to study the finite sample properties of the two-step estimator. Finally, Section 7 concludes.

1. LINEAR QUANTILE REGRESSION

The mean regression model assumes that the conditional mean of y given x is given by;

$$\mu_{y|X} = E(y|x_1, x_2, \dots, x_k) = X^T\beta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \quad (1)$$

Where $X = (1, x_1, \dots, x_k)^T$, and $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$, where $\beta \in \mathbb{R}^p (p = k + 1)$.

By solving via minimizing the L_2-squared distance, one obtains a least squares estimator for $\boldsymbol{\beta}$ from a random sample $(y_i, x_{i1}, x_{i2}, \dots, x_{ik})$ where i=1,2,…,n, given;

$$\hat{\beta}_{LS} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - X_i^T\beta)^2 \quad \text{and } \hat{\mu}_{LS} = \hat{\mu}_{y|X} = X^T\hat{\beta}_{LS} \quad (2)$$

the limitation as stated earlier is that this mean linear regression cannot be extended to non-central locations. This gave rise to the liner quantile regression. Similarly, linear Quantile regression estimates conditional quantiles. The τ^{th} sample quantile, $\hat{\alpha}(\tau)$, solves

$$\min_{\alpha \in \mathbb{R}} \sum_{i=1}^{n} \rho\tau(y_i - \alpha) \quad (3)$$

Where $\rho\tau(\mu) = (\tau - 1\{\mu < 0\})\mu$, this leads to specifying the τ^{th} conditional quantile function which is the inverse of the distribution function, where;

$$F_y(y|X) = \Pr[Y \le y|X] \quad (4)$$

The inverse is;

$$F_y^{-1}(\mu|X) = \{y: \Pr[Y \le y|X] = \mu\} \quad (5)$$

Thus:

$$Q_y(\tau|x) = F_y^{-1}(\tau|X) = X^T\beta(\tau) \quad (6)$$

And $\hat{\beta}(\tau)$ is estimated by solving;

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho\tau(y_i - X_i^T\beta) \quad (7)$$

(i.e minimizing the loss function at the desired quantile), Roger Koenker (2005).

PANEL QUANTILE REGRESSION:

The application of quantile regression methods to panel data analysis has proven to be especially challenging (e.g, Koenker 2005). The marriage of the panel data methodology with that of quantile regression methodology is a very immerging interesting area in Econometrics. Econometric panel data models allow controlling for (time invariant) unobserved individual heterogeneity while quantile regression in turn has the ability to accommodate heterogeneous effects (Abrevaya, 2001). However, the extension of quantile regression framework to panel data analysis is still somewhat limited mostly because of the difficulty in handling individual-specific heterogeneity at set quantiles. Two of the likely most popular methods of controlling for the unobserved heterogeneity are; the individual-specific intercept known as fixed effect that takes into account omitted variables that stay constant over time and the method that allows for heterogeneous responses across individuals to vary over time known as random coefficients models. In a paper by, Abrevaya and Dahl (2008) the extended the "correlated random effects" model by Chamberlain (1984) to a quantile regression framework. Arellano and Bonhomme (2013), developed correlated random effects estimators for panel quantile regression, they extended a method of Wei and Carroll (2009), which was developed for mis-measured regressors, to operationalize their identification results. In this work we assume that there is a correlation between the covariates and the unobserved heterogeneity. We proceeded thus; first we estimate this correlated random effect and secondly, we eliminate the unobserved heterogeneity and obtain a penalized fixed effect and thirdly we obtain the estimates using ordinary quantile regression. The purpose of this study is to employ the Panel quantile regression methodology using three models; Correlated Random effects model, Penalized Fixed Effects model and Quantile regression model in determining the effects of Trade and inflation on the GDP growth for six West-African Countries for the period from 1998 to 2007. The data for this study is a secondary data sourced from R inbuilt data in the Amelia package. The data is made up of 6 panels with seven variables studied over a period of 10 years, with 160 observations. Simulated data sets with two different sample sizes were also employed.

MODEL SETUP:

Describing the model, let i = 1,...,n denote individual units, and let t = 1,...,T denote time periods. The random-effects quantile regression model specifies the τ-specific conditional quantile of an outcome variable Y_{it}, given a sequence of strictly exogenous covariates $X_i = (X'_{i1},...,X'_{iT})'$ and unobserved heterogeneity υ_i, as follows:

$$Q(Y_{it}|X_i,\upsilon_i,\tau) = X_{it}^T\beta(\tau) + \upsilon_i\xi(\tau), for\ \tau\epsilon(0,1) \quad (8)$$

Note that υ_i does not depend on the percentile value τ. Model (8) specifies the conditional distribution of Y_{it} given X_{it} and υ_i. In order to complete the model, we specify the τ^{th} conditional distribution of υ_i given the sequence of covariates X_i as follows:

$$Q(\upsilon_i|X_i,\tau) = X_i^T\zeta(\tau), for\ \tau\epsilon(0,1) \quad (9)$$

Equations (8)-(9) provide a fully specified semiparametric model for the joint distribution of outcomes given the sequence of strictly exogenous covariates, in other to estimate parameters: β (τ), $\xi(\tau)$, and ζ(τ), for all τ, Abrevaya and Dahl (2008). Following Abrevaya and Dahl (2008) and Gamper Rabindran, Khan and Timmins (2008), we viewed the individual specific effect as a correlated random effect which is allowed to be correlated with the observed covariate $X_i = (X'_{i1},...,X'_{iT})'$ and $\{\upsilon_i\}_{i=1}^n$. The correlated random-effects model of Chamberlain (1982, 1984) views the unobservable ν_i as a linear projection onto the observables plus a disturbance, on this Abrevaya and Dahl (2008) extended their work. Considering

European Journal of Statistics and Probability
Vol.8, No.1, pp, 74-81, April 2020
Published by ECRTD-UK
Print ISSN: 2055-0154(Print), Online ISSN 2055-0162(Online)

the representation given by equations (8)-(9), under the assumptions given below, the quantity of interest, $\zeta(\tau)$, is identified for correlated random effects and the empirical criterion function is given by;

$$\left(\hat{\beta}(\tau), \hat{\zeta}(\tau)\right) = \underset{(\hat{\beta}(\tau), \hat{\zeta}(\tau))}{\operatorname{argmin}} \sum \sum \rho\tau_k [Y_{it} - X'_{it}\beta(\tau_k) - X_i^T \zeta(\tau)]$$

Assumptions;

- The function $\tau \longmapsto X'\beta(\tau)$ is assumed to be increasing strictly with $\tau \in (0,1)$.
- The individual specific effect follows a standard uniform distribution conditional on X_i, i.e $v|X \sim uniform(0,1)$

While for penalized fixed effect, following Koenker (2004), we treated $\xi(\tau)$ as fixed parameters to be jointly estimated with $\beta(\tau)$ for q different quantiles, using a penalized estimator

$$\left(\hat{\beta}(\tau), \hat{\xi}(\tau)\right) = \underset{(\hat{\beta}(\tau), \hat{\xi}(\tau))}{\operatorname{argmin}} \sum_{k=1}^{q} \sum_{i=1}^{n} \sum_{t=1}^{T} \rho\tau_k [Y_{it} - X'_{it}\beta(\tau_k) - v_i] + \gamma \sum_{i=1}^{n} |v_i|$$

Where $\rho_\tau(u) = u[\tau - 1(u < 0)]$, $I(.)$ denotes the indicator function and $\gamma \geq 0$ is a penalized parameter that shrinks the $\hat{v}s$ towards a common value. This estimator yields the penalized fixed effect.

DATA ANALYSIS

In this section we present analog estimators, based upon the Correlated Random effects model, Penalized Fixed Effects model and Quantile regression model in determining the effects of Trade and inflation on the GDP growth. We show the quantile regression with bootstrapped standard errors (Koenker, 2005) at 5% significant level. We also consider the conditional quantile coverage for $\tau = 0.25, 0.5, 0.75$. The results of panel quantile regression using real data are presented as below.

Table 1a reports the result of the estimates using correlated random effect (CRE) method at 25%, 50% and 75% quantile.

TABLE 1a: Estimates using real data for Correlated Random Effect method (CRE)

Coefficients	Estimates for each quantile		
	0.25	0.50	0.75
Intercept	-784.46457* (348.2856)	-945.96311 (736.85695)	-1077.7422 (1096.9426)
Trade	12.23579* (5.88907)	15.21511 (11.53662)	21.68483 (16.01675)
Inflation	-8.99504* (4.45523)	-5.96573 (8.44044)	-19.85935 (12.92047)

*sig at 5%, Bootstrap std error in bracket

This Table shows that the correlated effects was only significant at 25% quantile (τ =0.25) for both trade and inflation, that means that the effect of the correlation between the fixed effect and the explanatory variables (inflation and trade) was only significant when GDP is low (i.e. τ = 0.25 quantile).

We then consider the estimation using penalized fixed effect method and the result is shown in Table 1b.

Table 1b: Estimation Results using real data for Penalized Fixed Effect method (PFE)

Coefficients	Estimates for each quantile		
	0.25	0.50	0.75
Intercept	-30073.8167 (22126.6175)	-29883.5618 (22052.7202)	-29813.9062 (22025.0171)
Trade	17.23806* (1.69305)	16.52100* (2.17324)	20.50707* (4.13899)
Inflation	-5.59054* (2.65122)	-6.27131* (3.06682)	-8.67944* (3.15171)

*sig at 5%, Bootstrap std error in bracket

As we may observe from Table 1b, results of penalized fixed effect method show a significant effect of inflation and trade on GDP at all the tails of the distribution of the GDP considered here (i.e., τ = 0.25, 0.5 and 0.75). We can say that the correlated effect even though not significant at τ = 0.5 and 0.75 (i.e. 25% and 75% quantile) as shown in Table 1a, masked the marginal effect of the variables on GDP, but the elimination of this correlation effect using the Penalized Fixed Effect method, exposed the true effect of the variables, as can be seen in this Table 1b. Next is the estimation based on quantile regression method.

Table 1c: Estimation Results using real data for Quantile Regression method (QR)

Coefficients	Estimates for each quantile		
	0.25	0.50	0.75
Intercept	-88.90093 (52.66039)	-23.74875 (69.40818)	-63.20032 (149.82015)
Trade	15.45691* (0.99753)	-4.12272* (2.68376)	22.85117* (2.93843)
Inflation	-3.48632 (2.15363)	16.52007 (1.36027)	-7.69147* (2.67112)

*sig at 5%, Bootstrap std error in bracket

The quantile regression model here can be likened to the pooled regression model in panel data analysis. This table shows a significant effect of trade at 0.25, 0.5 and 0.75 quantiles. Inflation was seen to be significant only at 75% quantile (i.e. τ = 0.75), that means inflation only significantly affects GDP when it is more than average.

SIMULATION RESULTS

To further illustrate the performance of these three estimators we conduct a simulation study at n = 20, 200, 500 and 1000. Tables 2 and 3 summarize the results by reporting percentage bias (%Bias) and mean squared error (MSE) for each estimator considered. The simulated model is;

$$y_{it} = \alpha_{it} + X'_{it} + u_{it}$$

Where i is the cross sectional size and t is the time index, and u is assumed normally distributed.

Table 2: Result of Correlated Random Effect method (CRE), Penalized Fixed Effect method and Quantile Regression method (QR) at different sample sizes and levels of quantile

Samples		CRE			PFE			QR		
20		**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**
	β_0	-9.4252 (11.6713)	-0.2085 (2.5487)	-18.319* (0.1850)	2.7258* (0.4017)	1.9653* (0.7643)	2.2979* (0.3479)	2.4722* (0.8804)	1.4584 (1.0270)	2.1134 (1.1527)
	β_1	6.5958 (6.9208)	0.9815 (1.8488)	11.5821* (0.1103)	-3.1020* (0.5336)	-1.0209 (0.7996)	-0.4276* (0.0913)	-3.1020* (1.1073)	-1.0211 (1.1742)	-0.5139 (0.9696)
200		**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**
	β_0	1.0851* (0.4653)	1.5238* (0.3406)	2.2419 (0.5101)	0.6098 (0.6627)	0.9570 (0.5531)	1.4034 (0.56780)	0.6481 (0.5252)	0.9656* (0.3130)	1.4422* (0.3531)
	β_1	-0.4041* (0.2026)	-0.449* (0.1755)	-0.5771* (0.2377)	-1.8873 * (0.5102)	-0.9377 * (0.2976)	-0.2520 (0.2833)	-1.8939* (0.7061)	-0.9300* (0.3199)	-0.2556 (0.339)
500		**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**
	β_0	1.15064 (0.6463)	1.83547 (0.6879)	2.20521 (0.6829)	0.97520* (0.2005)	1.34086* (0.1985)	1.96346* (0.2302)	0.8727* (0.0689)	1.2408* (0.0561)	1.8508* (0.0744)
	β_1	-0.1807 (0.3631)	-0.3699 (0.4027)	-0.2191 (0.4004)	-2.0335* (0.0953)	-1.1755* (0.0564)	-0.5862* (0.0779)	-2.0475* (0.0883)	-1.1810* (0.0536)	-0.5937* (0.0777)
1000		**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**	**0.25**	**0.50**	**0.75**
	β_0	1.3404* (0.2414)	1.5577* (0.2285)	1.76519* (0.2592)	1.1856* (0.2317)	1.4355* (0.2009)	1.5199* (0.2200)	1.1962* (0.1664)	1.4512* (0.1293)	1.8412* (0.2066)
	β_1	-0.0966 (0.1409)	-0.0584 (0.1423)	-0.1400 (0.1339)	-2.4142* (0.1594)	-1.6372* (0.1312)	-0.5912* (0.1156)	-2.4133* (0.1798)	-1.6291* (0.1455)	-0.8881* (0.2568)

*sig at 5%, Bootstrap std error in bracket

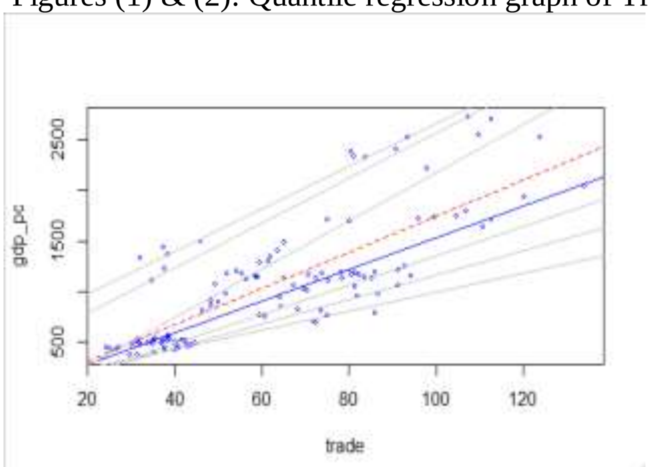
The table 2 above shows that for small samples, correlated significant effects occurred for only 0.75 quantile while PFE is significant for both 0.25 and 0.75 quantiles. For QR method, it has significant effects only for 0.25 quantile. As the sample increased to 200, CRE method shows a significant correlated effect at all the quantiles, PFE shows a significant effect of the β_1 at only 0.25 and 0.50 quantiles while QR method shows a significant effect of β_1 at the 0.25 and 0.50 quantiles like the penalized fixed effect. The table further revealed that the CRE method was insignificant for all selected levels of quantile when the sample increased to 500. However, the PFE and QR methods were found to be significant at all quantiles. Finally, we also considered the estimation result as the samples increased to 1000. The table shows an insignificant correlation effect of the covariate at all the quantiles unlike what was seen at lower sample size. We can infer from this that when sample size is very high/as sample size increases the correlated effect becomes insignificant. The penalized fixed effect and Quantile regression maintained significant effect of the covariates at all parts of the distribution considered in this work.

Table 3: Comparison of Mean Square Error (MSE) and Root Mean Square Error (RMSE) for all selected methods at different sample sizes and levels of quantile

Simulated Data	Estimator	CRE	PFE	QR
N = 20	MSE	19.8396	2.07184*	18.12158
	RMSE	4.45417	1.43939*	4.2569
N= 200	MSE	33.89295	4.124855*	34.50067
	RMSE	5.821765	2.030974*	5.873727
N= 500	MSE	25.16936	2.944731*	25.19814
	RMSE	5.016908	1.716022*	5.019725
N= 1000	MSE	22.83524	2.466775*	8.749531
	RMSE	4.778623	1.570597*	2.957961
REAL DATA	Estimator	CRE	PFE	QR
	MSE	139,693.2	15,596.34*	162,240.7
	RMSE	373.7555	124.8857*	402.7912

A closer look at table 3 above shows that the Penalized Fixed Effect (PFE) method maintained least mean square error and root mean square error in all the selected sample size (both small and large samples) and in both real and simulated data when compared with other methods (i.e. CRE and QR). This shows that Penalized Fixed Effect method is better than Correlated Random Effect and Quantile Regression methods. In order to verify statistical variation of coefficients along innovation conditional distribution, we depict a Figureic display of coefficients of interest. In Figures 1 and 2, we produce separate Figures for each explanatory variable (Trade and Inflation rate) of the estimated coefficient plotted against the quantile q. The doted lines are the OLS point estimates and confidence intervals (these do not vary with the quantile). The second plot shows that the coefficient on knowledge spillovers is positive, with a much larger effect at lower quantiles.

Figures (1) & (2): Quantile regression graph of Trade and Inflation on GDP



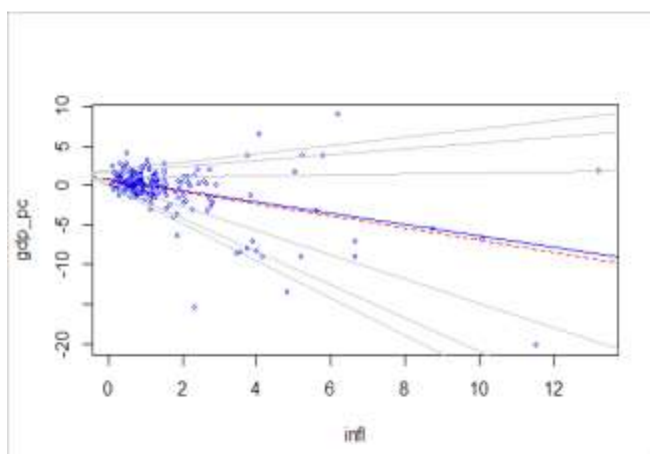


Fig. 1: Quantile Regression graph of Trade on GDP

Fig. 2: Quantile Regression graph of Inflation on GDP

Fig. 1 shows a positive correlation between GDP and Trade while Fig. 2 shows a negative correlation between GDP and Inflation. The figures display that the mean regression didn't capture all tails of the distribution but with the quantile regression captured the effects at all the tails of the distribution. The red dotted line represents the mean regression line, the blue line represents the median regression while the gray lines represent other quantiles (0.05, 0.1, 0.25, 0.75, 0.9 and 0.95).

CONCLUSIONS

The results showed significant correlated random effect for both inflation and trade only at the low level (0.25 quantile) of the GDP but when the unobserved heterogeneity was eliminated both variables were seen to significantly affect GDP at the 0.25, 0.5 and 0.75 quantiles of its distribution. The simulation studies also confirmed this. We also noticed that as the sample size increases in the simulation study the correlated random effects become insignificant, while the penalized fixed effect were evidently significant at all quantiles considered. The quantile regression results were seen to be similar to PFE results. Comparison of these methods showed that the penalized fixed effect had the least value for both MSE and RMSE. But this is not to say that it should be preferred over the CRE, it all depends on the aim of the study.

REFERENCES

Abrevaya, 2001. "The effects of demographics and maternal behavior on the distribution of births outcomes", Empirical Economics 26 (1): 247-257.

Abrevaya, J., Dahl, C., 2008. "The effects of birth inputs on birth weight: evidence from quantile estimation on panel data", Journal of Business & Economic Statistics, 26 (4): 379-397.

Arellano, M., and Bonhomme, S., (2013). "Random effects quantile regression", Mimeo,CEMFI.

Chamberlain, G., (1984)."Panel data", Handbook of Economics 2: 1247-1318 (Z. Griliches & Intriligator, Eds.). Amsterdam: North Holland.

Gamper-Rabindran, S., Khan, S., Timmins, C., 2008. The impact of piped water provision on infant mortality in Brazil: a quantile panel data approach. NBER Working Paper No.14365

Koenker, R. and Bassett, G., (1978). "Regression quantiles", Econometrica 46 (1): 33-50.

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

Koenker, R., (2005). "Quantile Regression", Cambridge University Press, New York.

Wei, Y. and Carroll, R., (2009). "Quantile regression and measurement error", Journal of American Statistical Association 104 (487): 1129-1143.