2019

Equal quantile order for two-stage quantile regression for dynamic panel data models: Monte-Carlo simulation study

Hossameldin Ahmed
hossameldin.ahmed@bue.edu.eg

aya afify Ms
The British University in Egypt, aya.afify@bue.edu.eg

Alaa Ibrahim Prof
The British University in Egypt, alaa.ahmed@bue.edu.eg

Recommended Citation

Ahmed, H., Ibrahim, A., & Mohamed, A. (2019). Equal quantile order for two-stage quantile regression for dynamic panel data models: Monte-Carlo simulation study. Advances and Applications in Statistics, 54(2), 43-56.

Advances and Applications in Statistics
© 2019 Pushpa Publishing House, Prayagraj, India
http://www.pphmj.com
http://dx.doi.org/10.17654/AS...
Volume …, Number …, 2019, Pages … **ISSN: 0972-3617**

EQUAL QUANTILE ORDER FOR TWO-STAGE QUANTILE REGRESSION FOR DYNAMIC PANEL DATA MODELS: MONTE-CARLO SIMULATION STUDY

Hossameldin A. Afify, Alaa A. Abdel-Aziz and Aya A. Mohamed

Department of Economics
Faculty of Business Administration, Economics
and Political Science
The British University in Egypt
Egypt

Abstract

The arising issue of gaining biased estimators clearly occurs in case of estimating dynamic panel data (DPD) models by quantile regression (QR) due to the addition of one or more of lagged value(s) for endogenous variable in DPD models as exogenous variable(s). A newly unique approach for reducing the resulting bias is theoretically and empirically proposed through the method of estimation of two-stage quantile regression for dynamic panel data (TSQRDPD). This new methodological framework of TSQRDPD is essentially dependent on evaluating the two stages by QR at a pre-specified rank of quantile. Besides, bias trait is investigated through a Monte-Carlo simulation under several quantile orders for the two stages, and panels' dimensions. Monte-Carlo simulation results provide a reliable

Received: May 18, 2019; Accepted: June 28, 2019

2010 Mathematics Subject Classification: 91G70.

Keywords and phrases: quantile regression, two-stage quantile regression, dynamic panel data model, endogeneity, biasedness, instrumental variable, Monte-Carlo simulation.

evidence that the new methodological technique of estimation (TSQRDPD) is capable of noticeably reducing estimators' biasedness.

1. Introduction

Estimating dynamic panel data (DPD) models by quantile regression (QR) approach has been increasingly considered in the literature over the last decade. It is a type of models which capture the dynamic effect through the inclusion of lagged value(s) for the response variable. As a result of that, it allows defining the dynamic effect for both the dimensions time and cross sections [1]. There is an individual main approach for dealing with the problem of endogeneity in quantile regression in dynamic panel data (QRDPD) models; it relies on adding instrumental variable(s) (IV) that is/ are not correlated with innovations which are similar to the estimation framework proposed by [2]. Galvao [3] utilized the process of [4] along with lagged instruments to minimize the bias in DPD with fixed effects (FE) models. Elhoussainy et al. [5] proposed a new framework for handling the issue of endogeneity in QRDPD which is by the implication of two-stage quantile regression for dynamic panel data (TSQRDPD) models. TSQRDPD works mainly on estimating the endogenous variable (lagged dependent variable) by its lagged values in the first stage. Afterwards, it plugs the resulted new estimated values in the original model as the second stage of estimation. Afify and Abdel-Aziz [6] showed that TSQRDPD is noticeably better than other estimation methods through different panel sizes when the variables of study are generated from different distributions at different levels of quantiles for both the stages. It should be taken into consideration that both the stages are estimated by quantile regression to maintain the level of robustness.

2. Estimation of TSQRDPD

For the purpose of handling the problem of endogeneity in DPD models due to the inclusion of lagged value of dependent variable, we extensively propose a new frame for estimating DPD models with fixed effects through two-stage quantile regression method.

TSQRDPD provides a much wider scope for interpreting the results at several quantiles of variables of study. Hence, the parameters along with their estimators are evaluated at different quantiles. Therefore, not only the mean relationships are presented but also the larger scale is provided according to each quantile level.

Several recent studies considered estimating the first step by different estimation methods other than quantile regression. However, our main concern is to estimate both the stages by using quantile regression which is dissimilar to the majority of the literature. Regarding our new approach for estimating TSQRDPD models, we will consider evaluating the parameters of interest at same levels of quantiles to assure the robustness.

The main aim is to estimate the structural parameter 1 which is presented as follows:

$$y_{it} = Y_{i,t-1}\Gamma_0 + X_{1it}\Lambda_0 + U_{it}, \quad i = 1, \ldots, N, \quad t = 2, \ldots, T. \tag{1}$$

It can also be represented in compressed form as follows:

$$\begin{aligned} y_{it} &= Z_{it}\omega_0 + U_{it}, \\ U_{it} &= \mu_i + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 2, \ldots, T. \end{aligned} \tag{2}$$

For ith cross section and tth time period, $[y_{it}, Y_{i,t-1}]$ is an $NT \times (\mathcal{W} + 1)$ matrix of endogenous variables, y_{it} is the variable of study, $Y_{i,t-1}$ is the first lagged value of the dependent variable, X'_{1it} is an $NT \times \mathcal{L}_1$ of exogenous variables, the set of explanatory variables that are included in the model $Z_{it} = [Y_{i,t-1}, X_{1it}]$, the structural parameter $\omega_0 = (\Gamma'_0, \Lambda'_0)'$, ε_{it} is an $NT \times 1$ vector, μ_i is a vector of $I_N \times 1$. X_{2it} is an $NT \times \mathcal{L}_2$ of exogenous variables that are not involved in the model in (1) and used to estimate $Y_{i,t-1}$. It should be taken into consideration that $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$.

X_{2it} is an instrumental variable that can be used in the first stage to estimate the endogenous variable $Y_{i,t-1}$. According to Koenker (2005), $Q_\theta(\cdot)$ is the quantile function at (θ) conditional on Y_{it}. The problem of endogeneity would exist if $Q_\theta(U_{it} \mid Y_{it}) \neq Q_\theta(U_{it})$. This inequality between $Q_\theta(U_{it} \mid Y_{it})$ and $Q_\theta(U_{it})$ can represent an explanation of endogeneity in quantile regression. τ and θ are quantile levels for the first stage and the second stage, respectively.

The *first stage* is to estimate the endogenous variable Y_{it-1}. By collecting the observations over time periods and cross sections, the first stage model can be proposed in the matrix form as follows:

$$Y_{it-1} = X_{it}\Phi_0 + V_{it-1}, \quad i = 1, \ldots, N, \quad t = 2, \ldots, T. \tag{3}$$

The model in (3) is considered to be specified correctly, where $Y_{i,t-1}$ is the lagged value of dependent variable and $X_{it} = [X_{1it}, X_{2it}]$ which is the matrix of size $NT \times \mathcal{L}$. Φ_0 is the matrix of the first stage parameters that are unknown of size $\mathcal{L} \times \mathcal{W}$. V_{it} is an $NT \times \mathcal{W}$ matrix of observations of error term over N and T. The *second stage* is to estimate the dependent variable y_{it}. By aggregating the observations over time periods and cross sections, the second stage model can be proposed as follows:

$$y_{it} = X_{it}\varphi_0 + v_{it}, \tag{4}$$

where $\left[\Phi_0, \begin{pmatrix} I_{\mathcal{L}_1} \\ 0 \end{pmatrix}\right]\omega_0 = \mathbb{N}(\Phi_0)\omega_0 = \varphi_0$ and the error term in (4) can be modeled as follows: $v_{it} = V_{it}\Gamma_0 + U_{it}$. Φ_0 which is obtained from the first stage will be included in φ_0 which represents the structural parameter in the second stage as shown in (4). The TSQRDPD estimator $\hat{\omega}_0 = (\hat{\Gamma}_0', \hat{\Lambda}_0')'$ for ω_0 can be obtained by solving the following minimization argument:

$$\min_{\omega_0} \mathfrak{D}_{NT}(\omega_0, \hat{\varphi}_0, \hat{\Phi}_0, \theta, \tau)$$

$$= \min_{\omega_0} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\theta(\tau y_{it} + (1-\tau)x'_{it}\hat{\varphi}_0 - x'_{it}\mathbb{N}(\hat{\Phi}_0)\omega_0), \quad (5)$$

where y_{it} and x'_{it} are the observations of the dependent and independent variables, respectively, for ith cross section and tth time period. Both τ and θ are quantile levels for the first stage and the second stage, respectively. They range from 0 to 1. $\rho_\theta(g)$ is defined to be the check function, where $\rho_\theta : R \to R^+$. The orders of quantiles are assumed to take the same value for consistency purposes $\rho_\theta(g) = g\psi_\theta(g)$, where $\psi_\theta(g) = \theta - I_{(g \leq 0)}$ and $I_{(\cdot)}$ is the indicator function defined by [6]. The formulation of the dependent variable in two stages can also take the form $\tau y_{it} + (1-\tau)x'_{it}\hat{\varphi}_0$. The formulation in (5) can be partitioned into two minimization arguments to obtain the two stages' estimators $\hat{\Phi}$ and $\hat{\varphi}$ based on the generalization provided by [7] as follows:

$$\min_{\varphi} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\theta(y_{it} - x_{it}\varphi)$$

$$= \min_{\Phi_r} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\theta(y_{rit} - x_{it}\Phi_r), \quad r = 1, 2, \ldots, \mathcal{W}, \quad (6)$$

where φ and Φ_r are $\mathcal{W} \times 1$ vectors, Y_{rit} is rth in (i, t)th elements of Y, Φ_r is the rth element in Φ which is obtained in the first stage of the estimation for the corresponding estimated endogenous variable.

3. Monte-Carlo Simulation Study

This section is devoted to present the undertaken steps to conduct simulation study. The numbers of cross sections are 400 and 600. The lengths of time series are assumed to be 5, 25 and 50. Both dependent and independent variables are generated from the following probability distributions of Chi (10), and t (3). The number of replications is set to be 10000 times which is greater than any simulation of study of the number of replications. The larger number of replications is determined to seek more accuracy for bias results. The quantile ranks are chosen to be 0.25, 0.50 and 0.75. A comparative study on bias values performance of exogenous variable coefficients is conducted. The values are obtained from the following: QRDPD model:

$$y_{it} = \beta_0 + \rho_1 y_{it-1} + \beta_1 x_{1i} + \Theta_{1it}, \tag{7}$$

where $i = 1, 2, ..., N$, $t = 2, 3, ..., T$ and Θ_{it} is the error term.

QRDPDFE model:

$$y_{it} = \beta_0 + \rho_1 y_{it-1} + \beta_1 x_{1i} + \mu_i + \mathfrak{H}_{it}, \tag{8}$$

where $i = 1, 2, ..., N$, $t = 2, 3, ..., T$, μ_i is the individual effect and $\mathfrak{H}_{it}$ is the error term. Quantile regression for dynamic panel data with fixed effects and instrumental variable (QRDPDFEIV) model:

$$y_{it} = \beta_0 + \rho_1 y_{it-1} + \rho_2 y_{it-2} + \beta_1 x_{1i} + \mu_i + \mathfrak{G}_{it}, \tag{9}$$

where $i = 1, 2, ..., N$, $t = 3, 4, ..., T$, μ_i is the individual effect, y_{it-2} is the instrumental variable for y_{it-1} and $\mathfrak{G}_{it}$ is the error term. TSQRDPD model, where the first stage model is as follows:

$$y_{it-1} = \rho y_{it-2} + \mathfrak{A}_{it-1}, \tag{10}$$

where $i = 1, 2, ..., N$, $t = 3, 4, ..., T$, y_{it-2} is instrumental variable for y_{it-1} and $\mathfrak{A}_{it-1}$ is the error term of the first stage, and the second stage model is given as:

$$y_{it} = \beta_0 + \rho_1 \hat{y}_{it-1} + \beta_1 x_{1it} + \mu_i + \mathfrak{D}_{it}, \tag{11}$$

where $i = 1, 2, ..., N$, $t = 2, 3, ..., T$, μ_i is the individual effect, $\hat{y}_{it-1}$ is the estimated value of y_{it-1} predicted by y_{it-2} and $\mathfrak{D}_{it}$ is the error term of the second stage. Both relative efficiency (RE) of mean squared error and relative bias (RB) are utilized to assess the performance of the bias of new method and compare it with the obtained value from the base method which is QRDPD. Parameter value is selected to be $\beta_1 = 0.6$ which represents the arithmetic mean of he values chosen by [8, 9]. The absolute value of ρ is restricted to be less than one for stationarity purposes.

4. Results and Discussion

This section is dedicated to present the main remarks that are obtained from simulation study results.

Table 1. RE and RB performance where X and $Y \sim t\ (3)$ where $n = 400$ and $t = 5, 25$ and 50

t	θ, τ	RE and RB	Quantile regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
5	0.25	RE	66.4110	37.1430	15.5360	5.7770
		RB	59.7058	36.7149	10.2140	0.8078
	0.5	RE	66.3540	36.6700	15.4750	5.7180
		RB	59.3677	36.6035	10.1956	0.7966
	0.75	RE	66.1560	36.1390	15.4410	5.3280
		RB	58.5750	36.3655	10.1796	0.7876
25	0.25	RE	66.1910	36.7990	15.5090	5.6750
		RB	58.8906	35.4791	10.1515	0.8054
	0.5	RE	66.0590	36.5330	15.4070	5.5910
		RB	58.2815	34.9805	10.1372	0.7687
	0.75	RE	65.9940	36.0610	15.4050	5.3110
		RB	57.9027	34.5723	10.1242	0.7353

50	0.25	RE	66.1270	36.7160	15.4430	5.6440
		RB	57.8755	35.3951	10.1358	0.7505
	0.5	RE	66.0100	36.4040	15.3860	5.5370
		RB	56.7150	34.8112	10.1328	0.7434
	0.75	RE	65.9500	35.9680	15.3800	5.3010
		RB	55.8456	34.4955	10.1225	0.7259

Table 2. RE and RB performance where X and $Y \sim t\ (3)$ where $n = 600$ and $t = 5, 25$ and 50

t	θ, τ	RE and RB	Quantile regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
5	0.25	RE	65.9760	36.5500	15.4110	5.5980
		RB	56.7895	34.2620	10.1279	0.7404
	0.5	RE	65.7830	36.2260	15.3700	5.4280
		RB	56.2836	34.0429	10.1170	0.7246
	0.75	RE	65.3970	35.8200	15.3330	5.2820
		RB	54.0080	32.7870	10.0819	0.7116
25	0.25	RE	65.6780	36.3520	15.3690	5.5890
		RB	56.1805	33.9643	10.1007	0.7190
	0.5	RE	65.3060	36.1000	15.3340	5.4230
		RB	55.8496	32.2961	10.0850	0.6371
	0.75	RE	65.1680	35.6390	15.2980	5.2680
		RB	53.7141	31.4255	10.0506	0.6076
50	0.25	RE	65.6280	36.3030	15.3600	5.5800
		RB	55.8967	33.3101	10.0415	0.6383
	0.5	RE	65.2530	35.9800	15.3240	5.4130
		RB	55.3711	32.1505	10.0285	0.6185
	0.75	RE	64.9850	35.6250	15.2790	5.2580
		RB	53.7084	30.8557	10.0058	0.6055

Table 3. RE and RB performance where X and $Y \sim \text{Chi}(8)$ where $n = 400$ and $t = 5, 25$ and 50

t	θ, τ	RE and RB	Quantile regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
5	0.25	RE	62.7430	34.0820	14.8730	4.8520
		RB	59.6882	36.8482	10.1882	0.8076
	0.5	RE	62.0380	33.4660	14.7050	4.8280
		RB	59.1872	36.5472	10.1758	0.8073
	0.75	RE	61.5430	33.4490	14.6750	4.5630
		RB	59.1252	35.8466	10.1718	0.7918
25	0.25	RE	62.6860	34.0300	14.7940	4.8160
		RB	58.5833	35.7727	10.1662	0.7737
	0.5	RE	61.9330	33.2610	14.6680	4.7710
		RB	57.5772	33.7276	10.1560	0.7726
	0.75	RE	61.1150	33.1950	14.6550	4.5550
		RB	56.8607	33.4594	10.1364	0.7487
50	0.25	RE	62.5320	33.9830	14.7750	4.7650
		RB	57.8533	35.4674	10.1639	0.7387
	0.5	RE	61.8890	33.1980	14.6290	4.7560
		RB	57.1827	33.2829	10.1416	0.7072
	0.75	RE	61.0250	33.1780	14.6090	4.5440
		RB	56.8400	31.7687	10.1176	0.6717

Table 4. RE and RB performance where X and $Y \sim \text{Chi}(8)$ where $n = 600$ and $t = 5, 25$ and 50

t	θ, τ	RE & RB	Quantile regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
5	0.25	RE	62.3190	33.9530	14.7030	4.7460
		RB	56.4840	34.5520	10.1142	0.6977
	0.5	RE	61.7540	33.0940	14.5680	4.7140
		RB	55.7487	31.9332	10.1061	0.6886
	0.75	RE	60.8450	32.9860	14.5470	4.4920
		RB	55.3491	31.3120	10.0990	0.6657

25	0.25	RE	62.1090	33.8130	14.6420	4.7160
		RB	55.7894	32.9370	10.0940	0.6794
	0.5	RE	61.4730	32.9070	14.5400	4.7000
		RB	54.6157	31.5077	10.0632	0.6508
	0.75	RE	60.6520	32.7400	14.5110	4.4790
		RB	53.0593	30.7495	10.0534	0.6050
50	0.25	RE	62.0540	33.7630	14.5520	4.7040
		RB	54.3190	32.0539	10.0752	0.6597
	0.5	RE	61.4350	32.8920	14.5050	4.5940
		RB	53.7000	31.4690	10.0631	0.6191
	0.75	RE	60.4460	32.5830	14.4920	4.4130
		RB	51.1598	30.5315	10.0162	0.6025

In reference to the obtained results in Tables 1-4, for the behavior of RE and RB over the methods of estimation of (QRDPD, QRDPDFE, QRDPDFEIV, TSQRDPD), and for every combination of t and n, the performance of RE and RB have a declining tendency from an estimation method to another where TSQRDPD has the least bias. Moreover, in each of the estimation methods, and for the same order of quantile and when t value increases for the same sizes of n, the RE and RB of each estimation method has decreased due to controlling for the effect of including a lagged dependent value and larger sample size. For different values of t and $\theta = \tau$, the bias takes a decreasing behavior. As the value of number of cross sections increases along with the time series lengths and quantile ranks, the RE and RB become less across the cases due to the better performance of the estimator of TSQRDPD in large samples.

5. Conclusion

A total new unique approach to solve the problem of endogeneity in DPD models due to the involvement of lagged endogenous variable(s) by TSQRDPD is critically presented where the two stages are estimated at the same order of quantile. TSQRDPD has revealed more reliability in terms of lower values of RE and RB than the other methods of estimation for each quantile order given a specific panel dimension.

References

[1] R. Moon and M. Weidner, Dynamic linear panel regression models with interactive fixed effects, Econometric Theory 33(1) (2017), 158-195.

[2] V. Chernozhukov and C. Hansen, Instrumental variable quantile regression: a robust inference approach, J. Econometrics 142(1) (2008), 379-398.

[3] A. Galvao, Quantile regression for dynamic panel data with fixed effects, J. Econometrics 164 (2011), 142-157.

[4] V. Chernozhukov and C. Hansen, Instrumental qunatile regression inference for structural and treatment effects models, J. Econometrics 132 (2006), 491-525.

[5] A. R. Elhoussainy, A. E. Ahmed and M. A. Hossameldin, Estimating quantile regression for dynamic panel data using two-stage approach, Far East J. Math. Sci. (FJMS) 103(2) (2018), 451-465.

[6] H. Afify and A. Abdel-Aziz, Two-stage quantile regression for dynamic panel data models with fixed effects: Monte-Carlo simulation study, Adv. Appl. Stat. 54(1) (2019), 21-30.

[7] J. Gu and R. Koenker, Unobserved heterogeneity in income dynamics: an empirical Bayes perspective, J. Bus. Econom. Statist. 35(1) (2017), 1-16.

[8] T. Amemiya, Two stage least absolute deviations estimators, Econometrica 50(3) (1982), 689-711.