2019

Monte-Carlo Simulation Study of Two-Stage Quantile Regression for Dynamic Panel Data

Hossameldin Ahmed
hossameldin.ahmed@bue.edu.eg

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

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

Logo
Advances and Applications in Statistics

http://www.pphmj.com
http://dx.doi.org/.../...
Volume …, Number …, 2018, Pages … **ISSN: 0972-0871**

Monte-Carlo Simulation Study of Two-Stage Quantile Regression for Dynamic Panel Data

Hossameldin A. Afify[1], Alaa A. Abdel-Aziz [1], Aya A. Mohamed[1]

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

Abstract

The issue of obtaining biased estimators gradually appears when dynamic panel data (DPD) models are principally estimated by quantile regression (QR) due to the addition of lagged value(s) for dependent variable in DPD models as independent variable(s). A new approach for minimizing the bias is theoretically presented through the method of estimation of two-stage quantile regression for dynamic panel data (TSQRDPD). This proposed framework of TSQRDPD is mainly based on estimating the two stages by QR at a pre-determined rank of quantile. Moreover, bias behavior is studies through a monte-carlo simulation under several quantile ranks for stages, probability distributions and panels' dimensions. Monte-Carlo simulation results reveal that the new method of estimation is able to reduce estimators' biasedness considerably.

Key words: Quantile regression, Two-stage quantile regression, Dynamic panel data model, Endogeneity, Biasedness, Instrumental variable, Monte Carlo simulation

Received: ….; Revised: ….; Accepted: ….
Keywords and phrases: Kindly provide.
Communicated by ….

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 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 is 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. Afify [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 [6] showed that TSQRDPD has 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 stages. It should be taken into consideration that, both stages are estimated by quantile regression to maintain 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.

Monte-Carlo Simulation Study of Two-Stage Quantile Regression for Dynamic Panel Data

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 the presented but also 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 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 parameter1which is presented as follows:

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

It can also be represented in compressed form as follows:

$$y_{it} = \mathrm{Z}_{it}\,\omega_0 + U_{it}$$
$$U_{it} = \mu_i + \varepsilon_{it} \quad , \quad i = 1, \dots, N \quad t = 2, \dots, T \tag{2}$$

For i^{th} cross section and t^{th} time period, $[y_{it}\,,\,Y_{i,t-1}]$ is a $NT \times (\mathcal{W} + 1)$ 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 $\mathrm{Z}_{it} = [Y_{i,t-1}\,,\,X_{1\,it}]$,the structural parameter

$\omega_0 = (\Gamma_0', \Lambda_0')'$, ε_{it} is an $NT \times 1$ vector, μ_i is a vector of $I_N \times 1$. $X_{2\,it}$ is an $NT \times \mathcal{L}_2$ 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 $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2$. $X_{2\,it}$ is instrumental variables that can be used in the first stage to estimate the endogenous variable $Y_{i,t-1}$. According to koenker (2005), $Q_\theta(\,.\,)$ is the quantile function at (θ) conditional on Y_{it}. The problem of endogeneity would exist if $Q_\theta(U_{it}|Y_{it}) \neq Q_\theta(U_{it})$. This inequality between $Q_\theta(U_{it}|Y_{it})$ and $Q_\theta(U_{it})$ can represent an explanation of endogeneity in quantile regression. τ and θ are quantile levels for the first and the second stages 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:

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

The model in (3) is considered to be specified correctly, where $Y_{i,t-1}$ is the lagged value of dependent variable and $X_{it} = [X_{1\,it}, \; X_{2\,it}]$ which is matrix of size $NT \times \mathcal{L}$. Φ_0 is matrix of the first stage parameters that are unknown of size $\mathcal{L} \times \mathcal{W}$. V_{it} is a $NT \times \mathcal{W}$ 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:

Monte-Carlo Simulation Study of Two-Stage Quantile Regression for Dynamic Panel Data

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

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 $\widehat{\omega}_0 = \left(\widehat{\Gamma}_0' , \widehat{\Lambda}_0'\right)'$ for ω_0 can be obtained by solving the following minimization argument:

$$\min_{\omega_0} \mathfrak{D}_{NT}\left(\omega_0, \widehat{\varphi}_0, \widehat{\Phi}_0, \theta, \tau\right) = \min_{\omega_0} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\theta\left(\tau\, y_{it} + (1-\tau)\, x_{it}'\, \widehat{\varphi}_0 - x_{it}'\, \mathbb{N}\left(\widehat{\Phi}_0\right)\omega_0\right) \tag{5}$$

Where y_{it} and x_{it}' are the observations of the dependent and independent variables respectively for i^{th} cross section for t^{th} time period. Both τ and θ are quantile levels for the first and second stages respectively they range from 0 to 1. $\rho_\theta(g)$ is defined to be check function, where $\rho_\theta : R \rightarrow R^+$. 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 $I_{(.)}$ 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:

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

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

Where φ and Φ_r are $\mathcal{W} \times 1$ vectors, $Y_{r\,it}$ 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 correspondent estimated endogenous variable.

3. Monte-Carlo Simulation Study

This section is devoted to present the undertaken steps to conduct simulation study. The number of cross sections is 50 and 250. The length of time series is assumed to be 10, 50 and 100. Both dependent and independent variables are generated from the following probability distributions of Chi (8), and Cauchy (2,4). The number of replications is set to be 10000 times which is greater than any simulation of study 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 $i = 1, 2, \dots N$, $t = 2, 3, \dots T$ and Θ_{it} is the error term.

Monte-Carlo Simulation Study of Two-Stage Quantile Regression for Dynamic Panel Data

QRDPDFE model

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

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

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

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

Second stage model

$$y_{it} = \beta_0 + \rho_1 \hat{y}_{it-1} + \beta_1 x_{1it} + \mu_i + \mathfrak{D}_{it} \tag{10}$$

Where $i = 1, 2, \ldots N$, $t = 2, 3, \ldots T$, 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. 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 the bias of new method and compare it to 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 & RB performance where X and Y ~ Cauchy (2,3) where n=50 and t=10, 50 and 100

t	θ,τ	RE & RB	Quantile Regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
10	0.25	RE	60.1720	31.8750	14.1300	4.2930
		RB	66.0745	41.6460	10.3499	1.0047
	0.5	RE	60.0730	31.6300	14.0730	4.1350
		RB	65.5064	41.5438	10.3442	0.9984
	0.75	RE	59.3840	30.8220	13.9760	4.0480
		RB	64.7917	40.6923	10.3417	0.9881
50	0.25	RE	60.0830	31.7440	13.9920	4.2510
		RB	65.7091	41.3751	10.3203	0.9730
	0.5	RE	59.1980	30.9760	13.8430	4.0620
		RB	65.1385	40.8924	10.3135	0.9373
	0.75	RE	59.1540	30.2520	13.7860	3.9830
		RB	64.1797	40.4024	10.2980	0.9326
100	0.25	RE	60.0360	31.6180	13.9870	4.2370
		RB	65.4739	41.1830	10.3159	0.9520
	0.5	RE	59.1980	30.7900	13.8150	4.0530
		RB	64.7057	40.8922	10.3036	0.9341
	0.75	RE	58.7620	30.0860	13.7760	3.8510
		RB	64.1020	40.1140	10.2932	0.9319

Table(2):RE & RB performance where X and Y ~ Cauchy (2,3) where n=250 and t=10, 50 and 100

t	θ,τ	RE & RB	Quantile Regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
10	0.25	RE	58.7600	30.0370	13.5740	3.9750
		RB	61.3349	37.5839	10.2134	0.8786
	0.5	RE	58.1810	29.3480	13.5400	3.8580
		RB	60.6699	37.0632	10.2086	0.8708
	0.75	RE	57.8830	28.8290	13.3620	3.5260
		RB	60.5203	36.7142	10.1966	0.8537
50	0.25	RE	58.4400	29.4520	13.5410	3.8330
		RB	60.6976	37.4452	10.1891	0.8722
	0.5	RE	58.1070	29.0850	13.4850	3.8140
		RB	60.0423	36.9085	10.1890	0.8579
	0.75	RE	57.6220	28.5150	13.3180	3.4800
		RB	59.7705	36.0430	10.1826	0.8314
100	0.25	RE	58.1810	28.7550	13.5280	3.8190
		RB	60.5141	36.3769	10.1707	0.8582
	0.5	RE	57.6360	28.7300	13.4710	3.6170
		RB	59.9579	36.1181	10.1690	0.8566
	0.75	RE	57.3810	27.8370	13.2430	3.3820
		RB	58.8743	35.8061	10.1558	0.8195

Monte-Carlo Simulation Study of Two-Stage Quantile Regression for Dynamic Panel Data

Table(3):RE & RB performance where X and Y ~ Chi(8) where n=50 and t=10, 50 and 100

t	θ,τ	RE & RB	Quantile Regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
10	0.25	RE	64.4240	35.1690	15.0970	5.3430
		RB	65.0476	42.9228	10.3625	0.9716
	0.5	RE	63.7260	34.8400	15.0200	5.1150
		RB	64.2152	42.2124	10.3611	0.9608
	0.75	RE	63.3830	34.5950	15.0140	4.9340
		RB	63.7269	41.7341	10.3459	0.9450
50	0.25	RE	64.2710	35.0330	15.0620	5.3350
		RB	64.1640	42.4027	10.3581	0.9655
	0.5	RE	63.6890	34.7980	14.9930	5.0320
		RB	63.7770	41.6195	10.3537	0.9522
	0.75	RE	63.2780	34.4870	14.9930	4.9140
		RB	63.4586	41.4179	10.3278	0.9321
100	0.25	RE	64.1010	34.8150	15.0530	5.2700
		RB	64.0808	41.5634	10.3450	0.9226
	0.5	RE	63.5580	34.5410	14.9910	5.0300
		RB	63.3995	41.4603	10.3356	0.9196
	0.75	RE	63.0090	34.3740	14.9860	4.8840
		RB	62.8630	40.4230	10.3243	0.9154

Table(4):RE & RB performance where X and Y ~ Chi(8) where n=250 and t=10, 50 and 100

t	θ,τ	RE & RB	Quantile Regression			
			QRDPD	QRDPDFE	QRDPDIVFE	TSQRDPD
10	0.25	RE	63.4370	34.2450	14.9020	4.9880
		RB	61.2621	38.1726	10.2584	0.8541
	0.5	RE	62.6150	33.9890	14.7810	4.8840
		RB	61.1479	37.9068	10.2499	0.8512
	0.75	RE	62.4670	33.9450	14.7470	4.7150
		RB	60.9643	37.3352	10.2269	0.8308
50	0.25	RE	63.3260	34.1940	14.8970	4.9340
		RB	60.1166	37.9999	10.2523	0.8342
	0.5	RE	62.1040	33.8670	14.7160	4.8490
		RB	59.9997	37.5114	10.2300	0.8241
	0.75	RE	61.6530	33.6880	14.6890	4.6040
		RB	59.9617	37.2778	10.2009	0.8074
100	0.25	RE	63.2330	34.0910	14.8890	4.8890
		RB	59.9813	37.6356	10.1973	0.8128
	0.5	RE	62.0580	33.6330	14.7140	4.8400
		RB	59.5716	37.4846	10.1873	0.8093
	0.75	RE	61.5810	33.6250	14.6850	4.5990
		RB	59.5305	36.9254	10.1776	0.8049

Based on the above results in tables 1-4, the performance of RE & RB across the different estimation methods (QRDPD, QRDPDFE, QRDPDFEIV, TSQRDPD), and in each combination of n & t the RE & RB have a decreasing pattern from an estimation method to another where TSQRDPD is the least. In addition to that, in each of the estimation methods, and for the same rank of quantile and when t value increases for the same sizes of n the RE & 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 declining pattern. As the value of cross sectional increases along with the lengths of time series and quantile values, the RE & RB decrease across the cases due to the controlling for estimator RE & RB of endogenous variable and larger sample size.

5. Conclusion

An entire new different approach to handle the problem of endogeneity in DPD models due to the inclusion of lagged dependent variable(s) by TSQRDPD is extensively presented where both stages are estimated at the same rank of quantile. TSQRDPD has empirically showed more efficiency in terms of decreasing RE & RB than the other estimation techniques in every quantile rank given a particular cross sectional & time dimensions.

References

[1] R. Moon and M. Weidner, "Dynamic linear panel regression models with interactive fixed effects," *Econometric Theory*, vol. 33, no. 1, pp. 158-195, 2017.

[2] V. Chernozhukov and C. Hansen, "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, vol. 142, no. 1, pp. 379-398, 2008.

[3] A. Galvao, "Quantile regression for Dynamic Panel Data with Fixed Effects," *Journal of Econometrics*, vol. 164, pp. 142-157, 2011.

[4] V. Chernozhukov and C. Hansen, "instrumental qunatile regession inference for structral and treatment effects models," *Journal of Econometrics*, vol. 132, pp. 491-525, 2006.

[5] H. Afify, A. Elsheikh, and E. Rady, "Estimating Qunatile Regression for Dynamic Panel Data Using two -Stage Approach," *Far East Journal of Mathematical Sciences*, vol. 103, no. 2, pp. 451-465, 2018.

[6] H. Afify and A. Abdel-Aziz, "Two-Stage Quantile Regression For Dynamic Panel Data Models With Fixed Effects: Monte Carlo Simulation Study," *Advances and Applications in Statistics*, vol. 54, no. 1, pp. 21 - 30, 2019.

[7] J. Gu and R. Koenker, "Unobserved heterogeneity in income dynamics: an empirical Bayes perspective," *Journal of Business & Economic Statistics*, vol. 35, no. 1, pp. 1-16, 2017.

[8] T. Amemiya, "Two Stage Least Absolute Deviations Estimators," *Econometrica*, vol. 50, no. 3, pp. 689-711, 1982.