Economics Letters 104 (2009) 133–135

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A quantile regression approach for estimating panel data models using instrumental variables☆

Matthew Harding [a,*], Carlos Lamarche [b,1]

[a] Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, United States
[b] Department of Economics, University of Oklahoma, 729 Elm Avenue, Norman, OK 73019, United States

ARTICLE INFO

Article history:
Received 7 August 2008
Received in revised form 23 April 2009
Accepted 28 April 2009
Available online 5 May 2009

Keywords:
Quantile regression
Instrumental Variables
Individual Effects

JEL classification:
C33
C31
I21

ABSTRACT

We introduce a quantile regression approach to panel data models with endogenous variables and individual effects correlated with the independent variables. We find newly developed quantile regression methods can be easily adapted to estimate this class of models efficiently.

1. Introduction

We consider the quantile regression estimation of a panel data model with endogenous independent variables, where we allow the endogenous variable to be correlated with unobserved factors affecting the response variable. The model is similar to the framework analyzed by Chernozhukov and Hansen (2008) on instrumental variables for quantile regression. It was applied by Chernozhukov and Hansen (2004) to the analysis of 401(k) plans, by Hausman and Sidak (2004) to study the cost of long-distance calls by less affluent customers, and by Lamarche (2008) to investigate educational attainment. Ignoring individual factors in panel data makes it difficult to infer the causal relation between the covariate of interest and the outcome. If these individual sources of variation are correlated with the endogenous variable, instrumental variable quantile regression may offer biased estimates. The approach presented in this paper allows the researcher to estimate covariate effects at different points of the distribution while controlling for individual factors that may be affecting the response and are correlated with the independent variables.

Our approach extends the recent work of Chernozhukov and Hansen (2008) on instrumental variables by allowing for "fixed effects" as introduced in Koenker (2004). We present evidence that the method performs well in finite samples and compare different methods in an empirical application.

2. Model, method and inferential procedure

This paper considers the following model,

$$y_{it} = \boldsymbol{d}_{it}'\boldsymbol{\delta} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it}, \quad i = 1, \ldots N; t = 1, \ldots T \tag{2.1}$$

$$\boldsymbol{d}_{it} = h(\boldsymbol{x}_{it}, \boldsymbol{w}_{it}, \upsilon_{it}) \tag{2.2}$$

$$\alpha_i = g(\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}, \boldsymbol{d}_{i1}, \ldots, \boldsymbol{d}_{iT}, \epsilon_i). \tag{2.3}$$

The first equation is the classical panel data model where y is the response variable for subject i at time t, $\boldsymbol{d}$ is a vector of endogenous variables, $\boldsymbol{x}$ is a vector of exogenous variables, and u is the error term. Eq. (2.2) defines the endogenous variable $\boldsymbol{d}$ related to a vector of instruments $\boldsymbol{w}$, which are stochastically independent of u. The variable υ is stochastically dependent on u. Eq. (2.3) considers the typical case of correlation between the covariate and the individual effects. The

☆ We are grateful to Cecilia Rouse for providing the MPCP data.
* Corresponding author. Tel.: +1 650 723 4116; fax: +1 650 725 5702.
E-mail addresses: mch@stanford.edu (M. Harding), lamarche@ou.edu (C. Lamarche).
[1] Tel.: +1 405 325 5857.

0165-1765/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2009.04.025

variable ϵ is assumed to be independent of v and u. The model has the following random coefficient representation,

$$y_{it} = \boldsymbol{d}_{it}'\boldsymbol{\delta}(u_{it}) + \boldsymbol{x}_{it}'\boldsymbol{\beta}(u_{it}) + \boldsymbol{z}_{it}'\alpha(u_{it}) \quad u_{it}|\boldsymbol{d}_{it},\boldsymbol{x}_{it},z_{it}\sim\mathcal{U}(0,1) \tag{2.4}$$

$$\tau \mapsto \boldsymbol{d}_{it}'\boldsymbol{\delta}(\tau) + \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \boldsymbol{z}_{it}'\alpha(\tau) \tag{2.5}$$

where $\boldsymbol{z}_{it}$ is an indicator variable for the individual effect α_i, $\mathcal{U}(\cdot)$ denotes a uniform distribution, and τ is the τ-th quantile of the conditional distribution of y.

Note that individual effects $\alpha(\tau)$'s which enter the above model are indexed by τ, the τ-th quantile of the conditional distribution of y. It is thus not fully appropriate to refer to these effects as "fixed effects" since these estimated α parameters are expected to change over the distribution of y. The use of quantile individual effects allows for the presence of individual factors which are correlated with the independent variables but the random coefficient representation shows that they are allowed to vary over the conditional distribution of y. Quantile individual effects are thus best characterized as a hybrid between fixed and random effects, allowing for a more flexible specification of econometric models. Alexander, Harding and Lamarche (2008) use quantile individual effects in a cross-country analysis of the relationship between economic and political development in order to capture institutional change as a result of political shocks.

Estimating $\boldsymbol{\delta}$ at different quantiles of the conditional distribution of the response, provides an opportunity for investigating how the treatment $\boldsymbol{d}$ impacts the location, scale and shape of the distribution. The procedure described below accommodates the instrumental variable method by incorporating individual effects possibly correlated with the independent variables.

Consider the objective function for the conditional instrumental quantile relationship:

$$R(\tau,\boldsymbol{\delta},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha}) = \sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau(y_{it} - \boldsymbol{d}_{it}'\boldsymbol{\delta} - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \boldsymbol{z}_{it}'\boldsymbol{\alpha} - \hat{\boldsymbol{w}}_{it}'\boldsymbol{\gamma}), \tag{2.6}$$

where $\rho_\tau = u(\tau - I(u\le 0))$ is the quantile regression loss function, and $\hat{\boldsymbol{w}}$ is the least squares projection of the endogenous variables $\boldsymbol{d}$ on the instruments $\boldsymbol{w}$, the exogenous variables $\boldsymbol{x}$, and the vector of individual effects $\boldsymbol{z}$. We follow the estimation procedure of Chernozhukov and Hansen that proceeds in two steps. First we minimize the objective function above for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$ as functions of τ and $\boldsymbol{\delta}$,

$$\left\{\hat{\boldsymbol{\beta}}(\tau,\boldsymbol{\delta}),\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\delta}),\hat{\boldsymbol{\alpha}}(\tau,\boldsymbol{\delta})\right\} = arg\ \min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha}} R(\tau,\boldsymbol{\delta},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\alpha}). \tag{2.7}$$

Then, we estimate the coefficient on the endogenous variable by finding the value of $\boldsymbol{\delta}$, which minimizes a weighted distance function defined on $\boldsymbol{\gamma}$:

$$\hat{\boldsymbol{\delta}}(\tau) = arg\min_{\boldsymbol{\delta}} \hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\delta})'\boldsymbol{A}\hat{\boldsymbol{\gamma}}(\tau,\boldsymbol{\delta}), \tag{2.8}$$

for a given positive definite matrix $\boldsymbol{A}$. For identification purposes, we estimate a model with an overall intercept dropping the first individual effect.

The covariance matrix has the standard sandwich formula representation $\hat{\boldsymbol{J}}(\tau)^{-1}\hat{\boldsymbol{S}}(\tau)\hat{\boldsymbol{J}}(\tau)^{-1}$ where,

$$\hat{\boldsymbol{S}}(\tau) = \frac{\tau(1-\tau)}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{\Psi}_{it}\boldsymbol{\Psi}_{it}'$$

$$\hat{\boldsymbol{J}}(\tau) = \frac{1}{2NTh_{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}I(|\hat{u}_{it}(\tau)| \le h_{NT})\boldsymbol{\Psi}_{it}\boldsymbol{\Phi}_{it}'$$

with $\boldsymbol{\Psi}_{it} = (\boldsymbol{w}_{it}',\boldsymbol{x}_{it}',\boldsymbol{z}_{it}')'$, $\boldsymbol{\Phi}_{it} = (\boldsymbol{d}_{it}',\boldsymbol{x}_{it}',\boldsymbol{z}_{it}')'$, $\hat{u}_{it}(\tau) = y_{it} - \boldsymbol{d}_{it}'\hat{\boldsymbol{\delta}}(\tau) - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{z}_{it}'\hat{\boldsymbol{\alpha}}(\tau)$ and h is a properly chosen bandwidth (see Koenker (2005) and Chernozhukov and Hansen (2008) for additional details including specific choices of h).

3. Empirical evidence

3.1. Monte Carlo

Consider,

$$y_{it} = \beta_0 + \delta d_{it} + \beta_1 x_{it} + \alpha_i + u_{it}$$

$$d_{it} = w_{it} + v_{it};\quad x_{it} = \mu_i + \pi_{it};\quad w_{it} = \theta_i + \psi_{it},$$

where μ_i, $\pi_{it}\sim\chi_3^2$, $\theta_i,\psi_{it}\sim\mathcal{N}(0,1)$, and the variables $(u_{it},v_{it})'\sim\mathcal{N}(0,\Omega)$. The parameters are: $\beta_0 = 0$, $\beta_1 = 0.5$, $\delta = 1$, $\Omega_{11} = \Omega_{22} = 1$, and $\Omega_{12} = \Omega_{21} = 0.8$. Two versions of $\alpha_i = \lambda\overline{d}_i + \epsilon_i$ are considered in the simulation experiments. First, we generate ϵ_i from a Gaussian distribution assuming that $\lambda = 0$. Lastly, we draw ϵ_i from a Gaussian distribution assuming that $\lambda = 0.5$ and $\overline{d}_i = (1/T)\sum_t d_{it}$.

In Table 1, we compare the bias and root MSE of the following estimators: (1) ordinary least squares (OLS); fixed effects (FE); instrumental variable (IV); pooled quantile regression (QR); fixed effects quantile regression – Koenker (2004) (FEQR);instrumental variable quantile regression – Chernozhukov and Hansen (2008) (IVQR); instrumental variable quantile regression with fixed effects (IVFEQR). In the simulations, we report results considering different sample sizes $N = \{100,250\}$ and $T = \{5,12\}$. The performance of the methods that ignore the endogeneity of the treatment d are unsatisfactory. When the individual effects are correlated with the independent variables, the IVQR estimates are biased, while IVFEQR produces unbiased results achieving the minimum root MSE in this class of estimators for panel data.

3.2. Application

This section uses data from the Milwaukee Parental Choice program (MPCP) previously analyzed by Rouse (1998). The MPCP provides vouchers to low-income students to attend private schools. It targets poor families living in the city of Milwaukee whose children were not attending private school that year.

To address the issue of lacking a valid control group, Rouse's study used a sample of students from the Milwaukee public schools as a comparison group and individual fixed effects to control for latent characteristics. We consider model 0.1–0.2, which is similar to the one used by Rouse. The variable y measures educational attainment, d is actual attendance to choice school, $\boldsymbol{x}$ is a vector of exogenous variables that includes grade level of the test and a dummy variable for whether the test was imputed. The instrument w is an indicator variable for whether the student was randomly selected to attend choice schools (Rouse, 1998).

Table 1
Performance of quantile regression estimators.

N	T		Least squares			Quantile regression			
			OLS	FE	IV	QR	FEQR	IVQR	IVFEQR
Gaussian individual effects									
100	5	Bias	0.4101	0.3922	−0.0073	0.4139	0.3922	−0.0080	−0.0008
100	5	RMSE	0.4126	0.3930	0.0776	0.4173	0.3936	0.0906	0.0611
250	12	Bias	0.3899	0.4000	−0.0006	0.3919	0.3999	−0.0018	−0.0012
250	12	RMSE	0.3907	0.4001	0.0439	0.3930	0.4001	0.0471	0.0242
Correlated individual effects									
100	5	Bias	0.6966	0.3922	0.2686	0.6994	0.3922	0.2638	−0.0008
100	5	RMSE	0.6985	0.3930	0.2816	0.7019	0.3936	0.2813	0.0611
250	12	Bias	0.6654	0.4000	0.2782	0.6647	0.3999	0.2774	−0.0012
250	12	RMSE	0.6661	0.4001	0.2824	0.6655	0.4001	0.2827	0.0242

Table 2
Estimates of the causal effect of choice schools on math test scores.

Covariate of interest	Method	Quantiles					Mean
		0.1	0.25	0.5	0.75	0.9	
Enrolled in choice school	Pooled	−0.800	−0.077	−0.425	−2.000	−2.357	−1.145
		(1.054)	(0.800)	(0.663)	(0.671)	(1.043)	(0.552)
	IV	−1.066	0.000	−0.190	−2.165	−1.379	−1.014
		(1.267)	(0.964)	(0.774)	(0.768)	(1.271)	(0.653)
	FE	1.000	1.000	−0.406	−1.043	−1.500	0.170
		(1.188)	(1.255)	(1.179)	(1.353)	(1.468)	(0.758)
	IVFE	3.969	4.126	2.783	1.138	0.500	3.315
		(1.661)	(1.625)	(1.429)	(1.730)	(1.980)	(1.073)

The classical OLS and IV estimates presented in Table 2 suggest that the students enrolled in the program earn approximately −1.15 additional percentile points per year relative to the students that were not attending the choice schools.

If instead of focusing on the mean effect we consider various quantiles, we see negative (significant) signs among the best-performers. IVFEQR allows us to control for unobserved individual heterogeneity and examine the causal effect of the program at different points of the educational attainment distribution. Looking at the estimates, we see that the effect is positive with a tendency to decrease as we go across the quantiles. This evidence suggests that the causal effect of the program is positive and large for low-performing students but small and insignificant for high-performing students.

Applied researchers estimating a panel data model often need to account for both the presence of individual factors possibly correlated with the independent variables and the presence of endogenous explanatory variables. In this paper we show that quantile regression methods allow for the estimation of such models with the same ease that can be expected from a simple linear panel data model.

References

Alexander, M., Harding, M., and Lamarche, C., 2008. The Political Economy of Heterogeneous Development: Quantile Effects of Income and Education. mimeo, Department of Economics, Stanford University.
Chernozhukov, V., Hansen, C., 2004. The effects of 401(k) participation on the wealth distribution: an instrumental quantile regression analysis. Review of Economics and Statistics 86 (3), 735–751.
Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: a robust inference approach. Journal of Econometrics 142 (1), 379–398.
Hausman, J., Sidak, J.G., 2004. Why do the poor and less-educated pay higher prices for long-distance calls? Contributions to Economic Analysis and Policy 3 (1) (Article 3).
Koenker, R., 2004. Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74–89.
Koenker, R., 2005. Quantile Regression. Cambridge University Press.
Lamarche, C., 2008. Private school vouchers and student achievement: a fixed effects quantile regression evaluation. Labour Economics 15, 575–590.
Rouse, C., 1998. Private school vouchers and student achievement: an evaluation of the Milwaukee parental choice program. Quarterly Journal of Economics 553–602.