Double-Stage Quantile Regression

Tae-Hwan Kim
University of Nottingham, School of Economics
University Park, NG7 2RD
Nottingham, United Kingdom
Tae-Hwan.Kim@nottingham.ac.uk

Christophe Muller
University of Nottingham, School of Economics
University Park, NG7 2RD
Nottingham, United Kingdom
Christophe.Muller@nottingham.ac.uk

1. Introduction

Two-Stage Least Square estimator and Two-Stage Least Absolute Deviation estimator have been studied in Amemiya (1982) and Powell (1983). Researchers interested in specific parts of the conditional distribution have estimated quantile regressions. Some of them attempt to correct the endogeneity of some explanatory variables by implementing a first stage of least-squares instrumental estimation [Arias et al (2001), Garcia et al (2001)]. In this paper we extend Amemiya (1982) and Powell (1983) by predicting the values of endogenous variables with first stage quantile regressions based on the same quantile than the final quantile regression (Double-Stage Quantile Regression or DSQR). We derive the asymptotic representation and normality of the DSQR estimator and provide a consistent estimator of the covariance matrix of the DSQR estimator. We finally study finite sample properties by using simulations.

2. The Model

We are interested in the structural parameter ($\boldsymbol{a}_0$) in an equation given in the following matrix form for a sample of T observations:

$$y = Y\boldsymbol{g}_0 + X_1\boldsymbol{b}_0 + u \equiv Z\boldsymbol{a}_0 + u \qquad (1)$$

where $[y, Y]$ is a $T \times (G+1)$ matrix of endogenous variables, X_1 is $T \times K_1$ matrix of exogenous variables, $Z \equiv [Y, X_1]$, $\boldsymbol{a}_0' \equiv [\boldsymbol{g}_0', \boldsymbol{b}_0']$, and u is a $T \times 1$ vector. We denote by X_2 the matrix of $K_2 \equiv (K - K_1)$ exogenous variables absent from equation (1). Let us assume that Y has a reduced-form representation:

$$Y = X\Pi_0 + V \qquad (2)$$

where $X \equiv [X_1, X_2]$ is a $T \times K$ matrix, Π_0 is a $K \times G$ matrix of unknown parameters and V is a $T \times G$ matrix of unknown error terms. Then, the reduced form representation is

$$y = X\boldsymbol{p}_0 + v \qquad (3)$$

where $\boldsymbol{p}_0 \equiv \left[\Pi_0, \begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}\right]\boldsymbol{a}_0 \equiv H(\Pi_0)\boldsymbol{a}_0$ and $v \equiv u + V\boldsymbol{g}_0$. The first stage estimation of equations (2) and (3) yields some estimators $\hat{\boldsymbol{p}}$ and $\hat{\Pi}$ respectively of $\boldsymbol{p}_0$ and Π_0 with $T^{1/2}(\hat{\boldsymbol{p}} - \boldsymbol{p}_0) = O_p(1)$ and $T^{1/2}(\hat{\Pi} - \Pi_0) = O_p(1)$.

Let be the "check function" $\boldsymbol{r}_q(z): R \to R^+$ for a given $\boldsymbol{q} \in (0,1)$ as $\boldsymbol{r}_q(z) \equiv z\boldsymbol{y}_q(z)$ where $\boldsymbol{y}_q(z) \equiv \boldsymbol{q} - 1_{[z \le 0]}$ where $1_{[\bullet]}$ is the Kronecker index. We define the Double-Stage Quantile Regression (DSQR($\boldsymbol{q}, q$)) estimator $\hat{\boldsymbol{a}}$ of $\boldsymbol{a}_0$ as a solution to the following minimisation programme:

$$\min_{\boldsymbol{a}} \; S_T(\boldsymbol{a}, \hat{\boldsymbol{p}}, \hat{\Pi}, q, \boldsymbol{q}) \equiv \sum_{t=1}^{T} \boldsymbol{r}_q(qy_t + (1-q)X_t'\hat{\boldsymbol{p}} - X_t'H(\hat{\Pi})\boldsymbol{a}) \qquad (4)$$

where y_t and X_t are the t^{th} elements in y and X respectively and q is a non-zero constant chosen in advance by the researcher. We assume that (i) $\{u_t, V_t\}$ are independent and identically distributed where u_t, V_t are the t^{th} elements in u, V respectively (ii) $E(\boldsymbol{y}_q(v_t)) = 0$ (iii)

$T^{-1}\sum_{t=1}^{T} X_t X_t' \to Q$ and Q is finite and positive definite and (iv) v_t has a continuous density f and $f(0) > 0$.

3. The Asymptotic Representation

We consider the following data generating process deduced from equation (3).

$$\tilde{y}_t = \tilde{X}_t' \boldsymbol{a}_0 + \tilde{\boldsymbol{e}}_t \quad (5)$$

where $\tilde{y}_t \equiv q y_t + (1-q) X_t' \boldsymbol{p}_0$, $\tilde{X}_t' \equiv X_t' H(\Pi_0)$ and $\tilde{\boldsymbol{e}}_t = q v_t$. Equation (5) provides us with insight on the estimation of $\boldsymbol{a}_0$ when the true auxiliary parameters ($\boldsymbol{p}_0$, Π_0) are known and allows the direct application of Bickel's (1975) results. We first define

$$M_T(\Delta) \equiv T^{-1/2} \sum_{t=1}^{T} X_t \boldsymbol{y}_q (\tilde{\boldsymbol{e}}_t - T^{-1/2} X_t' \Delta),$$

where Δ is a $K \times 1$ vector. The following lemma is a direct application of Bickel's lemma.

Lemma 1. Under our assumptions, for any $L > 0$,

$$\sup_{\|\Delta\| \le L} \| M_T(\Delta) - M_T(0) + \boldsymbol{w} Q \Delta \| = o_p(1),$$

where $\|\Delta\| \equiv (\Delta'\Delta)^{1/2}$ and $\boldsymbol{w} \equiv q^{-1} f(0)$.

Using Lemma 1 we obtain the following asymptotic representation for the DSQR($\boldsymbol{q}, q$).

Proposition 1. Under our assumptions, the DSQR($\boldsymbol{q}, q$) estimator has the asymptotic representation:

$$T^{1/2}(\hat{\boldsymbol{a}} - \boldsymbol{a}_0) = Q_{zz}^{-1} H(\mathrm{H}_0)' \{ T^{-1/2} \sum_{t=1}^{T} X_t q f(0)^{-1} \boldsymbol{y}_q(v_t)$$
$$+ (1-q) Q T^{1/2} (\hat{\boldsymbol{p}} - \boldsymbol{p}_0) - Q T^{1/2} (\hat{\Pi} - \Pi_0) \boldsymbol{g}_0 \} + o_p(1)$$

where $Q_{zz} \equiv H(\Pi_0)' Q H(\Pi_0)$.

4. The Asymptotic Normality and Covariance Matrix

We now substitute the asymptotic representations of $T^{1/2}(\hat{\boldsymbol{p}} - \boldsymbol{p}_0)$ and $T^{1/2}(\hat{\Pi} - \Pi_0)\boldsymbol{g}_0$ into the asymptotic representation of $T^{1/2}(\hat{\boldsymbol{a}} - \boldsymbol{a}_0)$. Here, using the LS estimation for $\hat{\boldsymbol{p}}$ and $\hat{\Pi}$ in the first step produces an asymptotic bias on the intercept coefficient caused by the nonvanishing difference between quantile and mean, $E(\boldsymbol{y}_q(v_t) - v_t)$. In contrast, using quantile regression to estimate $\hat{\boldsymbol{p}}$ and $\hat{\Pi}$ with the same quantile than the final stage will cause the bias to disappear.

Theorem 1. Under our assumptions and if there exists a positive constant Δ such that $\| X_t \| \le \Delta < \infty$ for all t, then

$$T^{1/2}(\hat{\boldsymbol{a}} - \boldsymbol{a}_0) \xrightarrow{d} N(0, C),$$

where $C \equiv \boldsymbol{s}_0^2 Q_{zz}^{-1}$, $\boldsymbol{s}_0^2 \equiv E(\boldsymbol{h}_t^2)$ and $\boldsymbol{h}_t \equiv f(0)^{-1} \boldsymbol{y}_q(v_t) - [g_1(0)^{-1} \boldsymbol{y}_q(V_{1t}), \ldots, g_G(0)^{-1} \boldsymbol{y}_q(V_{Gt})] \boldsymbol{g}_0$.

We now propose a consistent estimator of the asymptotic covariance matrix C. Using $E(\boldsymbol{y}_q(v_t)^2) = \boldsymbol{q}(1-\boldsymbol{q})$, an estimator for $\boldsymbol{s}_0^2$ is:

$$\hat{\boldsymbol{s}}^2 \equiv \boldsymbol{q}(1-\boldsymbol{q}) \hat{f}(0)^{-2} + \hat{\boldsymbol{g}}' \hat{\Omega} \hat{\boldsymbol{g}} - 2 \hat{f}(0)^{-1} \hat{\boldsymbol{g}}' \hat{\Gamma}.$$

The matrix $\hat{\Omega}$ has a typical element $\hat{\Omega}_{ij} \equiv \hat{g}_i(0)^{-2} \boldsymbol{q}(1-\boldsymbol{q}) 1_{[i=j]} + \hat{g}_i(0)^{-1} \hat{g}_j(0)^{-1} (\hat{\boldsymbol{d}}_{ij} - \boldsymbol{q}^2) 1_{[i \ne j]}$ where $\hat{\boldsymbol{d}}_{ij} \equiv T^{-1} \sum_{t=1}^{T} 1_{[\hat{V}_{it} \le 0]} 1_{[\hat{V}_{jt} \le 0]}$ and $\hat{V}_{jt}$ is the quantile regression residual from (2). The vector $\hat{\Gamma}$ has a typical element $\hat{\Gamma}_i \equiv \hat{g}_i(0)^{-1} (\hat{\boldsymbol{d}}_i - \boldsymbol{q}^2)$ where $\hat{\boldsymbol{d}}_i \equiv T^{-1} \sum_{t=1}^{T} 1_{[\hat{v}_t \le 0]} 1_{[\hat{V}_{it} \le 0]}$ and $\hat{v}_t$ is the quantile regression residual from (3). The density estimators $\hat{f}(0)$ and $\hat{g}_i(0)$ can be obtained by using standard kernel estimation methods. It can be easily shown that $\hat{\boldsymbol{s}}^2 \xrightarrow{p} \boldsymbol{s}_0^2$ which implies that our estimator given by $\hat{C} \equiv \hat{\boldsymbol{s}}^2 \hat{Q}_{zz}^{-1}$ where $\hat{Q}_{zz} \equiv T^{-1} \sum_{t=1}^{T} H(\hat{\Pi}) X_t X_t' H(\hat{\Pi})'$ is consistent.

5. Monte Carlo Simulation

We conduct simulation experiments to investigate the finite sample properties of the DSQR(q, q) estimator of the structural parameters (g_0, b_0) in two cases: (1) when the endogeneity problem is ignored and (2) when the problem is corrected using DSQR. The objective function (4) can be translated into a linear programming problem and the DSQR(q, q) is obtained by a modified version of simplex algorithm by Barrodale and Roberts (1974). The performance of the one-stage quantile regression estimator is displayed in Table 1 where the true structural parameters are given by g_0 = 0.5, b_{00} = 1 (intercept coefficient) and b_{10} = 0.2. As expected, this estimator is systematically biased in small and moderate size samples.

Table 1. Means and Standard Deviations of One Step Quantile Estimator in the Simulation

	q	0.05	0.25	0.50	0.75	0.95
T = 50						
$\tilde{g}$	Mean	0.1131	0.1043	0.1020	0.1070	0.1142
	Std	0.2879	0.1812	0.1708	0.1736	0.2841
$\tilde{b}_0$	Mean	1.9226	2.0333	2.1172	2.1808	2.2462
	Std	1.3145	0.6586	0.5264	0.4370	0.5007
$\tilde{b}_1$	Mean	0.4249	0.4374	0.4404	0.4353	0.4398
	Std	0.3598	0.2267	0.2110	0.2229	0.3548
T = 300						
$\tilde{g}$	Mean	0.1085	0.1135	0.1111	0.1120	0.1103
	Std	0.1087	0.0714	0.0670	0.0726	0.1098
$\tilde{b}_0$	Mean	1.9891	2.0940	2.1937	2.2838	2.4147
	Std	0.5351	0.2766	0.2188	0.1934	0.2129
$\tilde{b}_1$	Mean	0.3963	0.3903	0.3951	0.3940	0.3931
	Std	0.1236	0.0844	0.0759	0.0858	0.1294

The results for the DSQR(q, q) estimator are provided in Table 2. The means of ($\hat{g}$, $\hat{b}$) are much closer to the true parameters than the one-step quantile estimator over all values of q, although the corresponding standard deviations are generally greater. When T = 50, the DSQR(q, q) estimator is biased owing to the small sample size. The bias disappears and the standard deviation becomes smaller by half when we increase the sample size to 300. Also, since the standard deviations are symmetric quadratic functions of q about 0.5, they become larger as q becomes closer to 0 or 1.

Table 2. Means and Standard Deviations of $DSQR(q,q)$ ***in the Simulation***

	q	0.05	0.25	0.50	0.75	0.95
T = 50						
$\hat{g}$	Mean	-3.2926	0.7372	0.5570	0.5699	0.7038
	Std	119.6657	3.4276	1.1030	0.7669	3.5433
$\hat{b}_0$	Mean	14.1242	0.3456	0.8310	0.7923	0.3493
	Std	412.5810	9.6412	3.1042	2.2001	10.4769
$\hat{b}_1$	Mean	1.8155	0.0487	0.1792	0.1598	0.0799
	Std	49.6428	2.0736	0.7927	0.5350	2.1024
T = 300						
$\hat{g}$	Mean	0.5331	0.5119	0.5003	0.5056	0.5049
	Std	0.3428	0.1667	0.1505	0.1653	0.2888
$\hat{b}_0$	Mean	0.9012	0.9598	0.9978	0.9786	0.9775
	Std	1.0495	0.5264	0.4725	0.5197	0.9161
$\hat{b}_1$	Mean	0.1823	0.1976	0.2000	0.1994	0.2027
	Std	0.2220	0.1172	0.1059	0.1150	0.1915

REFERENCE

Amemiya, T. (1982). Two stage least absolute deviation estimators. *Econometrica*, **50**, 689-711.

Arias, O., Hallock, K.F. and Sosa-Escudero, W. (2001). Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics*, **26**, 7-40.

Barrodale, I. and Roberts, F.D.K. (1974). Algorithm 478: solution of an overdetermined system of equations in the L_1 norm. *Communications of the Association for Computing Machinery*, **17**, 319-320.

Bickel, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association*, **70**, 428-433.

Garcia, J., Hernandez, P.J. and Lopez-Nicolas, A. (2001). How wide is the gap? An investigation of gender wage differences using quantile regression. *Empirical Economics*, **26**, 149-168.

Powell, J. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica*, **51**, 1569-1575.

RESUME

We present in this paper the asymptotic properties of double-stage quantile regression estimators. These results permit valid inferences in models estimated using quantile regressions, in which the endogeneity of some explanatory variables is treated via ancilliary predictive equations estimated with quantile regressions.

Nous présentons dans ce papier les propriétés asymptotiques des estimateurs en deux étapes pour les régressions quantiles. Ces résultats permettent des inférences valides dans les modèles estimés à partir de régressions quantiles, dans lesquels l'endogénéité de certaines variables explicatives est traité via des équations prédictives auxilliaires.