Economics Bulletin

Volume 31, Issue 4

Quantile Regression with Classical Additive Measurement Errors

Gabriel Montes-Rojas
City University London

Abstract

This note derives the bias of the quantile regression estimator in the presence of classical additive measurement error, and show its connection to least squares models. The bias structure suggests that the instrumental variables estimator proposed for least squares can be applied to the quantile regression case.

The author would like to express his appreciation to two anonymous referees, Antonio Galvao, Anil Bera, John Conley, Roger Koenker, Arthur Lewbel, Steve Portnoy, participants in the seminars at Cass Business School, University of Alicante, University of Illinois at Urbana-Champaign, 2008 Midwest Econometrics Group Meeting, 15th Panel Data Conference (Bonn), 2009 Far East and South Asia Meeting of the Econometric Society, and 2010 Econometric Society World Congress for helpful comments and discussions. All the remaining errors are mine.
Citation: Gabriel Montes-Rojas, (2011) ''Quantile Regression with Classical Additive Measurement Errors'', *Economics Bulletin*, Vol. 31 No. 4 pp. 2863-2868.
Contact: Gabriel Montes-Rojas - gabriel.montes-rojas.1@city.ac.uk.
Submitted: May 24, 2011. **Published:** October 11, 2011.

1 Introduction

When the regressors are subject to measurement errors (ME), it is well known that the slope coefficient of the least squares (LS) regression estimator is inconsistent because the measurement error induces endogeneity in the model. In the one regressor case (or the multiple regressor case with uncorrelated regressors), under standard assumptions, the ordinary LS estimator is biased toward zero, a problem often denoted as attenuation. The most common remedy to reduce this bias caused by the endogeneity problem is to use either economic theory or intuition to find additional observable variables that can serve as instrumental variables (IV). Most of the literature on the estimation of models with ME is based on LS with IV. See for instance Hsiao and Taylor (1991), Wansbeek and Koning (1989), Griliches and Hausman (1986) and Wansbeek (2001).

Recently, the topic of ME in variables has also attracted considerable attention in the quantile regression (QR) literature. Chesher (2001) studies the impact of covariate ME on quantile functions using small variance approximation, and Schennach (2008) discusses identification and estimation issues for general quantile functions based on Fourier transforms and previous results for nonlinear models (see Schennach 2004,2008). Wei and Carroll (2009) proposes a method to correct measurement error induced bias by constructing joint estimating equations that simultaneously hold for all the quantile levels.

This paper derives the bias in the QR estimator in the presence of a classical additive ME in covariates using Angrist, Chernozhukov and Fernandez-Val (2006) omitted variables formula. This representation provides an explicit formulation for the bias in the slope coefficients and complements the results in Chesher (2001) and Wei and Carroll (2009). The QR representation of the ME problem as an omitted variable determines that the corresponding endogeneity bias has a similar structure to that in LS estimation, and therefore it suggests that LS-based IV strategies in the QR framework as in Chernozhukov and Hansen (2006,2008) solve the ME problem.

2 Additive measurement error bias in quantile regression

In this section we show that ME causes endogeneity bias in the QR estimator. Consider the following representation of a model with classical additive ME,

$$y_i = x_i^{*\prime}\beta + z_i^{\prime}\alpha + u_i \quad i = 1, ..., N, \tag{1}$$

where y_i is the response variable, x_i^* is a $dim(x) = p$-vector of the well-measured regressors, β is a $p \times 1$ vector of parameters of interest to be estimated, z_i is a $dim(z)$-vector of covariates without ME and coefficients α, and u_i is the residual. Suppose that we do not observe x_i^*, but rather x_i, which is a noisy measure of x_i^* subject to an additive ME ϵ_i,

$$x_i = x_i^* + \epsilon_i. \tag{2}$$

It is assumed that ϵ_i is independent and identically distributed (*iid*). Moreover, ϵ_i is independent of x_i^*, z_i and u_i. Using equation (2) we can express (1) in terms of the observed y and x as

$$y_i = x_i^{\prime}\beta + z_i^{\prime}\alpha + u_i - \epsilon_i^{\prime}\beta. \tag{3}$$

It follows that the observed regressor x_i in (3) will be correlated with the composite error, $u_i - \epsilon_i\beta$, inducing endogeneity in the model. This problem is of practical significance since the resulting bias may be large. The standard result for the LS estimator with ME can be seen as an omitted variables problem, where $-\epsilon_i$ is the omitted variable.

In the following paragraphs, we derive the bias in the QR estimator in the presence of ME using Angrist, Chernozhukov and Fernandez-Val (2006) (denoted ACFV hereafter) approach. As in LS, ME bias in QR can be derived analytically as an endogeneity bias, and provides a simpler representation than that in Chesher (2001).

Define $v^* = [x^{*\prime}, z^{\prime}]^{\prime}$, $\Lambda_\epsilon = [\epsilon^{\prime}, 0^{\prime}]^{\prime}$, $v = [x^{\prime}, z^{\prime}]^{\prime} = v^* + \Lambda_\epsilon$ and $\varphi = [\beta^{\prime}, \alpha^{\prime}]^{\prime}$ (here we omit the indexes i to simplify the notation). Let φ^* be the probability limit of the LS estimator without ME (i.e. using the well-measured covariate, x^*), and φ° be the probability limit of the LS estimator with ME (i.e. using x). It is well known that ME produce the following relation between these two estimators,

$$\varphi^\circ = \varphi^* - \left(E\left[(v^* v^{*\prime} + \Lambda_\epsilon \Lambda_\epsilon^{\prime})\right]\right)^{-1} E[\Lambda_\epsilon \epsilon^{\prime}\beta]. \tag{4}$$

The bias in the variable with measurement error parameter depends on the noise to signal ratio, thus generating attenuation bias.

Consider now the τth conditional quantile function of the response y,

$$Q_y(\tau|v^*) = x^{*\prime}\beta^*(\tau) + z^{\prime}\alpha^*(\tau). \tag{5}$$

Using equation (2) one can rewrite (5) as

$$Q_y(\tau|v^*) = (x-\epsilon)'\beta^*(\tau) + z'\alpha^*(\tau) = x'\beta^*(\tau) - \epsilon'\beta^*(\tau) + z'\alpha^*(\tau) = Q_y(\tau|v,\epsilon). \tag{6}$$

As in the standard LS case with ME, the QR counterpart can be seen as an omitted variable problem, where $-\epsilon$ is the omitted variable. We derive the approximate bias using the ACFV omitted variable bias formula. The QR estimator without ME solves

$$\varphi^*(\tau) = \underset{\varphi}{\operatorname{argmin}}\, E[\rho_\tau(y - v^{*\prime}\varphi)], \tag{7}$$

where $\rho_\tau(u) := u(\tau - I(u<0))$. However, the QR estimator, as in the LS case, is biased in the presence of the ME. In this case, in the problem of solving (6) omitting $-\epsilon$, the standard QR solves

$$\varphi^\circ(\tau) = \underset{\varphi}{\operatorname{argmin}}\, E[\rho_\tau(y - v'\varphi)]. \tag{8}$$

Here $\varphi^*(\tau)$ and $\varphi^\circ(\tau)$ are the parameters that solve the population minimization problem, defined in an analog way to ACFV paper.

The following Lemma shows that the ME bias in QR can be approximated to an expression similar to that in OLS.

Lemma 1 *Assume that: (i) the conditional density function $f_y(y|v,\epsilon)$ exists and is bounded a.s.; (ii) $E[y]$, $E[Q_y(\tau|v,\epsilon)^2]$, and $E\|[v',\epsilon']'\|^2$ are finite; (iii) $\varphi^*(\tau)$ and $\varphi^\circ(\tau)$ uniquely solves equations (7) and (8) respectively; (iv) ϵ is independent of (x^*,z,u). Then,*

$$\varphi^\circ(\tau) \;=\; \varphi^*(\tau) - \left(E\left[\omega_\tau(v,\epsilon)\cdot(vv')\right]\right)^{-1} E[\omega_\tau(v,\epsilon)\cdot v\epsilon'\beta^*(\tau)]. \tag{9}$$

where $\omega_\tau(v,\epsilon) := \int_0^1 f_{u(\tau)}\left(u\cdot\Delta_\tau(v,\epsilon;\varphi^\circ(\tau))|v,\epsilon\right)du/2$ is a weighting function, and $\Delta_\tau(v,\epsilon;\varphi^\circ(\tau)) = v'\cdot(\varphi^\circ(\tau)-\varphi^*(\tau))' + \epsilon'\beta^\circ(\tau)$ is the QR specification error.

Proof. The proof follows ACFV results for partial QR and omitted variables bias (p.545–548). Since the conditional quantile function is linear, $Q_y(\tau|v,\epsilon) = v'\varphi^*(\tau) - \epsilon'\beta^*(\tau)$, where $\varphi^*(\tau)$ is defined as in (7). Then, the conditional QR model in equation (5) can be seen as one with $[v',\epsilon']'$ as covariates. Moreover, the conditional QR with measurement error $Q_y(\tau|v) = v\varphi^\circ(\tau)$, obtaining the coefficient $\varphi^\circ(\tau)$ from equation (8), can be seen as a model with omitted variable ϵ.

Recall that $\omega_\tau(v,\epsilon) := \int_0^1 f_{u(\tau)}\left(u \cdot \Delta_\tau(v,\epsilon;\varphi^\circ(\tau))|v,\epsilon\right) du/2$ where $f_{u(\tau)}(.|.)$ is the conditional density function of $u(\tau) := y - Q_y(\tau|v,\epsilon)$, and $\Delta_\tau(v,\epsilon;\varphi) := v'\varphi - Q_y(\tau|v,\epsilon)$ is the bias in the estimated quantile function for a given φ. Then,

$$\Delta_\tau(v,\epsilon;\varphi^\circ(\tau)) = v^{*\prime} \cdot (\varphi^\circ(\tau) - \varphi^*(\tau)) + \epsilon'\beta^\circ(\tau).$$

Under the stated assumptions, by Theorem 2 in ACFV, $\varphi^\circ(\tau)$ uniquely solves the equation

$$\varphi^\circ(\tau) := \text{argmin}_\varphi E[\omega_\tau(v,\epsilon)\Delta^2_\tau(v,\epsilon;\varphi)].$$

Solving for $\varphi^\circ(\tau)$ we have

$$\varphi^\circ(\tau) \;\; = \;\; \varphi^*(\tau) - \left(E\left[\omega_\tau(v,\epsilon)\cdot(vv')\right]\right)^{-1} E[\omega_\tau(v,\epsilon)\cdot v\epsilon'\beta^*(\tau)].$$

■

Note that the weighting function $\omega(.)$ depends on both v and ϵ and it is a distinctinve feature of QR when compared with LS case. However, it can be shown that the leading term in the QR bias has the same form as that in LS. In order to show this, assume that $f_y(y|v,\epsilon)$ has a first derivative in y that is bounded in absolute value by $\bar{f}'(v,\epsilon)$ and consider a Taylor expansion of the weights as in ACFV, p.546,

$$\omega_\tau(v,\epsilon) \;\; = \;\; 1/2 \cdot f_y(Q_\tau(y|v,\epsilon)|v,\epsilon) + \varsigma(v,\epsilon),$$

where

$$|\varsigma(v,\epsilon)| \leq 1/4 \cdot |\Delta_\tau(v,\epsilon;\varphi^\circ(\tau))| \cdot \bar{f}'(v,\epsilon).$$

Note that by independence of y and ϵ, $f_y(Q_\tau(y|v,\epsilon)|v,\epsilon) = f_y(Q_\tau(y|v^*)|v^*)$ (with first derivative bounded by $\bar{f}'(v^*)$). Then, when either $\Delta_\tau(v,\epsilon;\varphi^\circ(\tau))$ or $\bar{f}'(v^*)$ is small,

$$\omega_\tau(v,\epsilon) \;\; \approx \;\; 1/2 \cdot f_y(Q_\tau(y|v^*)|v^*).$$

Then, the ACFV weighted LS approximation to QR implies that

$$\varphi^\circ(\tau) \approx \varphi^*(\tau) - \left(E\left[f_y(Q_\tau(y|v^*)|v^*)\left(v^*v^{*\prime} + \Lambda_\epsilon\Lambda_\epsilon'\right)\right]\right)^{-1} E[f_y(Q_\tau(y|v^*)|v^*)\Lambda_\epsilon\epsilon'\beta^*(\tau)]. \quad (10)$$

It is important to note that a key factor in the coefficient bias approximation given by (10) is the conditional density function $f_y(.|v^*)$. As in the LS case, the bias in the variable with measurement error parameter in the QR framework depends on the noise to signal ratio, thus generating attenuation bias (compare with equation (4)). However, in the QR case, this bias is weighted by the conditional density function $f_y(.|v^*)$.

3 Conclusions

This paper frames measurement error problems in quantile regression models. The results in this paper show that measurement error problems is a subset of endogeneity bias, which in turn can be solved using instrumental variables techniques.

References

Angrist, J., Chernozhukov, V. and Fernandez-Val, I. (2006) "Quantile regression under misspecification, with an application to the U.S. wage structure" *Econometrica* **74**, 539–573.

Chernozhukov, V. and Hansen, C. (2006) "Instrumental quantile regression inference for structural and treatment effects models" *Journal of Econometrics* **132**, 491–525.

Chernozhukov, V. and Hansen, C. (2008) "Instrumental variable quantile regression: A robust inference approach" *Journal of Econometrics* **142**, 379–398.

Chesher, A. (2001) "Parameter approximations for quantile regressions with measurement error" Working Paper CWP02/01, Department of Economics, University College London.

Griliches, Z. and Hausman, J. A. (1986) "Errors in variables in panel data" *Journal of Econometrics* **31**, 93–118.

Hsiao, C. and Taylor, G. (1991) "Some remarks on measurement errors and the identification of panel data models" *Statistica Neerlandica* **45**, 187–194.

Schennach, S. M. (2004) "Estimation of nonlinear models with measurement error" *Econometrica* **72**, 33–75.

Schennach, S. M. (2008) "Quantile regression with mismeasured covariates" *Econometric Theory* **24** 1010–1043.

Wansbeek, T. (2001) "GMM estimation in panel data models with measurement error" *Journal of Econometrics* **104**, 259–268.

Wansbeek, T. and Koning, R. H. (1989) "Measurement error and panel data" *Statistica Neerlandica* **45**, 85–92.

Wei, Y. and Carroll, R. J. (2009) "Quantile regression with measurement error" *Journal of the American Statistical Association* **104**, 1129–1143.