

# Performance of Various Estimators for Censored Response Models with Endogenous Regressors

(forthcoming in the *Pacific Economic Review*)

Changhui Kang

Dept. Economics, Chung-Ang Univ.

Heukseok-dong, Dongjak-gu

Seoul 156-756, South Korea

ckang@cau.ac.kr

Myoung-jae Lee\*

Dept. Economics, Korea Univ.

Anam-dong, Sungbuk-gu

Seoul 136-701, South Korea

myoungjae@korea.ac.kr

Often in censored response ( $y_1$ ) models, there are endogenous regressors ( $y_2$ ), for which, differently from linear models, the usual instrumental variable estimator (IVE) does not apply. As reviewed in Lee (2008,2009), there are (at least) four approaches for this case: (i) ‘substitution (SUB)’ replacing the endogenous regressors with their projected versions on exogenous regressors  $x$ , (ii) ‘control function (CF)’ adding an extra error as a regressor to control for the endogeneity source, (iii) ‘system reduced form (SRF)’ finding reduced form (RF) parameters first and then the structural form (SF) parameters, (iv) ‘artificial instrumental regressor (AIR)’ using instruments as artificial regressors with zero coefficients. Also depending on the model assumptions, these approaches can be parametric (tobit) or semiparametric. This paper briefly reviews these, and then compares them through a fairly large-scale simulation. To make the simulation realistic, we use a large real data set. The regressors are drawn from the data set and only the error term (thus the response) is simulated. Also we use the same data along with the actual  $y_1$  to conduct an empirical analysis. Based on our simulation and mean-squared-error type criteria, CF and AIR perform better than SUB and SRF; in terms of computational ease, however, SUB and CF are the easiest, closely followed by SRF—AIR is very time-consuming. While CF does well in both accounts, its assumptions are most restrictive, and CF gave much different results from the others when the actual  $y_1$  was used. Therefore, although the choice of an estimator among the four should be case-specific, for practitioners, we would recommend SUB.

Key Words: censored response, tobit, endogenous regressors, instruments.

\* Corresponding Author.

\*\* The authors are grateful to Hung-Jen Wang and an anonymous reviewer for their comments.

# 1 Introduction

Often in applied studies with censored responses, we face endogenous regressors. Differently from linear models, however, endogenous regressors are not easy to deal with in nonlinear models even if instruments are available. As reviewed in Lee (2008,2009), there are (at least) four different approaches to censored response models with endogenous regressors, and the four approaches can be further sub-classified as parametric (e.g., if tobit is used) or semiparametric (if a semiparametric estimator is used). They differ in terms of the ease in application and their asymptotic distributions. But their finite sample performances are yet to be known. That is, it is not clear which estimator works better than others in what kind of circumstances. The goal of this paper is to compare the estimators in a realistic simulation setting to shed light on this important issue so that practitioners can use the best estimator suiting their needs and circumstances. Also, we will apply the estimators to a real data set to see how much actual differences there can be in reality.

There are many examples of censored response models and their more “sophisticated” versions with endogenous regressors. Blundell and Smith (1986) examine a female labor supply model in which the non-labor household income is a key endogenous variable; their model is a “classic” censored model. Labeaga (1999) estimates a rational addiction model for tobacco consumption in which an endogenous regressor is personal income; he uses a double hurdle model that includes a censored response component. Kerkhofs et al. (1999) investigate the impact of health—an endogenous regressor—on retirement time and type decision within a competing risk framework where the work duration may be censored. Savage and Wright (2003) estimate the effect of health insurance on hospital-stay duration where the health insurance dummy is endogenous; as their duration variable is grouped, they use a grouped-duration estimation method where censoring is accommodated in the last group. More examples can be found in the references of Blundell and Smith (1993) and Angrist (2001).

Let  $1[A] = 1$  if  $A$  holds and 0 otherwise. With ‘SF’ standing for ‘structural form’, consider

$$\begin{aligned} \text{SF1} & : y_1 = \max(0, y_1^*) = \max(0, \alpha_1 y_2 + x_1' \beta_1 + u_1) = y_1^* 1[y_1^* > 0] \\ \text{SF2} & : y_2 = \alpha_2 y_1^* + x_2' \beta_2 + u_2 \end{aligned} \tag{1.1}$$

where  $y_1$  is the response variable of interest with its latent version  $y_1^*$ ,  $y_2$  is an endogenous

regressor,  $x_j$  is a  $k_j \times 1$  exogenous regressor vector in SF $j$ ,  $u_j$  is a mean-zero error term, and  $\alpha_j$  and  $\beta_j$  are the model parameters,  $j = 1, 2$ . Let  $x$  be a  $k \times 1$  vector collecting the elements in  $x_1$  and  $x_2$ ; i.e.,  $x$  is the system exogenous regressor vector. In the two SF equations system, the primary interest is on SF1. The left-censored model (1.1) can be rewritten to accommodate right-censored models that appear in duration analysis so long as the right-censoring point is known for all individuals or unknown but the same for everybody.

An instrumental variable (IV) for  $y_2$  should satisfy a number of conditions: (i) it should be exogenous to  $u_1$ , (ii) appear in the  $y_2$  SF ('inclusion restriction'), but (iii) not appear in the  $y_1$  SF ('exclusion restriction'). As will become clear later, all approaches need at least one IV one way or another. For the future reference, solving the  $y_1^*$  and  $y_2$  SF's, we can get the reduced form (RF) equations  $y_1^* = x'\eta_1 + v_1$  and  $y_2 = x'\eta_2 + v_2$ , where  $\eta_j$  is the RF parameter and  $v_j$  is the RF error,  $j = 1, 2$ . The RF's for  $y_1^*$  and  $y_2$  in turn give the RF's for  $y_1$  and  $y_2$ :

$$\begin{aligned} \text{RF1} & : y_1 = \max(0, y_1^*) = \max(0, x'\eta_1 + v_1) = y_1^* 1[y_1^* > 0] \\ \text{RF2} & : y_2 = x'\eta_2 + v_2. \end{aligned} \tag{1.2}$$

Although we presented two SF's and then the RF's, as we will focus on SF1 and its parameters  $\alpha_1$  and  $\beta_1$ , we will use mainly SF1 and RF2 in our discussion and estimation that follow:

$$\begin{aligned} \text{SF1} & : y_1 = \max(0, \alpha_1 y_2 + x'_1 \beta_1 + u_1) \\ \text{RF2} & : y_2 = x'\eta_2 + v_2. \end{aligned} \tag{1.3}$$

This is often called a 'recursive' or 'triangular' system. What is observed is

$$y_{1i}, y_{2i}, x_i, \quad i = 1, \dots, N, \text{ that are iid.}$$

In view of the iid assumption, we will often omit the subscript  $i$  indexing individuals as we have already done above.

Three remarks are in order before we proceed further. First, as all approaches are two-stage estimators, accounting for the first-stage estimation error on the second-stage is needed. Although this has been done in the original references and some econometric textbooks, it is a rather cumbersome task. Instead, we propose to use nonparametric bootstrap—resampling from the original sample with replacement—to construct confidence intervals/regions and

conduct tests with them. This will simplify our presentation and help stay focused on the main issue: finite sample performances of the estimators and their differences in a real data set; the appendix, however, shows the asymptotic variance and its estimator for some estimators. Second,  $y_1^*$ , not  $y_1$ , appears in SF2, which makes a light work of deriving the RF's. If  $y_1$ , not  $y_1^*$ , appears instead, getting the RF's is no longer straightforward and it brings up the issue 'coherency' and 'completeness' (see Lewbel (2007) and the references therein). We will stick to (1.3) in this paper and eschew these issues. Third, if tobit is used, then the error term standard deviation  $\sigma$  (a nuisance parameter) has to be estimated while no nuisance parameter appears in semiparametric methods. Hence, to handle both tobit and semiparametric methods together, we will not further mention the necessity of estimating  $\sigma$  for tobit in the main text of this paper.

The rest of this paper is organized as follows. Section 2 briefly reviews the aforementioned four approaches. Section 3 explains the data and our simulation designs. Section 4 presents the simulation results. Section 5 conducts a real data analysis. Finally, Section 6 concludes. Throughout the paper, we will denote the independence between two random variables  $z_1$  and  $z_2$  as ' $z_1 \perp z_2$ '.

## 2 Four Approaches

Each of the four approaches is a two-stage procedure, and depending on the model assumptions, various estimators are available for the second-stage. If we assume that the error term follows a normal distribution independently of  $x$  as traditionally done in the econometric literature, then tobit can be used. But many semiparametric estimators are available as well which perform nearly as well as tobit even when the normality assumption holds but ignored in estimation. See Powell (1994), Pagan and Ullah (1999), Blundell and Powell (2003,2004) and Lee (2009) for more.

The semiparametric estimators can be further classified as 'bandwidth-free' and 'bandwidth-dependent'. The bandwidth-free varieties are relatively scarce and can be seen, e.g., in Powell (1984,1986a,1986b) and Lee (1992). With recent popularity of quantiles, it seems that the censored least absolute deviation estimator (CLAD) in Powell (1984) and its quantile generalization in Powell (1986b) are most widely used among the semiparametric estimators. At *Google Scholar*, CLAD has as many as 456 citations as of December, 2008. Hence, in our

simulation study and empirical analysis later, we will use tobit and CLAD in the second-stage of the four approaches.

## 2.1 Substitution (SUB)

Substitution (SUB) approach replaces  $y_2$  in the SF1 with a ‘ $x$ -projected’ version. For instance, rewrite SF1 as

$$y_1 = \max[0, \alpha_1 E(y_2|x) + x_1' \beta_1 + u_1 + \alpha_1 \{y_2 - E(y_2|x)\}]$$

where  $u_1 + \alpha_1 \{y_2 - E(y_2|x)\}$  becomes the new error. Using  $E(y_2|x)$  at least implies that the new error has zero conditional mean. But  $E(y_2|x)$  is not the only choice, as other functions of  $x$  might be used. Whichever function of  $x$  is used, it is in essence a predicted value of  $y_2$  in RF2 because  $x$  gets used. A simpler choice than  $E(y_2|x)$  is the linear projection

$$L(y_2|x) \equiv x' E^{-1}(xx') E(xy_2) \quad \text{where } E^{-1}(\cdot) \equiv \{E(\cdot)\}^{-1}$$

and replacing  $y_2$  with the following sample version of  $L(y_2|x)$  is what is often done in practice:

$$L_N(y_2|x_i) \equiv x_i' \left( \frac{1}{N} \sum_i x_i x_i' \right)^{-1} \left( \frac{1}{N} \sum_i x_i y_{2i} \right).$$

Rewrite the model with  $L(y_2|x)$  (which entails  $y_2 - L(y_2|x) = y_2 - x' \eta_2 = v_2$ ):

$$y_1 = \max\{0, \alpha_1 L(y_2|x) + x_1' \beta_1 + (u_1 + \alpha_1 v_2)\}. \quad (2.1)$$

The SF1 parameters can be estimated in two-stages: obtain  $L_N(y_2|x)$  in the first stage, and estimate

$$\gamma_1 \equiv (\alpha_1, \beta_1')$$

in the second-stage using  $L_N(y_2|x_i)$  and  $x_1$  as the regressors. As  $L_N(y_2|x) - L(y_2|x) = o_p(1)$  while  $u_1 + \alpha_1 v_2$  is a non-degenerate random variable, the estimation error  $L_N(y_2|x) - L(y_2|x)$  can be ignored for the second-stage estimator consistency. It affects, however, the asymptotic distribution of the second-stage estimator as shown, e.g., in Lee (1996). This point—the first stage error ignorable for the second-stage consistency but not for the asymptotic distribution—is relevant for all four approaches, and will not be further mentioned.

## 2.2 Control Function (CF)

Although there are different versions of control function (CF) approaches as examined in Lee (2008), the typical CF approach assumes

$$u_1 = \lambda(v_2; \theta) + \varepsilon_1 \quad \text{where } v_2 \perp \varepsilon_1$$

and  $\lambda(v_2; \theta)$  is a parametric function of  $v_2$  such that  $\lambda(v_2; 0) = 0$ .

For instance, we may adopt a linear or quadratic model for  $\lambda(v_2; \theta)$ :

$$\lambda(v_2; \theta) = \theta_1 v_2 \quad \text{or} \quad \lambda(v_2; \theta) = \theta_1 v_2 + \theta_2 v_2^2.$$

If  $\theta = 0$ , then  $u_1 = \varepsilon_1$  and thus  $v_2 \perp u_1$ :  $y_2 (= x' \eta_2 + v_2)$  is not an endogenous regressor anymore in the  $y_1$  SF.

Substituting  $\lambda(v_2; \theta) = \theta_1 v_2$  into  $u_1$  yields

$$y_1 = \max(0, \alpha_1 y_2 + x'_1 \beta_1 + \theta_1 v_2 + \varepsilon_1). \quad (2.2)$$

As in SUB, this renders a two-stage procedure: estimate  $v_2$  in the first-stage with the residual in the least squares estimator (LSE) of  $y_2$  on  $x$ , and estimate  $(\gamma'_1, \theta_1)'$  in the second stage. A bonus of CF is that the t-value for  $v_2$  provides a built-in test for the  $y_2$  endogeneity with  $\theta_1 = 0$ .

## 2.3 System Reduced Form (SRF)

In system reduced form approach (SRF), not just the  $y_2$  RF, but the entire system RF's are needed, which is a minor drawback of SRF. Recall (1.2), and insert the RF's into  $y_1^* = \alpha_1 y_2 + x'_1 \beta_1 + u_1$  to get (see Lee (1995) and Lee and Kimhi (2005)):

$$\begin{aligned} x' \eta_1 + v_1 &= \alpha_1 (x' \eta_2 + v_2) + x'_1 \beta_1 + u_1 \\ \implies x' \eta_1 + v_1 &= x' (\eta_2 \alpha_1 + S_1 \beta_1) + (\alpha_1 v_2 + u_1), \quad \text{where } x' S_1 = x'_1; \end{aligned}$$

$S_1$  is a known  $k \times k_1$  'selection matrix' consisting of 0's and 1's. For instance, if  $x_i = (1, p_i, q_i)'$  with  $k = 3$  and  $x_{1i} = (1, q_i)'$  with  $k_1 = 2$ , then

$$S_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad 'x' S_1 = x'_1' \quad \text{is} \quad [1, p_i, q_i] \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = [1, q_i].$$

Pre-multiply both sides of the equation ‘ $x'\eta_1 + v_1 = \dots$ ’ by  $x$  and take  $E(\cdot)$  to get rid of the error terms. As  $E(xx')$  is assumed to be of full rank as usual, we obtain (the numbers below a matrix denote its dimension)

$$\underset{k \times 1}{\eta_1} = \underset{k \times 1}{\eta_2} \underset{k \times k_1}{\alpha_1} + \underset{k \times k_1}{S_1} \underset{k \times 1}{\beta_1}. \quad (2.3)$$

This yields a two-stage procedure: first, estimate  $\eta_1$  using tobit or CLAD of  $y_1$  on  $x$ , and  $\eta_2$  by the LSE of  $y_2$  on  $x$ ; second, estimate  $\gamma_1$  by the LSE of  $\hat{\eta}_1$  on  $(\hat{\eta}_2, S_1)$  where  $\hat{\eta}_1$  and  $\hat{\eta}_2$  are the estimators for  $\eta_1$  and  $\eta_2$ .

## 2.4 Artificial Instrumental Regressors (AIR)

Suppose that the IV vector is  $c$ , which is the components in  $x$  not appearing in  $x_1$ . That is,  $x = (x'_1, c')$  and  $c$  equals  $x_2$  if  $x_1$  and  $x_2$  do not overlap. Augment SF1 with  $c$  to get

$$\begin{aligned} y_1^* &= \alpha_1 y_2 + x'_1 \beta_1 + c' \psi + u_1 \\ \implies y_1 &= \max(0, \alpha_1 y_2 + x'_1 \beta_1 + c' \psi + u_1). \end{aligned} \quad (2.4)$$

Apply tobit (or CLAD) to this to estimate  $\beta_1$  and  $\psi$ , pretending that  $\alpha_1$  is known for a while. With  $s_{\beta_1}$  and  $s_{\psi}$  denoting the score functions, the first order condition is

$$\frac{1}{N} \sum_i s_{\beta_1}(\alpha_1, \beta_1, \psi) = 0 \quad \text{and} \quad \frac{1}{N} \sum_i s_{\psi}(\alpha_1, \beta_1, \psi) = 0.$$

We know that the true value of  $\psi$  is zero due to the exclusion restriction. Thus set  $\psi = 0$  in the last display to get

$$\frac{1}{N} \sum_i s_{\beta_1}(\alpha_1, \beta_1, 0) = 0 \quad \text{and} \quad \frac{1}{N} \sum_i s_{\psi}(\alpha_1, \beta_1, 0) = 0. \quad (2.5)$$

These are moment conditions, and Generalized Method of Moments (GMM) can be applied to estimate  $\gamma_1$ . The idea of augmenting SF1 with artificial regressors that are instruments—this gives the name ‘artificial instrumental regressors (AIR)’—can be seen, e.g., in Chernozhukov and Hansen (2008); see also the references therein.

Although the GMM is straightforward in principle, there is no guarantee that it will converge well, differently from the well-known nice convergence property of tobit—in fact, the GMM did not converge well in our experience. To avoid the convergence problem, take the part  $\alpha_1 y_2$  in (2.4) as a regressor with a known coefficient, or take  $-\alpha_1 y_2$  as a new censoring threshold in the rewritten model

$$y_1 - \alpha_1 y_2 = \max(-\alpha_1 y_2, x'_1 \beta_1 + c' \psi + u_1)$$

depending on what the software in use allows. Then estimate  $\beta_1$  and  $\psi$  with  $\alpha_1$  fixed at a given value to denote the resulting estimator as

$$\hat{\beta}_1(\alpha_1), \hat{\psi}(\alpha_1).$$

Repeat this for all possible values of  $\alpha_1$ . Then choose  $\hat{\alpha}_1$  as the one that makes

$$\hat{\psi}(\alpha_1) = 0, \quad \text{which then yields } \hat{\beta}_1(\hat{\alpha}_1).$$

A reasonable range for  $\alpha_1$  may be obtained by initially estimating SF1 ignoring the  $y_2$  endogeneity. As tobit (or CLAD) has to be repeated for each possible value of  $\alpha_1$ , AIR is quite time-consuming.

### 3 Data and Simulation Designs

The real data set used for our simulation is drawn from Kang (2008) who examines the effect of sibling size on educational investments for a second-born girl. Specifically, the model in Kang (2008) is

$$y_1^* = \alpha_1 y_2 + x_1' \beta_1 + u_1$$

where  $y_1^*$  is the logarithm of monthly expenditure on private tutoring for a second-born girl,  $y_2$  is the number of siblings,  $x_1$  is a vector of child characteristics including the parents' average age and average education level, family income in log, residential region and the child's grade level. In order to deal with zero expenditures in the log transformation, 10 is added to every child's raw value of expenditure before log is taken; 10 ( $\simeq$  \$10) is chosen because it is the smallest accounting unit reported in the survey. The regressor  $y_2$  can be endogenous, because, for example, parents with higher socioeconomic status or cognitive abilities may have smaller families and invest more in education of each child than parents with lower socioeconomic status or cognitive abilities. Among  $N = 1,893$  observations, 559 observations (29.5%) have zero expenditures on private tutoring. We set the censoring point of  $y_1^*$  at  $\tau \equiv \ln(10) \simeq 2.3026$  to get  $y_1 = \max(\tau, y_1^*)$ .

In simulations, we set  $\alpha_1 = -0.3$  following the real-data finding of Kang (2008); also the true value of  $\beta_1$  in the simulation was chosen to be the ordinary tobit estimates ignoring the  $y_2$  endogeneity. Given  $\alpha_1$  and  $\beta_1$ , our simulation designs are as follows. We list the base design first, and for the other designs, we list only the difference from the base design:

- Simulation 0 (Base Design):

1. Generate  $c$ , which is an IV for  $y_2$ , from  $N(0, 1)$ .
2. Generate  $v_2$  from  $N(0, 1)$ .
3. Given  $c$  and  $v_2$ , generate  $y_2 = 0.3 \times c + v_2$ .
4. Given  $v_2$ , generate  $u_1 = -0.5 \times v_2 + 0.866 \times \zeta$  where  $\zeta \sim N(0, 1)$  so that  $(u_1, v_2)$  is jointly normal with unit marginal variances and  $COR(u_1, v_2) = -0.5$ .
5. Given  $y_2$ ,  $u_1$  and  $x_1$  ( $x_1$  from the real data), generate  $y_1^* = -0.3 \times y_2 + x_1' \beta_1 + u_1$ , and finally  $y_1 = \max(\tau, y_1^*)$ .

- Simulation I (Varying Strength of IV): set

Simulation I-1 :  $y_2 = \underline{0.1} \times c + v_2$  in Step 3 (weaker IV)

Simulation I-2 :  $y_2 = \underline{0.6} \times c + v_2$  in Step 3 (stronger IV).

- Simulation II (Dummy Endogenous Regressor  $y_2$ ): in step 3, use  $y_2 = 1[0.3 \times c + v_2 > 0]$ .
- Simulation III (Varying  $COR(u_1, v_2)$ ): set

Simulation III-1 :  $u_1 = -0.1 \times v_2 + 0.995 \times \zeta$  in Step 4

to get  $COR(u_1, v_2) = -0.1$ ;  $(u_1, v_2)$  is still jointly normal with unit marginal variances. Also set

Simulation III-2 :  $u_1 = -0.9 \times v_2 + 0.436 \times \zeta$  in Step 4

to get  $COR(u_1, v_2) = -0.9$ .

- Simulation IV (Heteroskedastic  $u_1$ ): set

Simulation IV-1 :  $u_1 = -0.5 \times v_2 + \pi \zeta \cdot \exp(\delta_1 y_2 + x_1' \delta_2)$  in Step 4

where  $\zeta \sim N(0, 1)$  and  $\pi$  is a constant to make  $V(u_1) \simeq 1$ . For the values of  $\delta_1$  and  $\delta_2$ , we use a heteroskedastic tobit estimates with real data:

$$y_1 = \max(\tau, \alpha_1 y_2 + x_1' \beta_1 + \xi_1) \quad \text{with } V(\xi_1 | y_2, x) = \exp(\delta_1 y_2 + x_1' \delta_2).$$

Also we use an heteroskedastic  $u_1$  design with  $\delta_2 = 0$ : set

Simulation IV-2 :  $u_1 = -0.5 \times v_2 + \pi \zeta \cdot \exp(\delta_1 y_2)$  in Step 4.

- Simulation V (Varying Marginal Distributions of  $u_1$  and  $v_2$ ): Set

Simulation V-1 :  $v_2 = \ln(\chi_3^2) - 0.73$  in Step 2 and  $u_1 = -0.5 \times v_2 + U[-1.5, 1.5]$  in Step 4.

where  $U[a, b]$  is the uniform distribution on  $[a, b]$ . The marginal distributions of  $u_1$  and  $v_2$  are asymmetric with mean close to 0 and variance close to 1, and  $COR(u_1, v_2) \simeq -0.5$ . The skewness of  $u_1$  is around 0.1; the kurtosis is around 2.4. Also tried is

Simulation V-2 :  $v_2 = \ln(\chi_3^2) - 0.73$  in Step 2 and  $u_1 = -0.5 \times v_2 - 0.69 \times \ln(\epsilon)$  in Step 4

where  $\epsilon$  follows the exponential distribution with parameter 1.8. The marginal distributions of  $u_1$  and  $v_2$  are asymmetric with mean close to 0 and variance close to 1, and  $COR(u_1, v_2) \simeq -0.5$ . The skewness of  $u_1$  is around 0.9; the kurtosis is around 4.5.

## 4 Simulation Results

For each simulation design, the four endogeneity-correction methods are applied with tobit or CLAD. As performance measures, we use average bias (BIAS), root mean squared error (RMSE) and mean absolute error (MAE) across 500 simulations. Specifically, with  $s$  indexing simulations

$$\begin{aligned} \text{BIAS} & : \bar{\hat{\alpha}}_1 - \alpha_1 = \left( \frac{1}{500} \sum_{s=1}^{500} \hat{\alpha}_1^s \right) - \alpha_1 \\ \text{RMSE} & : \left( \frac{1}{500} \sum_{s=1}^{500} (\hat{\alpha}_1^s - \alpha_1)^2 \right)^{1/2} \\ \text{MAE} & : \frac{1}{500} \sum_{s=1}^{500} |\hat{\alpha}_1^s - \alpha_1|. \end{aligned}$$

The simulation results are reported in Table 1 for tobit, and in Table 2 for CLAD. In each row of the tables, we denote the smallest absolute values of BIAS, RMSE and MAE among the four methods with ‘\*’ and the largest absolute values with ‘†’.

It is hardly a surprise that ordinary tobit and CLAD ignoring the endogeneity yield the largest absolute BIAS, RMSE and MAE in all simulation designs. The BIAS of ordinary tobit and CLAD dominates RMSE in all designs. The BIAS of the four endogeneity-correction methods is quite small in most designs. The absolute values of BIAS for SUB, CF, SRF and AIR are typically less than 5 percent of those of the ordinary tobit and CLAD. An exception

Table 1: Tobit Results of Simulated Data Sets

Simulation design:	Estimation method				Ordinary tobit
	SUB	CF	SRF	AIR	
Simulation 0:					
BIAS	-0.0013*	-0.0020	-0.0013*	0.0026 <sup>†</sup>	-0.4598
RMSE	0.0751 <sup>†</sup>	0.0736	0.0751 <sup>†</sup>	0.0732*	0.4603
MAE	0.0594 <sup>†</sup>	0.0584	0.0594 <sup>†</sup>	0.0577*	0.4598
Simulation I - 1:					
BIAS	0.0122	0.0098*	0.0122	0.0174 <sup>†</sup>	-0.4960
RMSE	0.2428 <sup>†</sup>	0.2379*	0.2428 <sup>†</sup>	0.2427	0.4965
MAE	0.1850 <sup>†</sup>	0.1815*	0.1850 <sup>†</sup>	0.1825	0.4960
Simulation I - 2:					
BIAS	0.0003	0.0001*	0.0003	0.0043 <sup>†</sup>	-0.3677
RMSE	0.0415 <sup>†</sup>	0.0411	0.0415 <sup>†</sup>	0.0410*	0.3682
MAE	0.0329 <sup>†</sup>	0.0325	0.0329 <sup>†</sup>	0.0324*	0.3677
Simulation II:					
BIAS	0.1810*	0.1817	0.1810*	0.1821 <sup>†</sup>	-0.2761
RMSE	0.1959*	0.1962	0.1959*	0.1966 <sup>†</sup>	0.2769
MAE	0.1812*	0.1818	0.1812*	0.1822 <sup>†</sup>	0.2761
Simulation III - 1:					
BIAS	0.0036*	0.0038	0.0036*	0.0041 <sup>†</sup>	-0.0913
RMSE	0.0845	0.0843*	0.0845	0.0849 <sup>†</sup>	0.0940
MAE	0.0677*	0.0677*	0.0677*	0.0681 <sup>†</sup>	0.0913
Simulation III - 2:					
BIAS	0.0011*	0.0017	0.0011*	0.0088 <sup>†</sup>	-0.8262
RMSE	0.0798 <sup>†</sup>	0.0777	0.0798 <sup>†</sup>	0.0768*	0.8263
MAE	0.0622 <sup>†</sup>	0.0616	0.0622 <sup>†</sup>	0.0591*	0.8262

(continued)

Table 2: Tobit Results of Simulated Data Sets (continued)

Simulation design:	Estimation method				Ordinary
	SUB	CF	SRF	AIR	tobit
Simulation IV - 1:					
BIAS	0.0084	0.0082*	0.0084	0.0123 <sup>†</sup>	-0.4485
RMSE	0.0775 <sup>†</sup>	0.0765*	0.0775 <sup>†</sup>	0.0765	0.4489
MAE	0.0621 <sup>†</sup>	0.0609*	0.0621 <sup>†</sup>	0.0609	0.4485
Simulation IV - 2:					
BIAS	-0.0001*	-0.0008	-0.0001*	0.0039 <sup>†</sup>	-0.4608
RMSE	0.0778 <sup>†</sup>	0.0771	0.0778 <sup>†</sup>	0.0763*	0.4613
MAE	0.0621 <sup>†</sup>	0.0616	0.0621 <sup>†</sup>	0.0607*	0.4608
Simulation V - 1:					
BIAS	-0.0044 <sup>†</sup>	-0.0013	-0.0044 <sup>†</sup>	0.0004*	-0.4593
RMSE	0.0809 <sup>†</sup>	0.0792	0.0809 <sup>†</sup>	0.0788*	0.4598
MAE	0.0641 <sup>†</sup>	0.0625	0.0641 <sup>†</sup>	0.0619*	0.4593
Simulation V - 2:					
BIAS	-0.0058 <sup>†</sup>	-0.0037	-0.0058 <sup>†</sup>	-0.0015*	-0.4701
RMSE	0.0869 <sup>†</sup>	0.0861	0.0869 <sup>†</sup>	0.0852*	0.4707
MAE	0.0681 <sup>†</sup>	0.0672	0.0681 <sup>†</sup>	0.0664*	0.4701

Notes: “\*” denotes the smallest absolute values of BIAS, RMSE and MAE among four endogeneity-correction methods; “<sup>†</sup>” denotes the largest absolute values.

Table 3: CLAD Results of Simulated Data Sets

Simulation design:	Estimation method				
	SUB	CF	SRF	AIR	Ordinary CLAD
Simulation 0:					
BIAS	0.0011*	-0.0019	0.0014	0.0038 <sup>†</sup>	-0.4593
RMSE	0.1062 <sup>†</sup>	0.0933	0.1059	0.0800*	0.4602
MAE	0.0841 <sup>†</sup>	0.0735	0.0837	0.0628*	0.4593
Simulation I - 1:					
BIAS	0.0177 <sup>†</sup>	0.0094*	0.0177 <sup>†</sup>	0.0100	-0.4954
RMSE	0.3381	0.3081	0.3394 <sup>†</sup>	0.2414*	0.4962
MAE	0.2654	0.2329	0.2666 <sup>†</sup>	0.1841*	0.4954
Simulation I - 2:					
BIAS	0.0022	0.0001*	0.0022	0.0074 <sup>†</sup>	-0.3669
RMSE	0.0541	0.0497	0.0543 <sup>†</sup>	0.0472*	0.3677
MAE	0.0432	0.0397	0.0434 <sup>†</sup>	0.0376*	0.3669
Simulation II:					
BIAS	0.1611*	0.1799 <sup>†</sup>	0.1611*	0.1766	-0.2781
RMSE	0.1893*	0.2020 <sup>†</sup>	0.1894	0.1928	0.2795
MAE	0.1656*	0.1827 <sup>†</sup>	0.1656	0.1771	0.2781
Simulation III - 1:					
BIAS	0.0055	0.0062 <sup>†</sup>	0.0058	0.0048*	-0.0907
RMSE	0.1017	0.1018 <sup>†</sup>	0.1013	0.0874*	0.0949
MAE	0.0799	0.0802 <sup>†</sup>	0.0796	0.0685*	0.0907
Simulation III - 2:					
BIAS	0.0071	0.0002*	0.0067	0.0117 <sup>†</sup>	-0.8259
RMSE	0.1090	0.0816*	0.1090 <sup>†</sup>	0.0827	0.8261
MAE	0.0851	0.0632*	0.0853 <sup>†</sup>	0.0643	0.8259

(continued)

Table 4: CLAD Results of Simulated Data Sets (continued)

Simulation design:	Estimation method				
	SUB	CF	SRF	AIR	Ordinary CLAD
Simulation IV - 1:					
BIAS	0.0074	0.0063*	0.0071	0.0110 <sup>†</sup>	-0.4575
RMSE	0.1020 <sup>†</sup>	0.0890	0.1015	0.0813*	0.4581
MAE	0.0821 <sup>†</sup>	0.0720	0.0817	0.0644*	0.4575
Simulation IV - 2:					
BIAS	-0.0026	-0.0015*	-0.0027	0.0053 <sup>†</sup>	-0.4606
RMSE	0.1091 <sup>†</sup>	0.0985	0.1090	0.0837*	0.4613
MAE	0.0851 <sup>†</sup>	0.0795	0.0847	0.0670*	0.4606
Simulation V - 1:					
BIAS	-0.0049	-0.0061 <sup>†</sup>	-0.0047	-0.0021*	-0.4597
RMSE	0.1190	0.1284 <sup>†</sup>	0.1190	0.0958*	0.4611
MAE	0.0950	0.1017 <sup>†</sup>	0.0951	0.0766*	0.4597
Simulation V - 2:					
BIAS	0.0024	0.0009*	0.0026	0.0032 <sup>†</sup>	-0.4606
RMSE	0.1079	0.0874*	0.1082 <sup>†</sup>	0.0893	0.4612
MAE	0.0850	0.0695*	0.0855 <sup>†</sup>	0.0720	0.4606

Notes: ‘\*’ denotes the smallest absolute values of BIAS, RMSE and MAE among four endogeneity-correction methods; ‘<sup>†</sup>’ denotes the largest absolute values.

is simulation II in which a dummy  $y_2$  is used instead of a continuous  $y_2$ ; in this design, the absolute BIAS for all endogeneity-correction methods is around 60 percent of that of the ordinary tobit and CLAD. Also, in simulation II, BIAS accounts for more than 70 percent of RMSE. This suggests that caution is required when  $y_2$  is a dummy rather than a continuous variable and a dummy  $y_2$  seems to need a further study.

Among the four endogeneity-correction methods, AIR performs worst in terms of BIAS. AIR yields the largest absolute BIAS in 8 (6) out of 10 simulation designs for tobit (CLAD). In contrast, CF appears to perform best among the four methods. Although it produces the largest absolute BIAS with CLAD estimates in 3 designs, CF yields the smallest absolute BIAS in 3 and 6 designs when tobit and CLAD are applied, respectively. SUB and SRF show intermediate levels of performance for tobit and CLAD.

While AIR performs relatively poorly in BIAS, it performs best for both tobit and CLAD in terms of RMSE and MAE. For a majority of the cases—6 designs in tobit and 7 designs in CLAD, AIR yields the smallest RMSE and MAE. In contrast, SUB and SRF perform worst for tobit in terms of RMSE and MAE with the largest RMSE and MAE in 8 designs, but perform comparably to CF for CLAD. SUB and SRF returned the same estimates for tobit, but different ones for CLAD; SUB does somewhat better than SRF for CLAD. In some models, equivalences of certain estimators can be established under some assumptions—e.g., for linear models, Lee (2008) shows SUB=AIR—Table 1 suggests SUB=SRF for tobit, but proving this goes beyond the scope of this paper.

To summarize, in terms of BIAS, CF seems to perform best and AIR performs worst among the four endogeneity-correction methods. In terms of RMSE and MAE, however, AIR performs best, and SUB and SRF performs worst. Because RMSE and MAE are in general regarded as more comprehensive performance measures than BIAS, these findings favor AIR over the other methods. But AIR takes hundreds times longer than the others in computation because a grid search is required for each possible value of  $\alpha_1$ . Also CF requires most restrictive assumptions. Hence the choice of an estimator seems case-specific. This conclusion notwithstanding, it should be noted that the performance differences were fairly small across the four approaches in our simulation designs.

## 5 Empirical Analysis

To highlight practical differences in the four endogeneity-correcting procedures, in this section, we apply the four methods as well as the ordinary tobit and CLAD to the real data set with the real  $y_1$ . Following an idea in Kang (2008) and other studies therein, we use the gender of the first-born as an IV for the number of siblings  $y_2$ . A rationale is that son preference shown by Korean parents increases the number of children in a family that has a daughter in parity one, whereas they let the nature decide the sex of the first-born. (See Kang (2008) for merits and demerits of the gender of the first-born as an IV for the number of siblings.)

Table 5: Tobit and CLAD Results of the Raw Data Set

---

Panel A: OLS results for RF2

Variables in $x$	Estimate	SD	t-value
First-born son (Yes=1)	0.410	0.019	21.81
Parents average edu.	0.006	0.004	1.28
Parents average age	-0.034	0.003	-10.53
Family income (in log)	-0.028	0.013	-2.17
Other regressors	suppressed		

  

Panel B: Coefficient for the number of siblings ( $\alpha_1$ )

Estimation methods:	Tobit			CLAD		
	Estimate	SD	t-value	Estimate	SD	t-value
SUB	-0.376	0.178	-2.11	-0.253	0.139	-1.82
CF	-0.230	0.208	-1.10	-0.222	0.179	-1.24
SRF	-0.376	0.190	-1.98	-0.253	0.153	-1.65
AIR	-0.376	0.187	-2.01	-0.343	0.172	-1.99
Ordinary tobit/CLAD	-0.384	0.087	-4.41	-0.274	0.073	-3.74

---

In Panel A of Table 3, we report the LSE of the key variables in RF2. The panel shows

that the IV is strong as the t-value for the IV is 21.8. Panel B of Table 3 presents the estimates for  $\alpha_1$ , their standard deviation (SD) and t-values for tobit and CLAD. The SD of SUB, SRF and AIR were obtained by nonparametric bootstrap with the bootstrap repetition number 500.

As for tobit, SUB, SRF and AIR yield nearly the same estimates for  $\alpha_1$  ( $-0.376$ ) whereas CF noticeably differs ( $-0.230$ ). The range  $[-0.230, -0.376]$  formed by these numbers falls in the tightest bounds for  $\alpha_1$  estimated by Kang (2008). Except CF, the other three methods SUB, SRF and AIR yield fairly precise estimates for  $\alpha_1$  with significant t-values. We noted already that SUB might be equivalent to SRF in tobit, and this holds again in Table 3; the bootstrap SD and t-values are not the same, however, since different bootstrap samples were drawn for different estimators.

As for CLAD, SUB, CF and SRF yield similar estimates for  $\alpha_1$  ( $-0.253$  to  $-0.222$ ), whereas AIR is a little different ( $-0.343$ ). The range produced by these numbers more or less agrees with the preceding range with tobit. Again, the absolute t-values for SUB, SRF and AIR are higher than that for CF.

## 6 Conclusions

We reviewed four different approaches (SUB, CF, SRF and AIR) that can deal with endogeneity in censored response models. Using simulations based on a large real data set, we then compared the finite-sample performance of the four approaches applied to tobit and CLAD. Based on our simulation and mean-squared-error type criteria, CF and AIR perform better than SUB and SRF; in terms of computational ease, however, SUB and CF are the easiest, closely followed by SRF—AIR is very time-consuming. While CF does well in both accounts, its assumptions are most restrictive, and CF gave much different results from the others when the actual  $y_1$  was used. Therefore, although the choice of an estimator among the four should be case-specific, for practitioners, we would recommend SUB which differs little from SRF. This recommendation of SUB is based on the followings: (i) although SUB did worse than AIR and CF, the differences were rather small; (ii) SUB is the simplest to use; (iii) SUB is the easiest to understand with its intuitive appeal.

## APPENDIX

For the readers who want to do asymptotic inference instead of bootstrap inference, we present the asymptotic variance and its estimator for tobit under SUB, CF and AIR in this appendix. SRF is omitted as it is easy to do. Lee (1995) provides details on SRF approach where three-equation system is used, one of which is censored; a GAUSS program for two-equation system (one continuous and one censored) is also provided in the appendix of Lee (1995). We can present the asymptotic variance and its estimator for CLAD, but its asymptotic variance depends on the error term density, whose estimation requires a bandwidth (i.e., smoothing parameter). Hence we omit the CLAD case as bootstrap inference has a definite advantage without any bandwidth choice. Recall that we did not mention estimating the error term SD for tobit in the main text, but we will do that in this appendix.

Before we proceed, we quickly show how to obtain numerical derivatives. For a function  $q(b)$  of a  $k \times 1$  vector  $b$ , its numerical first derivative  $\nabla q(\beta)$  at  $b = \beta$  consists of

$$\nabla q_j(\beta) = \frac{q(\beta + \varepsilon e_j) - q(\beta - \varepsilon e_j)}{2\varepsilon}, \quad j = 1, \dots, k$$

where  $e_j$  is the  $k \times 1$  vector with 1 in its  $j$ th row and 0 elsewhere and  $\varepsilon$  is a small positive constant, say  $10^{-7}$ . For instance, with  $k = 2$  and  $\beta = (\beta_1, \beta_2)'$ , the numerical first derivatives are

$$\nabla q_1(\beta) = \frac{q\{\beta + \varepsilon \begin{pmatrix} 1 \\ 0 \end{pmatrix}\} - q\{\beta - \varepsilon \begin{pmatrix} 1 \\ 0 \end{pmatrix}\}}{2\varepsilon}, \quad \nabla q_2(\beta) = \frac{q\{\beta + \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\} - q\{\beta - \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\}}{2\varepsilon}.$$

Going further, consider  $q(b, a)$  where both  $a$  and  $b$  are  $2 \times 1$ , and we want to obtain the  $2 \times 2$  cross derivative matrix  $\partial q(b, a) / \partial b \partial a'$  at  $\beta$  and  $\alpha$ . The term at row 1 and column 2 is

$$\begin{aligned} & \frac{1}{2\varepsilon} [\nabla q_1\{\beta, \alpha + \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\} - \nabla q_1\{\beta, \alpha - \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\}] \\ = & \frac{1}{2\varepsilon} \left[ \frac{q\{\beta + \varepsilon \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \alpha + \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\} - q\{\beta - \varepsilon \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \alpha + \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\}}{2\varepsilon} \right. \\ & \left. - \frac{q\{\beta + \varepsilon \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \alpha - \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\} - q\{\beta - \varepsilon \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \alpha - \varepsilon \begin{pmatrix} 0 \\ 1 \end{pmatrix}\}}{2\varepsilon} \right]. \end{aligned}$$

### Asymptotic Variance of Tobit with SUB

Recall that, in SUB, we have

$$y_1 = \max\{0, \alpha_1(x'\eta_2) + x'_1\beta_1 + u_1 + \alpha_1 v_2\}$$

where  $\eta_2$  is the first-stage parameter estimated by the LSE of  $y_2$  on  $x$ . To make estimating the error term SD explicit, redefine  $\gamma_1$  as

$$\gamma_1 \equiv (\alpha_1, \beta'_1, \sigma_{12})' \quad \text{where} \quad \sigma_{12} \equiv SD(u_1 + \alpha_1 v_2).$$

Let  $s_{\gamma_1}(\hat{\gamma}_1, \hat{\eta}_2)$  be the  $(k_1 + 2) \times 1$  tobit score function for  $\gamma_1$  using  $(x' \hat{\eta}_2, x'_1)$  as the regressors evaluated at  $\hat{\gamma}_1$  and  $\hat{\eta}_2$ , where  $\hat{\gamma}_1$  is the tobit and  $\hat{\eta}_2$  is the LSE for  $\eta_2$ . Clearly, random variables are included in  $s_{\gamma_1}$ , which are omitted to simplify the presentation.

The following asymptotic expansion holds (see, e.g., Lee (1996, 2009)): omitting the arguments in  $s_{\gamma_1}(\gamma_1, \eta_2)$  when the arguments are the true values,

$$\begin{aligned} \sqrt{N}(\hat{\gamma}_1 - \gamma_1) &= -E^{-1}(s_{\gamma_1} s_{\gamma_1}') \cdot \left\{ \frac{1}{\sqrt{N}} \sum_i s_{\gamma_1} + E\left(\frac{\partial s_{\gamma_1}}{\partial \eta_2'}\right) \sqrt{N}(\hat{\eta}_2 - \eta_2) \right\} + o_p(1) \\ &= -E^{-1}(s_{\gamma_1} s_{\gamma_1}') \cdot \frac{1}{\sqrt{N}} \sum_i \{s_{\gamma_1} + E\left(\frac{\partial s_{\gamma_1}}{\partial \eta_2'}\right) \zeta_i\} + o_p(1) \quad \text{because} \\ \sqrt{N}(\hat{\eta}_2 - \eta_2) &= \frac{1}{\sqrt{N}} \sum_i \zeta_i + o_p(1) \quad \text{where} \quad \zeta_i = E^{-1}(xx') \cdot x_i v_{2i}. \end{aligned}$$

Hence, with ' $\rightsquigarrow$ ' denoting convergence in law,

$$\begin{aligned} \sqrt{N}(\hat{\gamma}_1 - \gamma_1) &\rightsquigarrow N\{0, E^{-1}(s_{\gamma_1} s_{\gamma_1}') \cdot C_{\gamma_1} \cdot E^{-1}(s_{\gamma_1} s_{\gamma_1}')\} \quad \text{where} \\ C_{\gamma_1} &\equiv E[\{s_{\gamma_1} + E\left(\frac{\partial s_{\gamma_1}}{\partial \eta_2'}\right) \zeta\} \{s_{\gamma_1} + E\left(\frac{\partial s_{\gamma_1}}{\partial \eta_2'}\right) \zeta\}']. \end{aligned}$$

As for estimating the asymptotic variance,

$$\begin{aligned} \frac{1}{N} \sum_i s_{\gamma_1}(\hat{\gamma}_1, \hat{\eta}_2) s_{\gamma_1}(\hat{\gamma}_1, \hat{\eta}_2)' &\rightarrow^p E(s_{\gamma_1} s_{\gamma_1}') \quad \text{and} \quad \frac{1}{N} \sum_i \hat{\delta}_i \hat{\delta}_i' \rightarrow^p C_{\gamma_1} \quad \text{where} \\ \hat{\delta}_i &\equiv s_{\gamma_1}(\hat{\gamma}_1, \hat{\eta}_2) + \left\{ \frac{1}{N} \sum_i \frac{\partial s_{\gamma_1}(\hat{\gamma}_1, \hat{\eta}_2)}{\partial \eta_2'} \right\} \left( \frac{1}{N} \sum_i x_i x_i' \right)^{-1} x_i (y_{2i} - x_i' \hat{\eta}_2). \end{aligned}$$

The  $(k_1 + 2) \times k$  derivative matrix  $\partial s_{\gamma_1} / \partial \eta_2'$  can be found by numerical gradients; deriving its analytical form will be rather involved.

### Asymptotic Variance of Tobit with CF

In CF, we have  $u_1 = \theta_1 v_2 + \varepsilon_1$  with  $v_2 \perp \varepsilon_1$ , and thus

$$y_1 = \max\{0, \alpha_1 y_2 + x'_1 \beta_1 + \theta_1 (y_2 - x' \eta_2) + \varepsilon_1\}.$$

Let

$$\mu_1 \equiv (\alpha_1, \beta'_1, \theta_1, \sigma_\varepsilon)' \quad \text{where} \quad \sigma_\varepsilon \equiv SD(\varepsilon_1)$$

and  $s_{\mu_1}(\hat{\mu}_1, \hat{\eta}_2)$  be the  $(k_1 + 3) \times 1$  score function for  $\mu_1$  using  $(y_2, x'_1, \hat{v}_2)$  as the regressors, where  $\hat{\mu}_1$  is the tobit and  $\hat{v}_2 \equiv y_2 - x'_1 \hat{\eta}_2$ ;  $\eta_2$  is still the first-stage parameter.

Analogously to the SUB case, the following asymptotic expansion holds:

$$\sqrt{N}(\hat{\mu}_1 - \mu_1) = -E^{-1}(s_{\mu_1} s_{\mu_1}') \cdot \frac{1}{\sqrt{N}} \sum_i \{s_{\mu_1} + E(\frac{\partial s_{\mu_1}}{\partial \eta_2'}) \zeta_i\} + o_p(1).$$

Hence

$$\begin{aligned} \sqrt{N}(\hat{\mu}_1 - \mu_1) &\rightsquigarrow N\{0, E^{-1}(s_{\mu_1} s_{\mu_1}') \cdot C_{\mu_1} \cdot E^{-1}(s_{\mu_1} s_{\mu_1}')\} \quad \text{where} \\ C_{\mu_1} &\equiv E[\{s_{\mu_1} + E(\frac{\partial s_{\mu_1}}{\partial \eta_2'}) \zeta\} \{s_{\mu_1} + E(\frac{\partial s_{\mu_1}}{\partial \eta_2'}) \zeta\}']. \end{aligned}$$

As for estimating the asymptotic variance,

$$\begin{aligned} \frac{1}{N} \sum_i s_{\mu_1}(\hat{\mu}_1, \hat{\eta}_2) s_{\mu_1}(\hat{\mu}_1, \hat{\eta}_2)' &\rightarrow^p E(s_{\mu_1} s_{\mu_1}') \quad \text{and} \quad \frac{1}{N} \sum_i \tilde{\delta}_i \tilde{\delta}_i' \rightarrow^p C_{\mu_1} \quad \text{where} \\ \tilde{\delta}_i &\equiv s_{\mu_1}(\hat{\mu}_1, \hat{\eta}_2) + \left\{ \frac{1}{N} \sum_i \frac{\partial s_{\mu_1}(\hat{\mu}_1, \hat{\eta}_2)}{\partial \eta_2'} \right\} \left( \frac{1}{N} \sum_i x_i x_i' \right)^{-1} x_i \hat{v}_{2i}. \end{aligned}$$

### Asymptotic Variance of Tobit with AIR

For AIR, denote the tobit score function for  $(\beta_1', \sigma_1, \psi)'$  with  $\sigma_1 \equiv SD(u_1)$  as

$$s_{-\alpha}(\alpha_1, \beta_1, \sigma_1, \psi) = s_{-\alpha}(\gamma_1, \psi) \quad \text{where} \quad \gamma_1 \equiv (\alpha_1, \beta_1', \sigma_1)';$$

although we use the same notation  $\gamma_1$ ,  $\gamma_1$  here differs from the above  $\gamma_1$  for SUB, because the error term SD's are different. Recalling (2.5), the AIR estimator  $\hat{\gamma}_1$  satisfies the moment condition

$$\frac{1}{N} \sum_i s_{-\alpha}(\hat{\gamma}_1, 0) = 0.$$

Consider the “just-identified” case where  $\psi$  is a scalar so that the dimension of  $(\beta_1', \sigma_1, \psi)'$  is the same as that of  $\gamma_1$ . Taylor-expand the last display around  $\gamma_1$  to get

$$\begin{aligned} 0 &= \frac{1}{\sqrt{N}} \sum_i s_{-\alpha}(\gamma_1, 0) + \frac{1}{N} \sum_i \frac{\partial s_{-\alpha}(\gamma_1, 0)}{\partial \gamma_1'} \sqrt{N}(\hat{\gamma}_1 - \gamma_1) + o_p(1) \\ \implies \sqrt{N}(\hat{\gamma}_1 - \gamma_1) &= \left\{ \frac{1}{N} \sum_i \frac{\partial s_{-\alpha}(\gamma_1, 0)}{\partial \gamma_1'} \right\}^{-1} \frac{-1}{\sqrt{N}} \sum_i s_{-\alpha}(\gamma_1, 0) + o_p(1) \\ \rightsquigarrow N[0, E^{-1} \left\{ \frac{\partial s_{-\alpha}(\gamma_1, 0)}{\partial \gamma_1'} \right\} E \{ s_{-\alpha}(\gamma_1, 0) s_{-\alpha}(\gamma_1, 0)' \} E^{-1} \left\{ \frac{\partial s_{-\alpha}(\gamma_1, 0)}{\partial \gamma_1'} \right\}]. \end{aligned}$$

As for estimating the asymptotic variance,

$$\begin{aligned} \frac{1}{N} \sum_i \frac{\partial s_{-\alpha}(\hat{\gamma}_1, 0)}{\partial \gamma_1} &\rightarrow^p E\left\{ \frac{\partial s_{-\alpha}(\gamma_1, 0)}{\partial \gamma_1} \right\} \\ \frac{1}{N} \sum_i s_{-\alpha}(\hat{\gamma}_1, 0) s_{-\alpha}(\hat{\gamma}_1, 0)' &\rightarrow^p E\{s_{-\alpha}(\gamma_1, 0) s_{-\alpha}(\gamma_1, 0)'\}. \end{aligned}$$

In the “over-identified” case, i.e., when  $\psi$  is not a scalar,  $E\{\partial s_{-\alpha}(\gamma_1, 0)/\partial \gamma_1\}$  is not square, and thus not invertible. In this case, it is possible to proceed as in GMM over-identified case using the above moment condition. But this case raises the issue of which function of  $c$  to use; e.g., not just  $c$ , its polynomial functions can be used as well. This aspect is a shortcoming of AIR. Instead of mulling over what to do, it is much simpler in practice to turn the instrument vector  $c$  into a single variable, and one good choice would be using  $L_N(y_2|c)$ , following the usual practice of two-stage LSE. Using  $L_N(y_2|c)$  instead of  $L(y_2|c)$  does not affect the asymptotic distribution of AIR, following the fact that estimating instruments is innocuous (whereas estimating regressors is not); see Lee (1996,2009). In the moment condition with  $\psi = 0$ ,  $L_N(y_2|c)$  appears only as an instrument.

## REFERENCES

Angrist, J.D., 2001, Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice, *Journal of Business and Economic Statistics* 19, 2-16.

Blundell, R.W. and J.L. Powell, 2003, Endogeneity in nonparametric and semiparametric regression models, *Advances in Economics and Econometrics: Theory and Applications*, Eighth World Congress, Vol.II, edited by M. Dewatripont, L.P. Hansen, and S.J. Turnovsky, Cambridge University Press.

Blundell, R.W. and J.L. Powell, 2004, Endogeneity in semiparametric binary response models, *Review of Economic Studies* 71, 655-679.

Blundell, R.W. and R.J. Smith, 1986, An exogeneity test for a simultaneous equation tobit model with an application to labor supply, *Econometrica* 54, 679-685.

Blundell, R.W. and R.J. Smith, 1993, Simultaneous microeconomic models with censored or qualitative dependent variables, in *Handbook of Statistics* 11, edited by G.S. Maddala, C.R. Rao and H.D. Vinod, North Holland.

Chernozhukov, V. and C. Hansen, 2008, Instrumental variable quantile regression: a robust inference approach, *Journal of Econometrics* 142, 379-398.

Kang, C., 2008, Family size and educational investments in children: evidence from private tutoring expenditures in South Korea, unpublished paper.

Kerkhofs, M., M. Lindeboom and J. Theeuwes, 1999, Retirement, financial incentives and health, *Labour Economics* 6, 203-227.

Labeaga, J.M., 1999, A double-hurdle rational addiction model with heterogeneity: estimating the demand for tobacco, *Journal of Econometrics* 93, 49-72.

Lee, M.J., 1992, Winsorized mean estimator for censored regression, *Econometric Theory* 8, 368-382.

Lee, M.J., 1995, Semiparametric estimation of simultaneous equations with limited dependent variables: a case study of female labor supply, *Journal of Applied Econometrics* 10, 187-200.

Lee, M.J., 1996, Nonparametric two stage estimation of simultaneous equations with limited endogenous regressors, *Econometric Theory* 12, 305-330.

Lee, M.J., 2008, Semiparametric estimators for limited dependent variable (LDV) models with endogenous regressors, unpublished paper.

Lee, M.J., 2009, Micro-econometrics: methods of moments and limited dependent variables, Springer, forthcoming.

Lee, M.J. and A. Kimhi, 2005, Simultaneous equations in ordered discrete responses with regressor-dependent thresholds, *Econometrics Journal* 8, 176-196.

Lewbel, A., 2007, Coherency and completeness of structural models containing a dummy endogenous variable, *International Economic Review* 48, 1379-1392.

Pagan, A.R. and A. Ullah, 1999, *Nonparametric econometrics*, Cambridge University Press.

Powell, J.L., 1984, Least absolute deviations estimation for the censored regression model, *Journal of Econometrics* 25, 303-325.

Powell, J.L., 1986a, Symmetrically trimmed least squares estimation for Tobit models, *Econometrica* 54, 1435-1460.

Powell, J.L., 1986b, Censored regression quantiles, *Journal of Econometrics* 32, 143-155.

Powell, J.L., 1994, Estimation of semiparametric models, in *Handbook of Econometrics IV*, edited by R.F. Engle and D.L. McFadden, Elsevier.

Savage, E. and D.J. Wright, 2003, Moral hazard and adverse selection in Australian private hospitals: 1989–1990, *Journal of Health Economics* 22, 331–359.