

RECENTERED AND RESCALED INSTRUMENTAL VARIABLE ESTIMATION OF TOBIT AND PROBIT MODELS WITH ERRORS IN VARIABLES

Shigeru Iwata

Department of Economics
University of Kansas
Lawrence, KS 66045, U.S.A.

Key Words and Phrases: instrumental variables; GMM estimator; nonlinear errors in variables; elliptically symmetric distribution.

Journal of Economic Literature classification: C24, C25.

ABSTRACT

Since Durbin (1954) and Sargan (1958), instrumental variable (IV) method has long been one of the most popular procedures among economists and other social scientists to handle linear models with errors-in-variables. A direct application of this method to nonlinear errors-in-variables models, however, fails to yield consistent estimators.

This article restricts attention to Tobit and Probit models and shows that simple recentering and rescaling of the observed dependent variable may restore consistency of the standard IV estimator if the true dependent variable and the IV's are jointly normally distributed. Although the required condition seems rarely to be satisfied by real data, our Monte Carlo experiment suggests that the proposed estimator may be quite robust to the possible deviation from normality.

1. INTRODUCTION

It is widely known that when the explanatory variables are measured with errors the least square regression will yield an inconsistent estimator. For linear models the instrumental variable method has provided an important solution to this problem. However, in non-linear models with errors in variables, the instrumental variable method fails to provide a consistent estimator. The importance of this problem has grown with the increasing use of non-linear models for economic analysis. The purpose of this paper is to show that, for Probit, Tobit, and many other non-linear models, simple re-centering and re-scaling of the observed dependent variable will restore consistency of the instrumental variable estimators, if the true dependent variable and the instrumental variables are jointly normally distributed. Furthermore, under a more relaxed distributional assumption, the proposed instrumental variable estimator of the slope is shown to be consistent up to a multiplicative constant.

The errors-in-variables (EIV) literature has a long history. Measurement errors result in a lack of orthogonality between regressors and regression equation disturbance terms and hence cause the least square estimator to be inconsistent. In fact, in the presence of such errors the model is generally not identified without additional information and consistent estimation is not possible. [Bekker (1986) provides a necessary and sufficient distributional condition for model identification]. A major breakthrough in this literature came with the recognition of the importance of a postulated set of variables that are correlated with the true explanatory variables but uncorrelated with the regression equation errors. It is now well known that the instrumental variable (IV) estimator that exploits this assumption is consistent [Durbin (1954), Sargan (1958), Leamer (1978), Carter and Fuller (1980), Anderson (1984), Fuller (1987, Ch.1.4 and 2.4), Iwata (1992a)].

Nonlinear models have recently become important for economic investigation and this has stimulated new interest in the problem of errors-in-variables [Wolter and Fuller (1982), Carroll et al. (1982), Hausman et al. (1991), Hsiao (1989, 1991, 1992), Iwata (1992b)]. The IV estimation method simply does not work in nonlinear EIV models. This observation is somewhat surprising because IV estimation does work in nonlinear simultaneous equations models. Because of the similarity between the use of IV methods in linear EIV models and linear simultaneous equations models, we might expect a corresponding similarity between the use of IV methods in the two nonlinear models. However, as Amemiya (1985) pointed out, there is a fundamental difference between EIV models and simultaneous equations models. In EIV models the parameters to be estimated characterize the relationship between unobserved true variables, while in the simultaneous equations model they pertain to the relationship between observed variables. In linear regression models, measurement errors cause the regressors to be correlated with equation disturbances (observationally equivalent to endogeneity) but the regression coefficients correctly describe the true relationship. In nonlinear regression models, however, measurement errors are no longer additively separable from the true regressors, and hence, the true relationship breaks down with error in variables. As a result, the IV estimator fails to be consistent for nonlinear EIV models.

Given the inherent inconsistency of the IV estimators in nonlinear models, the literature has proceeded under the assumption that the error variances shrink as the sample size increases. Along this line, the asymptotic behavior of the IV estimator has been analyzed [Wolter and Fuller (1982), Amemiya (1985), Stefanski and Carroll (1985) etc.] and the 'approximate consistency' of IV estimators has been investigated [Carroll and Stefanski (1990), Stefanski and Buzas (1995), Carroll et al. (1995)]. However, as Hsiao (1989) noted, the assumption of shrinking variances will not fit many econometric modeling situations.

In this paper we take a different approach and show that, for a variety of nonlinear models, under certain distributional assumptions, the consistency of IV estimators can be restored by a simple re-centering and re-scaling of the observed dependent variable.

The organization of the rest of the paper is as follows: the next section introduces the EIV Tobit and Probit models. Section 3 shows that re-centering and re-scaling the dependent variable yields a consistent IV estimator. In section 4, we relax the distributional assumption and examine the properties of the estimator. Section 5 conducts a Monte Carlo experiment to investigate the small sample behavior of the proposed estimator. Section 6 presents an application of the proposed estimation method to the female labor supply study. A brief conclusion is given in the final section.

2. TOBIT AND PROBIT EIV MODELS

Consider the linear regression model

$$y_i^* = \alpha_0 + \mathbf{x}_i^{*'} \boldsymbol{\beta}_0 + u_i \quad i = 1, \dots, n, \quad (1)$$

where y_i^* is the dependent variable, \mathbf{x}_i^* is a $k \times 1$ vector of explanatory variables, α_0 is a scalar, $\boldsymbol{\beta}_0$ is a $k \times 1$ vector of unknown parameters, and u_i is an unknown disturbance with mean zero and variance σ^2 . Unlike in the standard case in which the observations on (y_i^*, \mathbf{x}_i^*) are available, we assume here that neither y_i^* nor \mathbf{x}_i^* are directly observable. We observe instead y_i and \mathbf{x}_i , which are defined as

$$y_i = \tau(y_i^*) \quad (2)$$

$$\mathbf{x}_i = \mathbf{x}_i^* + \mathbf{v}_i \quad (3)$$

$$\mathbf{x}_i^* = \boldsymbol{\mu}_x + \mathbf{w}_i \quad (4)$$

where τ stands for some known, possibly non-bijective, transformation, \mathbf{v}_i is a $k \times 1$ vector of measurement errors with mean zero and variance $\boldsymbol{\Sigma}_{\mathbf{v}\mathbf{v}}$, and $\boldsymbol{\mu}_x$ is a vector of constants. The model given by (1)–(4) is general enough to cover the Tobit and Probit models [the Tobit model corresponds to $\tau(y^*) = 1(y^* > 0) \cdot y^*$, while the Probit model corresponds to $\tau(y^*) = 1(y^* > 0)$] as well as many familiar one-to-one transformations of the dependent variable. The error \mathbf{v}_i is assumed to be correlated neither with \mathbf{x}_i^* nor u_i . We include (4) to make sure that we consider the “structural” (as opposed to “functional”) errors-in-variables models. We also assume the availability of an $m \times 1$ vector ($m \geq k$) of instrumental variables (IV), \mathbf{z}_i , such that \mathbf{z}_i is uncorrelated with \mathbf{v}_i and u_i but correlated with \mathbf{x}_i^* . Not all components of \mathbf{x}_i need to contain measurement errors. If the j -th explanatory variable is error free, i.e. $\mathbf{x}_{ji} = \mathbf{x}_{ji}^*$ and $\mathbf{v}_{ji} = \mathbf{0}$, then \mathbf{x}_{ji} is included in \mathbf{z}_i . We denote by $\mathbf{y}, \mathbf{y}^*, \mathbf{X}, \mathbf{X}^*, \mathbf{u}, \mathbf{V}$, and \mathbf{Z} , the matrices whose i -th row is equal to $y_i, y_i^*, \mathbf{x}_i', \mathbf{x}_i'^*, u_i, \mathbf{v}_i'$ and \mathbf{z}_i' , respectively for $i = 1, \dots, n$.

If we can observe (y_i, \mathbf{x}_i^*) , (1) and (2) give the standard Tobit and Probit models. If we instead observe (y_i^*, \mathbf{x}_i) , then (1) and (3) give the standard classical errors-in-variables regression model and it is well known that the IV estimator [commonly referred to the two stage least square (2SLS) estimator] $\tilde{\boldsymbol{\theta}}_{IV} \equiv (\bar{\mathbf{X}}' \mathbf{P}_{\bar{\mathbf{Z}}} \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}' \mathbf{P}_{\bar{\mathbf{Z}}} \mathbf{y}^*$ yields a consistent and optimal estimator of $\boldsymbol{\theta}_0 = (\alpha_0, \boldsymbol{\beta}_0')$, where $\bar{\mathbf{X}} = [\mathbf{1}, \mathbf{X}]$, $\bar{\mathbf{Z}} = [\mathbf{1}, \mathbf{Z}]$, $\mathbf{P}_{\bar{\mathbf{Z}}} = \bar{\mathbf{Z}}(\bar{\mathbf{Z}}' \bar{\mathbf{Z}})^{-1} \bar{\mathbf{Z}}'$ and $\mathbf{1}$ denotes the

$n \times 1$ vector of 1's (Sargan 1958). When the observation is on (y_i, \mathbf{x}_i) , the usual orthogonality condition is clearly violated, i.e., $E[\mathbf{z}_i(y_i - \alpha_0 - \mathbf{x}_i' \boldsymbol{\beta}_0)] \neq \mathbf{0}$ and hence, the standard IV estimator $\tilde{\boldsymbol{\theta}}_{IV} \equiv (\bar{\mathbf{X}}' \mathbf{P}_{\bar{\mathbf{Z}}} \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}' \mathbf{P}_{\bar{\mathbf{Z}}} \mathbf{y}$ is inconsistent. Since the IV estimator can be written as $\tilde{\boldsymbol{\theta}}_{IV} \equiv (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \mathbf{y}$ where $\tilde{\mathbf{X}} = \mathbf{P}_{\bar{\mathbf{Z}}} \bar{\mathbf{X}}$, it is equivalent to applying OLS with $\bar{\mathbf{X}}^*$ replaced by $\tilde{\mathbf{X}}$ (hence the name 2SLS). Maximizing the Tobit likelihood with this replacement, however, does not result in a consistent estimator of $\boldsymbol{\theta}_0$ either. As shown by Amemiya (1985), the IV estimation of the type described above fails, in general, to produce a consistent estimator.

In the next section I shall provide a new IV approach that leads to a consistent estimator of $\boldsymbol{\theta}_0$ in the model given by (1)–(4) under the assumption that y_i^* and \mathbf{z}_i are jointly normally distributed. The idea is that, by re-centering and re-scaling y_i , the model may be expressed as a linear form to which the standard IV technique can apply. Before getting into the discussion, we summarize the first and second moments assumption as follows:

$$\begin{bmatrix} u_i \\ \mathbf{v}_i \\ \mathbf{w}_i \\ \mathbf{z}_i \end{bmatrix} \sim i.i.d. \left(\begin{bmatrix} 0 \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\mu}_z \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{vv} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{ww} & \boldsymbol{\Sigma}_{wz} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{zw} & \boldsymbol{\Sigma}_{zz} \end{bmatrix} \right), \quad (5)$$

which implies that $(y_i^*, \mathbf{x}_i, \mathbf{z}_i)$ is distributed according to

$$\begin{bmatrix} y_i^* \\ \mathbf{x}_i \\ \mathbf{z}_i \end{bmatrix} \sim i.i.d. \left(\begin{bmatrix} \alpha_0 + \boldsymbol{\mu}_x' \boldsymbol{\beta}_0 \\ \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{bmatrix}, \begin{bmatrix} \sigma_u^2 + \boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{ww} \boldsymbol{\beta}_0 & \boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{ww} & \boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{wz} \\ \boldsymbol{\Sigma}_{ww} \boldsymbol{\beta}_0 & \boldsymbol{\Sigma}_{ww} + \boldsymbol{\Sigma}_{vv} & \boldsymbol{\Sigma}_{wz} \\ \boldsymbol{\Sigma}_{zw} \boldsymbol{\beta}_0 & \boldsymbol{\Sigma}_{zw} & \boldsymbol{\Sigma}_{zz} \end{bmatrix} \right). \quad (6)$$

3. RECENTERED AND RESCALED IV ESTIMATOR

3.1 General model

We now show that re-centering and re-scaling y_i yields a consistent IV estimator of θ_0 . To motivate this procedure in a general framework, consider the linear projection of y_i on y_i^* :

$$E_*(y_i | y_i^*) = E(y_i) + \frac{\text{Cov}(y_i, y_i^*)}{\text{Var}(y_i^*)} [y_i^* - E(y_i^*)]$$

which can be solved for y_i^* as

$$y_i^* = \frac{\text{Var}(y_i^*)}{\text{Cov}(y_i, y_i^*)} [E_*(y_i | y_i^*) - E(y_i)] + E(y_i^*)$$

where $E_*(y | x)$ stands for the linear projection of y on x . Replacing $E_*(y_i | y_i^*)$ with $E(y_i | y_i^*)$, the conditional expectation of y_i given y_i^* , and noticing $E(y_i | y_i^*) = E[\tau(y_i^*) | y_i^*] = \tau(y_i^*) = y_i$, we obtain an ‘estimate’ of y_i^* as

$$\tilde{y}_i^* = \psi_1^{-1}(y_i - \psi_2) \tag{7}$$

where

$$\psi_1 = \text{Cov}(y_i, y_i^*) / \text{Var}(y_i^*) \tag{8}$$

$$\psi_2 = E(y_i) - \psi_1 E(y_i^*). \tag{9}$$

The \tilde{y}_i^* is simply a recentered and rescaled y_i to mimic the y_i^* that is unobservable. A key idea here is that $E_*(y_i | y_i^*) - E(y_i | y_i^*)$ is orthogonal to z_i when z_i and y_i^* are jointly normal [which is a direct implication of the key relation $Cov(z_i, y_i) = \psi_1 Cov(z_i, y_i^*)$]. The expression (7) still contains unobservable ψ_1 and ψ_2 but these are estimable for the Tobit and Probit models, as shown below. Denote any consistent estimators of ψ_1 and ψ_2 by $\hat{\psi}_1$ and $\hat{\psi}_2$, respectively. Define $\hat{y}_i^* = \hat{\psi}_1^{-1}(y_i - \hat{\psi}_2)$ and let $\hat{\mathbf{y}}^*$ be an $n \times 1$ vector with the i -th component equal to \hat{y}_i^* .

Proposition 1: Let $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\boldsymbol{\beta}})' = (\bar{\mathbf{X}}' \bar{\mathbf{Z}} \hat{\mathbf{H}} \bar{\mathbf{Z}}' \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}' \bar{\mathbf{Z}} \hat{\mathbf{H}} \bar{\mathbf{Z}}' \hat{\mathbf{y}}^*$, where $\hat{\mathbf{H}}$ is an $m \times m$ symmetric positive definite matrix. Then $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ in the model given by (1)–(4).

Proof: Note first that $\text{plim}(1/n) \bar{\mathbf{Z}}' (\hat{\mathbf{y}}^* - \tilde{\mathbf{y}}^*) = \mathbf{0}$ implies

$$\text{plim } \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = (\boldsymbol{\Sigma}_{\bar{\mathbf{z}}} \mathbf{H} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}})^{-1} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}} \mathbf{H} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}e}$$

where $\text{plim } \hat{\mathbf{H}} = \mathbf{H}$, $\mathbf{e} = \tilde{\mathbf{y}}^* - \bar{\mathbf{X}} \boldsymbol{\theta}_0$,

$$\boldsymbol{\Sigma}_{\bar{\mathbf{z}}} = \begin{bmatrix} \mathbf{1} & \boldsymbol{\mu}_z' \\ \boldsymbol{\mu}_x & \boldsymbol{\Sigma}_{\mathbf{wz}} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{\bar{\mathbf{z}}e} = \begin{bmatrix} E(e_i) \\ \boldsymbol{\Sigma}_{ze} \end{bmatrix}.$$

Consistency of estimator $\hat{\boldsymbol{\theta}}$ then follows from $\boldsymbol{\Sigma}_{\bar{\mathbf{z}}e} = \mathbf{0}$. To show the latter claim, note that (Brillinger 1982)

$$\begin{aligned} Cov(\mathbf{z}_i, \tilde{y}_i^*) &= \psi_1^{-1} Cov(\mathbf{z}_i, y_i) = \psi_1^{-1} E(E[(\mathbf{z}_i - E(\mathbf{z}_i)) y_i | y_i^*]) \\ &= \psi_1^{-1} E(E[\mathbf{z}_i - E(\mathbf{z}_i) | y_i^*] y_i) = \psi_1^{-1} E \left(\frac{Cov(\mathbf{z}_i, y_i^*)}{Var(y_i^*)} [y_i^* - E(y_i^*)] y_i \right) \quad (10) \\ &= \psi_1^{-1} Cov(\mathbf{z}_i, y_i^*) \cdot \psi_1 = Cov(\mathbf{z}_i, y_i^*) \end{aligned}$$

which implies

$$Cov(\mathbf{z}_i, e_i) = Cov(\mathbf{z}_i, y_i^* - \alpha_0 - \mathbf{x}_i' \boldsymbol{\beta}_0) = Cov(\mathbf{z}_i, u_i) = \mathbf{0}$$

by the IV assumption. The claim follows by noting that

$$E(e_i) = E(\tilde{y}_i^*) - E(\alpha_0 + \mathbf{x}_i' \boldsymbol{\beta}_0) = \boldsymbol{\psi}_1^{-1} [E(y_i) - \boldsymbol{\psi}_2] - E(y_i^*) = 0.$$

We refer to the above estimator $\hat{\boldsymbol{\theta}}$ as the Recentered and Rescaled IV (or GMM) estimator (RR estimator, in short) of $\boldsymbol{\theta}_0$. To find the limiting distribution of $\hat{\boldsymbol{\beta}}$, consider the case in which the true value of $\boldsymbol{\psi}_1$ is known. Write the GMM estimator for the latter case as $\tilde{\boldsymbol{\beta}} \equiv \boldsymbol{\beta}_{GMM}(\boldsymbol{\psi}_{10}) = (\mathbf{X}' \bar{\mathbf{Z}} \hat{\mathbf{G}} \bar{\mathbf{Z}}' \mathbf{X})^{-1} \mathbf{X}' \bar{\mathbf{Z}} \hat{\mathbf{G}} \bar{\mathbf{Z}}' \tilde{\mathbf{y}}^*$, which is distinguished from $\hat{\boldsymbol{\beta}} \equiv \boldsymbol{\beta}_{GMM}(\hat{\boldsymbol{\psi}}_{10}) = (\mathbf{X}' \bar{\mathbf{Z}} \hat{\mathbf{G}} \bar{\mathbf{Z}}' \mathbf{X})^{-1} \mathbf{X}' \bar{\mathbf{Z}} \hat{\mathbf{G}} \bar{\mathbf{Z}}' \hat{\mathbf{y}}^*$, where $\hat{\mathbf{G}} = \hat{\mathbf{H}}[\hat{\mathbf{H}}^{-1} - \bar{\mathbf{Z}} \mathbf{1} \mathbf{1}' \bar{\mathbf{Z}} / (\mathbf{1}' \bar{\mathbf{Z}} \hat{\mathbf{H}} \bar{\mathbf{Z}}' \mathbf{1})] \hat{\mathbf{H}}$. The limiting distribution of $\tilde{\boldsymbol{\beta}}$ follows from the standard result of GMM estimation (Hansen 1982), which is given by

$$\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Sigma}_*)$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_* &= (\boldsymbol{\Sigma}_{\bar{\mathbf{z}}} \mathbf{G} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}\mathbf{x}})^{-1} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}} \mathbf{G} \mathbf{A} \mathbf{G} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}\mathbf{x}} (\boldsymbol{\Sigma}_{\bar{\mathbf{z}}} \mathbf{G} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}\mathbf{x}})^{-1} \\ \mathbf{A} &= \text{plim} (1/n) E \left(\sum_{i=1}^n \omega_i^2 \mathbf{z}_i \mathbf{z}_i' \right) \\ \omega_i^2 &= E[(\tilde{y}_i^* - \alpha_0 - \mathbf{x}_i' \boldsymbol{\beta}_0)^2 | \mathbf{z}_i] \end{aligned} \quad (11)$$

with $\boldsymbol{\Sigma}_{\bar{\mathbf{z}}} = [\boldsymbol{\mu}_{\mathbf{x}} \quad \boldsymbol{\Sigma}_{\mathbf{wz}}]$, $\mathbf{G} = \mathbf{H}[\mathbf{H}^{-1} - \boldsymbol{\mu}_{\bar{\mathbf{z}}} \boldsymbol{\mu}_{\bar{\mathbf{z}}}'] / (\boldsymbol{\mu}_{\bar{\mathbf{z}}} \mathbf{H} \boldsymbol{\mu}_{\bar{\mathbf{z}}})] \mathbf{H}$ and $\boldsymbol{\mu}_{\bar{\mathbf{z}}} = [\mathbf{1} \quad \boldsymbol{\mu}_{\mathbf{z}}']$.

The limiting distribution of $\hat{\boldsymbol{\beta}}$ is then given by the following result.

Proposition 2: *Suppose*

$$\sqrt{n} \begin{bmatrix} \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\psi}}_1 - \boldsymbol{\psi}_{10} \end{bmatrix} \xrightarrow{L} N \left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_* & \boldsymbol{\sigma}_* \\ \boldsymbol{\sigma}_*' & \sigma_{11} \end{bmatrix} \right). \quad (12)$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, \tilde{\boldsymbol{\Sigma}})$$

where

$$\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_* - \boldsymbol{\beta}_0 \boldsymbol{\sigma}_* ' / \psi_{10} - \boldsymbol{\sigma}_* \boldsymbol{\beta}_0 ' / \psi_{10} + (\sigma_{11} / \psi_{10}^2) \boldsymbol{\beta}_0 \boldsymbol{\beta}_0 ' . \quad (13)$$

Proof: By Taylor's expansion,

$$\sqrt{n}\hat{\boldsymbol{\beta}} = \sqrt{n}\boldsymbol{\beta}_{GMM}(\hat{\psi}_1) = \sqrt{n}\boldsymbol{\beta}_{GMM}(\psi_{10}) + \frac{\partial}{\partial \psi_1} \boldsymbol{\beta}_{GMM}(\psi_{10}) \cdot \sqrt{n}(\hat{\psi}_1 - \psi_{10}) + o_p(1)$$

It follows from

$$\frac{\partial}{\partial \psi_1} \boldsymbol{\beta}_{GMM}(\psi_{10}) = -\boldsymbol{\beta}_0 / \psi_{10}$$

that

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \approx \sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - (\boldsymbol{\beta}_0 / \psi_{10}) \sqrt{n}(\hat{\psi}_1 - \psi_{10}) \xrightarrow{L} N(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}) .$$

The optimal GMM estimator is obtained by setting $\hat{\mathbf{H}}$ such that $\text{plim } \hat{\mathbf{H}} = \mathbf{A}^{-1}$. For example, one may estimate \mathbf{A} by $\hat{\mathbf{A}} = (1/n) \sum \hat{e}_i^2 \bar{\mathbf{z}}_i \bar{\mathbf{z}}_i'$ where $\hat{e}_i = \hat{y}_i^* - \hat{\alpha} - \mathbf{x}_i' \hat{\boldsymbol{\beta}}$ and set $\hat{\mathbf{H}} = \hat{\mathbf{A}}^{-1}$. Then we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, (\boldsymbol{\Sigma}_{\bar{\mathbf{z}}} \mathbf{B} \boldsymbol{\Sigma}_{\bar{\mathbf{z}}})^{-1}),$$

where $\mathbf{B} = \mathbf{A}^{-1}[\mathbf{A} - \boldsymbol{\mu}_{\bar{z}}\boldsymbol{\mu}_{\bar{z}}' / (\boldsymbol{\mu}_{\bar{z}}' \mathbf{A} \boldsymbol{\mu}_{\bar{z}})] \mathbf{A}^{-1}$. Propositions 1 and 2 cover the Tobit and Probit models as special cases. To discuss those specific cases, we denote $\sigma_y^{*2} = \sigma_u^2 + \boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{ww} \boldsymbol{\beta}_0$, $\Phi = \Phi(E(y_i^*) / \sigma_y^*)$, and $\phi = \phi(E(y_i^*) / \sigma_y^*)$, where $\Phi(\cdot)$ and $\phi(\cdot)$ stand for the distribution function and the density function of a standard normal r.v. and $E(y_i^*) = \alpha_0 + \boldsymbol{\mu}_x' \boldsymbol{\beta}$.

3.2 Tobit model

For the Tobit model, $\tau(y^*) = 1(y^* \geq 0) \cdot y^*$ and the rescaling and recentering parameters ψ_1 and ψ_2 are given by $\psi_1 = \Phi(\delta)$ and $\psi_2 = \sigma_y^* \phi(\delta)$, where $\delta = (\alpha_0 + \boldsymbol{\mu}_x' \boldsymbol{\beta}_0) / \sigma_y^*$. The estimators $\hat{\psi}_1$ and $\hat{\psi}_2$ are given by $\hat{\psi}_1 = \hat{\Phi} = (1/n) \sum_{i=1}^n 1(y_i > 0)$ and $\hat{\psi}_2 = \hat{\sigma}_y^* \hat{\phi}$ where $\hat{\phi} = \phi(\Phi^{-1}(\hat{\Phi}))$ and $\hat{\sigma}_y^{*2} = (1/n) \sum_{i=1}^n (y_i - \bar{y})^2 / \{\hat{\Phi} - [\hat{\phi} - \Phi^{-1}(\hat{\Phi}) \cdot (1 - \hat{\Phi})][\hat{\phi} + \Phi^{-1}(\hat{\Phi}) \cdot \hat{\Phi}]\}$ [see Greene (1981)]. Now we check Proposition 2 for the Tobit model. In this case, the covariance matrix (12) is given by

$$\begin{bmatrix} \boldsymbol{\Sigma}_* & \boldsymbol{\sigma}_* \\ \boldsymbol{\sigma}_*' & \sigma_{11} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_* & (1 - \Phi)\boldsymbol{\beta}_0 \\ (1 - \Phi)\boldsymbol{\beta}_0' & \Phi(1 - \Phi) \end{bmatrix}.$$

which implies that the limit distribution of $\hat{\boldsymbol{\beta}}$ is given by

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Sigma}_* - [\Phi(1 - \Phi)]\boldsymbol{\beta}_0\boldsymbol{\beta}_0') \quad (14)$$

where $\boldsymbol{\Sigma}_*$ is given in (11). If $\Phi = 1$ or there is no censored observation, (14) implies that the covariance matrix of the limit distribution of $\hat{\boldsymbol{\beta}}$ reduces to

$(\sigma^2 + \boldsymbol{\beta}_0' \boldsymbol{\Sigma}_{vv} \boldsymbol{\beta}_0)(\boldsymbol{\Sigma}_{\bar{z}} \mathbf{G} \boldsymbol{\Sigma}_{\bar{z}})^{-1}$, the covariance matrix for the standard GMM estimator. The matrix $\boldsymbol{\Sigma}_*$ is estimated by

$$\hat{\boldsymbol{\Sigma}}_* = n(\mathbf{X}' \bar{\mathbf{Z}} \hat{\mathbf{G}} \bar{\mathbf{Z}}' \mathbf{X})^{-1} \mathbf{X}' \bar{\mathbf{Z}} \hat{\mathbf{G}} \left(\sum_{i=1}^n \hat{e}_i^2 \bar{\mathbf{z}}_i \bar{\mathbf{z}}_i' \right) \hat{\mathbf{G}} \bar{\mathbf{Z}}' \mathbf{X} (\mathbf{X}' \bar{\mathbf{Z}} \hat{\mathbf{G}} \bar{\mathbf{Z}}' \mathbf{X})^{-1}$$

where $\hat{e}_i = \hat{y}_i^* - \hat{\alpha} - \mathbf{x}_i' \hat{\boldsymbol{\beta}}$.

3.3 Probit model

For the Probit model, $\tau(y^*) = 1(y^* \geq 0)$. Note that

$$\begin{aligned} \text{Cov}(y_i, y_i^*) &= P(y_i^* \geq 0) E([y_i^* - E(y_i^*)] y_i \mid y_i^* \geq 0) \\ &= \Phi(\delta) \{E(y_i^*) + \sigma_y^* [\phi(\delta) / \Phi(\delta)] - E(y_i^*)\} = \sigma_y^* \phi(\delta), \end{aligned}$$

which implies that, by (8) and (9),

$$\begin{aligned} \psi_1 &= \phi(\delta) / \sigma_y^* \\ \psi_2 &= \Phi(\delta) - \psi_1 \sigma_y^* \delta = \Phi(\delta) - \phi(\delta) \delta. \end{aligned}$$

Since some normalization is required in any case to identify the Probit model, we set $\sigma_y^* = 1$. Then ψ_1 and ψ_2 are consistently estimated by $\hat{\psi}_1 = \hat{\phi}$ and $\hat{\psi}_2 = \hat{\Phi} - \hat{\phi} \hat{\delta}$, where $\hat{\Phi} = \bar{y} = (1/n) \sum y_i$, $\hat{\delta} = \Phi^{-1}(\hat{\Phi})$ and $\hat{\phi} = \phi(\hat{\delta})$.

It is not difficult to find that the covariance matrix in (12) is given by

$$\begin{bmatrix} \Sigma_* & \sigma_* \\ \sigma_*' & \sigma_{11} \end{bmatrix} = \begin{bmatrix} \Sigma_* & -\delta(1-\Phi)\beta_0 \\ -\delta(1-\Phi)\beta_0' & \delta^2\Phi(1-\Phi) \end{bmatrix}$$

which implies

$$\tilde{\Sigma} = \Sigma_* + (\delta/\phi^2)(1-\Phi)(2\phi + \delta\Phi)\beta_0\beta_0'.$$

4. EXTENSION

4.1 Errors in dependent variable

On some occasions we encounter the situation in which not only the explanatory variables \mathbf{x}_i but also the dependent variable y_i are measured with errors. Consider, for example, the Tobit model where the observation rule (2) is replaced by

$$y_i = \begin{cases} y_i^* + \eta_i & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where η_i is zero mean measurement errors. It is assumed that $y_i > 0$ when $y_i^* > 0$, which is a reasonable assumption for wage data of married women, for example (Stapleton and Young, 1984). There are no measurement errors in classification between working and not-working women, but the reported wages for working women are likely to be mismeasured.

The problem with the above Tobit model given by (1), (3), (4) and (15) is that, while $\sigma_y^{*2} + \sigma_\eta^2$ is consistently estimable, σ_y^{*2} cannot be estimated separately.

As a result, the recentered rescaled IV estimator of β is consistent up to a multiplicative constant.

4.2 Non-normality case

In many instances the multivariate normality assumption is too restrictive. In particular, it automatically rules out any discrete variables. A closer look at equation (10), however, reveals that what is needed for consistency of $\hat{\beta}$ defined in Proposition 1 is only the linearity of the conditional expectation of \mathbf{z} given y^* , or

$$E(\mathbf{z}_i | y_i^*) = \boldsymbol{\mu}_z + \boldsymbol{\gamma}[y_i^* - E(y_i^*)] \quad (16)$$

for some $m \times 1$ vector $\boldsymbol{\gamma}$ [Chung and Goldberger (1984)]. Under condition (16) together with finiteness and non-block-diagonality of the covariance matrix of $(\mathbf{x}_i, \mathbf{z}_i)$, $\hat{\beta}$ is still consistent up to a scale factor. A sufficient condition for (10) is that (y_i^*, \mathbf{z}_i') has an elliptically symmetric distribution, which may include either discrete or continuous distributions.

5. MONTE CARLO SIMULATION

A series of Monte Carlo simulations was conducted to see how successfully the proposed estimator removes bias in small samples. The simulations are based on the model given by (1)–(5) with $\alpha_0 = 0.5$, $\beta_0 = 1.0$ and a scalar x_i^* . The artificial data are generated from the multivariate normal distribution [cases (a)–(d) in Tables 1 and 2]:

$$\begin{bmatrix} u_i \\ v_i \\ w_i \\ z_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 1/r & 0 & 0 \\ 0 & 0 & 1 & \rho \\ 0 & 0 & \rho & 1 \end{bmatrix} \right) \quad (17)$$

and the fraction p of the sample is then made censored or truncated, where r is the signal to noise ratio (σ_w^2 / σ_v^2) and ρ is the correlation between the true regressor and the IV. In our baseline model, we set $\rho = 0.6$, $r = 1.5$, and $p = 0.30$ with sample size $n = 200$. Variations from the baseline model are indexed by ρ , r , and p in Tables 1 and 2. Tables 1E and 2E report the case of non-normal regressors. In each case, 1,000 samples are generated, the (Tobit or Probit) maximum likelihood estimates (MLE), the simple IV estimates (IV), and the Recentered and Rescaled IV estimates (RR) are calculated for each sample, and the biases, standard errors, and the root mean squared errors are reported.

The simulation results are summarized as follows: First, the MLEs suffer quite substantial bias in both the Tobit and Probit models, when the explanatory variables are measured with error. The bias of the estimated slope coefficient is more severe in the Tobit case (Tables 1a, 1b) than in the Probit case (Tables 2a, 2b). Second, the simple IV estimators partially remove measurement error bias in the Tobit case (Table 1a), while the bias gets even worse in the Probit case (Table 2a). Third, the Recentered and Rescaled IV estimator are quite successful in removing bias, especially in the Tobit case (Table 2a, 2b). Fourth, improvement by the RR estimators of the slope coefficient is quite uniform and consistent over the range of the parameters ρ , r , and p (Tables 1b, 1c, 1d and 2b, 2c, 2d). Furthermore, such bias reduction appears to be robust to certain types of deviation from normality of (x_i^*, z_i) , i.e., when x_i^* and z_i are distributed as a t-distribution and a Chi-squared distribution (Tables 1e and 2e) and when \mathbf{z} consists of dummies indicating group membership (Tables 1f and 2f).

TABLE 1: Monte Carlo Comparison of Estimators for the Tobit Errors-in-Variables Model

(a) Different sample size when $\rho=0.6$, $r=1.5$ and $p=30\%$.

	n=100			n=200			n=500		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.579	0.059	0.582	-0.578	0.042	0.579	-0.581	0.026	0.582
IV	-0.283	0.165	0.327	-0.288	0.103	0.306	-0.296	0.066	0.303
RR	0.026	0.239	0.241	0.019	0.153	0.154	0.007	0.098	0.098

(b) Different percentage of censored when $n=200$, $\rho=0.6$ and $r=1.5$.

	p=15%			p=30%			p=50%		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.488	0.046	0.490	-0.578	0.042	0.579	-0.698	0.038	0.699
IV	-0.137	0.115	0.179	-0.288	0.103	0.306	-0.490	0.086	0.498
RR	0.017	0.138	0.139	0.019	0.153	0.154	0.022	0.179	0.180

(c) Different values of ρ when $n=200$, $r=1.5$ and $p=30\%$

	$\rho=0.40$			$\rho=0.60$			$\rho=0.90$		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.578	0.042	0.579	-0.578	0.042	0.579	-0.578	0.042	0.579
IV	-0.274	0.168	0.321	-0.288	0.103	0.306	-0.295	0.068	0.303
RR	0.039	0.244	0.248	0.019	0.153	0.154	0.011	0.104	0.105

(d) Different signal-noise ratios when $n=200$, $\rho=0.6$ and $p=30\%$

	r=1.00			r=1.50			r=2.00		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.648	0.040	0.649	-0.578	0.042	0.579	-0.531	0.043	0.533
IV	-0.285	0.116	0.307	-0.288	0.103	0.306	-0.289	0.097	0.305
RR	0.024	0.170	0.172	0.019	0.153	0.154	0.017	0.143	0.144

TABLE 1: (continued)

(e) Non-normal distributions of w and z when $\rho=0.6$ and $r=1.5$

	t(5)					
	n=200			n=500		
	bias	se	rmse	bias	se	rmse
MLE	-0.592	0.055	0.595	-0.594	0.037	0.595
IV	-0.307	0.118	0.329	-0.318	0.078	0.327
RR	-0.029	0.169	0.172	-0.045	0.110	0.119

	$\chi^2(10)$					
	n=200			n=500		
	bias	se	rmse	bias	se	rmse
MLE	-0.541	0.061	0.545	-0.542	0.036	0.543
IV	-0.208	0.233	0.312	-0.228	0.131	0.263
RR	0.197	0.364	0.414	0.164	0.202	0.261

(f) Dummy IV when $n=500$, $\rho=0.6$, $r=1.5$ and $p=30\%$

	Dummy (h=4)			Dummy (h=6)			Dummy (h=8)		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.581	0.026	0.582	-0.581	0.026	0.582	-0.581	0.026	0.582
IV	-0.301	0.075	0.310	-0.308	0.071	0.316	-0.312	0.068	0.320
RR	-0.017	0.116	0.117	-0.027	0.111	0.114	-0.034	0.107	0.112

Notes:

(1) The data are generated by

$$y_i^* = \alpha + \beta x_i^* + u_i \quad \text{where } \alpha = 0.5 \text{ and } \beta = 1.0$$

$$y_i = I(y_i^* > 0), \quad x_i = x_i^* + v_i, \quad x_i^* = \mu_x + w_i, \quad \text{where } \mu_x = 1.0$$

the variance-covariance matrix of (u_i, v_i, w_i, z_i) is given in equation (17) of the text.

- (2) “ ρ ” stands for the correlation between w and z, “ p ” the percentage of truncation, “ n ” the sample size, “ r ” the signal-noise ratio= $\text{var}(w)/\text{var}(v)$, and “ h ” the number of groups in the categorical IV.
- (3) “bias”, “se”, and “rmse” stand for the bias, the standard error, and the root mean square error of each estimates, respectively.
- (4) Non-normal w and z in (e) are generated from a bivariate Student t-distribution in case of t(5), and a bivariate Wishart distribution in case of $\chi^2(10)$, after adjusted for having the first and second moments given in (17). Unlike other cases, the percentage of censoring or truncation is not necessarily 30% in (e).
- (5) Dummy IV case (f) is based on the baseline simulation model (with $n=500$, $\rho=0.6$, $r=1.5$, and $p=30\%$) with z replaced by a $h \times 1$ vector of dummy variables $\tilde{\mathbf{Z}}$. The latter is created as follows: First, the real line (the support of z) is partitioned into h intervals. “h” stands for the number of the intervals. The partition is given by $(-\infty, -1.5, 0, 1.5, \infty)$ for $h=4$, $(-\infty, -2.0, -1.0, 0, 1.0, 2.0, \infty)$ for $h=6$, and $(-\infty, -2.1, -1.4, -0.7, 0, 0.7, 1.5, \infty)$ for $h=8$. Second, $\tilde{\mathbf{Z}}$ is obtained by setting its j-th element equal to unity and all other elements equal to zeroes when z falls into the j-th interval. The actual correlation in each case is not necessarily equal to 0.6.

TABLE 2: Monte Carlo Comparison of Estimators for the Probit Errors-in-Variables Model

(a) Different sample size when $\rho=0.6$, $r=1.5$ and $p=30\%$.

	n=100			n=200			n=500		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.333	0.160	0.370	-0.358	0.105	0.373	-0.364	0.064	0.370
IV	-0.710	0.076	0.714	-0.713	0.048	0.715	-0.714	0.033	0.715
RR	-0.156	0.217	0.270	-0.171	0.134	0.217	-0.176	0.091	0.198

(b) Different percentage of censored when $n=200$, $\rho=0.6$ and $r=1.5$.

	p=15%			p=30%			p=50%		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.354	0.128	0.376	-0.358	0.105	0.373	-0.353	0.101	0.367
IV	-0.809	0.043	0.810	-0.713	0.048	0.715	-0.669	0.053	0.671
RR	-0.177	0.167	0.243	-0.171	0.134	0.217	-0.167	0.132	0.213

(c) Different values of ρ when $n=200$, $r=1.5$ and $p=30\%$

	$\rho=0.40$			$\rho=0.60$			$\rho=0.90$		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.356	0.107	0.371	-0.358	0.105	0.373	-0.358	0.102	0.372
IV	-0.711	0.080	0.715	-0.713	0.048	0.715	-0.715	0.030	0.715
RR	-0.163	0.226	0.279	-0.171	0.134	0.217	-0.178	0.078	0.194

(d) Different signal-noise ratios when $n=200$, $\rho=0.6$ and $p=30\%$

	r=1.00			r=1.50			r=2.00		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.493	0.089	0.501	-0.358	0.105	0.373	-0.258	0.117	0.283
IV	-0.712	0.053	0.714	-0.713	0.048	0.715	-0.714	0.045	0.715
RR	-0.167	0.147	0.223	-0.171	0.134	0.217	-0.172	0.127	0.214

TABLE 2: (continued)

(e) Non-normal distributions of w and z when $\rho=0.6$ and $r=1.5$

	t(5)					
	n=200			n=500		
	bias	se	rmse	bias	se	rmse
MLE	-0.410	0.104	0.423	-0.417	0.062	0.422
IV	-0.737	0.054	0.739	-0.741	0.034	0.742
RR	-0.221	0.155	0.270	-0.236	0.101	0.257

	$\chi^2(10)$					
	n=200			n=500		
	bias	se	rmse	bias	se	rmse
MLE	-0.535	0.087	0.542	-0.541	0.054	0.544
IV	-0.765	0.109	0.772	-0.770	0.056	0.772
RR	-0.353	0.296	0.461	-0.370	0.151	0.399

(f) Dummy IV when $n=500$, $\rho=0.6$, $r=1.5$ and $p=30\%$

	Dummy (h=4)			Dummy (h=6)			Dummy (h=8)		
	bias	se	rmse	bias	se	rmse	bias	se	rmse
MLE	-0.364	0.064	0.370	-0.364	0.064	0.370	-0.364	0.064	0.370
IV	-0.716	0.036	0.717	-0.719	0.034	0.720	-0.721	0.033	0.722
RR	-0.203	0.109	0.231	-0.208	0.101	0.232	-0.215	0.101	0.237

Notes: See the notes of TABLE 1.

In summary, the RR estimator is proved to be very effective in reducing bias of the estimated slope coefficient caused by measurement errors in the regressors.

6. EXAMPLE

In this section we apply the estimation procedure developed in the previous sections to a simple model of the labor supply behavior of married women. There

have been extensive studies by economists on this subject in the 1970s and early 1980s [see Killingsworth (1983) for a comprehensive survey]. Here we take up the estimation problem of women's labor force participation decisions. A simplified version of the model postulates that the i -th woman works if and only if her market wage W_i is higher than or equal to her reservation wage W_i^r where the latter is defined as the minimum level of wage at which she is willing to work. Of course the market wage is observed only when the woman works, while the reservation wage is not directly observable in any case. We can, however, relate these wages to the observable variables via

$$\begin{aligned} W_i &= \bar{\mathbf{x}}_{1i}' \boldsymbol{\theta}_1 + u_{1i} \\ W_i^r &= \bar{\mathbf{x}}_{2i}' \boldsymbol{\theta}_2 + u_{2i} \end{aligned}$$

where $\bar{\mathbf{x}}_{ji} = [\mathbf{1}, \mathbf{x}_{ji}]'$ for $j = 1, 2$, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are vectors of unknown parameters and u_1 and u_2 are disturbances. The components of \mathbf{x}_1 are the variables included in the standard wage generating equation such as age, education, and experience, while \mathbf{x}_2 includes variables such as the number of children and the family income other than the wife's, as well as age and education. The woman's labor force participation decision is, therefore, described as follows: the i -th woman works if

$$W_i > W_i^r \Leftrightarrow u_i > -\bar{\mathbf{x}}_i^* \boldsymbol{\theta} = -\alpha - \mathbf{x}_i^* \boldsymbol{\beta}$$

where $u_i = u_{1i} - u_{2i}$ and $\bar{\mathbf{x}}_i^* \boldsymbol{\theta} = \bar{\mathbf{x}}_{1i}' \boldsymbol{\theta}_1 - \bar{\mathbf{x}}_{2i}' \boldsymbol{\theta}_2$. Assuming the joint normality of the distributions of u_1 and u_2 leads to the standard Probit likelihood $L(\boldsymbol{\theta}, \sigma) = \Phi_i^{y_i} (1 - \Phi_i)^{1-y_i}$ where $\Phi_i = \Phi((\alpha + \mathbf{x}_i^* \boldsymbol{\beta}) / \sigma)$, σ is the standard deviation of u_i (usually set equal to unity for normalization), and y_i is the labor

force participation variable ($y_i = 1$ if the i -th woman works and $y_i = 0$ otherwise). The model can also be described in a latent variable model form as (1) and $y_i = 1(y_i^* > 0)$. Our problem is to estimate α and β when some of the variables in \mathbf{x}_i^* are mismeasured in the manner given in (3).

The data used here contain 753 observations on married white women aged 30 to 60, extracted from the 1975 Panel Study of Income Dynamics data (Mroz 1987). Table 3 provides definitions and simple descriptive statistics for the variables used. Table 4 reports the estimation results. For comparison purpose, we first present three sets of estimates in columns 2 to 4, ignoring any measurement errors in \mathbf{x}_i . They are (i) ordinary least squares ('OLS') which regresses y on \mathbf{x} directly, (ii) the recentered and rescaled version of OLS ('RROLS'), and (iii) the Probit maximum likelihood estimates ('MLE'). RROLS uses $\mathbf{z} = \mathbf{x}$ as instruments. It is observed that RROLS is much closer to MLE than OLS is despite the apparent actual deviation from the normality assumption of the explanatory variables. The last two columns assume that the education variable contains measurement errors and report the instrumental variables estimate ('IV') and the recentered and rescaled generalized method of moments estimate ('RRGMM') given in Proposition 1 with $\hat{\mathbf{H}} = \hat{\mathbf{A}}^{-1}$. Both estimates use the education variables of the woman's mother and father as instruments. The latter two variables might also be measured with errors but can still serve as instruments as long as those errors are uncorrelated with hers. The absolute values of the estimated coefficients are found, except for the intercept, to be all larger with RRGMM than with IV. This is the same pattern as is displayed between OLS and RROLS. Comparison of RRGMM with RROLS or MLE leads to the conclusion that the positive effect of a woman's education on her working decision might be smaller than that which is estimated when measurement errors in education variables are not considered.

TABLE 3: List of Variables for Married Women Data

variable	description	mean	std	max	min
LFP	Labor force participation (1 if woman works, and 0 otherwise)	0.568	0.496	1	0
Age	Woman's age	42.5	8.07	60	30
Education	Woman's years of schooling	12.3	2.28	17	5
Kids under 6	Number of children of age lower than 6 years	0.238	0.524	3	0
Kids 6-18	Number of children of age between 6 and 18	1.35	1.32	8	0
Mother Ed	Years of schooling of woman's mother	9.3	3.37	17	0
Father Ed	Years of schooling of woman's father	8.8	3.57	17	0

Notes: The labels "mean," "std," "max" and "min" stand for, respectively, the mean, standard deviation, maximum, and minimum values of observation for each variable.

TABLE 4: Estimation of Woman's Labor Force Participation

	OLS	RROLS	MLE	IV	RRGGM
Constant	0.712 (0.155)	0.539 (0.395)	0.624 (0.465)	0.869 (0.248)	0.942 (0.633)
Age	-0.013 (0.002)	-0.034 (0.006)	-0.038 (0.007)	-0.014 (0.002)	-0.035 (0.006)
Education	0.042 (0.007)	0.107 (0.018)	0.120 (0.022)	0.031 (0.015)	0.078 (0.039)
Kids under 6	-0.307 (0.033)	-0.783 (0.083)	-0.886 (0.113)	-0.304 (0.033)	-0.774 (0.084)
Kids 6-18	-0.018 (0.014)	-0.045 (0.036)	-0.056 (0.040)	-0.020 (0.014)	-0.049 (0.036)

Notes: The figures in parentheses are the heteroskedasticity consistent standard errors (Eicker 1963, White 1980).

7. CONCLUSION

This article provides a recentered and rescaled instrumental variable estimation procedure for the errors-in-variables Tobit and Probit models that guarantees consistency when the true dependent variable and IV's are jointly normal.

Our Monte Carlo experiment seems to suggest that the proposed estimator is very effective in reducing bias of parameter estimates in small samples under normality and is quite robust to certain types of deviation from normality. If this is generally the case, its simplicity and ease of construction makes the proposed estimator practically useful for quick and preliminary investigations of Tobit and Probit models involving errors in variables.

ACKNOWLEDGMENTS

I would like to thank an anonymous referee, De-Min Wu and Joe Sicilian for their helpful comments and Yi Geng for his excellent computational assistance.

REFERENCES

- Amemiya, Y. (1985), "Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model," *Journal of Econometrics*, 28, 273--289.
- Anderson, T. W. (1984), "Estimating Linear Statistical Relationship," *The Annals of Statistics*, 12, 1--45.
- Bekker, P. A. (1986), "Comments on Identification in the Linear Errors in Variables Model," *Econometrica*, 54, 215--217.
- Brillinger, D. R. (1982), "A Generalized Linear Model with 'Gaussian' Regressor Variables," in *A Festschrift for Erich L. Lehmann*, eds. P. J. Bickel, K. A. Doksum and J. L. Hodges, Belmont, CA: Woodsworth International Group, pp. 97--114.
- Buzas, J. S. and Stefanski, L. A. (1996), "Instrumental Variable Estimation in a Probit Measurement Error Model," *Journal of Statistical Planning and Inference*, 55, 47--62.
- Carroll, R. J., Spiegelman, C. H., Lan, K. K., Bailey, K. T., and Abbot, R. D. (1982), "On Errors-in-Variables for Binary Regression Models," *Biometrika*, 71, 19--25.

- Carroll, R. J., and Stefanski, L. A. (1990), "Approximate Quasi-Likelihood Estimation in Models with Surrogate Predictors," *Journal of the American Statistical Association*, 85, 652--663.
- Carroll, R. J., D. Ruppert, and L. A. Stefanski (1995), *Measurement Error in Nonlinear Models*, London: Chapman and Hall.
- Carter, R. L. and Fuller, W. A. (1980), "Instrumental Variables Estimation of the Simple Errors-in-Variables Model," *Journal of the American Statistical Association*, 75, 687--692.
- Chung, C.-F., and Goldberger, A. S. (1984), "Proportional Projections in Limited Dependent Variable Models," *Econometrica*, 52, 531--534.
- Durbin, J. (1954), "Errors in Variables," *International Statistical Review*, 45, 117--127.
- Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Annals of Mathematical Statistics*, 34, 447--456.
- Fuller, W. A. (1987), *Measurement Error Models*, New York, NY: Wiley.
- Greene, W. H. (1981), "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," *Econometrica*, 49, 505--513.
- Greene, W. H. (1983), "Estimation of the Limited Dependent Variable Models by the Method of Moments," *Journal of Econometrics*, 21, 195--212.
- Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029--1054.
- Hausman, J., Ichimura, H., Newey, W., and Powell, J. (1991), "Identification and Estimation of Polynomial Errors-in-Variables Models," *Journal of Econometrics*, 50, 273--295.
- Hsiao, C. (1989), "Consistent Estimation for Some Nonlinear Errors-in-Variables Models," *Journal of Econometrics*, 41, 159--185.
- Hsiao, C. (1991), "Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data," *Review of Economic Studies*, 58, 717--731.

- Hsiao, C. (1992), "Nonlinear Latent Variable Models," in *The Econometrics of Panel Data*, eds. Matyas, L., and Sevestre, P. Kluwer Academic Publishers, pp. 242--261.
- Iwata, S. (1992a), "Instrumental Variables Estimation in Errors-in-Variables Models When Instruments are Correlated with Errors," *Journal of Econometrics*, 53, 297--322.
- Iwata, S. (1992b), "Errors-in-Variables Regression Using Estimated Latent Variables," *Econometric Reviews*, 11, 195--200.
- Kapteyn, A., and Wansbeek, T. (1984), "Errors in Variables: Consistent Adjusted Least Squares (CALS) Estimator," *Communications in Statistics--Theory and Methods*, 13, 1811--1837.
- Killingsworth, M. R. (1983), *Labor Supply*, New York: Cambridge University Press.
- Leamer, E.E. (1978), "Least Squares Versus Instrumental Variables Estimation in a Simple Errors in Variables Model," *Econometrica*, 46, 961--968.
- Mroz, T. A. (1987), "The Sensitivity of an Empirical Model of Married Woman's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765--799.
- Sargan, J. D. (1958), "The Estimation of Economic Relationship Using Instrumental Variables," *Econometrica*, 26, 393--415.
- Stapleton, D. C., and Young, D. J. (1984), "Censored Normal Regression with Measurement Error on the Dependent Variable," *Econometrica*, 52, 737--760.
- Stefanski, L. A., and Carroll, R. J. (1985), "Covariate Measurement Error in Logistic Regression," *The Annals of Statistics*, 13, 1335--1351.
- Stefanski, L. A., and Buzas, J. S. (1995), "Instrumental Variable Estimation in Binary Regression Measurement Error Models," *Journal of the American Statistical Association*, 90, 541--550.
- Wolter, K. M., and Fuller, W. A. (1982), "Estimation of Nonlinear Errors-in-Variables Models," *The Annals of Statistics*, 10, 539--548.

White, H. (1980), "A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817--838.