Journal of Econometrics 39 (1988) 347–366. North-Holland

LIMITED INFORMATION ESTIMATORS AND EXOGENEITY TESTS FOR SIMULTANEOUS PROBIT MODELS*

Douglas RIVERS

University of California, Los Angeles, CA 90069, USA

Quang H. VUONG

University of Southern California, Los Angeles, CA 90089, USA

Received December 1984, final version received January 1988

A two-step maximum likelihood procedure is proposed for estimating simultaneous probit models and is compared to alternative limited information estimators. Conditions under which each estimator attains the Cramer–Rao lower bound are obtained. Simple tests for exogeneity based on the new two-step estimator are proposed and are shown to be asymptotically equivalent to one another and to have the same local asymptotic power as classical tests based on the limited information maximum likelihood estimator. Finite sample comparisons between the new and alternative estimators are presented based on some Monte Carlo evidence. The performance of the proposed tests for exogeneity is also assessed.

1. Introduction

In this paper we investigate the properties of various estimators for probit models where some or all of the explanatory variables may be endogenous. A special case of this problem, which has received considerable attention, occurs when one endogenous variable in a simultaneous equation model with normally distributed errors is observed only with respect to sign. Heckman (1978) observed that maximum likelihood estimation of the structural parameters is computationally difficult and proposed a two-stage least squares analog which can be computed using standard probit and regression programs. Amemiya (1978) suggested alternative estimators based on a general method of obtaining structural parameter estimates from reduced form parameter estimates. Amemiya also showed that Heckman's estimator could be interpreted as a member of this class, but that another member of the class (G2SP) improves on the efficiency of the Heckman estimator though it involves an increase in the computational burden. Lee (1981) suggested a more straightforward version of the Heckman estimator (IVP), but showed that it is less efficient than G2SP, though somewhat easier to compute.

*This research was partially supported by National Science Foundation Grants SES-8309994, SES-8410593, and SES-23638973. An earlier version was presented at the World Congress of the Econometric Society, Boston, 1985. We are grateful to two referees for helpful remarks. This paper is dedicated to some of our former colleagues at Caltech.

At the risk of confusing matters further, we propose another estimator, two-stage conditional maximum likelihood (2SCML), which was introduced in Vuong (1984), and which was independently considered by Smith and Blundell (1986) for the Tobit model. The new estimator has several advantages over the Heckman and Amemiya estimators. In particular, it is easier to compute than G2SP and in some cases asymptotically more efficient, too, though a general efficiency ordering between the estimators is not possible. In small samples, our Monte Carlo evidence suggests that the new estimator performs favorably relative to its competitors. Another substantial advantage is that the 2SCML procedure incorporates simple and asymptotically optimal tests for the exogeneity of the explanatory variables.

A unifying perspective on the various estimators is provided by placing the estimation problem, which has previously been viewed in an *ad hoc* fashion, in a likelihood framework. This allows us to go beyond the Smith and Blundell (1986) paper and to compare the new estimator to existing limited information estimators which do not impose any restrictions on the reduced form equations for the explanatory variables. In particular, we derive the Cramer–Rao bound for limited information estimators and we give conditions under which each estimator attains this bound. The question of efficient estimation is closely related to the construction of optimal tests for exogeneity in probit models. Specifically, the efficiency properties of the 2SCML estimator enables us to construct analogs of the Wald, likelihood ratio, score and Hausman tests based on the conditional likelihood function which have the same asymptotic properties as the classical LIML tests under the null hypothesis of exogeneity and local alternatives.

The notation and assumptions used in the paper are stated in section 2. Section 3 describes limited information estimation methods for the model, including the new two-stage conditional maximum likelihood estimator. The relative asymptotic efficiency of the estimators is compared in section 4. Tests for exogeneity of the explanatory variables are proposed and compared in section 5. Some Monte Carlo evidence on the finite sample properties of the various estimators as well as of some tests for exogeneity are presented in section 6. Section 7 concludes the paper. Proofs can be found in Rivers and Vuong (1984).

2. Model and regularity conditions

The model is composed of a structural equation that is of primary interest and a set of reduced form equations for the endogenous explanatory variables:

$$y_i^* = Y_i'\gamma + X_{1i}'\beta + u_i, \tag{2.1}$$

$$i = 1, \ldots, n,$$

$$Y_i = \Pi' X_i + V_i, \tag{2.2}$$

where Y_i, X_{1i}, and X_i are $m \times 1$, $k \times 1$, and $p \times 1$ vectors, respectively, with X_i and X_{1i} related by the identity

$$X_{1i} = J'X_i, \tag{2.3}$$

where J is the appropriate selection matrix. Only the sign of y_i^* is observed:

$$\begin{aligned} y_i &= 1 \quad \text{if} \quad y_i^* > 0, \\ &= 0 \quad \text{if} \quad y_i^* \leq 0. \end{aligned} \tag{2.4}$$

The following assumptions are made:

Assumption 1. (X_i, u_i, V_i) is i.i.d. with X_i having finite positive definite covariance matrix Σ_{xx} and u_i and V_i having, conditional on X_i, a joint normal distribution with mean zero and finite positive definite covariance matrix:[1]

$$\Omega \equiv \begin{bmatrix} \sigma_{uu} & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_{vv} \end{bmatrix}.$$

Assumption 2. (Identification) rank $(\Pi, J) = m + k$.

Assumption 3. (Parameter space) $(\gamma, \beta, \Pi, \Omega)$ is known to lie in the interior of a compact convex subset of a Euclidean space, denoted Θ.

The model presented here contains, as a special case, Heckman's (1978) hybrid model without structural shift. Heckman treats the case $m = 1$ and writes eq. (2.2) in structural form rather than reduced form. For the purpose of limited information estimation we ignore any *a priori* restrictions on the structural form parameters which might imply some restrictions on Π or Σ_{uv}.

3. Limited information estimation

Four methods of estimation for eqs. (2.1)–(2.2) will be considered. The first three methods (LIML, IVP, and G2SP) have been discussed elsewhere and are reviewed only briefly. The other method (2SCML) is new and is described in greater detail.

The parameters in (2.1) are not identified without a further normalization. Any normalization will be arbitrary, so the choice of normalization is made on the grounds of convenience for the proposed estimation technique. For likelihood methods, the most convenient normalization is $V(y_i^* | X_i, Y_i) = 1$. Rewrite (2.1) in the form

$$y_i^* = Y_i'\gamma + X_{1i}'\beta + V_i'\lambda + \eta_i, \tag{3.1}$$

[1] The joint normality of (u_i, V_i) given X_i is standard in this type of model. This assumption is stronger than necessary for the 2SCML estimator considered below which only requires that u_i given (V_i, X_i) be normal with mean linear in V_i and constant variance.

where $\lambda = \Sigma_{\nu\nu}^{-1}\Sigma_{\nu u}$ and $\eta_i = u_i - V_i'\lambda$. Conditional on X_i and Y_i, $\eta_i \sim \mathrm{N}(0, \sigma_{uu} - \lambda'\Sigma_{\nu\nu}\lambda)$ so the appropriate normalization is

$$\sigma_{uu} - \lambda'\Sigma_{\nu\nu}\lambda = 1. \tag{3.2}$$

The normalization (3.2) is different from that imposed by Heckman (1978), Amemiya (1978), and Lee (1981) who use the normalization $\mathrm{var}(u_i + V_i'\gamma) = 0$ [see eq. (3.5) below]. Unlike the Tobit case considered by Smith and Blundell (1986), this leads to a slight complication in the asymptotic and finite sample comparison of G2SP, IVP, and 2SCML estimators.

3.1. Limited information maximum likelihood (LIML)

The joint density for y_i and Y_i conditional on X_i is given by

$$\begin{aligned}
&h(y_i, Y_i|X_i; \gamma, \beta, \lambda, \Pi, \Sigma_w)\\
&= (2\pi)^{-(m+1)/2}|\Omega|^{-1/2}\left[\int_{c_i}^{\infty}\exp\left\{-\tfrac{1}{2}(u, V_i')\Omega^{-1}(u, V_i')'\right\}\mathrm{d}u\right]^{y_i}\\
&\quad\times\left[\int_{-\infty}^{c_i}\exp\left\{-\tfrac{1}{2}(u, V_i')\Omega^{-1}(u, V_i')'\right\}\mathrm{d}u\right]^{1-y_i}\\
&= (2\pi)^{-(m+1)/2}|\Sigma_{\nu\nu}|^{-1}\\
&\quad\times\left[\int_{c_i}^{\infty}\exp\left\{-\tfrac{1}{2}\left[u^2 - 2\lambda'V_iu + V_i'\left(\Sigma_{\nu\nu}^{-1} + \lambda\lambda'\right)V_i\right]\right\}\mathrm{d}u\right]^{y_i}\\
&\quad\times\left[\int_{-\infty}^{c_i}\exp\left\{-\tfrac{1}{2}\left[u^2 - 2\lambda'V_iu + V_i'\left(\Sigma_{\nu\nu}^{-1} + \lambda\lambda'\right)V_i\right]\right\}\mathrm{d}u\right]^{1-y_i},
\end{aligned} \tag{3.3}$$

where $c_i = -(Y_i'\gamma - X_{1i}'\beta)$ and the second equation follows from the normalization (3.1) and formulae for the inverse and determinant of the partitioned matrix Ω. An alternative expression for the joint density is given below in terms of the joint and conditional densities that takes a somewhat simpler form.

The limited information maximum likelihood estimates are obtained by maximizing the sample log likelihood,

$$L_n(\gamma, \beta, \lambda, \Pi, \Sigma_{\nu\nu}) = \sum_{i=1}^{n} \log h(y_i, Y_i|X_i; \gamma, \beta, \lambda, \Pi, \Sigma_{\nu\nu}), \tag{3.4}$$

with respect to $(\gamma, \beta, \lambda, \Pi, \Sigma_{\nu\nu})$. This approach to limited information is adapted from Godfrey and Wickens (1982). We denote the LIML estimator as

$(\hat{\gamma}^L, \hat{\beta}^L, \hat{\lambda}^L, \hat{\Sigma}^L)$. LIML suffers from a number of computational disadvantages, especially in large models. As a consequence, the LIML estimator has generally been avoided in favor of less efficient but computationally simpler estimation methods which, if desired, could provide good starting values for a LIML maximization routine.

3.2. Instrumental variables probit (IVP)

Lee (1981) suggested writing (2.1) in reduced form:

$$y_i^* = (\Pi' X_i)'\gamma + X_{1i}'\beta + u_i + V_i'\gamma. \tag{3.5}$$

Then the marginal log likelihood for y given X is

$$L_n^*(\gamma_*, \beta_*, \Pi) = \sum_{i=1}^{n} y_i \log \Phi[(\Pi' X_i)\gamma_* + X_{1i}'\beta_*] + (1 - y_i)\log[1 - \Phi((\Pi' X_i)\gamma_* + X_{1i}'\beta_*)], \tag{3.6}$$

where $\Phi(\cdot)$ denotes a standardized normal cdf and

$$\gamma_* = \gamma/\omega, \qquad \beta_* = \beta/\omega, \tag{3.7}$$

$$\omega^2 = V(u_i + V_i'\gamma) = 1 + (\gamma + \lambda)'\Sigma_{\nu\nu}(\gamma + \lambda). \tag{3.8}$$

The second equality follows from the normalization (3.2). Given consistent estimates $\hat{\Pi}$ obtained by applying ordinary least squares to (2.2), one then maximizes $L_n^*(\gamma_*, \beta_*, \hat{\Pi})$ with respect to γ_* and β_*. The resulting estimators $(\hat{\gamma}_*^{IVP}, \hat{\beta}_*^{IVP})$ are consistent and straightforward to compute, requiring m linear regressions followed by a standard probit estimation.[2]

3.3. Generalized two-stage simultaneous probit (G2SP)

Instead of maximizing (3.6) with respect to γ_* and β_* conditional on $\Pi = \hat{\Pi}$, Amemiya (1978) proposed estimating the reduced form (3.5) without imposing any constraints by maximizing

$$L_n^*(\tau_*) = \sum_{i=1}^{n} y_i \log \Phi(X_i'\tau_*) + (1 - y_i)\log[1 - \Phi(X_i'\tau_*)], \tag{3.9}$$

with respect to τ_*. Let $\hat{\tau}_*$ denote the corresponding estimator of τ_* and, as before, $\hat{\Pi}$ the least squares estimator of Π. From (3.6) and (3.9),

$$\tau_* = \Pi\gamma_* + J\beta_*. \tag{3.10}$$

[2] Heckman originally proposed solving (2.1) for one of the observed endogenous variables, replacing y_i^* by an estimate of $E(y_i^* \mid X_i)$ [e.g., $X_i'\hat{\tau}_*$ from (3.9) below] and applying least squares.

Replacing τ_* and Π by their sample estimates in (3.10) yields

$$\hat{\tau}_* = (\hat{\Pi}, J)\begin{bmatrix}\gamma_* \\ \beta_*\end{bmatrix} + (\hat{\tau}_* - \tau_*) - (\hat{\Pi}_* - \Pi)\gamma_* = \hat{H}\begin{bmatrix}\gamma_* \\ \beta_*\end{bmatrix} + e, \quad (3.11)$$

where $\hat{H} = (\hat{\Pi}, J)$ and $e = (\hat{\tau}_* - \tau_*) - (\hat{\Pi} - \Pi)\gamma_*$. The estimation problem has been recast in the form of a linear regression. Ordinary least squares applied to (3.11) gives consistent estimates of γ_* and β_*, but more efficient estimates can be obtained via generalized least squares. Let $\hat{V}$ denote a consistent estimator of the asymptotic covariance matrix of e.[3] The Amemiya G2SP estimator is defined by

$$\begin{bmatrix}\hat{\gamma}_*^A \\ \hat{\beta}_*^A\end{bmatrix} = (\hat{H}'\hat{V}^{-1}\hat{H})^{-1}\hat{H}'\hat{V}^{-1}\hat{\tau}_*. \quad (3.12)$$

Since the covariance matrix of e depends on γ_* and $\lambda_* = \lambda/\omega$, to compute $\hat{V}$ Amemiya (1978, p. 1200) suggests some preliminary consistent estimates of these parameters. Then G2SP requires one more computational step [(3.12) in addition to the m reduced form regressions and one probit calculation] than two-step estimators such as IVP and 2SCML (described below). While none of these computational difficulties are prohibitive, they do appear to have made G2SP less attractive to empirical workers than IVP despite its advantage in efficiency.

3.4. Two-stage conditional maximum likelihood (2SCML)

When the joint density for a set of endogenous variables factors into a conditional distribution for one and a marginal distribution for the remaining variables, each of which takes a convenient form, then frequently estimation can be simplified by using the method of conditional maximum likelihood [Vuong (1984)]. As for the Tobit model with endogenous explanatory variables [see Smith and Blundell (1986)], the present problem is a case in point. The joint density (3.3) for y_i and Y_i factors into a probit likelihood and a normal density:

$$h(y_i, Y_i | X_i; \gamma, \beta, \lambda, \Pi, \Sigma_{vv})$$
$$= f(y_i | Y_i X_i; \gamma, \beta, \lambda, \Pi) g(Y_i | X_i; \Pi, \Sigma_{vv}), \quad (3.13)$$

[3] If V denotes the asymptotic covariance matrix of e, i.e., $n^{1/2}e \xrightarrow{D} \mathrm{N}(0, V)$, then $\hat{V} \xrightarrow{\text{a.s.}} V$ element by element. The matrix V is the matrix in brackets in (4.15).

where

$$f(y_i|Y_iX_i;\gamma,\beta,\lambda,\Pi)$$

$$=\Phi(Y_i'\gamma+X_{1i}'\beta+V_i'\lambda)^{y_i}[1-\Phi(Y_i'\gamma+X_{1i}'\beta+V_i'\lambda)]^{1-y_i}, \tag{3.14}$$

$$g(Y_i|X_i;\Pi,\Sigma_{vv})$$

$$=(2\pi)^{-m/2}|\Sigma_{vv}|^{-1/2}\exp\{-1/2(Y_i-\Pi'X_i)'\Sigma_{vv}^{-1}(Y_i-\Pi'X_i)\}. \tag{3.15}$$

The 2SCML estimator is computed in two steps. First, estimators $\hat{\Pi}$ and $\hat{\Sigma}_{vv}$ are obtained by maximizing the marginal log likelihood for Y_i,

$$L_n^g(\Pi,\Sigma_{vv})=\sum_{i=1}^n \log g(Y_i|X_i;\Pi,\Sigma_{vv}), \tag{3.16}$$

with respect to Π and Σ_{vv}. Second, the conditional log likelihood for y_i, setting $\Pi=\hat{\Pi}$, is maximized with respect to the remaining parameters:

$$L_n^f(\gamma,\beta,\lambda,\hat{\Pi})=\sum_{i=1}^n \log f\big(y_i|Y_i,X_i;\gamma,\beta,\lambda,\hat{\Pi}\big). \tag{3.17}$$

Both of these steps can be easily carried out with standard regression and probit programs:

(1) Regress Y_i on X_i to obtain $\hat{\Pi}$. $\hat{\Sigma}_{vv}$ is estimated in the usual way by $n^{-1}\sum_{i=1}^n\hat{V}_i\hat{V}_i'$ where $\hat{V}_i=Y_i-\hat{\Pi}X_i$ denotes the least squares residuals.

(2) Probit analysis of y_i with Y_i, X_{1i}, and $\hat{V}_i$ as explanatory variables provides estimates $(\hat{\gamma},\hat{\beta},\hat{\lambda})$.

In addition, a convenient feature of the procedure is that it provides an estimate of λ that can be used to construct tests for exogeneity (see section 5).

4. Asymptotic properties of limited information estimators

Each of the estimators described in the previous sections is strongly consistent and asymptotically normally distributed [Amemiya (1978), Lee (1981), Vuong (1984)]. In general, however, only the LIML estimator will attain the Cramer–Rao bound, which is given in Proposition 1 below. Let $\theta'=(\gamma',\beta',\lambda')$ and $\delta'=(\gamma',\beta')$ with the corresponding LIML estimators denoted $\hat{\theta}^L$ and $\hat{\delta}^L$, respectively. Let $Z_i'=(Y_i',X_{1i}')$.

Proposition 1 (Cramer–Rao bound for limited information estimators). *Under Assumptions 1–3,*

$$n^{1/2}(\hat{\theta}^L - \theta) \xrightarrow{\mathrm{D}} \mathrm{N}(0, V(\hat{\theta}^L)), \tag{4.1}$$

where

$$V(\hat{\theta}^L) = \left\{ \tilde{H}' \left[\tilde{\Sigma}^{-1} + \lambda' \Sigma_{vv} \lambda \begin{bmatrix} \Sigma_{xx}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right]^{-1} \tilde{H} \right\}^{-1}, \tag{4.2}$$

$$\tilde{H} = \begin{bmatrix} \Pi & J & 0 \\ I_m & 0 & I_m \end{bmatrix}, \tag{4.3}$$

$$\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{xx} & \tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vx} & \tilde{\Sigma}_{vv} \end{bmatrix}$$

$$= \mathrm{E}\left[\frac{\phi(Z_i'\delta + V_i'\lambda)^2}{\Phi(Z_i'\delta + V_i'\lambda)[1 - \Phi(Z_i'\delta + V_i'\lambda)]} \begin{bmatrix} X_i \\ V_i \end{bmatrix} \begin{bmatrix} X_i \\ V_i \end{bmatrix}' \right], \tag{4.4}$$

and ϕ is the density of the standard normal distribution.

For comparisons with other estimators, we need the asymptotic covariance matrix of $\hat{\delta}^L$ which is the upper left-hand block of $V(\hat{\theta}^L)$. It can be shown that

$$V(\hat{\delta}^L) = \left\{ H' \left[\tilde{\Sigma}^{xx} + (\lambda' \Sigma_{vv} \lambda) \Sigma_{xx}^{-1} \right]^{-1} H \right\}^{-1}, \tag{4.5}$$

where $H = [\Pi; J]$ is a submatrix of $\tilde{H}$, and $\tilde{\Sigma}^{xx}$ is the upper left-hand block of $\tilde{\Sigma}^{-1}$.

Next, we derive the asymptotic covariance matrix of the 2SCML estimator of θ.

Proposition 2 (Asymptotic properties of the 2SCML estimator). *Under Assumptions 1–3,*

$$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{\mathrm{D}} \mathrm{N}(0, V(\hat{\theta})), \tag{4.6}$$

where

$$V(\hat{\theta}) = \left\{ \tilde{H}' [\tilde{\Sigma}^{-1} + M]^{-1} \tilde{H} \right\}^{-1}, \tag{4.7}$$

where M is a matrix which is null except for its upper left-hand block which is

$$M_{xx} = (\lambda' \otimes I_p) \left[\Sigma_{vv}^{-1} \otimes \Sigma_{xx} - \lambda\lambda' \otimes S_{xx} \right]^{-1} (\lambda \otimes I_p), \tag{4.8}$$

and S_{xx} is the upper left-hand block of

$$S = \tilde{\Sigma} - \tilde{\Sigma}\tilde{H}(\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1}\tilde{H}'\tilde{\Sigma}. \tag{4.9}$$

The asymptotic covariance matrix for the 2SCML estimator of δ, denoted $\hat{\delta}$, is obtained from the upper left-hand block of $V(\hat{\theta})$. As in eq. (4.5), we obtain

$$V(\hat{\delta}) = \Big\{ H' \Big[\tilde{\Sigma}^{xx} + (\lambda' \otimes I_p) \big[\Sigma_{vv}^{-1} \otimes \Sigma_{xx} - \lambda\lambda' \otimes S_{xx} \big]^{-1} \times (\lambda \otimes I_p) \Big]^{-1} H \Big\}^{-1}. \tag{4.10}$$

Expressions (4.7) and (4.10) are directly comparable to the Cramer–Rao bounds (4.2) and (4.5) for θ and δ, respectively. In practice, the following expression for the asymptotic covariance matrix of $\hat{\theta}$ is more useful:

$$V(\hat{\theta}) = (\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1} + (\lambda'\Sigma_{vv}\lambda)(\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1} \times \tilde{H}'\tilde{\Sigma} \begin{pmatrix} \Sigma_{xx}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tilde{\Sigma}\tilde{H}(\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1}. \tag{4.11}$$

The matrix $(\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1}$ is the asymptotic covariance for θ in eq. (3.17) as if $\hat{\Pi}$ were not estimated. This matrix is therefore ordinarily estimated by any probit program. Eq. (4.11) shows that $V(\hat{\theta})$ is the sum of $(\hat{H}'\tilde{\Sigma}\tilde{H})^{-1}$ and a positive semi-definite matrix.

With these results, it is a simple matter to determine under which conditions the 2SCML estimator will be asymptotically efficient. If $\lambda = 0$, then it is evident from (4.7) that the variance of the 2SCML estimator $\hat{\theta}$ is given by

$$V_0(\hat{\theta}) = (\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1}, \tag{4.12}$$

which, by inspection of (4.2), is equal to $V(\hat{\theta}^L)$ evaluated for $\lambda = 0$. In this case 2SCML is efficient for the entire vector θ (including λ) which will turn out to be important in constructing optimal tests of exogeneity.

Another condition which is sufficient for $V(\hat{\delta}) = V(\hat{\delta}^L)$ is that $k = p - m$. In this case, H is square and, by Assumption 2, non-singular. It follows that when $k = p - m$, we have $S = 0$ and (4.7) reduces to (4.2) and $V(\hat{\theta}) = V(\hat{\theta}^L)$. Thus we have:

Corollary 1 (Efficiency of 2SCML estimators). *Under Assumptions 1–3, $V(\hat{\theta}) \geq V(\hat{\theta}^L)$ with equality if one of the following conditions holds*:

$$\text{(i)} \quad \lambda = 0 \quad or \quad \text{(ii)} \quad k = p - m.$$

When $\lambda = 0$, then Y and u are independent conditional on X, so that Y may be treated as exogenous in (2.1). Tests for $\lambda = 0$ are discussed in section 5. The other condition for efficiency of 2SCML in the simultaneous probit model is that the number of excluded exogenous variables be equal to the number of endogenous variables appearing on the right-hand side of (2.1), i.e., that eq. (2.1) be just identified. In fact, in this case $\hat{\theta}$ and $\hat{\theta}^L$ are *numerically* equal. To see this, note that $Y_i'\gamma + X_{1i}'\beta + V_i'\lambda = Y_i'\mu + X_i'\nu$ with $\mu = \gamma + \lambda$ and $\nu = J\beta - \Pi\lambda$. So given Π, the maximization of the conditional likelihood (3.14) with respect to (γ, β, λ) is equivalent to maximization with respect to (μ, ν) subject to ν being in the column space of $H = [\Pi, J]$. When $k = p - m$, H is square (and non-singular) so that there are no restrictions on ν. Reparameterization in terms of $(\mu, \nu, \Pi, \Sigma_{\nu\nu})$ immediately shows that 2SCML and LIML coincide.[4] Given that in much empirical work identification is a serious issue, the efficiency of 2SCML estimators in the just identified case is a useful property.

The IVP and G2SP estimators will also generally involve some inefficiency, though conditions under which one or both of these estimators fail to attain the Cramer–Rao bound and the amount of the efficiency loss have not been investigated previously. The asymptotic distributions of G2SP and IVP have been derived by Amemiya (1978) and Lee (1981):

$$n^{1/2}(\hat{\delta}_*^A - \delta_*) \xrightarrow{\mathrm{D}} \mathrm{N}(0, V(\hat{\delta}_*^A)), \tag{4.13}$$

$$n^{1/2}(\hat{\delta}_*^{IVP} - \delta_*) \xrightarrow{\mathrm{D}} \mathrm{N}(0, V(\hat{\delta}_*^{IVP})), \tag{4.14}$$

where

$$V(\hat{\delta}_*^A) = \left\{ H'\left[\bar{\Sigma}_{xx}^{-1} - (\gamma_*'\Sigma_{\nu\nu}\gamma_* + 2\gamma_*'\Sigma_{\nu\nu}\lambda_*)\Sigma_{xx}^{-1}\right]^{-1} H \right\}^{-1}, \tag{4.15}$$

$$\begin{aligned} V(\hat{\delta}_*^{IVP}) = {} & (H'\bar{\Sigma}_{xx}H)^{-1} \\ & \times H'\bar{\Sigma}_{xx}\left[\bar{\Sigma}_{xx}^{-1} - (\gamma_*'\Sigma_{\nu\nu}\gamma_* + 2\gamma_*'\Sigma_{\nu\nu}\lambda_*)\Sigma_{xx}^{-1}\right] \\ & \times \bar{\Sigma}_{xx}H(H'\bar{\Sigma}_{xx}H)^{-1}, \end{aligned} \tag{4.16}$$

$$\bar{\Sigma}_{xx} = \mathrm{E}\left[\frac{\phi(X_i'\tau_*)^2}{\Phi(X_i'\tau_*)[1 - \Phi(X_i'\tau_*)]} X_i X_i' \right]. \tag{4.17}$$

Note that $\bar{\Sigma}_{xx}$, which is Amemiya's (1978) matrix A, is not in general equal to $\tilde{\Sigma}_{xx}$ which appears in the Cramer–Rao bound (4.2).

Lee (1981) showed that IVP is never more efficient than G2SP, though there are cases in which the estimators are asymptotically equivalent. In particular,

[4] See also Newey (1987).

when eq. (2.1) is just identified, IVP and G2SP are *numerically* identical.[5] In general, however, G2SP will not be efficient. Since the G2SP estimator of δ^* is not optimal within the limited information class, a comparison between IVP and G2SP is less interesting than one between G2SP (or IVP) and LIML. Unfortunately, the normalization used by Amemiya and Lee does not allow a direct comparison of $V(\hat{\delta}_*^A)$ and $V(\hat{\delta}_*^{IVP})$ with the Cramer–Rao bound for δ derived in Proposition 1. The bound for δ^*, denoted $V(\hat{\delta}_*^L)$, is given in Proposition 3.

Proposition 3. Let $\hat{\delta}_*^L = \hat{\delta}^L/[1 + (\hat{\lambda}^L + \hat{\gamma}^L)'\hat{\Sigma}_{\nu\nu}^L(\hat{\lambda}^L + \hat{\gamma}^L)]^{1/2}$ *denote the LIML estimator of* δ_*. *Then* $n^{1/2}(\hat{\delta}_*^L - \delta_*) \xrightarrow{D} \mathrm{N}(0, V(\hat{\delta}_*^L))$, *where*

$$\begin{aligned} V(\hat{\delta}_*^L) = & \frac{1}{\omega^2} V\left[\hat{\delta}^L - (\gamma_*' + \lambda_*')\Sigma_{\nu\nu}(\hat{\gamma}^L + \hat{\lambda}^L)\delta_*\right] \\ & + \tfrac{1}{4}((\gamma_*' + \lambda_*') \otimes (\gamma_*' + \lambda_*'))V(\operatorname{vec}\hat{\Sigma}_{\nu\nu}^L) \\ & \times ((\gamma_* + \lambda_*) \otimes (\gamma_* + \lambda_*))\delta_*\delta_*'. \end{aligned} \tag{4.18}$$

The first term in eq. (4.18) is the asymptotic covariance matrix of a particular linear combination of $\hat{\theta}^L$, and hence can be obtained from the Cramer–Rao bound (4.2). The second term uses the Cramer–Rao bound for $\operatorname{vec}\hat{\Sigma}_{\nu\nu}^L$ which can be found in Richard (1975).

When $\gamma + \lambda = 0$, it follows that the Cramer–Rao bound for δ_* reduces to $V(\hat{\delta}^L)/\omega^2$, i.e., the bound for δ divided by ω^2. It turns out that under this condition, Amemiya's G2SP estimator is asymptotically efficient as the following corollary states.

Corollary 2. If $\gamma + \lambda = 0$, *then* $V(\hat{\delta}_*^A) = V(\hat{\delta}_*^L)$.

It should be noted that this condition is not sufficient for the efficiency of IVP.

If, however, $\lambda = 0$ but $\gamma \neq 0$, $\hat{\delta}_*^A$ does not necessarily attain the Cramer–Rao bound for δ_*, while under the same conditions, as shown in Corollary 1, the 2SCML estimator $\hat{\delta}$ will be efficient for δ. Similarly, when the model is just-identified, G2SP will generally not attain the Cramer–Rao bound while 2SCML will be efficient. On the other hand, if $\gamma + \lambda = 0$ but $\lambda \neq 0$, then 2SCML may fail to attain the Cramer–Rao bound while G2SP will be efficient. Hence, no general efficiency ordering between the estimators is

[5]If (2.1) is just identified, $\hat{H}$ is (almost surely) non-singular, so $\hat{\delta}_*^A = \hat{H}^{-1}\hat{\tau}_*$. Also the columns of $X\hat{\Pi}$ and X_1 span the column space of X, so $\hat{\delta}^{IVP}$ is also a non-singular transformation of $\hat{\tau}_*$. Hence, $\hat{\delta}_*^A = \hat{\delta}_*^{IVP}$

possible. In contrast, in the case of linear simultaneous equations (i.e., if y^* were observed), Holly and Sargan (1982) have shown that G2SP, IVP, and 2SCML are all numerically equivalent. Since each estimator may be viewed as an extension of the two-stage least squares estimation principle to the simultaneous probit model, our analysis shows that they may not even maintain their asymptotic equivalence when non-linearity is present.

5. Tests of exogeneity

When Y_i and u_i are correlated, the usual probit estimator of (2.1) is inconsistent for γ and β so it is necessary to resort to one of the estimators discussed above. If $\Sigma_{vu} = 0$ or, equivalently, $\lambda = 0$, then Y_i can be treated as exogenous in (2.1). In this section we propose three tests for exogeneity of Y_i based on classical principles as well as three of the Hausman (1978) variety.[6] All tests are based on the 2SCML estimator rather than the LIML estimator so, strictly speaking, these are not classical tests. However, as Corollary 1 indicated, under the null hypothesis H_0: $\lambda = 0$, 2SCML is asymptotically equivalent to LIML. This enables us to show that, under H_0, the 2SCML-based tests are asymptotically equivalent to the classical tests. The 2SCML-based test statistics might be preferred over test statistics based on LIML or other estimators on the grounds of computational convenience. In fact, the tests statistics discussed below are readily calculated from information routinely produced by probit programs.

Analogous to the usual Wald, likelihood ratio, and gradient tests based on the joint likelihood (3.4) for (y_i, Y_i), we construct similar tests based on the conditional likelihood (3.17) for y_i given Y_i. The modified Wald statistic is given by

$$MW = n\hat{\lambda}'\hat{V}_0(\hat{\lambda})^{-1}\hat{\lambda}, \tag{5.1}$$

where $\hat{V}_0(\hat{\lambda})$ is a consistent estimator of the lower righthand block of $V_0(\hat{\theta}) = (\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1}$ corresponding to $\hat{\lambda}$. The modified Wald statistic differs from the usual Wald statistic in two respects. First, λ is estimated by 2SCMLE instead of LIML. Second, the covariance matrix of $\hat{\lambda}$ is estimated under the null rather than the alternative. See (4.7) and (4.11). The conditional score statistic is given by

$$CS = \frac{1}{n}\frac{\partial L_n^f(\tilde{\gamma},\tilde{\beta},0,\hat{\Pi})}{\partial\lambda'}\hat{V}_0(\hat{\lambda})\frac{\partial L_n^f(\tilde{\gamma},\tilde{\beta},0,\hat{\Pi})}{\partial\lambda}, \tag{5.2}$$

[6] See also Newey (1985) and Smith and Blundell (1986). Smith and Blundell (1986) only considered the MW and CS statistics for constructing tests for exogeneity in the Tobit case. Newey (1985, sec. 4) considered the score test in the probit case.

where $(\tilde{\gamma}, \tilde{\beta})$ is the usual maximum likelihood probit estimator of (2.1). The conditional likelihood ratio statistic is given by

$$CLR = 2\left[L_n^f(\hat{\gamma}, \hat{\beta}, \hat{\lambda}, \hat{\Pi}) - L_n^f(\tilde{\gamma}, \tilde{\beta}, 0, \hat{\Pi})\right]. \tag{5.3}$$

We also consider some Hausman type tests. Let $\hat{\delta}' = (\hat{\gamma}', \hat{\beta}')$ and $\tilde{\theta}' = (\tilde{\gamma}', \tilde{\beta}', 0')$. Under H_0, θ is efficiently estimated by the usual probit estimator $\tilde{\theta}$, while under the alternative the 2SCML estimator $\hat{\theta}$ will be consistent though possibly inefficient.[7] Three different Hausman statistics can be formed using various subsets of the parameters:

$$M_1 = n(\hat{\gamma} - \tilde{\gamma})'\hat{G}_1(\hat{\gamma} - \tilde{\gamma}), \tag{5.4}$$

$$M_2 = n(\hat{\delta} - \tilde{\delta})'\hat{G}_2(\hat{\delta} - \tilde{\delta}), \tag{5.5}$$

$$M_3 = n(\hat{\theta} - \tilde{\theta})'\hat{G}_3(\hat{\theta} - \tilde{\theta}), \tag{5.6}$$

where $\hat{G}_i$ is a consistent estimator of a g-inverse [Rao and Mitra (1971)] of the appropriate submatrix of $V_0(\hat{\theta}) - V_0(\tilde{\theta})$, $V_0(\hat{\theta})$ is given by (4.12),

$$V_0(\tilde{\theta}) = \begin{bmatrix} (\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{5.7}$$

and $\tilde{H}_1$ is composed of the first $m + k$ columns of $\tilde{H}$ [see (4.3)].

This method for constructing quadratic forms based on consistent estimators of g-inverses is discussed in Vuong (1987).[8] In the case of the M_1 test, $V_0(\hat{\gamma}) - V_0(\tilde{\gamma})$ is non-singular, so we can choose $\hat{G}_1 = [\hat{V}_0(\hat{\gamma}) - V_0(\tilde{\gamma})]^{-1}$. For M_2,

$$V_0(\hat{\delta} - \tilde{\delta}) = V_0(\hat{\delta})\left[\tilde{H}_1'\tilde{\Sigma}_\nu\tilde{\Sigma}_{\nu\nu}^{-1}\tilde{\Sigma}_\nu'\tilde{H}_1\right]V_0(\tilde{\delta}), \tag{5.8}$$

where $\tilde{\Sigma}_\nu = [\tilde{\Sigma}_{\nu x}, \tilde{\Sigma}_{\nu\nu}]'$. The matrix in square brackets is singular. A g-inverse of $V_0(\hat{\delta} - \tilde{\delta})$ is

$$G_2 = V_0(\tilde{\delta})^{-1}\left[\tilde{H}_1'\tilde{\Sigma}_\nu\tilde{\Sigma}_{\nu\nu}^{-1}\tilde{\Sigma}_\nu'\tilde{H}_1\right]^+ V_0(\hat{\delta})^{-1}, \tag{5.9}$$

where $[\cdot]^+$ denotes the Moore–Penrose inverse. If $H_1'\Sigma_{\nu\nu}$ has full column rank,

$$\left[\tilde{H}_1'\tilde{\Sigma}_\nu\tilde{\Sigma}_{\nu\nu}^{-1}\tilde{\Sigma}_\nu'\tilde{H}_1\right]^+ = N'\tilde{\Sigma}_{\nu\nu}N, \tag{5.10}$$

[7]Note that when $\lambda = 0$, then $\sigma_{uu} = 1$ [see (3.2)]. Thus $(\tilde{\gamma}, \tilde{\beta})$ is indeed an estimator of (γ, β).

[8]The more usual method which is based on g-inverses of consistent estimators requires additional assumptions [Andrews (1987)].

where $N = V_0(\hat{\lambda})[C_0(\hat{\lambda}, \hat{\delta})V_0(\hat{\delta})^{-1}C_0(\hat{\delta}, \hat{\lambda})]^{-1}C_0(\hat{\lambda}, \hat{\delta})$ and $C_0(\cdot, \cdot)$ denotes an asymptotic covariance matrix. Thus an appropriate choice for $\hat{G}_2$ is

$$\hat{G}_2 = \hat{V}_0(\tilde{\delta})^{-1}\hat{N}'\tilde{S}_{vv}\hat{N}V_0(\hat{\delta})^{-1}, \tag{5.11}$$

where $\hat{N}$ and $\tilde{S}_{vv}$ are consistent estimators of N and $\tilde{\Sigma}_{vv}$, respectively. For M_3, a g-inverse of $V_0(\hat{\theta}) - V_0(\tilde{\theta})$ is $[V_0(\hat{\theta})]^{-1}$, so $\hat{G}_3 = -(\sqrt{n})\partial^2 L_n^f(\hat{\theta}, \hat{\Pi})/\partial\theta\partial\theta'$ can be used.

Under the null hypothesis each of the test statistics (5.1)–(5.3) and (5.6) has an asymptotic central chi-square distribution with m degrees of freedom, where m is the number of endogenous variables included in the probit equation (2.1). In fact, under the null hypothesis, the four tests are asymptotically equivalent.

Proposition 4. Under Assumptions 1 and 2 and H_0,

(i) $\operatorname*{plim}_{n\to\infty}(MW - CLR) = 0,$

(ii) $\operatorname*{plim}_{n\to\infty}(MW - CS) = 0,$

(iii) $\operatorname*{plim}_{n\to\infty}(MW - M_3) = 0.$

Next we consider the behavior of these tests under a sequence of local alternatives of the form H_n: $\lambda = n^{-1/2}b$, where b is an arbitrary $m \times 1$ vector. It is known that the classical LIML-based tests have the same Pitman efficiency [Wald (1943)]. It is straightforward to show that the same result also holds here:

Proposition 5. Under Assumptions 1 and 2 and the sequence of local alternatives H_n: $\lambda = n^{-1/2}b$, each of the tests statistics MW, CLR, CS, and M_3 has a limiting non-central chi-square distribution with m degrees of freedom and non-centrality parameter $b'V_0(\hat{\lambda})^{-1}b$.[9]

In fact, the limiting distribution of the 2SCML-based statistics is the same as the classical LIML-based statistics under local alternatives. This result follows from the fact that under the null hypothesis, $V_0(\hat{\lambda})$ is also the asymptotic covariance matrix of the LIML estimator. Estimators without this property, such as IVP or G2SP, will have smaller local asymptotic power than either the LIML or 2SCML-based tests.

Note that the estimate $\hat{V}_0(\hat{\lambda})$ used to calculate the modified Wald statistic can be obtained from the information matrix associated with the conditional likelihood (3.17) as if $\hat{\Pi}$ was not estimated [see eq. (4.11)]. This matrix would

[9] If $u_1, \ldots, u_m$ are independent normal random variables with $\mathrm{E}(u_j) = \mu_j$ and $\mathrm{V}(u_j) = 1$, then $\Sigma_{j=1}^m u_j^2$ has a non-central chi-square distribution with non-centrality parameter $\Sigma_{j=1}^m \mu_j^2$.

ordinarily be computed by any probit program used to perform the second stage of the 2SCML estimation procedure (i.e., the uncorrected information matrix associated with the conditional likelihood function).

The test based on the Hausman statistic M_3 is consistent against all alternatives. This is not necessarily the case for the tests based on the first two Hausman statistics M_1 and M_2 [see Holly (1982), Holly and Monfort (1986)]. In fact, it can be shown that the test based on M_1 is equivalent to the previous tests if and only if $\Pi'\tilde{\Sigma}_{xv} + \tilde{\Sigma}_{vv}$ is non-singular. For the test based on M_2, the condition is weaker and is that $\tilde{H}_1'\tilde{\Sigma}_v$ has full column rank.

6. Monte Carlo evidence

A series of Monte Carlo simulations were performed to evaluate the small sample performance of the various estimators and tests described in sections 3 and 5. The simulations are based on a probit model with a single endogenous variable which is either just-identified or over-identified:

$$y_{1i}^* = \gamma y_{2i} + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i. \tag{6.1}$$

In the just-identified case,

$$y_{2i} = \pi_1 x_{1i} + \pi_2 x_{2i} + \pi_3 x_{3i} + v_i, \tag{6.2}$$

while in the over-identified case,

$$y_{2i} = \pi_1 x_{1i} + \pi_2 x_{2i} + \pi_3 x_{3i} + \pi_4 x_{4i} + v_i. \tag{6.3}$$

The variable x_{1i} is a constant term. In all simulations the following parameter values were chosen: $\gamma = 1$, $\beta_1 = 0$, $\beta_2 = -1$, $\pi_1 = 0$, $\pi_2 = 1$, $\pi_3 = 1$, $\pi_4 = -1$.

Each simulation was based on a random sample of 100 observations and was replicated 1000 times. The exogenous variables (x_{2i}, x_{3i}, x_{4i}) were drawn from an equicorrelated multivariate normal distribution with zero means, unit variances, and covariances of 0.5. To generate values for the disturbances, a pair of independent standard normal variates (v_i, η_i) were drawn and $u_i = \lambda v_i + \eta_i$. Values on λ in the interval $[-2, 2]$, as shown in table 1, were chosen. This includes points at which each of the various estimators were shown to be efficient (see Corollaries 1 and 2), as well as various intermediate cases. Table 1 also shows the correlation between (u, v), the standard error ω of the probit reduced form error, and the true values of the normalized parameters estimated by G2SP and IVP.

As mentioned previously, since the IVP and G2SP estimators employ a different normalization from 2SCML and, hence, estimate different parameters, some adjustments were required to make the results comparable. One

Table 1

Parameters.

	λ						
	2.0	1.0	0.5	0.0	−0.5	−1.0	−2.0
corr(u, v)	0.894	0.707	0.447	0.0	−0.447	−0.707	−0.894
ω	3.162	2.236	1.803	1.414	1.118	1.000	1.414
γ_*	0.316	0.447	0.555	0.707	0.894	1.000	0.707
β_{1*}	0.0	0.0	0.0	0.0	0.0	0.0	0.0
β_{2*}	−0.316	−0.447	−0.555	−0.707	−0.894	−1.000	−0.707

approach takes γ and β to be the parameters of interest and utilizes an estimator $\hat{\omega}$ of $\omega = 1 + (\gamma+\lambda)^2\sigma_v^2$ to obtain estimates $(\hat{\gamma}^{IVP}, \hat{\beta}^{IVP})$ and $(\hat{\gamma}^A, \hat{\beta}^A)$ of (γ, β): $\hat{\gamma}^{IVP} = \tilde{\omega}\hat{\gamma}_*^{IVP}$, $\hat{\gamma}^A = \tilde{\omega}\hat{\gamma}_*^A$, etc., where[10]

$$\tilde{\omega} = \left(\frac{s_v^2}{|s_v^2 - s_{uv}|}\right)^{1/2}, \tag{6.4}$$

$$s_v^2 = (1/n)\sum_{i=1}^{n} \hat{v}_i^2, \tag{6.5}$$

$$s_{uv} = (1/n)\sum_{i=1}^{n} \frac{y_{1i}\hat{v}_i}{\phi(\hat{\tau}_*' x_i)}. \tag{6.6}$$

With this normalization, the 2SCML estimator outperforms IVP and G2SP. The efficiency gain is substantial (typically between 30% and 60%), but this is due primarily to the inaccuracy in estimating ω using $\tilde{\omega}$. To save space, these results are not reported here. Alternatively, one can focus on the parameters γ_* and β_* that result from the other normalization. In this case, an estimator $\hat{\omega}$ based on the 2SCML estimator $(\hat{\gamma}, \hat{\beta})$ is required. We used

$$\hat{\omega} = \left(1 + (\hat{\gamma} + \hat{\lambda})^2 s_v^2\right)^{1/2}, \tag{6.7}$$

to obtain $\hat{\gamma}_* = \hat{\gamma}/\hat{\omega}$ and $\hat{\beta}_* = \hat{\beta}/\hat{\omega}$. These results are reported in tables 2 and 3.

In the case of a just-identified model, 2SCML is asymptotically efficient for (γ, β) or (γ_*, β_*). The simulation results indicate that 2SCML performs somewhat better than G2SP in relatively small samples. As mentioned earlier, IVP is numerically equivalent to G2SP in this case. When $\gamma + \lambda$ is close to

[10] The estimator s_{uv} is that proposed by Amemiya (1978). The absolute value in (6.4) ensures that the argument in the square-root is always positive in small samples.

Table 2

Just-identified case.

		γ^*		β_1^*		β_2^*	
λ	Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE
2.0	2SCML	0.001	0.117	−0.006	0.095	−0.001	0.193
	G2SP/IVP	0.006	0.131	−0.008	0.109	−0.006	0.220
1.0	2SCML	0.013	0.129	−0.003	0.101	−0.007	0.217
	G2SP/IVP	0.012	0.146	−0.001	0.111	−0.002	0.233
0.5	2SCML	0.025	0.141	−0.001	0.107	−0.023	0.223
	G2SP/IVP	0.029	0.152	0.001	0.111	−0.027	0.238
0.0	2SCML	0.034	0.178	0.000	0.122	−0.036	0.271
	G2SP/IVP	0.041	0.190	−0.001	0.126	−0.045	0.284
−0.5	2SCML	0.057	0.231	−0.001	0.153	−0.065	0.345
	G2SP/IVP	0.059	0.236	−0.001	0.153	−0.066	0.351
−1.0	2SCML	0.070	0.290	−0.008	0.189	−0.088	0.440
	G2SP/IVP	0.071	0.289	−0.008	0.189	−0.089	0.440
−2.0	2SCML	0.034	0.249	−0.000	0.190	−0.035	0.411
	G2SP/IVP	0.038	0.256	−0.002	0.195	−0.041	0.417

Table 3

Over-identified case.

		γ^*		β_1^*		β_2^*	
λ	Estimators	Bias	RMSE	Bias	RMSE	Bias	RMSE
2.0	2SCML	0.012	0.107	−0.003	0.096	−0.010	0.141
	G2SP	0.012	0.122	−0.005	0.108	−0.004	0.162
	IVP	0.014	0.123	−0.005	0.109	−0.007	0.163
1.0	2SCML	0.025	0.116	−0.000	0.099	−0.027	0.155
	G2SP	0.025	0.125	0.002	0.107	−0.031	0168
	IVP	0.029	0.127	0.002	0.108	−0.035	0.170
0.5	2SCML	0.025	0.130	−0.002	0.103	−0.024	0.165
	G2SP	0.024	0.154	−0.004	0.114	−0.024	0.187
	IVP	0.028	0.141	−0.003	0.110	−0.027	0.177
0.0	2SCML	0.040	0.157	−0.002	0.119	−0.042	0.201
	G2SP	0.036	0.193	−0.002	0.126	−0.036	0.253
	IVP	0.045	0.166	−0.001	0.122	−0.047	0.241
−0.5	2SCML	0.040	0.194	−0.002	0.146	−0.040	0.247
	G2SP	0.038	0.216	−0.002	0.151	−0.037	0.264
	IVP	0.040	0.197	−0.001	0.148	−0.039	0.249
−1.0	2SCML	0.062	0.253	0.006	0.193	−0.061	0.312
	G2SP	0.071	0.270	0.006	0.195	−0.070	0.329
	IVP	0.063	0.252	0.006	0.192	−0.062	0.312
−2.0	2SCML	0.028	0.209	0.000	0.191	−0.020	0.282
	G2SP	0.023	0.346	−0.004	0.200	−0.015	0.416
	IVP	0.031	0.215	−0.003	0.193	−0.023	0.289

Table 4

Power and size of exogeneity tests.

Asymptotic significance level	Tests	λ						
		2.0	1.0	0.5	0.0	−0.5	−1.0	−2.0
Just-identified								
5%	Wald	99.6	90.6	41.0	6.0	56.2	99.1	100.0
	LR	99.6	92.4	45.6	7.6	57.7	99.2	100.0
	Score	100.0	94.6	51.6	8.9	59.9	99.3	100.0
	Hausman 1	32.3	54.1	37.1	3.9	46.1	99.2	100.0
	Hausman 2	46.3	78.7	48.0	4.4	46.3	99.2	100.0
	Hausman 3	99.8	92.1	42.7	5.8	55.2	99.5	100.0
100%	Wald	99.8	95.2	56.8	11.8	67.5	99.5	100.0
	LR	99.6	95.8	59.4	12.4	68.9	99.5	100.0
	Score	100.0	97.0	63.8	14.4	71.7	99.6	100.0
	Hausman 1	38.2	64.5	51.4	8.1	62.3	99.8	100.0
	Hausman 2	57.2	85.6	61.6	8.8	62.4	99.8	100.0
	Hausman 3	99.9	95.4	58.2	10.0	68.0	99.9	100.0
Over-identified								
5%	Wald	99.4	93.6	50.8	5.0	61.0	99.8	100.0
	LR	100.0	95.0	53.7	6.2	62.9	99.9	100.0
	Score	100.0	96.2	59.2	7.8	64.3	99.9	100.0
	Hausman 1	16.5	38.6	33.1	2.6	52.1	99.4	100.0
	Hausman 2	24.9	66.8	53.0	4.1	52.1	99.4	100.0
	Hausman 3	100.0	93.6	51.3	5.7	64.1	99.7	100.0
10%	Wald	99.6	97.6	63.8	11.2	73.2	100.0	100.0
	LR	100.0	97.8	64.7	12.3	73.8	100.0	100.0
	Score	100.0	97.8	69.7	14.2	75.2	99.9	100.0
	Hausman 1	21.2	48.7	46.9	7.9	68.4	99.8	100.0
	Hausman 2	33.2	74.0	64.8	10.7	68.6	99.8	100.0
	Hausman 3	100.0	97.3	64.0	10.4	75.2	99.9	100.0

zero (i.e., $\lambda \simeq -1$), both estimators are asymptotically efficient and perform equally well in the simulations. For large values of λ (e.g., $\lambda = 2$), however, 2SCML produces a RMSE between 9% and 16% smaller than G2SP.

The results for the over-identified case are displayed in table 3. Although G2SP is asymptotically efficient when $\lambda = -1$ (since $\gamma = 1$), its small sample performance is slightly worse than the other two estimators. In general, G2SP and IVP perform equally well. For $\lambda \geq 0$, 2SCML has a RMSE between 3% and 12% smaller than either G2SP or IVP, with the larger $\gamma + \lambda$ the greater the efficiency gain. In no simulation did G2SP or IVP out-perform 2SCML.

We also evaluated the performance of the tests (5.1)–(5.6) for exogeneity in these simulations. Table 4 displays the simulated probabilities for rejecting the null hypothesis of exogeneity ($\lambda = 0$) at the 5% and 10% significance levels in the just- and over-identified cases. In both sets of simulations, the asymptotic significance levels were remarkably close to, though slightly lower than, the exact sizes for the Wald, Hausman M_3, likelihood ratio, and score tests. Adjusted for sizes, these four tests perform comparably and have equally high power for $|\lambda| \geq 1$. On the other hand, the Hausman M_1 and M_2 tests did not perform well, especially for positive values of λ where the type II error probabilities remain quite high.[11]

7. Conclusion

In this paper we have compared alternative limited information estimators for simultaneous probit models. Four considerations guiding the choice between estimators are: (1) asymptotic efficiency, (2) small sample properties, (3) usefulness for testing, and (4) computational convenience. On grounds other than computational convenience, the limited information maximum likelihood estimator would be preferred. Of the computationally simpler estimators, however, the proposed two-stage conditional maximum likelihood estimator was shown to have attractive properties in large and small samples. Although no general asymptotic efficiency ordering between these estimators is possible, conditions were stated under which these estimators attain the Cramer–Rao bound. In small samples, our Monte Carlo results suggest that the new estimator performs favorably relative to its competitors. Moreover, the efficiency of the 2SCML estimator under the null hypothesis of exogeneity allowed us to construct several simple exogeneity tests for probit models. Most of these tests were shown to be asymptotically equivalent to one another and have the same local asymptotic power as the classical LIML-based tests.

References

Amemiya, T., 1978, The estimation of a simultaneous equation generalized probit model, Econometrica 46, 1193–1205.

Andrews, D.W.K., 1987, Asymptotic results for generalized Wald tests, Econometric Theory 3, 348–358.

Godfrey, L.G. and M.R. Wickens, 1982, A simple derivation of the limited information maximum likelihood estimator, Economics Letters 10, 277–283.

Hausman, J.A., 1978, Specification tests in econometrics, Econometrica 46, 1251–1272.

Heckman, J., 1978, Dummy endogenous variables in a simultaneous equation system, Econometrica 46, 931–959.

Holly, H., 1982, A remark on Hausman's specification tests, Econometrica 50, 749–759.

Holly, H. and A. Monfort, 1986, Some useful equivalence properties of Hausman's test, Economics Letters 20, 39–43.

Holly, H. and J.D. Sargan, 1982, Testing for exogeneity within a limited information framework, Cahier no. 820 (Université de Lausanne, Lausanne).

[11] In fewer than 2% of our replications were the M_1 and M_2 statistics negative. In addition, the Hausman M_1 and M_2 tests are consistent against local alternatives in our simulations as it can be shown that $\Pi'\tilde{\Sigma}_{xv} + \tilde{\Sigma}_{vv}$ is non-singular when x has a normal distribution.

Lee, L.F., 1981, Simultaneous equation models with discrete and censored dependent variables, in: C. Manski and D. McFadden, eds., Structural analysis of discrete data with economic applications (MIT Press, Cambridge, MA).
Newey, W., 1985, Maximum likelihood specification testing and conditional moment tests, Econometrica 53, 1047–1070.
Newey, W., 1987, Efficient estimation of limited dependent variable models with endogenous explanatory variables, Journal of Econometrics 36, 230–251.
Rao, C.R. and S.K. Mitra, 1971, Generalized inverse of matrices and its applications (Wiley, New York).
Richard, J.F., 1975, A note on the information matrix of the multivariate normal distribution, Journal of Econometrics 3, 57–60.
Rivers, D. and Q.H. Vuong, 1984, Limited information estimators and exogeneity tests for simultaneous probit models, Social science working paper no. 539 (California Institute of Technology, Pasadena, CA).
Smith, R.J. and R.W. Blundell, 1986, An exogeneity test for a simultaneous equation tobit model with an application to labor supply, Econometrica 54, 679–685.
Vuong, Q.H., 1984, Two-stage conditional maximum likelihood estimation of econometric models, Social science working paper no. 538 (California Institute of Technology, Pasadena, CA).
Vuong, Q.H., 1987, Generalized inverses and asymptotic properties of Wald tests, Economics Letters 24, 343–347.
Wald, A., 1943, Tests of statistical hypotheses concerning several parameters when the number of observations is large, Transactions of the American Mathematical Society 54, 426–482.