DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

LIMITED INFORMATION ESTIMATORS AND EXOGENEITY TESTS FOR SIMULTANEOUS PROBIT MODELS

Douglas Rivers and Quang H. Vuong

SOCIAL SCIENCE WORKING PAPER 539

November 1984

ABSTRACT

A two-step maximum likelihood procedure is proposed for estimating simultaneous probit models and is compared to alternative limited information estimators. Conditions under which these estimators attain the Cramer-Rao lower bound are stated. Simple tests of exogeneity 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 limitedd information maximum likelihood estimator.

LIMITED INFORMATION ESTIMATORS AND EXOGENEITY TESTS FOR SIMULTANEOUS PROBIT MODELS

Douglas Rivers and Quang H. Vuong
California Institute of Technology

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 quite difficult computationally 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. These estimators are also applicable to the general problem considered here.

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 has several advantages over the Heckman and Amemiya estimators. 2SCML is easier to compute than G2SP and in some cases more efficient, too, though contrary to the linear case (Rivers and Vuong (1984)) a general efficiency ordering between the estimators is no longer possible. Another advantage is that the 2SCML procedure incorporates a simple test 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. We restrict our attention to limited information estimators which do not impose any restrictions on the reduced form equations for the explanatory variables. The Cramer-Rao bound for limited information estimators is derived and conditions under which the estimators attain this bound are discussed. 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 enable us to construct analogs of the Wald, likelihood ratio, and score 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, and the relative efficiency of the estimators is compared in Section 4. Tests for exogeneity of the explanatory variables are proposed and compared in Section 5. Section 6 concludes the paper. Proofs are relegated to the appendix.

2. MODEL

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:

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

$$(i = 1,\dots,n)$$

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

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:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \le 0 \end{cases} \tag{2.4}$$

The following assumptions are made:

<u>Assumption 1</u>: (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:

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

<u>Assumption 2</u>: (Identification) Rank $(\Pi, J) = m + k$.

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 equation (2.2) in structural form rather than reduced form. For the purpose of limited information estimation we ignore any <u>a priori</u> 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 equations (2.1-2) will be considered. The first three methods (LIML, IVP, and G2SP) are known 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:

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

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

$$\sigma_{uu} - \lambda'\textstyle\sum_{vv}\lambda = 1 \qquad (3.2)$$

The normalization (3.2) is different from that imposed by Heckman (1978), Amemiya (1978), and Lee (1981) and leads to a slightly different parameterization than theirs.

Limited Information Maximum Likelihood (LIML)

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

$$h(y_i,Y_i|X_i;\gamma,\beta,\lambda,\Pi,\textstyle\sum_{vv})$$

$$= (2\pi)^{-(m+1)/2}(\sigma_{uu} - \lambda'\textstyle\sum_{vv}\lambda)^{1/2}|\textstyle\sum_{vv}|^{-1/2}. \qquad (3.3)$$

$$[\int_{c_i}^{\infty} \exp\{-1/2[u^2 - 2\lambda'V_iu + V_i'(\textstyle\sum_{vv}^{-1} + \lambda\lambda')V_i]\}du]^{y_i}$$

$$\times [\int_{-\infty}^{c_i} \exp\{-1/2[u^2 - 2\lambda'V_iu + V_i'(\textstyle\sum_{vv}^{-1} + \lambda\lambda')V_i]\}du]^{1-y_i}$$

where $c_i = -(Y_i'\gamma + X_{1i}'\beta + V_i'\lambda)$. The limited information maximum likelihood estimates are obtained by maximizing the sample log likelihood,

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

with respect to $(\gamma,\beta,\lambda,\Pi,\sum_{vv})$. (This approach to limited information is adapted from Godfrey and Wickens (1982).) We denote the LIML estimator as $(\hat{\gamma}^L,\hat{\beta}^L,\hat{\lambda}^L,\hat{\Pi}^L,\hat{\sum}^L)$. Since iterative maximization of (3.4) requires a fairly messy numerical integration (involving all the parameters) at each step, the LIML estimator has generally been avoided in favor of less efficient but computationally more tractable estimation methods.

Instrumental Variables Probit (IVP)

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

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

Then the marginal log likelihood for y given X is:

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

$$+ (1 - y_i) \log[1 - \Phi((\Pi'X_i)\gamma_* + X_{1i}'\beta_*)] \qquad (3.6)$$

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

$$\gamma_* = \gamma/(\sigma_{uu} + 2\gamma'\textstyle\sum_{vu} + \gamma'\sum_{vv}\gamma)^{1/2}$$

$$= \gamma/\omega \qquad (3.7)$$

$$\beta_* = \beta/(\sigma_{uu} + 2\gamma'\textstyle\sum_{uv} + \gamma'\sum_{vv}\gamma)^{1/2} \qquad (3.8)$$

$$= \beta/\omega$$

where

$$\omega^2 = 1 + (\gamma + \lambda)' \textstyle\sum_{vv}(\gamma + \lambda)$$

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.[1]

<u>Generalized Two-Stage Simultaneous Probit (G2SP)</u>

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

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

with respect to τ_* where:

$$\tau_* = \Pi\gamma_* + J\beta_* \qquad (3.10)$$

Let $\hat{\tau}_*$ denote the corresponding estimator of τ_* and, as before, $\hat{\Pi}_*$ the least squares estimator of Π. Replacing τ_* and Π by their sample estimates in (3.10) yields:

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

$$= \hat{H} \begin{bmatrix} \gamma_* \\ \beta_* \end{bmatrix} + e \qquad (3.11)$$

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.[2] The Amemiya G2SP estimator is defined by:

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

Since the covariance matrix of e depends on γ_*, to compute $\hat{V}$ one needs a preliminary estimate of γ_*. 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). The GLS regression in (3.12) is also a bit awkward to perform with standard statistical software. 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.

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)). 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:

$$h(y_i, Y_i | X_i; \gamma, \beta, \lambda, \Pi, \Sigma_{vv})$$

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

where:

$$f(y_i | Y_i, X_i; \gamma, \beta, \lambda, \Pi) =$$

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

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

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

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:

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

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 $(\hat{\gamma}, \hat{\beta}, \hat{\lambda})$.

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.

<u>Proposition 1</u>. (Cramer-Rao Bound for Limited Information Estimators) Under Assumptions 1 and 2,

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

where:

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

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

$$\tilde{\sum} = \begin{bmatrix}\tilde{\sum}_{xx} & \tilde{\sum}_{xv}\\ \tilde{\sum}_{vx} & \tilde{\sum}_{vv}\end{bmatrix} = E\left[\frac{\phi(Z_i'\delta + V_i'\lambda)^2}{\Phi(Z_i'\delta + V_i'\lambda)[1 - \Phi(Z_i'\delta + V_i'\lambda)]}\begin{bmatrix}X_i\\ V_i\end{bmatrix}\begin{bmatrix}X_i\\ V_i\end{bmatrix}'\right] \tag{4.4}$$

For comparisons with other estimators, we need the asymptotic covariance matrix of $\hat{\delta}^L$ which is the upper lefthand block of $V(\hat{\theta}^L)$: It is shown in the appendix that:

$$V(\hat{\delta}^L) = \{H'[\tilde{\sum}_{xx}^{-1} + \tilde{\sum}_{xx}^{-1}\tilde{\sum}_{xv}(\tilde{\sum}_{vv} - \tilde{\sum}_{vx}\tilde{\sum}_{xx}^{-1}\tilde{\sum}_{xv})^{-1}\tilde{\sum}_{vx}\tilde{\sum}_{xx}^{-1} \tag{4.5}$$

$$+ (\lambda'\sum_{vv}\lambda)\sum_{xx}^{-1}]^{-1}H\}^{-1}$$

where $H = [\Pi; J]$ is a submatrix of $\tilde{H}$.

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

<u>Proposition 2</u>. (Asymptotic Properties of the 2SCML Estimator) Under Assumptions 1 and 2, $\hat{\theta} \overset{a.s.}{\rightarrow} \theta$ and

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

where:

$$V(\hat{\theta}) = \{\tilde{H}'[\tilde{\sum} - \tilde{\sum}(\lambda' \otimes (I_p;0)')M^{-1}(\lambda \otimes (I_p;0))\tilde{\sum}]\tilde{H}\}^{-1} \tag{4.7}$$

$$M = \sum_{vv}^{-1} \otimes \sum_{xx} + [\lambda \otimes (I_p;0)]\tilde{\sum}\tilde{H}(\tilde{H}'\tilde{\sum}\tilde{H})^{-1}\tilde{H}'\tilde{\sum}[\lambda' \otimes (I_p;0)'] \tag{4.8}$$

As before, the asymptotic covariance matrix for the 2SCML estimator of δ, denoted $\hat{\delta}$, is obtained from the upper left hand block of $V(\hat{\theta})$:

$$V(\hat{\delta}) = \{H'[\tilde{\sum}^{xx} + (\lambda' \otimes I_p)[\sum_{vv}^{-1} \otimes \sum_{xx} \tag{4.9}$$

$$- (\lambda \otimes (I_p;0))S\tilde{\sum}^{-1}S(\lambda \otimes (I_p;0))']^{-1}(\lambda \otimes I_p)]^{-1}H\}^{-1}$$

where:

$$S = \tilde{\sum} - \tilde{\sum}\tilde{H}(\tilde{H}'\tilde{\sum}\tilde{H})^{-1}\tilde{H}'\tilde{\sum} \tag{4.10}$$

$$\tilde{\Sigma}^{xx} = \tilde{\Sigma}_{xx}^{-1} + \tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xv}(\tilde{\Sigma}_{vv} - \tilde{\Sigma}_{vx}\tilde{\Sigma}_{xx}^{-1}\tilde{\Sigma}_{xv})^{-1}\tilde{\Sigma}_{vx}\tilde{\Sigma}_{xx}^{-1} \qquad (4.11)$$

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 equation (4.7) that the variance of the 2SCML estimator $\hat{\theta}$ is given by:

$$V_o(\hat{\theta}) = (\tilde{H}'\tilde{\Sigma}\tilde{H})^{-1} \qquad (4.12)$$

which, by inspection of equation (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, nonsingular. It follows that when $k = p - m$, we have $S = 0$ so (4.9) reduces to (4.5) and $V(\hat{\delta}) = V(\hat{\delta}^L)$. Thus we have:

<u>Corollary 1</u>. (Efficiency of 2SCML Estimators) Under Assumption (1) and (2), $V(\hat{\delta}) \geq V(\hat{\delta}^L)$ with equality if one of the following conditions holds:

(i) $\lambda = 0$

(ii) $k = p - m$

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 the next section. 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 righthand side of (2.1), i.e. that equation (2.1) be just identified.

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):

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

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

where:

$$V(\hat{\delta}_*^A) = \omega\{H'[\bar{\Sigma}_{xx}^{-1} - (\gamma_*'\Sigma_{vv}\gamma_* + 2\gamma_*'\Sigma_{vv}\lambda_*)\Sigma_{xx}^{-1}]^{-1}H\}^{-1} \qquad (4.15)$$

$$V(\hat{\delta}_*^{IVP}) = \omega(H'\bar{\Sigma}_{xx}H)^{-1}H'\bar{\Sigma}_{xx}[\bar{\Sigma}_{xx}^{-1} - (\gamma_*'\Sigma_{vv}\gamma_* + 2\gamma_*'\Sigma_{vv}\lambda_*)\Sigma_{xx}^{-1}]\bar{\Sigma}_{xx}H(H'\bar{\Sigma}_{xx}H)^{-1} \qquad (4.16)$$

$$\bar{\Sigma}_{xx} = E\left[\frac{\phi(X_i'\alpha_*)^2}{\Phi(X_i'\alpha_*)[1 - \Phi(X_i'\alpha_*)]}X_iX_i'\right] \qquad (4.17)$$

Lee (1981) showed that IVP is never more efficient than G2SP, though there are cases in which the estimators are asymptotically equivalent, or, even, when Equation (2.1) is just identified, numerically identical.[3] Since, however, the G2SP estimator of δ^* is not optimal within the limited information class, this comparison 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 following Proposition 1. The bound for δ^*, denoted $V(\hat{\delta}_*^L)$, is given in Proposition 3.

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

$$V(\hat{\delta}_*^L) = V(\hat{\delta}^L)/\omega^2 + [(\gamma_*' + \lambda_*')\Sigma_{vv}V(\hat{\gamma}^L + \hat{\lambda}^L)\Sigma_{vv}(\gamma_* + \lambda_*)/\omega^2$$

$$+ \frac{1}{4}((\gamma_*' + \lambda_*') \otimes (\gamma_*' + \lambda_*'))V(\text{vec}\,\hat{\Sigma}_{vv}^L)((\gamma_* + \lambda_*) \otimes (\gamma_* + \lambda_*))]\delta_*\delta_*'$$

$$- \delta_*(\gamma_*' + \lambda_*')\Sigma_{vv}\text{Cov}(\hat{\gamma}^L + \hat{\lambda}^L, \hat{\delta}^L)/\omega^2$$

$$- \text{Cov}(\hat{\delta}^L, \hat{\gamma}^L + \hat{\lambda}^L)\Sigma_{vv}(\gamma_* + \lambda_*)\delta_*'/\omega^2. \qquad (4.18)$$

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.

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

If, however, $\lambda = 0$ but $\gamma \neq 0$ evidently $\hat{\delta}_*^A$ does not 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. On the other hand, if $\gamma + \lambda = 0$ but $\lambda \neq 0$, then 2SCML will fail to attain the Cramer-Rao bound while G2SP will be efficient. Hence, no general efficiency ordering between the estimators is possible. In contrast, in the case of linear simultaneous equations (i.e., if y^* were observed), Rivers and Vuong (1984) 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, it is somewhat surprising to find that they do not even maintain their asymptotic equivalence in this case.

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 one of the Hausman (1978) variety. All tests are based on the 2SCML estimator rather than the LIML estimator so, strictly speaking, these are not classical tests. However, as the remarks preceding 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 most 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[4] is given by:

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

where $\hat{V}_o(\hat{\lambda})$ is a consistent estimator of the lower righthand block of $V_o(\hat{\theta}) = (\tilde{H}'\tilde{\sum}\tilde{H})^{-1}$ corresponding to $\hat{\lambda}$. The conditional likelihood ratio statistic is given by:

$$CLR = 2[L_n^f(\hat{\gamma},\hat{\beta},\hat{\lambda},\hat{\Pi}) - L_n^f(\tilde{\gamma},\tilde{\beta},0,\hat{\Pi})] \tag{5.2}$$

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

$$CS = \frac{1}{n}\frac{\partial L_n^f(\tilde{\gamma},\tilde{\beta},0,\hat{\Pi})}{\partial\lambda'}\hat{V}_o(\hat{\lambda})\frac{\partial L_n^f(\tilde{\gamma},\tilde{\beta},0,\hat{\Pi})}{\partial\lambda} \tag{5.3}$$

Under H_0, δ is efficiently estimated by the usual probit estimator $\tilde{\delta} = (\tilde{\gamma}',\tilde{\beta}')'$ while under the alternative the 2SCML estimator $\hat{\delta}$ will be consistent though possibly inefficient. This enables us to form the Hausman statistic:

$$H = n(\hat{\delta} - \tilde{\delta})'[\hat{V}_o(\hat{\delta}) - \hat{V}(\tilde{\delta})]^-(\hat{\delta} - \tilde{\delta}) \tag{5.4}$$

where $[\cdot]^-$ denotes a generalized inverse and $\hat{V}(\tilde{\delta})$ is obtained from the estimated information matrix for LIML under the null:

$$\hat{V}(\tilde{\delta}) = n\left[\sum_{i=1}^{n}\frac{\phi_i(Y_i'\tilde{\gamma} + X_i'\tilde{\beta})^2}{\Phi(Y_i'\tilde{\gamma} + X_i'\tilde{\beta})[1 - \Phi(Y_i'\tilde{\gamma} + X_i'\tilde{\beta})]}\begin{bmatrix}Y_i\\X_i\end{bmatrix}\begin{bmatrix}Y_i\\X_i\end{bmatrix}'\right]^{-1} \tag{5.5}$$

Under the null hypothesis each of the test statistics (5.1)-(5.4) 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, as shown by the following proposition.

<u>Proposition 4</u>. 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) $\operatorname{plim}_{n\to\infty} MW - H = 0$

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 mX1 vector. It is known that the classical LIML-based tests have the same Pitman efficiency (Wald (1943), Chandra and Joshi (1983)). The next proposition shows that the same result holds for the 2SCML-based tests.

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 H has a limiting noncentral chi-square distribution with m degrees of freedom and noncentrality parameter[5] $b'V_o(\hat{\lambda})^{-1}b$.

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, $V_o(\hat{\lambda})$ 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 $\hat{V}_o(\hat{\lambda})$ used to calculate the modified Wald statistic can be obtained from the information matrix associated with the conditional likelihood (3.17). This matrix would 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). In view of the above result, the modified Wald test might be recommended since it reduces the computational burden.

6. CONCLUSION

In this paper we have compared alternative limited information estimators for simultaneous probit models. Three considerations guiding the choice between estimators are: (1) efficiency, (2) computational convenience, and (3) usefulness for testing. 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. Although no general efficiency ordering between these estimators is possible, conditions were stated under which these estimators attain the Cramer-Rao bound. In particular, the efficiency of the 2SCLM estimator under the null hypothesis of exogeneity allowed us to construct several simple exogeneity tests for probit models. These tests were shown to be asymptotically equivalent to one another and have the same local asymptotic power as the classical LIML-based tests.

APPENDIX

From Equation (3.14) the first partial derivatives of the log-density, $\log f(.|.;.)$, of y_i given (Y_i, X_i) are:

$$\partial \log f/\partial \delta = m_i(y_i - \Phi_i)Z_i,$$
$$\partial \log f/\partial \lambda = m_i(y_i - \Phi_i)V_i,$$
$$\partial \log f/\partial \mathrm{vec}\Pi = - m_i(y_i - \Phi_i)(\lambda \otimes X_i),$$
$$\partial \log f/\partial \ \mathrm{vech} \sum_{vv} = 0,$$

where vec and vec h are the operators that stack the columns of a matrix and a symmetric matrix (see, e.g., Henderson and Searle (1979)), and $m_i = \phi_i/[\Phi_i(1 - \Phi_i)]$ with ϕ_i and Φ_i being defined in the text.

It follows that the second partial derivatives of $\log f(.|.;)$ are:

$$\partial^2 \log f/\partial\delta\partial\delta' = - d_i Z_i Z_i' + (y_i - \Phi_i)Z_i(\partial m_i/\partial\delta'),$$
$$\partial^2 \log f/\partial\delta\partial\lambda' = - d_i Z_i V_i' + (y_i - \Phi_i)Z_i(\partial m_i/\partial\lambda'),$$
$$\partial^2 \log f/\partial\delta\partial \mathrm{vec}\ \Pi' = d_i(\lambda' \otimes Z_i x_i') + (y_i - \Phi_i)Z_i(\partial m_i/\partial \mathrm{vec}\ \Pi')$$
$$\partial^2 \log f/\partial\delta\partial \mathrm{vech} \sum_{vv} = 0$$
$$\partial^2 \log f/\partial\lambda\partial\lambda' = - d_i V_i V_i' + (y_i - \Phi_i)V_i(\partial m_i/\partial\lambda')$$
$$\partial^2 \log f/\partial\lambda\partial \mathrm{vec}\ \Pi' = d_i(\lambda' \otimes V_i x_i') + (y_i - \Phi_i)V_i(\partial m_i/\partial \mathrm{vec}\ \Pi') - m_i(y_i - \Phi_i)(I_m \otimes x_i')$$
$$\partial^2 \log f/\partial\lambda\partial \mathrm{vech} \sum_{vv} = 0$$
$$\partial^2 \log f/\partial \mathrm{vec}\ \Pi\ \partial \mathrm{vec}\ \Pi' = - d_i(\lambda\lambda' \otimes X_i X_i') - (y_i - \Phi_i)(\lambda \otimes X_i)(\partial m_i/\partial \mathrm{vec}\ \Pi)$$
$$\partial^2 \log f/\partial \mathrm{vec}\ \Pi\ \partial \mathrm{vech} \sum_{vv}' = 0$$

$$\partial^2 \log f/\partial \mathrm{vech} \sum_{v} \partial \mathrm{vech} \sum_{vv}' = 0$$

where $d_i = \phi_i^2/(\Phi_i(1 - \Phi_i))$

Since $E(y_i|Y_i, X_i) = \Phi_i$, then $E(y_i - \Phi_i) = 0$. Using the fact that

$$Z_i = \begin{bmatrix} Y_i \\ X_{1i} \end{bmatrix} = \begin{bmatrix} \Pi' & I_m \\ J' & 0 \end{bmatrix} \begin{bmatrix} X_i \\ V_i \end{bmatrix}$$

and the definitions of $\tilde{H}$ and $\tilde{\sum}$ given in the text, it follows that the matrix of expectations of second partial derivatives of $\log f(.|.;.)$ with respect to $\alpha \equiv (\theta, \mathrm{vec}\ \Pi, \mathrm{vech} \sum_{vv})$ is given by:

$$A^f(\alpha) = - \begin{bmatrix} \tilde{H}'\tilde{\sum}\tilde{H} & * & 0 \\ -[\lambda \otimes (I_p 0)]\tilde{\sum}\tilde{H} & \lambda\lambda' \otimes \tilde{\sum}_{xx} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{matrix} \text{columns: } \theta,\ \mathrm{vec}\ \Pi,\ \mathrm{vech} \sum_{vv} \end{matrix} \tag{A.1}$$

From Richard (1975), the matrix of expectations of second partial derivatives of $\log g(.|.;.)$ with respect to α is:

$$A^g(\alpha) = - \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sum_{vv}^{-1} \otimes \sum_{xx} & 0 \\ 0 & 0 & \frac{1}{2}R(\sum_{vv}^{-1} \otimes \sum_{vv}^{-1})R' \end{bmatrix} \quad \begin{matrix} \text{columns: } \theta,\ \mathrm{vec}\ \Pi,\ \mathrm{vech} \sum_{vv} \end{matrix} \tag{A.2}$$

where R is a matrix such that $\mathrm{vech} \sum_{vv} = R\ \mathrm{vec} \sum_{vv}$.

The Cramer-Rao lower bound for α is therefore $-[A^f + A^g]^{-1}$.

PROOF OF PROPOSITION 1: To obtain the Cramer-Rao lower bound for θ, it suffices to compute to upper-left block of $-[A^f + A^g]^{-1}$. Using the formula for a partitioned inverse, we have:

$$\text{Var } \hat{\theta}^L = [\tilde{H}'\{\tilde{\Sigma} - \tilde{\Sigma}(\lambda' \otimes (I_p;0)')[(\lambda\lambda' \otimes \tilde{\Sigma}_{xx}) + \Sigma_{vv}^{-1} \otimes \Sigma_{xx})]^{-1}(\lambda \otimes \Sigma_{xx})]^{-1}(\lambda \otimes (I_p;0))\tilde{\Sigma}\}]\tilde{H}]^{-1}$$

Taking the inverse of the matrix in braces, we obtain after some matrix algebra:

$$\text{Var } \hat{\theta}^L = [\tilde{H}'\{\tilde{\Sigma}^{-1} + (\lambda' \otimes (I_p;0))(\Sigma_{vv} \otimes \Sigma_{xx}^{-1})(\lambda \otimes (I_p;0))\}^{-1}\tilde{H}]^{-1}$$

which gives the desired result.

Q.E.D.

To establish Equation 4.5, we use the following identity which is straightforward to prove:

$$N - N(0;I_m)'[(0;I_m)N(0;I_m)']^{-1}(0;I_m)N = \begin{bmatrix} N_{11} - N_{12}N_{22}^{-1}N_{21} & 0 \\ 0 & 0 \end{bmatrix} \qquad (A.3)$$

where N is any partitioned matrix $[N_{ij}]$, $i,j = 1,2$, and N_{22} is a $m \times m$ non-singular matrix,

Let N be chosen so that $V(\hat{\theta}^L) = \{\tilde{H}'N\tilde{H}\}^{-1}$ (see Equation (4.2)). In what follows the subscript 1 indicates partitioning with respect to

θ, (or X) while the subscript 2 indicates partitioning with respect to (vec Π, vech Σ_{vv}) (or V) depending on the context. Then using the formula for a partitioned inverse, the definition of $\tilde{H}$, and the above identity, we obtain for the lower bound of $\hat{\delta}^L$:

$$V(\hat{\delta}^L) = \{H'(N_{11} - N_{12}(N_{22})^{-1}N_{21})H\}^{-1}$$
$$= \{H'[\tilde{\Sigma}^{xx} + (\lambda'\Sigma_{vv}\lambda)\Sigma_{xx}^{-1}]^{-1}H\}^{-1}$$

where $H = [\Pi;J]$ and $\tilde{\Sigma}^{xx}$ is the top-left block of $\tilde{\Sigma}^{-1}$. Equation (4.5) now follows from the formula for a partitioned inverse.

PROOF OF PROPOSITION 2: From Vuong (1984) the asymptotic covariance matrix of the 2SCML estimator $\hat{\theta}$ is:

$$V(\hat{\theta}) = -\{A^f_{11} - A^f_{12}\tilde{M}^{-1}A^f_{21}\}^{-1}$$

where $\tilde{M} = A^g_{22} + A^f_{21}(A^f_{11})^{-1}A^f_{12}$.

From the formulas (A.1) and (A.2) for A^f and A^g derived earlier, we have:

$$A^f_{11} = -\tilde{H}'\tilde{\Sigma}\tilde{H}$$
$$A^f_{12} = [\tilde{H}'\tilde{\Sigma}(\lambda' \otimes (I_p;0)');0]$$

Since A^g_{22} is block diagonal, it is easy to see that $\tilde{M}$ is also block diagonal with first block equal to M as given by Equation (4.8). Equation (4.7) follows.

Q.E.D.

To establish (4.9), we use the partitioned inverse formula. Let G be chosen so that $V(\hat{\theta}) = \{\tilde{H}'G^{-1}\tilde{H}\}$ (see Equation (4.7)). We obtain using the definition of $\tilde{H}$ and the identity (A.3) where N is taken to be G^{-1}:

$$V(\hat{\delta}) = \{H'[G^{11} - G^{12}(G^{22})^{-1}G^{21}]H\}^{-1}$$
$$= \{H'(G_{11})^{-1}H\}^{-1}$$

where $G^{-1} = [G^{ij}]$. It now suffices to compute G_{11}. We have:

$$G = [\tilde{\Sigma} - \tilde{\Sigma}(\lambda' \otimes (I_p;0)')M^{-1}(\lambda \otimes (I_p;0))\tilde{\Sigma}]^{-1}$$

$$= \tilde{\Sigma}^{-1} + (\lambda' \otimes (I_p;0)')[(\Sigma_{vv}^{-1} \otimes \Sigma_{xx}) - (\lambda \otimes (I_p;0))S\tilde{\Sigma}^{-1}S(\lambda' \otimes (I_p;0)')]^{-1}(\lambda \otimes (I_p;0))$$

where the second equality follows from the formula for the inverse of P - R'QR, and the definition (4.11) of S. Equation (4.9) follows.

To derive the lower bound for S_*, we need the Jacobian of the transformation from $(\theta, \text{vec}\, \Pi, \text{vech}\, \Sigma_{vv})$ to δ_*. We have $\omega = [1 + (\gamma + \lambda)'\Sigma_{vv}(\gamma + \lambda)]^{1/2}$ so that:

$$\partial\omega/\partial\gamma = (1/\omega)\Sigma_{vv}(\lambda + \gamma),$$
$$\partial\omega/\partial\lambda = (1/\omega)\Sigma_{vv}(\lambda + \gamma),$$
$$\partial\omega/\partial\, \text{vech}\Sigma_{vv} = (1/2\omega)Q'[(\lambda + \gamma) \otimes (\lambda + \gamma)],$$

where $\text{vec}\Sigma_{vv} = Q\, \text{vech}\, \Sigma_{vv}$. Thus:

$$\partial\gamma_*/\partial\gamma' = (1/\omega)I_m - (1/\omega^3)\gamma(\lambda' + \gamma')\Sigma_{vv},$$
$$\partial\gamma_*/\partial\lambda' = -(1/\omega^3)\gamma(\lambda' + \gamma')\Sigma_{vv},$$
$$\partial\gamma_*/\partial\text{vech}\, \Sigma_{vv}' = -(1/2\omega^3)\gamma[(\lambda + \gamma)' \otimes (\lambda + \gamma)']Q,$$
$$\partial\beta_*/\partial\beta' = (1/\omega)I_k,$$
$$\partial\beta_*/\partial\gamma' = -(1/\omega^3)\beta(\lambda' + \gamma')\Sigma_{vv},$$
$$\partial\beta_*/\partial\lambda' = -(1/\omega^3)\beta(\lambda' + \gamma')\Sigma_{vv},$$
$$\partial\beta_*/\partial\text{vech}\, \Sigma_{vv}' = -(1/2\omega^3)\beta[(\lambda + \gamma)' \otimes (\lambda + \gamma)']Q,$$

all the remaining partial derivatives being zero. The Jacobian of the transformation,

$$J = [\partial\delta_*/\partial\theta';\ \partial\delta_*/\partial\text{ vec}\, \Pi';\ \partial\delta_*/\partial\text{vech}\Sigma_{vv}'],$$

is readily obtained.

PROOF OF PROPOSITION 3: The lower bound for δ_* is readily derived from the lower bound for δ since:

$$V(\hat{\delta}_*^L) = J \cdot V(\hat{\delta}^L) \cdot J'.$$

Using the fact that $\partial\delta_*/\partial\text{vech}\Pi = 0$ and that $V(\hat{\delta}^L)$ is block diagonal in vech Σ_{vv} (see Equations (A.1)-(A.2)), we have:

$$V(\hat{\delta}_*^L) = (\partial\delta_*/\partial\theta')V(\hat{\delta}^L)(\partial\delta_*/\partial\theta) + (\partial\delta_*/\partial\text{vech}\, \Sigma_{vv}')V(\text{vech}\, \Sigma_{vv}^L)(\partial\delta_*/\partial\text{vech}\, \Sigma_{vv}).$$

From the above partial derivatives, it follows that

$$\omega \cdot \partial\delta_*/\partial\theta' = [I_{m+k};0] - (1/\omega^2)\delta(\lambda' + \gamma')\Sigma_{vv}[I_m;0;I_m].$$

In addition, we have:

$$Q(V(\text{vech } \hat{\sum}^{L}_{vv})Q' = V(\text{vec } \hat{\sum}^{L}_{vv}).$$

Equation (4.18) then follows using the fact that $\delta_* = \delta/\omega$.

Q.E.D.

To prove Corollary 2, we note that the following properties hold when $\gamma + \lambda = 0$:

(i) $\omega^2 = 1$, $\gamma_* = \gamma$, $\lambda_* = \lambda$,

(ii) $Z_i'\delta + V_i'\lambda = X_i'\alpha_*$,

(iii) $\tilde{\sum}_{xx} = \bar{\sum}_{xx}, \tilde{\sum}_{xv} = 0$.

Part (i) follows from the definition of ω (see Equation (3.7) and (3.8)). Part (ii) follows from Equation (3.10). The first equation of (iii) follows from Equations (4.4) and (4.16), while the second equation follows from Equation (4.4) by taking expectations conditional on X_i and by invoking Assumption 1.

PROOF OF COROLLARY 2: From Equation (4.18) it follows that, when $\gamma + \lambda = 0$, we have:

$$\begin{aligned} V(\hat{\delta}^L_*) &= V(\hat{\delta}^L)/\omega^2, \\ &= \{H'[\tilde{\sum}_{xx}^{-1} + (\lambda'\sum_{vv}\lambda)\sum_{xx}^{-1}]^{-1}H\}^{-1}, \end{aligned}$$

where the second equality follows from Parts (i) and (iii) above, and Equation (4.5).

On the other hand, from Equation (4.15) we have:

$$V(\hat{\delta}^A_*) = \{H'[\bar{\sum}_{xx}^{-1} - ((\gamma' + \lambda')\sum_{vv}(\gamma + \lambda) - (\lambda'\sum_{vv}\lambda)\sum_{xx}^{-1}]^{-1}H\}^{-1}$$

where we have used $\omega^2 = 1$. The desired property follows from Part (iii).

Q.E.D.

PROOF OF PROPOSITION 4: First we derive some useful identities and establish some additional notation. Expanding the partial derivatives of the conditional log likelihood function around $(\tilde{\delta},0)$ gives:

$$\begin{aligned} 0 &= n^{-1/2}\frac{\partial L_n^f(\hat{\delta},\hat{\lambda})}{\partial\theta} \\ &= n^{-1/2}\frac{\partial L_n^f(\tilde{\delta},0)}{\partial\theta} + n^{-1}\frac{\partial^2 L_n^f(\tilde{\delta},0)}{\partial\theta\partial\theta'}n^{1/2}\begin{bmatrix}\hat{\delta} - \tilde{\delta} \\ \hat{\lambda}\end{bmatrix} + o_p(1) \\ &= n^{-1/2}\frac{\partial L_n^f(\tilde{\delta},0)}{\partial\theta} - (\tilde{H}'\tilde{\sum}\tilde{H})n^{1/2}\begin{bmatrix}\hat{\delta} - \tilde{\delta} \\ \hat{\lambda}\end{bmatrix} + o_p(1) \end{aligned}$$

Since $\partial L_n^f(\tilde{\delta},0)/\partial\delta = 0$:

$$n^{1/2}\begin{bmatrix}\hat{\delta} - \tilde{\delta} \\ \hat{\lambda}\end{bmatrix} = (\tilde{H}'\tilde{\sum}\tilde{H})^{-1}n^{-1/2}\frac{\partial L_n^f(\tilde{\delta},0)}{\partial\theta} + o_p(1) \qquad \text{(A4)}$$

$$= \begin{bmatrix} -(\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1}\tilde{H}_1'\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix} \\ I_m \end{bmatrix} V_o(\hat{\lambda})\; n^{-1/2}\,\frac{\partial L_n^f(\tilde{\delta},0)}{\partial\lambda} + o_p(1)$$

where:

$$\tilde{H}_1 = \begin{bmatrix} \Pi & J \\ I_m & 0 \end{bmatrix}$$

Therefore:

$$n^{1/2}(\hat{\delta} - \tilde{\delta}) = -\,(\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1}\tilde{H}_1'\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix} n^{1/2}\hat{\lambda} + o_p(1) \qquad \text{(A5)}$$

$$n^{1/2}\frac{\partial L_n^f(\tilde{\delta},0)}{\partial\lambda} = V_o(\hat{\lambda})^{-1} n^{1/2}\hat{\lambda} + o_p(1) \qquad \text{(A6)}$$

To prove (i) and (ii), observe that:

$$\mathrm{CLR} = 2n^{-1/2}\frac{\partial L_n^f(\tilde{\delta},0)}{\partial\theta'} n^{1/2}\begin{bmatrix}\hat{\delta} - \tilde{\delta} \\ \hat{\lambda}\end{bmatrix}$$

$$+\; n\begin{bmatrix}\hat{\delta} - \tilde{\delta} \\ \hat{\lambda}\end{bmatrix}'\left[n^{-1}\frac{\partial^2 L_n^f(\tilde{\delta},0)}{\partial\theta\partial\theta'}\right]\begin{bmatrix}\hat{\delta} - \tilde{\delta} \\ \hat{\lambda}\end{bmatrix} + o_p(1)$$

$$= n^{-1}\frac{\partial L_n^f(\hat{\delta},0)}{\partial\lambda'} V_o(\hat{\lambda})\frac{\partial L_n^f(\tilde{\delta},0)}{\partial\lambda} + o_p(1)$$

$$= \mathrm{CS} + o_p(1) = n\hat{\lambda}'V_o(\hat{\lambda})^{-1}\hat{\lambda} + o_p(1) = \mathrm{MW} + o_p(1) \qquad \text{(A7)}$$

where the second and fourth equalities follow from (A4) and (A6), respectively.

The proof of (iii) follows an argument similar to Holly (1983). By Hausman's (1978) lemma and (A5):

$$H = n\hat{\lambda}'\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix}'\tilde{H}_1(\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1}V(\hat{\delta} - \tilde{\delta})^{-}(\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1}\tilde{H}_1'\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix}\hat{\lambda} + o_p(1)$$

$$= n\hat{\lambda}\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix}'\tilde{H}_1(\tilde{H}_1'\tilde{\Sigma}_1)^{-1}\left[(\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1}\tilde{H}_1'\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix}V_o(\hat{\lambda})\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix}'\tilde{H}_1(\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1}\right]^{-}$$

$$\times\,(\tilde{H}_1'\tilde{\Sigma}\tilde{H}_1)^{-1}\tilde{H}_1'\begin{bmatrix}\tilde{\Sigma}_{xv} \\ \tilde{\Sigma}_{vv}\end{bmatrix}\hat{\lambda} + o_p(1)$$

$$= n\hat{\lambda}'P[P'V_o(\hat{\lambda})P]^{-}P'\hat{\lambda} + o_p(1)$$

in an obvious notation. Since

$$\mathrm{rank}[P'V_o(\hat{\lambda})P] = \mathrm{rank}[V_o(\hat{\lambda})]$$

and $V_o(\hat{\lambda})$ is nonsingular, lemma 2.2.5(c) in Rao and Mitra (1971) implies that, for any choice of generalized inverse,

$$P[P'V(\hat{\lambda})P]^{-}P' = V_o(\hat{\lambda})^{-1}$$

Therefore:

$$H = n\hat{\lambda}'V_o(\hat{\lambda})^{-1}\hat{\lambda} + o_p(1) = MW + o_p(1). \quad \text{(A8)}$$

Q.E.D.

PROOF OF PROPOSITION 5: Let $\alpha \equiv (\theta, \text{vec}\ \Pi, \text{vech}\ \sum_{vv})$ denote again the complete set of parameters. Then it can be verified that, under assumptions 1 and 2, the regularity conditions of Parzen (1954) or Sweeting (1980) are satisfied so that:

$$n^{-1/2}\ \partial L_n(\alpha)/\partial\alpha \xrightarrow{D} N(0, -A^f(\alpha) - A^g(\alpha))$$

uniformly in α. It follows that Theorem 2 of Vuong (1984) holds uniformly in α, i.e.:

$$n^{1/2}(\hat{\theta}-\theta) \xrightarrow{D} N(0, [A_{11}^f(\alpha) - A_{12}^f(\alpha)\tilde{M}(\alpha)^{-1}A_{21}^f(\alpha)]^{-1})$$

uniformly in α. (Here the subscript 1 indicates partitioning with respect to θ and the subscript 2 indicates partitioning with respect to the remaining parameters.) Under the sequence of local alternatives H_n: $\lambda = n^{-1/2}b$, it follows that:

$$n^{1/2}\hat{\lambda} \xrightarrow{D} N(b, V_o(\hat{\lambda}))$$

The expansions (A5) and (A6) are still valid under local alternatives so that (A7) and (A8) also hold. The desired result follows immediately.

Q.E.D.

FOOTNOTES

1. Heckman originally proposed solving (2.1) for one of the observed endogenous variables, replacing y_i^* by an estimate of $E(y_i^*|X_i)$ (e.g., $X_i'\hat{\delta}_n^*$ from equation (3.9) below), and applying least squares.

2. Let V denote the asymptotic covariance matrix of e, i.e., $n^{1/2}e \xrightarrow{D} N(0,V)$ and $\hat{V} \xrightarrow{a.s.} V$ element by element.

3. If (2.1) is just identified, $\hat{H}$ is almost surely nonsingular, 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 a nonsingular transformation of $\hat{\tau}_*$. Hence, $\hat{\delta}_*^A = \hat{\delta}_*^{IVP}$.

4. 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.

5. If $u_1, \ldots, u_m$ are independent normal random variables with $E(u_j) = \mu_j$ and $V(u_j) = 1$, then $\sum_{j=1}^{m} u_j^2$ has a noncentral chi-square distribution with noncentrality parameter $\sum_{j=1}^{m} \mu_j^2$.

BIBLIOGRAPHY

Amemiya, T. "The Estimation of a Simultaneous Equation Generalized Probit Model." Econometrica 46 (1978):1193-1205.

Chandra, T.K. and Joshi, S. N. "Comparison of the Likelihood Ratio, Rao's and Wald's Tests and a Conjecture of C. R. Rao." Sankhya 45 (1983), Series A:226-246.

Godfrey, L. G. and Wickens, M. R. "A Simple Derivation of the Limited Information Maximum Likelihood Estimator." Economic Letters 10 (1982):277-283.

Hausman, J. A. "Specification Tests in Econometrics." Econometrica 46 (1978):1251-1272.

Heckman, J. "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica 46 (1978):931-959.

Henderson, H. V. and Searle, S. R. "Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics." Canadian Journal of Statistics 7 (1979):65-81.

Holly, A. "A Simple Procedure for Testing whether a Subset of Endogenous Variables is Independent of the Disturbance Term in a Structural Equation." Discussion Paper no. 8306. Universite de Lausanne, 1983.

Lee, L. F. "Simultaneous Equation Models with Discrete and Censored

Dependent Variables." In Structural Analysis of Discrete Data with Economic Applications, edited by C. Manski and D. McFadden. Cambridge: MIT Press, 1981.

Parzen, E. "On Uniform Convergence of Sequences of Random Variables." University of California Publications in Statistics 2 (1954):23-53.

Rao, C. R. and Mitra, S. K. Generalized Inverse of Matrices and Its Applications New York: John Wiley, 1971.

Richard, J. F. "A Note on the Information Matrix of the Multivariate Normal Distribution." Journal of Econometrics 3 (1975):57-60.

Rivers, D. and Vuong, Q. H. "Two-Stage Conditional Maximum Likelihood Estimation of Linear Simultaneous Equations." 1984, mimeo.

Sweeting, T.J. "Uniform Asymptotic Normality of the Maximum Likelihood Estimator." Annals of Statistics 8 (1980):1375-1381.

Vuong, Q. H. "Two-Stage Conditional Maximum Likelihood Estimation of Econometric Models." Social Sciance Working Paper no. 538. California Institute of Technology, 1984.

Wald, A. "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large." Transactions of the American Mathematical Society 54 (1943):426-482.