AMEMIYA'S GENERALIZED LEAST SQUARES
AND TESTS OF OVERIDENTIFICATION IN
SIMULTANEOUS EQUATION MODELS WITH QUALITATIVE
OR LIMITED DEPENDENT VARIABLES

by

Lung-Fei Lee

Discussion Paper No. 262, May 1991

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, MN   55455

Amemiya's Generalized Least Squares and Tests of Overidentification in Simultaneous Equation Models with Qualitative or Limited Dependent Variables

by

Lung-Fei Lee
Department of Economics
University of Minnesota
Minneapolis, MN 55455

ABSTRACT

Amemiya's generalized least squares method for the estimation of simultaneous equation models with qualitative or limited dependent variables is known to be efficient relative to many popular two-stage estimators. This note points out that test statistics for overidentification restrictions can be obtained as by-products of Amemiya's generalized least squares procedure. Amemiya's procedure is shown to be a minimum chi-square method. The Amemiya procedure is valuable both for efficient estimation and for model evaluation of such models.

CORRESPONDENCE ADDRESS:
Lung-Fei Lee, 1035 Management and Economics Building, Department of Economics, University of Minnesota, 271 19th Avenue S., Minneapolis, MN 55455

Amemiya's Generalized Least Squares and Tests of Overidentification

in Simultaneous Equation Models with Qualitative or Limited Dependent Variables

by

Lung-Fei Lee*

1. Introduction

Amemiya (1978, 1979) introduced a general method for the estimation of structural equation parameters from estimated reduced form parameters in simultaneous equation models. Amemiya (1978, 1979, 1983), Lee (1981) and Newey (1987) have shown that the Amemiya generalized least squares (AGLS) procedure provides estimators that are more efficient relative to many popular two-stage estimators in simultaneous equation models with qualitative or limited dependent variables. Amemiya's estimation method has been applied, for example, in Lee (1982) for the estimation of a simultaneous equation health and wage model, among others.

Lee (1982) and Newey (1987) have pointed out that for the estimation of a system of equations, Amemiya's estimation procedure is closely related to the classical minimum distance methods [Neyman (1949), Taylor (1953) and Ferguson (1958)].[1] In general, Amemiya's procedure has some common features with existing minimum distance estimation procedures, but they are different. This is so, in particular, for single-equation estimation.[2] In the econometric literature, minimum distance methods have been introduced in Malinvaud (1970) and Rothenberg (1973) for the system estimation of classical simultaneous equation models. With frequency data, they are well-known alternative estimation procedures to the method of maximum likelihood for the estimation of qualitative response models [see, e.g., Amemiya (1981) and Hsiao (1984) for surveys]. For the estimation of panel data models, they provide a systematic way for the estimation of constrained parameters [Chamberlain (1982)]. The classical minimum distance methods are also known as the minimum chi-square methods. The minimized distance function is asymptotically chi-square distributed and can be used as a goodness-of-fit test statistic [see, e.g., Neyman (1949) and Chamberlain (1982)].

---

[1] For a brief survey of several classical minimum distance methods, see Hsiao (1984).

[2] This may explain why Amemiya (1978, 1979) did not name his method as a minimum distance method.

In this note, we point out that the AGLS procedure provides not only relatively efficient estimates but, as a by-product, its minimized distance function provides a convenient statistic for the test of overidentification restrictions. Therefore, Amemiya's GLS method is also a minimum chi-square method. AGLS provides a systematic estimation and specification test procedure for simultaneous equation models with qualitative or limited dependent variables. Such features of the procedure will be valuable for empirical researchers.

2. Amemiya's GLS and Minimum Distance Estimation

Amemiya (1978, 1979) introduced a general method for the estimation of a simultaneous equation probit model and a simultaneous equation tobit model. Some slight generalizations to other models can be found in Lee (1981). Amemiya's method is a systematic procedure for the estimation of structural parameters from estimated reduced form parameters. In many parametric or semiparametric models with qualitative or limited dependent variables, reduced form equations can be estimated by some conventional procedures. With the estimated reduced form parameters, structural parameters can then be derived from Amemiya's procedure. As an illustration, consider the following system of equations model:

$$y^* = y^* B + x\Gamma + \epsilon, \tag{2.1}$$

where y^* is a $1 \times G$ row vector of (latent) endogenous variables, x is a $1 \times K$ vector of exogenous variables and ϵ is a vector of disturbances. The observed sample y is generated by a complicated transformation of y^* [see, e.g., Lee (1981, p. 347)]. Equation (2.1) implies the following reduced form equations:

$$y^* = x\Pi + u, \tag{2.2}$$

where $\Pi = \Gamma(I - B)^{-1}$, and $x = \epsilon(I - B)^{-1}$. Amemiya's method is a two stage procedure. In the first stage, the reduced form parameters are estimated. Each equation in (2.2) can be estimated by some conventional method such as the probit likelihood method or the tobit likelihood method, depending on the nature of the dependent variable and the distributional assumption of the disturbances. Without parametric distributional assumptions, various semiparametric methods are also available for the estimation of such equations. The second stage is to estimate the structural equations of (2.1). Consider the first structural equation

$$y_1^* = y_{(1)}^* \beta_1 + x_1 \gamma_1 + \epsilon_1, \tag{2.3}$$

where $y_{(1)}^*$ is a subvector of endogenous variables other than y_1^* in y^*, and x_1 is a subvector of exogenous variables of x included in the equation. Let J_1 and J_2 be selection matrices such that $y_{(1)}^* = y^* J_1$ and $x_1 = x J_2$. It follows from (2.2) and (2.3) that

$$\begin{aligned} y_1^* &= y^* J_1 \beta_1 + x J_2 \gamma_1 + \epsilon_1 \\ &= x(\Pi J_1 \beta_1 + J_2 \gamma_1) + u_1, \end{aligned} \tag{2.4}$$

3

where $u_1 = uJ_1\beta_1 + \epsilon_1$. Let π_1 denote the first column of Π. Comparing (2.2) with (2.4), we have

$$\pi_1 = \Pi J_1 \beta_1 + J_2 \gamma_1. \tag{2.5}$$

Let $\hat{\Pi}$ be the first stage estimate of Π. Amemiya's generalized least squares (AGLS) method is to apply a generalized least squares (GLS) procedure to estimate the equation:

$$\hat{\pi}_1 = \hat{\Pi} J_1 \beta_1 + J_2 \gamma_1 + \xi_1, \tag{2.6}$$

where

$$\xi_1 = \hat{\pi}_1 - \pi_1 - (\hat{\Pi} - \Pi) J_1 \beta_1.$$

Suppose that $\sqrt{n}\xi_1 \overset{D}{\to} N(0, \Omega_1)$ and $\hat{\Omega}_1$ is a consistent estimate of Ω_1.[3] The AGLS estimator $\hat{\theta}$ of $\theta = (\beta_1', \gamma_1')'$ is derived from

$$\min_{\beta_1, \gamma_1} (\hat{\pi}_1 - \hat{\Pi} J_1 \beta_1 - J_2 \gamma_1)' \hat{\Omega}_1^{-1} (\hat{\pi}_1 - \hat{\Pi} J_1 \beta_1 - J_2 \gamma_1). \tag{2.7}$$

Instead of single equation estimation, the Amemiya procedure can be generalized to the estimation of systems of equations by stacking up the G equations which are similar to (2.6).[4] Under some general regularity conditions, Amemiya (1978, 1979) has shown that the estimators are consistent and asymptotically normal. Furthermore, the AGLS estimators are sysmptotically efficient relative to many conventional two stage probit and tobit estimators [see also Lee (1981), Amemiya (1983) and Newey (1987)].

The minimum distance methods in Taylor (1953), Ferguson (1958), Malinvaud (1970) and Chamberlain (1982), among others, are methods for imposing restrictions in estimation. Let μ be a q-dimensional vector of parameters which depend on a p-dimensional vector θ of unknown parameters:

$$\mu = g(\theta), \tag{2.8}$$

where g is a known function and $p \leq g$. Suppose that μ can be consistently estimated by $\hat{\mu}$ and $\sqrt{n}(\hat{\mu} - \mu) \overset{D}{\to} N(0, \Psi)$, where Ψ is a positive definite matrix. Let $\hat{\Psi}$ be a consistent estimate of Ψ. A minimum distance estimator of θ is $\hat{\theta}$, derived from

$$\min_{\theta} [\hat{\mu} - g(\theta)]' \hat{\Psi}^{-1} [\hat{\mu} - g(\theta)]. \tag{2.9}$$

---

[3] The construction of $\hat{\Omega}_1$, in general, may require some initial consistent estimates of β_1 which can, for example, be derived by the application of the ordinary least square procedure to (2.6).

[4] A systematic estimation procedure has been used in Lee (1982) for an empirical study.

The asymptotic properties such as consistency and asymptotic normality of $\hat{\theta}$ have been developed in the literature [see, e.g., Ferguson (1958) or Chamberlain (1982)]. It is also known [see, e.g., Neyman (1949) or Chamberlain (1982)] that

$$n[\hat{\mu} - g(\hat{\theta})]'\hat{\Psi}^{-1}[\hat{\mu} - g(\hat{\theta})] \xrightarrow{D} \chi^2(q - p), \tag{2.10}$$

which provides a goodness-of-fit test statistic for the restrictions. Comparing (2.7) and (2.9), the AGLS procedure and the minimum distance procedure have some common features. However, the formulation (2.7) is, in general, not exactly the minimum distance formulation in (2.9).[5] In the subsequent section we will show that $n(\hat{\pi}_1 - \hat{\Pi}J_1\beta_1 - J_2\gamma_1)'\hat{\Omega}_1(\hat{\pi}_1 - \hat{\Pi}J_1\beta_1 - J_2\gamma_1)$ is asymptotically chi-square distributed with the number of overidentification restrictions as the degrees of freedom of the statistic. Amemiya's procedure is therefore also a mimimum chi-square method.

---

[5] For system estimation, the AGLS procedure is more directly related to some classical minimum chi-square formulations [see Lee (1982) and Newey (1987)].

3. A Generalized Minimum Chi-square Method and Overidentification Tests

The AGLS procedure and the minimum distance method can be regarded as special cases of a more general minimum distance method. Let $f(\pi, \theta)$ be a continuously twice differentiable function from $R^k \times R^p$ to R^q, where $p \leq q$. Suppose that $\hat{\pi}$ is a consistent estimate of the k-dimensional vector π_0 with an asymptotic normal distribution, specifically, $\sqrt{n}(\hat{\pi} - \pi_0) \xrightarrow{D} N(0, \Sigma)$. By the δ-method,

$$\sqrt{n}(f(\hat{\pi}, \theta_0) - f(\pi_0, \theta_0)) \xrightarrow{D} N(0, \Omega), \tag{3.1}$$

where

$$\Omega = \frac{\partial f(\pi_0, \theta_0)}{\partial \pi'} \Sigma \frac{\partial f'(\pi_0, \theta_0)}{\partial \pi}. \tag{3.2}$$

Supposing $\hat{\Omega}$ is a consistent estimate of Ω, a generalized minimum distance method for the estimation of θ is

$$\min_{\theta \in \Theta} f'(\hat{\pi}, \theta) \hat{\Omega}^{-1} f(\hat{\pi}, \theta), \tag{3.3}$$

where Θ is the parameter space of θ. Let $\hat{\theta}$ be the estimator derived from (3.3). We will show that $\hat{\theta}$ is asymptotically normally distributed and $n f'(\hat{\pi}, \hat{\theta}) \hat{\Omega}^{-1} f(\hat{\pi}, \hat{\theta})$ is asymptotically chi-square distributed with $q - p$ degrees of freedom.

To justify the asymptotic results, we assume the following regularity conditions:

Assumption 1. $f(\pi_0, \theta_0) = 0$; $\hat{\pi} \xrightarrow{P} \pi_0$, and $\sqrt{n}(\hat{\pi} - \pi_0) \xrightarrow{D} N(0, \Sigma)$, where Σ is positive definite; Θ is a compact subset of R^p that contains θ_0; f is continuous on $R^k \times \Theta$, and $f(\pi_0, \theta) = f(\pi_0, \theta_0)$ for $\theta \in \Theta$ implies that $\theta = \theta_0$; $\hat{\Omega} \xrightarrow{P} \Omega$, where $\Omega = \frac{\partial f(\pi_0, \theta_0)}{\partial \pi'} \Sigma \frac{\partial f'(\pi_0, \theta_0)}{\partial \pi}$; and rank $\left(\frac{\partial f(\pi_0, \theta_0)}{\partial \pi'} \right) = q$.

Assumption 2. θ_0 is an interior point of Θ; f is differentiable with π and twice differentiable with θ at a neighborhood of (π_0, θ_0); and rank $\left(\frac{\partial f(\pi_0, \theta_0)}{\partial \theta'} \right) = p$.

The following Proposition 1 establishes the consistency of $\hat{\theta}$, and Proposition 2 derives its asymptotic distribution. Proposition 3 shows that (3.3) is a minimum chi-square method. The proofs are simple but the results are useful.

Proposition 1: Under Assumption 1, $\hat{\theta} \xrightarrow{P} \theta_0$.

6

Proof: Let $D(\theta) = f'(\pi_o, \theta)\Omega^{-1}f(\pi_o, \theta)$ and $D_n(\theta) = f'(\hat{\pi}, \theta)\hat{\Omega}^{-1}f(\hat{\pi}, \theta)$. $D_n(\theta)$ converges in probability to $D(\theta)$ uniformly on Θ. Since $D(\theta)$ has the unique minimum at $\theta = \theta_o$, $\hat{\theta}$ converges in probability to θ_o [see, e.g., Jennrich (1969)].

<div align="right">Q.E.D.</div>

Proposition 2: Under Assumptions 1 and 2, $\sqrt{n}(\hat{\theta} - \theta_o) \xrightarrow{D} N(0, \Lambda)$, where

$$\Lambda = \left(\frac{\partial f'(\pi_o, \theta_o)}{\partial \theta} \Omega^{-1} \frac{\partial f(\pi_o, \theta_o)}{\partial \theta'} \right)^{-1}. \tag{3.4}$$

Proof: Since $\hat{\theta} \xrightarrow{P} \theta_o$, for n sufficiently large, $\frac{\partial f'(\hat{\pi}, \hat{\theta})}{\partial \theta}\hat{\Omega}^{-1}f(\hat{\pi}, \hat{\theta}) = 0$ with probability arbitrarily close to one. By the mean value theorem,

$$0 = \frac{\partial f'(\hat{\pi}, \theta_o)}{\partial \theta}\hat{\Omega}^{-1}f(\hat{\pi}, \theta_o) + \left\{ \frac{\partial f'(\hat{\pi}, \bar{\theta})}{\partial \theta}\hat{\Omega}^{-1}\frac{\partial f(\hat{\pi}, \bar{\theta})}{\partial \theta'} + (f'(\hat{\pi}, \bar{\theta})\hat{\Omega}^{-1} \otimes I) \cdot \frac{\partial}{\partial \theta'}vec\left(\frac{\partial f'(\hat{\pi}, \bar{\theta})}{\partial \theta} \right) \right\}(\hat{\theta} - \theta_o), \tag{3.5}$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and θ_o. By the δ-method,

$$\begin{aligned}
\sqrt{n}f(\hat{\pi}, \theta_o) &= \frac{\partial f(\bar{\pi}, \theta_o)}{\partial \pi'} \cdot \sqrt{n}(\hat{\pi} - \pi_o) \\
&= \frac{\partial f(\pi_o, \theta_o)}{\partial \pi'} \cdot \sqrt{n}(\hat{\pi} - \pi_o) + o_p(1) \\
&\xrightarrow{D} N(0, \Omega).
\end{aligned} \tag{3.6}$$

By the continuous mapping theorem, $f(\hat{\pi}, \bar{\theta}) \xrightarrow{P} 0$, $\frac{\partial f'(\hat{\pi}, \bar{\theta})}{\partial \theta} \xrightarrow{P} \frac{\partial f'(\pi_o, \theta_o)}{\partial \theta}$, and

$$\frac{\partial}{\partial \theta'}vec\left(\frac{\partial f'(\hat{\pi}, \bar{\theta})}{\partial \theta} \right) \xrightarrow{P} \frac{\partial}{\partial \theta'}vec\left(\frac{\partial f'(\pi_o, \theta_o)}{\partial \theta} \right).$$

Hence by the Slutsky lemma,

$$\sqrt{n}(\hat{\theta} - \theta_o) = -\left[\frac{\partial f'(\pi_o, \theta_o)}{\partial \theta}\Omega^{-1}\frac{\partial f(\pi_o, \theta_o)}{\partial \theta'} \right]^{-1} \frac{\partial f'(\pi_o, \theta_o)}{\partial \theta}\Omega^{-1} \cdot \sqrt{n}f(\hat{\pi}, \theta_o) + o_p(1). \tag{3.7}$$

The result follows from (3.6) and (3.7).

<div align="right">Q.E.D.</div>

Proposition 3: Under Assumptions 1 and 2,

$$nf'(\hat{\pi}, \hat{\theta})\hat{\Omega}^{-1}f(\hat{\pi}, \hat{\theta}) \xrightarrow{D} \chi^2(q - p).$$

<div align="center">7</div>

Proof: By the mean value theorem, for sufficiently large n and with probability close to one,

$$f(\hat{\pi}, \hat{\theta}) = \frac{\partial f(\bar{\pi}, \bar{\theta})}{\partial \pi'}(\hat{\pi} - \pi_0) + \frac{\partial f(\bar{\pi}, \bar{\theta})}{\partial \theta'}(\hat{\theta} - \theta_0), \tag{3.8}$$

because $f(\pi_0, \theta_0) = 0$. With (3.6) and (3.7),

$$\begin{aligned}
&\sqrt{n} f(\hat{\pi}, \hat{\theta}) \\
&= \frac{\partial f(\pi_0, \theta_0)}{\partial \pi'} \cdot \sqrt{n}(\hat{\pi} - \pi_0) + \frac{\partial f(\pi_0, \theta_0)}{\partial \theta'} \cdot \sqrt{n}(\hat{\theta} - \theta_0) + o_p(1) \\
&= \left\{ I - \frac{\partial f(\pi_0, \theta_0)}{\partial \theta'} \left[\frac{\partial f'(\pi_0, \theta_0)}{\partial \theta} \Omega^{-1} \frac{\partial f(\pi_0, \theta_0)}{\partial \theta'} \right]^{-1} \frac{\partial f'(\pi_0, \theta_0)}{\partial \theta} \Omega^{-1} \right\} \cdot \sqrt{n} f(\hat{\pi}, \theta_0) + o_p(1).
\end{aligned} \tag{3.9}$$

Let C be a nonsingular matrix such that $CC' = \Omega^{-1}$. It follows from (3.6) that

$$C' \cdot \sqrt{n} f(\hat{\pi}, \theta_0) \xrightarrow{D} N(0, I). \tag{3.10}$$

Denote

$$M = I - C' \frac{\partial f(\pi_0, \theta_0)}{\partial \theta'} \left[\frac{\partial f'(\pi_0, \theta_0)}{\partial \theta} CC' \frac{\partial f(\pi_0, \theta_0)}{\partial \theta'} \right]^{-1} \frac{\partial f'(\pi_0, \theta_0)}{\partial \theta} C, \tag{3.11}$$

which is an idempotent matrix with rank $q - p$. Equation (3.9) implies that

$$C' \cdot \sqrt{n} f(\hat{\pi}, \hat{\theta}) = M \cdot C' \sqrt{n} f(\hat{\pi}, \theta_0) + o_p(1). \tag{3.12}$$

It follows from (3.9)-(3.12) that

$$\begin{aligned}
n f'(\hat{\pi}, \hat{\theta}) \hat{\Omega}^{-1} f(\hat{\pi}, \hat{\theta}) &= n f'(\hat{\pi}, \hat{\theta}) \Omega^{-1} f(\hat{\pi}, \hat{\theta}) + o_p(1) \\
&= n f'(\hat{\pi}, \hat{\theta}) CC' f(\hat{\pi}, \hat{\theta}) + o_p(1) \\
&= n f'(\hat{\pi}, \theta_0) CMC' f(\hat{\pi}, \theta_0) + o_p(1) \\
&\xrightarrow{D} \chi^2(q - p).
\end{aligned} \tag{3.13}$$

Q.E.D.

The generalized minimum distance method in (3.3) has also an interest property on the testing of nested restrictions as the minimum distance method in (2.9). Consider imposing additional restrictions, which can be expressed as $\theta = g(\alpha)$, where g is a known function from R^s to R^p with $s \leq p$.

Assumption 3: The parameter space $\mathcal{A}$ of α is a compact subset of R^s; the true parameter vector α_0 is in the interior of $\mathcal{A}$; g is a continuous mapping on $\mathcal{A}$; $g(\alpha) = \theta_0$ for $\alpha \in \mathcal{A}$ implies $\alpha = \alpha_0$; g is twice continuously differentiable at α_0; and $\text{rank}\left(\frac{\partial g(\alpha_0)}{\partial \alpha'} \right) = s$.

8

The generalized minimum distance estimator $\hat{\alpha}$ is derived from

$$\min_{\alpha \in \mathcal{A}} f'(\hat{\pi}, g(\alpha)) \hat{\Omega}^{-1} f(\hat{\pi}, g(\alpha)). \tag{3.14}$$

Proposition 4: Under Assumptions 1, 2 and 3, $d_1 - d_2 \xrightarrow{D} \chi^2(p - s)$, where

$$d_1 = n f'(\hat{\pi}, g(\hat{\alpha})) \hat{\Omega}^{-1} f(\hat{\pi}, g(\hat{\alpha})), \tag{3.15}$$

and

$$d_2 = n f'(\hat{\pi}, \hat{\theta}) \hat{\Omega}^{-1} f(\hat{\pi}, \hat{\theta}). \tag{3.16}$$

Furthermore, $d_1 - d_2$ is asymptotically independent of d_2, and $d_1 - d_2$ is asymptotically equivalent to d_3.

Proof: Similar to the derivation of (3.13),

$$n f'(\hat{\pi}, g(\hat{\alpha})) \hat{\Omega}^{-1} f(\hat{\pi}, g(\hat{\alpha})) = u' M_\alpha u + o_p(1), \tag{3.17}$$

where $u = C' \sqrt{n} f(\hat{\pi}, \theta_o)$,

$$M_\alpha = I - D' \frac{\partial g(\alpha_o)}{\partial \alpha'} \left[\frac{\partial g'(\alpha_o)}{\partial \alpha} DD' \frac{\partial g(\alpha_o)}{\partial \alpha'} \right]^{-1} \frac{\partial g'(\alpha_o)}{\partial \alpha} D, \tag{3.18}$$

and $D = \frac{\partial f'(\pi_o, \theta_o)}{\partial \theta} C$. It follows from (3.13) and (3.17) that

$$d_1 - d_2 = u'(M_\alpha - M) u + o_p(1). \tag{3.19}$$

Since DD' is positive definite, there exists a nonsingular matrix A such that $AA' = DD'$. Equations (3.11) and (3.18) imply that

$$M_\alpha - M = D' \left\{ (AA')^{-1} - \frac{\partial g(\alpha_o)}{\partial \alpha'} \left[\frac{\partial g'(\alpha_o)}{\partial \alpha} AA' \frac{\partial g(\alpha_o)}{\partial \alpha'} \right]^{-1} \frac{\partial g'(\alpha_o)}{\partial \alpha} \right\} D$$
$$= D' A'^{-1} Q A^{-1} D, \tag{3.20}$$

where $Q = I - A' \frac{\partial g(\alpha_o)}{\partial \alpha'} \left[\frac{\partial g'(\alpha_o)}{\partial \alpha} AA' \frac{\partial g(\alpha_o)}{\partial \alpha'} \right]^{-1} \frac{\partial g'(\alpha_o)}{\partial \alpha} A$ is an idempotent matrix with rank $p - s$. Since $A^{-1} D (A^{-1} D)' = I$, it follows from (3.10) that $A^{-1} D u \xrightarrow{D} N(0, I)$. Hence $d_1 - d_2 \xrightarrow{D} \chi^2(p - s)$. From (3.13),

$$d_2 = u'[I - D'(DD')^{-1} D] u + o_p(1). \tag{3.21}$$

9

Since $[I - D'(DD')^{-1}D]D' = 0$, it follows that $[I - D'(DD')^{-1}D](M_\alpha - M) = 0$. Hence, it is evident from (3.19) and (3.21) that $d_1 - d_2$ is asymptotically independent of d_2.

<div align="right">Q.E.D.</div>

For AGLS, $\theta = (\beta_1, \gamma_1)$, $f(\hat{\pi}, \theta) = \hat{\pi}_1 - [\hat{\Pi}J_1, J_2]\theta$, $q = K$ and $p = m_1 + k_1$, where K is the total number of exogenous variables in the system, m_1 is the number of endogenous variables, and k_1 is the number of exogenous variables included on the right-hand side of (2.3). $q - p = K - m_1 - k_1$ is the number of overidentification restrictions in (2.3). Proposition 3 implies that

$$n(\hat{\pi}_1 - \hat{\Pi}J_1\hat{\beta}_1 - J_2\hat{\gamma}_1)'\hat{\Omega}_1^{-1}(\hat{\pi}_1 - \hat{\Pi}J_1\hat{\beta}_1 - J_2\hat{\gamma}_1) \xrightarrow{D} \chi^2(K - m_1 - k_1), \tag{3.22}$$

which provides a statistic for the testing of overidentification restrictions.

The overidentification test statistic above is related to some overidentification test statistics in the literature under some situations. For the single equation estimation of the classical simultaneous equation model, AGLS estimator is identical to Theil's two stage least squares (2LS) estimator [Amemiya (1978)]. It can be easily shown that for the classical model, the overidentification test statistic (3.22) corresponds to the statistic:

$$(Y_1 - Y_{(1)}\hat{\beta}_1 - X_1\hat{\gamma}_1)'X(X'X)^{-1}X'(Y_1 - Y_{(1)}\hat{\beta}_1 - X_1\hat{\gamma}_1)/\hat{\sigma}_1^2, \tag{3.23}$$

where Y_1, $Y_{(1)}$, X_1 and X are the data matrices of y_1, $y_{(1)}$, x_1 and x, respectively, $(\hat{\beta}_1, \hat{\gamma}_1)$ is the 2SL estimate of (β_1, γ_1), and $\hat{\sigma}_1^2$ is a consistent estimate of the variance of ϵ_1. The statistic (3.23) has been derived as a modification of the Anderson-Rubin (1949) overidentification test statistic in Hausman (1983, p. 433). The test statistic (3.23) can also be derived from Hansen's generalized method of moments (GMM) procedure [Hansen(1982) and Newey (1987)]. For the special cases where $\hat{\pi}_1$ and $\hat{\Pi}$ in the first stage estimation can be expressed as sample moments of the data, the test statistic (3.22) will correspond to Hansen's overidentification test [see Hansen (1982)].

References

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

2. Amemiya, T., 1979, The estimation of a simultaneous equation tobit model, International Economic Review 20, 169-181.

3. Amemiya, T., 1981, Qualitative response models: a survey, Journal of Economic Literature 19, 1483-1536.

4. Amemiya, T., 1983, A comparison of the Amemiya GLS and the Lee-Maddala-Trost G2SLS in a simultaneous-equations tobit model, Journal of Econometrics 23, 295-300.

5. Anderson, T.W. and H. Rudin, 1949, Estimation of the parameters of a single equation in a complete system of stochastic equations, Annals of Mathematical Statistics 20, 46-63.

6. Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics 18, 5-46.

7. Ferguson, T.S., 1958, A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities, Annals of Mathematical Statistics 29, 1046-1062.

8. Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

9. Hausman, J.A., 1983, Specification and estimation of simultaneous equation models, in: Z. Griliches and M.D. Intrilligator, eds., Handbook of econometrics, Vol. 1 (North-Holland, Amsterdam).

10. Hsiao, C., 1984, Minimum chi-square, MRG working paper #M8401, (Department of Economics, University of Southern California, CA).

11. Jennrich, R.I., 1969, Asymptotic properties of non-linear least squares estimators, The Annals of Mathematical Statistics 40, 633-643.

12. Lee, L.F., 1981, Simultaneous equations models with discrete and censored variables, in: C.F. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT press, Cambridge, MA).

13. Lee, L.F., 1982, Health and wage: a simultaneous equation model with multiple discrete indicators, International Economic Review 23, 199-221.

14. Malinvaud, E., 1970, Statistical methods of econometrics (North Holland, Amsterdam).

15. Neyman, J., 1949, Contribution to the theory of χ^2 test, in: J. Neyman, ed., Proceedings of the first Berkeley symposium on mathematical statistics and probability (University of California Press, Berkeley) 239-273.

16. Newey, W.K., 1982, Maximum likelihood specification testing and conditional moment tests, Econometrica 53, 1047-1069.

17. Newey, W.K., 1987, Efficient estimation of limited dependent variable models with endogenous explanatory variables, Journal of Econometrics 36, 231-250.

18. Taylor, W.F., 1953, Distance functions and regular best asymptotically normal estimates, Annals of Mathematical Statistics 24, 85-92.

19. Rothenberg, T., 1973, Efficient estimation with a priori information (Yale University Press, New Haven).