DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

PARAMETERIZATION AND TWO-STAGE CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION

Da-Hsiang Donald Lien and Quang H. Vuong
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 578

July 1985

PARAMETERIZATION AND TWO-STAGE CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION

Da-Hsiang Donald Lien and Quang H. Vuong
California Institute of Technology

ABSTRACT

This paper considers the case where, after appropriate reparameterization, the probability density function can be factorized into a marginal density function and a conditional density function such that one of them involves fewer parameters. Then, two types of two-stage conditional maximum-likelihood estimators, 2SCMLEI and 2SCMLEII, can be considered according to whether the marginal or the conditional density has fewer parameters. Our first result indicates that, under some identification assumptions, there is a connection between the number of parameters in the marginal (or conditional) density functions under the two reparameterizations. Moreover, conditions for asymptotic equivalence and numerical equivalence between these two-stage estimators and the FIML estimator are obtained. Finally, examples are provided to illustrate our results.

PARAMETERIZATION AND TWO-STAGE CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION

Da-Hsiang Donald Lien and Quang H. Vuong*

1. INTRODUCTION

Ever since Wald (1949) and Lecam (1953), maximum likelihood estimation has been widely applied to non-linear models due to its nice asymptotic properties, such as strong consistency and asymptotic efficiency. In general, however, MLE's (Maximum Likelihood Estimators) are difficult to compute. In addition, when the log-likelihood function is not globally concave in the parameter, computation of the MLE's heavily relies on good initial estimator. Thus, more tractable estimators that are consistent but not as efficient as MLE's are often desired.

In this paper, we consider the case where the probability density function can be factorized into a marginal density function and a conditional density function such that one of them involves fewer number of parameters. In such a situation, two-stage conditional maximum likelihood estimators (2SCMLE's) can be constructed. In Vuong (1984), one such estimator was carefully studied; this estimator that we call 2SCMLEI, used the fact that fewer parameters appear in the marginal density than in the conditional density. Necessary and sufficient conditions for asymptotic efficiency were derived under general conditions. In the alternative situation where fewer parameters appear in the conditional density than in the marginal density, we can consider another 2SCML estimator, namely 2SCMLEII. Due to the similarity of the two 2SCMLE's, 2SCMLEII is expected to possess the same statistical properties as 2SCMLEI.

At this point, it is then natural to investigate the relationship between 2SCMLEI and 2SCMLEII, when, after suitable reparameterizations of the model of interest, both methods can be carried out. The first result of this paper indicates that there is a connection between the number of parameters in the marginal [or conditional] density functions under standard identification assumptions. Then, we study conditions under which 2SCMLEI and 2SCMLEII are asymptotically equivalent. We show that, when a certain condition holds on the number of parameters in the marginal or conditional densities, then these two two-stage estimators are asymptotically equivalent if and only if they are both asymptotically efficient. If in addition 2SCMLEI and 2SCMLEII are both unique, then we can establish a stronger result, namely the numerical equivalence between 2SCMLEI, 2SCMLEII, and FIML (Full Information Maximum Likelihood) estimators. As an illustration, we consider the seemingly unrelated regression model of Zellner (1962) with some exclusion restrictions. Our results then state that, for this particular model 2SCMLEI is numerically equal to FIML and hence asymptotically efficient. Moreover, we also show that the property holds even when one of the variables is observed only discretely.

The structure of this paper is as follows. Section 2 presents the definitions and basic framework. Section 3 compares the parameterizations for 2SCMLEI and 2SCMLEII. A theorem relating the number of parameters in both parameterizations is derived. Section 4 states two equivalence theorems and Section 5 presents some numerical equivalence examples. Section 6 concludes the paper. All the proofs are collected in the Appendix.

2. NOTATIONS AND BASIC ASSUMPTIONS

Let X_t be an $m \times 1$ observed random vector defined on an Euclidean measurable space (X, σ_x, ν_x), while the process generating the observations X_t, $t = 1,2,\ldots$ satisfies the following assumption:

Assumption A1: The random vectors X_t, $t = 1,2,\ldots$ are independent and identically distributed with common true cummulative distribution function H^0 on (X, σ_x, ν_x).

As in Vuong (1984), we now partition X_t into $(Y_{1t}', Y_{2t}', Z_t')'$ where Y_{1t}, Y_{2t} and Z_t are respectively p_1, p_2 and q dimensional vectors with $m = p_1 + p_2 + q$. Furthermore, let $Y_t = (Y_{1t}', Y_{2t}')'$ and denote the true (but unknown) conditional distribution of Y_t given Z_t by $F^0_{Y|Z}(\cdot|\cdot)$. To estimate $F^0_{Y|Z}$, we specify a parametric family of conditional distributions $F^\theta_{Y|Z}(\cdot|\cdot;\theta)$ where $\theta \in \Theta \subset \mathbb{R}^k$. Given $F^\theta_{Y|Z}(\cdot|\cdot;\theta)$, we can derive the conditional distribution of Y_{1t} given (Y_{2t}, Z_t), $F^\theta_1(y_{1t}|y_{2t}, z_t;\theta)$ and the conditional distribution of Y_{2t} given Z_t, $F^\theta_2(y_{2t}|z_t;\theta)$.

Assumption A2: Θ is a compact subset of $\mathbb{R}^k$ such that (a) for every $\theta \in \Theta$, and for all z, $F^\theta_{Y|Z}(\cdot|\cdot;\theta)$ has a density with respect to ν_Y (derived from ν_X): $f^\theta(\cdot|z;\theta) = dF^\theta_{Y|Z}(\cdot|z;\theta)/d\nu_Y$; (b) the conditional densities $f^\theta_1(y_1|y_2, z;\theta)$ and $f^\theta_2(y_2|z;\theta)$ are strictly positive functions that are measurable in (y,z) for any θ, and continuous in θ for all (y,z).

Assumption A2-(a) ensures that the density functions f^θ_1 and f^θ_2 exist, while assumption A2-(b) requires in particular that the conditional models for Y_{1t} given (Y_{2t}, Z_t) and Y_{2t} given Z_t are homogeneous (see, e.g., Lehman (1957), Monfort (1982)). To apply our two stage estimation procedures, we require that either f^θ_1 or f^θ_2 contains fewer number of parameters. A direct approach will be imposing these conditions on f^θ_1 or f^θ_2 as in Vuong (1984). Alternatively, we may employ appropriate reparameterizations to incorporate these necessities.

Definition 2.1: A parameter $\alpha \in A \subset \mathbb{R}^k$ is said to be a (proper) reparameterization of $\theta \in \Theta$ if and only if there exists a mapping $\alpha(\cdot)$ from Θ to A such that $\alpha(\cdot)$: $\theta \to \alpha(\theta) = \alpha$ satisfies: (i) $\alpha(\cdot)$ is bijective; (ii) $\alpha(\cdot)$ and $\alpha^{-1}(\cdot)$ are C^0.[1]

Now given the parametric probability family $\{f^\theta(\cdot,\cdot|z;\theta);\theta \in \Theta\}$, to apply our two stage estimation procedure, we require a reparameterization α (or β) of $\theta \in \Theta$ such that $\{f^\theta(\cdot,\cdot|z;\theta);\theta \in \Theta \subset \mathbb{R}^k\} = \{f^\alpha(\cdot,\cdot|z;\alpha);\alpha \in A \subset \mathbb{R}^k\}$ [or $\{f^\beta(\cdot,\cdot|z;\beta);\beta \in B \subset \mathbb{R}^k\}$] and f^α_2 only depends on a subset parameter vector of α [or f^β_1 only depends on a subset parameter vector of β]. Formally, 2SCMLEI requires assumption A3-I; 2SCMLEII requires assumption A3-II.

<u>Assumption A3-I</u>: Given $\{f^\theta(\cdot,\cdot|z;\theta);\theta \in \Theta\}$, (a) there exists a reparameterization $\alpha(\cdot)$: $\theta \to \alpha(\theta) = \alpha$ such that $\alpha = (\alpha_1', \alpha_2')'$ with $\alpha_i \in A_i \subset \mathbb{R}^{k_i}$, $k_i > 0$, $\forall i = 1,2$, $k_1 + k_2 = k$; (b) for every $\theta \in \Theta$, $f^\theta(y_1,y_2|z;\theta) = f^\alpha(y_1,y_2|z;\alpha)$, $f_1^\theta(y_1|y_2,z;\theta) = f_1^\alpha(y_1|y_2,z;\alpha)$, and $f_2^\theta(y_2|z;\theta) = f_2^\alpha(y_2|z;\alpha_2)$.

<u>Assumption A3-II</u>: Given $\{f^\theta(\cdot,\cdot|z;\theta);\theta \in \Theta\}$, (a) there exists a reparameterization $\beta(\cdot)$: $\theta \to \beta(\theta) = \beta$ such that $\beta = (\beta_1', \beta_2')'$ with $\beta_i \in B_i \subset \mathbb{R}^{\ell_i}$, $\ell_i > 0$, $\forall i = 1,2$, $\ell_1 + \ell_2 = k$; (b) for every $\theta \in \Theta$, $f^\theta(y_1,y_2|z;\theta) = f^\beta(y_1,y_2|z;\beta)$, $f_1^\theta(y_1|y_2,z;\theta) = f_1^\beta(y_1|y_2,z;\beta_1)$, and $f_2^\theta(y_2|z;\theta) = f_2^\beta(y_2|z;\beta)$.

In other words, after reparameterizations, Assumption A3-I ensures that the marginal density (with respect to "conditional" models) involves fewer number of parameters while A3-II ensures that the conditional density involves fewer number of parameters. Given assumptions A2, A3-I, A3-II, we can define (almost surely) the conditional log-likelihood function in the following three ways:

$$\text{(i)} \quad L_n^\theta(Y_1,Y_2|Z;\theta) = \sum_{t=1}^n \log f^\theta(y_{1t},y_{2t}|z_t;\theta) \tag{2.1}$$

$$\text{(ii)} \quad L_n^\alpha(Y_1,Y_2|Z;\alpha) = \sum_{t=1}^n \log f^\alpha(y_{1t},y_{2t}|z_t;\alpha)$$

$$= L_{1n}^\alpha(Y_1|Y_2,Z;\alpha) + L_{2n}^\alpha(Y_2|Z;\alpha_2) \tag{2.2a}$$

$$\text{where} \quad L_{1n}^\alpha(Y_1|Y_2,Z;\alpha) = \sum_{t=1}^n \log f_1^\alpha(y_{1t}|y_{2t},z_t;\alpha) \tag{2.3a}$$

$$\text{and} \quad L_{2n}^\alpha(Y_1|Y_2,Z;\alpha_2) = \sum_{t=1}^n \log f_2^\alpha(y_{2t}|z_t;\alpha_2) \tag{2.4a}$$

$$\text{(iii)} \quad L_n^\beta(Y_1,Y_2|Z;\beta) = \sum_{t=1}^n \log f^\beta(y_{1t},y_{2t}|z_t;\beta)$$

$$= L_{1n}^\beta(Y_1|Y_2,Z;\beta_1) + L_{2n}^\beta(Y_2|Z;\beta) \tag{2.2b}$$

$$\text{where} \quad L_{1n}^\beta(Y_1|Y_2,Z;\beta_1) = \sum_{t=1}^n \log f_1^\beta(y_{1t}|y_{2t},z_t;\beta_1) \tag{2.3b}$$

$$\text{and} \quad L_{2n}^\beta(Y_2|Z;\beta) = \sum_{t=1}^n \log f_2^\beta(y_{2t}|z_t;\beta) \tag{2.4b}$$

Obviously, by assumptions we know that $L_n^\theta \equiv L_n^\alpha \equiv L_n^\beta$.

<u>Definition 2.2</u>: A CMLE (Conditional Maximum Likelihood Estimator) is a σ_x^n-measurable function $\hat{\theta}_n$ of $X_1,X_2,\ldots,X_n$ such that:

$$L_n^\theta(Y_1,Y_2|Z;\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n^\theta(Y_1,Y_2|Z;\theta) \tag{2.5}$$

<u>Definition 2.3</u>: A 2SCMLEI (Two Stage Conditional Maximum Likelihood Estimator I) is a σ_x^n-measurable function $\hat{\alpha}_n = (\hat{\alpha}_{1n}', \hat{\alpha}_{2n}')'$ of $(X_1,X_2,\ldots,X_n)$ such that:

$$L_{2n}^\alpha(Y_2|Z;\hat{\alpha}_{2n}) = \sup_{\alpha_2 \in A_2} L_{2n}^\alpha(Y_2|Z;\alpha_2) \tag{2.6a}$$

$$L^{\alpha}_{1n}(Y_1|Y_2,Z;\hat{\alpha}_{1n},\hat{\alpha}_{2n}) = \sup_{\alpha_1 \varepsilon A_1(\hat{\alpha}_{2n})} L^{\alpha}_{1n}(Y_1|Y_2,Z;\alpha_1,\hat{\alpha}_{2n}) \qquad (2.7a)$$

where A_2 is the projection of A (i.e. $\alpha(\theta)$) on the α_2-hyperplane and $A_1(\alpha_2)$ is the section of A at α_2.

Definition 2.4: A 2SCMLEII (Two Stage Conditional Maximum Likelihood Estimator II) is a σ^n_x-measurable function $\hat{\beta}_n = (\hat{\beta}'_{1n},\hat{\beta}'_{2n})'$ of $(X_1,X_2,\ldots,X_n)$ such that:

$$L^{\beta}_{1n}(Y_1|Y_2,Z;\hat{\beta}_{1n}) = \sup_{\beta_1 \varepsilon B_1} L^{\beta}_{1n}(Y_1|Y_2,Z;\beta_1) \qquad (2.6b)$$

$$L^{\beta}_{2n}(Y_2|Z;\hat{\beta}_{1n},\hat{\beta}_{2n}) = \sup_{\beta_2 \varepsilon B_2(\hat{\beta}_{1n})} L^{\beta}_{2n}(Y_2|Z;\hat{\beta}_{1n},\beta_2) \qquad (2.7b)$$

where B_1 is the projection of B (i.e. $\beta(\theta)$) on the β_1-hyperplane, and $B_2(\beta_1)$ is the section of B at β_1.

From the above definitions, we can easily see that these two-stage conditional estimators are easier to compute than the CMLE, due to the advantage of having fewer parameters in either conditional density or marginal density. On the other hand, CMLE is actually FIML in the conditional model. Moreover, since we are interested in the estimations of θ, once $\hat{\alpha}_n$ or $\hat{\beta}_n$ are obtained, $\alpha^{-1}(\cdot)$ or $\beta^{-1}(\cdot)$ must be applied to get 2SCMLEI for θ as $\theta_{2SCMLEI} = \alpha^{-1}(\hat{\alpha}_n)$ or 2SCMLEII for θ as $\theta_{2SCMLEII} = \beta^{-1}(\hat{\beta}_n)$.

As shown in Vuong (1983, 1984), both CMLE and 2SCMLEI are consistent estimators under appropriate regularity conditions. The

asymptotic variance-covariance matrix for 2SCMLEI under correct "conditional" model specification (i.e. $F^{\alpha}_{Y|Z}(\cdot|\cdot;\alpha^0) = F^0_{Y|Z}(\cdot|\cdot)$ for some α^0 in A) was shown to be:

$$\textstyle\sum^{\alpha}(\alpha^0) = \begin{bmatrix} B^1_{\alpha_1\alpha_1}(\alpha^0) & -A^1_{\alpha_1\alpha_2}(\alpha^0) \\ -A^1_{\alpha_2\alpha_1}(\alpha^0) & B^2_{\alpha_2\alpha_2}(\alpha^0_2) + A^1_{\alpha_2\alpha_1}(\alpha^0)[B^1_{\alpha_1\alpha_1}(\alpha^0)]^{-1}A^1_{\alpha_1\alpha_2}(\alpha^0) \end{bmatrix}^{-1} \qquad (2.8a)$$

$$\text{with } A^1_{\alpha_1\alpha}(\alpha^0) = E^0\left[\frac{\partial^2 \log f^{\alpha}_1(y_1|y_2,z;\alpha^0)}{\partial\alpha_1\partial\alpha'}\right] \qquad (2.9a)$$

$$B^1_{\alpha_1\alpha_1}(\alpha^0) = E^0\left[\frac{\partial \log f^{\alpha}_1(y_1|y_2,z;\alpha^0)}{\partial\alpha_1}\cdot\frac{\partial \log f^{\alpha}_1(y_1|y_2,z;\alpha^0)}{\partial\alpha'_1}\right] \qquad (2.10a)$$

$$B^2_{\alpha_2\alpha_2}(\alpha^0_2) = E^0\left[\frac{\partial \log f^{\alpha}_2(y_2|z;\alpha^0_2)}{\partial\alpha_2}\cdot\frac{\partial \log f^{\alpha}_2(y_2|z;\alpha^0_2)}{\partial\alpha'_2}\right] \qquad (2.11a)$$

where $E^0[\cdot]$ is the expectation with respect to the true c.d.f. $H^0(\cdot)$, $A^1_{\alpha_1\alpha_2}(\cdot)$ is the $k_1 \times k_2$ matrix obtained from $A^1_{\alpha_1\alpha}(\cdot)$ by deleting its first k_1 columns, and $A^1_{\alpha_2\alpha_1}(\cdot) = [A^1_{\alpha_1\alpha_2}(\cdot)]'$.

From (2.8a), necessary and sufficient conditions for 2SCMLEI to be asymptotically efficient under correct "conditional" model specification were established. Furthermore, exogeneity tests of the Holly and Sargan (1982), Holly (1983), and Rivers and Vuong (1984) type, and model specification tests along the lines of Hausman (1978) and White (1982) can be constructed from 2SCMLEI. Due to the similar structure of both two-stage estimators, it is expected that 2SCMLEII possess similar statistical properties.[2] These properties can actually be derived following Vuong (1984) under appropriate additional assumptions. For example, assume $F^{\beta}_{Y|Z}(\cdot|\cdot;\beta^0) = F^0_{Y|Z}(\cdot|\cdot)$ for some β^0 in B, then it can easily be shown that the asymptotic variance-covariance matrix of 2SCMLEII is:

$$\sum{}^{\beta}(\beta^0) = \begin{bmatrix} D^1_{\beta_1\beta_1}(\beta^0_1) + C^2_{\beta_1\beta_2}(\beta^0)[D^2_{\beta_2\beta_2}(\beta^0)]^{-1}C^2_{\beta_2\beta_1}(\beta^0) & -C^2_{\beta_1\beta_2}(\beta^0) \\ -C^2_{\beta_2\beta_1}(\beta^0) & D^2_{\beta_2\beta_2}(\beta^0) \end{bmatrix}^{-1} \tag{2.8b}$$

with

$$D^1_{\beta_1\beta_1}(\beta^0_1) = E^0\left[\frac{\partial \log f^{\beta}_1(y_1|y_2,z;\beta^0_1)}{\partial\beta_1}\cdot\frac{\partial \log f^{\beta}_1(y_1|y_2,z;\beta^0_1)}{\partial\beta_1'}\right] \tag{2.9b}$$

$$C^2_{\beta_2\beta}(\beta^0) = E^0\left[\frac{\partial^2 \log f^{\beta}_2(y_2|z;\beta^0)}{\partial\beta_2\partial\beta'}\right] \tag{2.10b}$$

$$D^2_{\beta_2\beta_2}(\beta^0) = E^0\left[\frac{\partial \log f^{\beta}_2(y_2|z;\beta^0)}{\partial\beta_2}\cdot\frac{\partial \log f^{\beta}_2(y_2|z;\beta^0)}{\partial\beta_2'}\right] \tag{2.11b}$$

where, again, $E^0[\cdot]$ is the expectation with respect to the true c.d.f. $H^0(\cdot)$, $C^2_{\beta_2\beta_1}(\cdot)$ is the $\ell_2 \times \ell_1$ matrix obtained from $C^2_{\beta_2\beta}(\cdot)$ by deleting its last ℓ_2 columns, and $C^2_{\beta_1\beta_2}(\cdot) = [C^2_{\beta_2\beta_1}(\cdot)]'$. All the other properties of 2SCMLEI can also be established for 2SCMLEII.

3. A GENERAL RESULT ON PARAMETERIZATION

Now, we turn to the problem of comparing 2SCMLEI and 2SCMLEII since, as the examples below illustrate, it is often possible to find two reparametizations of a given parametric model that will satisfy Assumptions A3-I and A3-II respectively. First, note that from Definitions 2.3-2.4, the partitions of these two estimators to which two-stage estimation procedures apply are different if $k_1 \neq \ell_1$ or $k_2 \neq \ell_2$. In case they are different, some parameter estimators will use information contained in the marginal density under 2SCMLEI while information contained in the conditional density will be used under 2SCMLEII. Therefore, any kind of equivalence relationship between 2SCMLEI and 2SCMLEII will be difficult to establish. Hence, a preliminary problem relating k_1 to ℓ_1 or k_2 to ℓ_2 has to be considered before comparing these two two-stage estimators.

The main purpose of this section is to show that, given the partitions (α_1,α_2) and (β_1,β_2) of Assumption 3, then one has $k_1 \leq \ell_1$ or equivalently $k_2 \geq \ell_2$ under some identification conditions. To establish this result, we need some preliminary definitions. Following Matsushima (1972), we define the dimension of a set as follows.

<u>Definition 3.1</u>: A set $X \subset \mathbb{R}^p$ is said to be of (Euclidean) dimension k at $x \in X$ if and only if there exists a (relative) neighborhood of x, N(x), and a C^0-function ϕ from N(x) to an open set U of $\mathbb{R}^k$ such

that d is bijective and d^{-1} is C^0, (i.e., d is a homeomorphism).

The dimension of a set is then defined as follows.

Definition 3.2. A set $X \subset \mathbb{R}^p$ is said to be of (Euclidean) dimension k, denoted by dim X = k, if and only if for every $x \in X$, the dimension of X at x is k.

Assumption A4: dim Θ = k.[3]

Assumption A4 then requires that at each point of Θ, which is assumed to be compact (Assumption A2), the dimension is well-defined. In addition, it requires that the dimension be constant and equal to k at every point of Θ. Thus Assumption A4 implies some restrictions on the parameter space Θ.

To prove the desired result, $k_1 \le \ell_1$ or $k_2 \ge \ell_2$, we need four lemmas which are now presented. The first lemma relates the dimension of A, A_i, and B_i to k, k_i, and ℓ_i.

Lemma 3.1: Given A3-I, A3-II and A4, dim A = dim B = dim Θ = k, where $A = \alpha(\Theta)$, $B = \beta(\Theta)$. Moreover, dim $A_i = k_i$ and dim $B_i = \ell_i$, i = 1,2.

The above lemma states that, although we may use the reparameterizations α and β to reduce the numbers of parameters in the marginal density or conditional density, yet the whole probability density functions after reparameterizations maintain the same number of parameters. The result is obvious, because otherwise Assumption A4 will be violated.

Furthermore, as shown in Vuong (1984), to derive the statistical properties of 2SCMLEI, some identification assumptions must be imposed, namely that α_1 be identified in $f_1^\alpha(Y_1|Y_2,Z;\alpha_1,\alpha_2)$ given α_2. Similarly, we require that β_2 be identified in $f_2^\beta(Y_2|Z;\beta_1,\beta_2)$ given β_1 to derive the statistical properties of 2SCMLEII. Formally, following Barankin (1960):

Definition 3.3. Given a collection of probability density functions $\{p(\cdot;\theta), \theta \in \Theta \subset \mathbb{R}^p\}$, a sub-vector of θ, $\theta_1 \in \Theta_1(\bar{\theta}_2)$ is identified in $p(\cdot;\theta)$ given $\bar{\theta}_2 \in \Theta_2$ if and only if for any $\tilde{\theta}_1 \in \Theta_1(\bar{\theta}_2)$ with $\tilde{\theta}_1 \neq \theta_1$, $p(\cdot;\theta_1,\bar{\theta}_2) \neq p(\cdot;\tilde{\theta}_1,\bar{\theta}_2)$, where Θ_2 is the projection of Θ on the θ_2-hyperplane and $\Theta_1(\bar{\theta}_2)$ is the section of Θ at $\bar{\theta}_2$.

Assumption A5-I: Given $\{f_1^\alpha(y_1|y_2,z;\alpha_1,\alpha_2);\alpha \in A\}$ from assumption A3-I, α_1 is identified in $f_1^\alpha(\cdot|\cdot,\cdot;\alpha)$ given α_2 for any α_2 in A_2.

Assumption A5-II: Given $\{f_2^\beta(y_2|z;\beta_1,\beta_2);\beta \in B\}$ from assumption A3-II, β_2 is identified in $f_2^\beta(\cdot|\cdot;\beta)$ given β_1 for any β_1 in B_1.

In order to derive the relationship between ℓ_2 and k_2 (or equivalently, ℓ_1 and k_1 from Lemma 3.1), we contrast dim $B_2(\beta_1)$ with dim A_2 by constructing a bijective mapping from $B_2(\beta_1)$ to a subset of A_2 such that the mapping, together with its inverse, is C^0. This mapping is explicitly established in Lemma 3.2.

Lemma 3.2: Pick any $\beta_1^0 \in B_1$. For every $\beta_2^0 \in B_2(\beta_1^0)$, define $\alpha_2^0 = \alpha_2(\beta^{-1}(\beta_1^0,\beta_2^0)) \in \alpha_2[\beta_1^{-1}(\beta_1^0)]$, where $\beta_1^{-1}(\beta_1^0)$ is the pre-image of β_1^0. Then the mapping $d_{\beta_1^0}(\cdot)$: $B_2(\beta_1^0) \to \alpha_2[\beta_1^{-1}(\beta_1^0)]$ such that

$\phi_{\beta_1^0}(\beta_2^0) = \alpha_2^0$ is well-defined. Furthermore, given assumption A5-II, $\phi_{\beta_1^0}$ is a bijective mapping.

The "well-definedness" of $\phi_{\beta_1^0}$ is straightforward by the fact that $\beta(\cdot)$ is bijective. Thus, given β_1^0, any vector (β_1^0,β_2) where $\beta_2 \varepsilon B_2(\beta_1^0)$ will lead to a unique θ, which, in turn, determines a unique α_2^0. It is also easy to establish that $\phi_{\beta_1^0}$ is onto. However, the identification assumption plays an important role in establishing the one-to-one property of $\phi_{\beta_1^0}$. Specifically, given any α_2, a marginal density function f_2^{α} can be derived, which can also be expressed as f_2^{β}. Now, since β_1^0 is given, by the assumption that β_2 is identified given β_1, the equivalence of f_2^{β} (implied by the equivalence of f_2^{α}) will imply the equivalence of β_2. Hence $\phi_{\beta_1^0}$ is one-to-one. Without the identification assumption, the injectiveness of $\phi_{\beta_1^0}$ will not hold.

Since $\phi_{\beta_1^0}$ is a bijective mapping from $B_2(\beta_1^0)$ onto $\alpha_2[\beta_1^{-1}(\beta_1^0)]$ then $\phi_{\beta_1^0}^{-1}$ exists. The next lemma considers some properties of $\phi_{\beta_1^0}$ and $\phi_{B_1^0}^{-1}$.

<u>Lemma 3.3</u>: For any β_1^0 in B_1, the mapping $\phi_{\beta_1^0}(\cdot)$ from $B_2(\beta_1^0)$ onto $\alpha_2[\beta_1^{-1}(\beta_1^0)]$, and its inverse mapping $\phi_{\beta_1^0}^{-1}(\cdot)$ are both continuous.

From Lemma 3.3, it follows that $\dim B_2(\beta_1^0) = \dim \alpha_2(\beta_1^{-1}(\beta_1^0))$. Since $\alpha_2(\beta_1^{-1}(\beta_1^0)) \subset A_2$, then $\dim \alpha_2(\beta_1^{-1}(\beta_1^0)) \leq \dim A_2$ which is equal to k_2 from Lemma 3.1. The next lemma gives the dimension of $B_2(\beta_1^0)$.

<u>Lemma 3.4</u>: If $\dim B = k$, then $\dim B_2(\beta_1^0) = \ell_2$, $\forall \beta_1^0 \varepsilon \overset{\circ}{B}_1$ (i.e., the interior of B_1).

Note that the underlying assumption of Lemma 3.4 was in fact established by Lemma 3.1. Therefore, under appropriate assumptions, letting $B_1^0 \varepsilon \overset{\circ}{B}_1^0$ we have $\ell_2 = \dim B_2(\beta_1^0) = \dim \alpha_2[\beta_1^{-1}(\beta_1^0)] \leq \dim A_2 = k_2$, which is the desired result. Formally, we state the result as follows:

<u>Theorem 1</u>: Given A2, A3, and A4, then $\ell_2 \leq k_2$ (or equivalently $\ell_1 \geq k_1$) if either A5-I or A5-II holds.

A relationship between ℓ_1 and k_1 (or ℓ_2 and k_2) that is more precise than $\ell_1 \geq k_1$ cannot be obtained since one may have $\ell_1 > k_1$ or $\ell_1 = k_1$. As an example for the case $\ell_1 > k_1$, consider the following statistical model (another example with $\ell_1 = k_1$ is given at the end of the next section):

$$Y_1 = Z_{11}\gamma_{11} + Z_{12}\gamma_{12} + U_1 \tag{3.1}$$

$$Y_2 = Z_{12}\gamma_{21} + Z_{22}\gamma_{22} + U_2 \tag{3.2}$$

where Y_1, Y_2, Z_{11}, Z_{12} and Z_{22} are all scalars, and

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right) \tag{3.3}$$

Hence, we can characterize $Y_1|Y_2$ and Y_2 as:

$$Y_1|Y_2 \sim N(Z_{11}\gamma_{11} + Z_{12}\gamma_{12} + \frac{\sigma_{12}}{\sigma_{22}}(Y_2 - Z_{12}\gamma_{21} - Z_{22}\gamma_{22}), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}})$$

$$= N(Z_{11}\gamma_{11} + Z_{12}(\gamma_{12} - \frac{\sigma_{12}}{\sigma_{22}}\gamma_{21}) - \frac{\sigma_{12}}{\sigma_{22}}Z_{22}\gamma_{22} + \frac{\sigma_{12}}{\sigma_{22}}Y_2, \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}), \quad (3.4)$$

and $Y_2 \sim N(Z_{12}\gamma_{21} + Z_{22}\gamma_{22}, \sigma_{22})$ (3.5)

Let $\theta = (\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \sigma_{11}, \sigma_{12}, \sigma_{22})'$, and define $\alpha(\theta) = (\alpha_1'(\theta), \alpha_2'(\theta))'$ with $\alpha_1(\theta) = (\gamma_{11}, \gamma_{12}, \sigma_{11}, \sigma_{12})'$, $\alpha_2(\theta) = (\gamma_{21}, \gamma_{22}, \sigma_{22})'$, then this function will construct a reparameterization of θ for 2SCMLEI with $k_1 = 4$, $k_2 = 3$. Alternatively define $\beta(\theta) = (\beta_1'(\theta), \beta_2'(\theta))'$ with

$$\beta_1(\theta) = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15})'$$

$$= (\gamma_{11}, \gamma_{12} - \frac{\sigma_{12}}{\sigma_{22}}\gamma_{21}, -\frac{\sigma_{12}}{\sigma_{22}}\gamma_{22}, \frac{\sigma_{12}}{\sigma_{22}}, \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}})'$$

$$\beta_2(\theta) = (\beta_{21}, \beta_{22})' = (\gamma_{21} - \frac{\sigma_{12}}{\sigma_{12}}\gamma_{12}, \sigma_{22})'$$

then we may characterize $Y_1|Y_2$ and Y_2 by:

$$Y_1|Y_2 \sim N(Z_{11}\beta_{11} + Z_{12}\beta_{12} + Z_{22}\beta_{13} + Y_2\beta_{14}, \beta_{15}) \quad (3.6)$$

and

$$Y_2 \sim N(Z_{12}\frac{(\beta_{12} + \beta_{13}\beta_{21})(\beta_{15} + \beta_{14}^2\beta_{22}) - \beta_{12}\beta_{15}}{\beta_{14}\beta_{15}} - \frac{Z_{22}\beta_{13}}{\beta_{14}}, \beta_{22}) \quad (3.7)$$

Hence, β is an appropriate reparameterization of θ for 2SCMLEII with

$\ell_1 = 5$, $\ell_2 = 2$. Therefore, for this example we have $\ell_1 > k_1$, $\ell_2 < k_2$ and $\ell_1 + \ell_2 = k_1 + k_2 = k = 7$.

4. COMPARISONS OF 2SCMLEI AND 2SCMLEII

In this section, we shall investigate under what conditions the equivalence relationship between 2SCMLEI and 2SCMLEII can be established, particularly asymptotic equivalence and numerical equivalence. Since asymptotic equivalence requires identical asymptotic variance-covariance matrix, the assumption that $\alpha(\cdot)$ and $\beta(\cdot)$ be C^2 is imposed. Moreover, as shown in Theorem 1, in general we may only have $\ell_1 \geq k_1$ which renders the comparisons of 2SCMLEI and 2SCMLEII much more difficult. To tackle this problem, an assumption which implies $\ell_1 = k_1$ will be made. In fact, condition (iii) of Theorem 2 suffices this purpose.

Theorem 2: Given assumptions A1 - A4, suppose the "conditional" model specifications are correct for α and β with

(i) $\alpha(\cdot) \in C^2, \beta(\cdot) \in C^2$;

(ii) $[\partial\beta(\theta^{-1}(\alpha))/\partial\alpha]$ is non-singular over $\alpha \in A$;

(iii) $[\partial\beta_1/\partial\alpha_1]$ has rank ℓ_1 over $\alpha_1 \in A_1$, $[\partial\alpha_2/\partial\beta_2]$ has rank k_2 over $\beta_2 \in B_2$,

then 2SCMLEI (for θ) is asymptotically equivalent to 2SCMLEII (for θ) if and only if both estimators are asymptotically efficient in the conditional model.

In view of Theorem 1, the most stringent requirement for

Theorem 2 to hold is that ℓ_1 be equal to k_1. If, however $\ell_1 = k_1$, as the examples given below illustrate, then Theorem 2 states that 2SCMLEI and 2SCMLEII are not asymptotically equivalent if and only if one of them is asymptotically inefficient. Since in general two-stage estimators are not asymptotically efficient (see Vuong (1984)), it follows that 2SCMLEI and 2SCMLEII are not in general asymptotically equivalent. As a practical consequence, this implies that if either 2SCMLEI or 2SCMLEII is asymptotically inefficient, one may gain in efficiency by reparameterizing the model so as to apply the other two-stage estimation procedure. A definitive answer on which procedure is preferable must then rest on the direct comparison of the asymptotic covariance matrices given in Equations (2.8c) and (2.8b).

Theorem 3: In addition to the assumptions in Theorem 2, assume further that

(iv) there exists a unique 2SCMLEI (for α) and a unique 2SCMLEII (for β),

then 2SCMLEI (for θ) = 2SCMLEII (for θ) = CMLE (for θ).

Although Theorem 2 indicated that the asymptotic equivalence impinges on asymptotic efficiency, once uniqueness of 2SCMLEI and 2SCMLEII is satisfied, Theorem 3 establishes the numerical equivalence of 2SCMLEI, 2SCMLEII and CMLE. Thus, asymptotic equivalence and asymptotic efficiency as stated in Theorem 2 are both ensured. In general, it appears that uniqueness is not a strong assumption. In particular such an assumption is satisfied when the conditional likelihood functions (2.3.ab)-(2.4.ab) are globally concave in their parameters.

The import of Theorem 3 arises from the numerical equality between the 2SCML estimators and the CML estimator, which is the FIML estimator in conditional models. As a first practical consequence, Theorem 3 provides an easy way to check the asymptotic efficiency of 2SCML estimators. Indeed, suppose that the natural parameterization (i.e., θ) of the model satisfies Assumption A3-I so that the 2SCMLEI estimator can be obtained. Then to determine if this 2CMLEI estimator is asymptotically efficient, it essentially suffices to find another parameterization ($\beta(\theta)$) that satisfies Assumption A3-II and to verify that $\ell_1 = k_1$. If such a parameterization can be found then 2SCMLEI is asymptotically efficient.[4] As a second practical consequence, it follows that, when the assumptions of Theorem 3 hold, one can <u>numerically</u> obtain the FIML estimator by either one of the two 2SCML procedures. This is a definitive advantage when the computation of the FIML requires the maximization of a complicated joint likelihood function, while the computation of the 2SCML estimators uses only standard computer programs for maximization of univariate likelihood functions.

As a simple example for Theorem 3, consider the following statistical model:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right), \tag{4.1}$$

where Y_1 and Y_2 are both scalars.[5] Given (4.1), $Y_1|Y_2$ and Y_2 can be characterized as:

$$Y_1|Y_2 \sim N(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(Y_2 - \mu_2), \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}) \tag{4.2}$$

and

$$Y_2 \sim N(\mu_2, \sigma_{22}) \tag{4.3}$$

Let $\theta = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})'$ and define $\alpha_1(\theta) = (\mu_1, \sigma_{11}, \sigma_{12})'$, $\alpha_2(\theta) = (\mu_2, \sigma_{22})'$, then we have $f^\theta(y_1, y_2; \theta) = f^\alpha(y_1, y_2; \alpha)$, $f_1^\theta(y_1|y_2;\theta) = f_1^\alpha(y_1|y_2;\alpha)$ and $f_2^\theta(y_2;\theta) = f_2^\alpha(y_2;\alpha_2)$, which constructs the framework for 2SCMLEI. Alternatively, define

$$\beta_1'(\theta) = (\beta_{11}, \beta_{12}, \beta_{13}) = (\mu_1 - \frac{\sigma_{12}}{\sigma_{22}}\mu_2, \frac{\sigma_{12}}{\sigma_{22}}, \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}),$$

$$\beta_2'(\theta) = (\beta_{21}, \beta_{22}) = (\mu_2 - \frac{\sigma_{12}}{\sigma_{11}}\mu_1, \sigma_{22}),$$

then

$$Y_1|Y_2 \sim N(\beta_{12}Y_2 + \beta_{11}, \beta_{13}) \tag{4.4}$$

and

$$Y_2 \sim N(\frac{(\beta_{11} + \beta_{12}\beta_{21})(\beta_{13} + \beta_{12}^2\beta_{22}) - \beta_{11}\beta_{13}}{\beta_{12}\beta_{13}}; \beta_{22}). \tag{4.5}$$

Hence, $f^\theta(y_1, y_2;\theta) = f^\beta(y_1, y_2;\beta)$, $f_1^\theta(y_1|y_2;\theta) = f_1^\beta(y_1|y_2;\beta_1)$ and $f_2^\theta(y_2;\theta) = f_2^\beta(y_2;\beta)$ which constructs the framework for 2SCMLEII.

Assume $\theta \in \Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+ \times (\mathbb{R}_+ - \{0\}) \times \mathbb{R}_+$. To derive 2SCMLEI, we maximize L_{2n}^α over $\alpha_2 \in A_2 = \mathbb{R} \times \mathbb{R}_+$ and L_{1n}^α over $\alpha_1 \in A_1(\hat{\alpha}_2) = \mathbb{R} \times (\mathbb{R} - \{0\}) \times \mathbb{R}_+$, respectively, where $\hat{\alpha}_2 = \underset{\alpha_2 \in A_2}{\text{argmax}}\, L_{2n}^\alpha$, and

$$L_{2n}^\alpha = \sum_{t=1}^n \{- 1/2 \log 2\pi - 1/2 \log \sigma_{22} - \frac{(y_{2t} - \mu_2)^2}{2\sigma_{22}}\} \tag{4.6}$$

$$L_{1n}^\alpha = \sum_{t=1}^n \{- 1/2 \log 2\pi - 1/2 \log (\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}})$$

$$- \frac{(y_{1t} - \mu_1 - \frac{\sigma_{12}}{\sigma_{22}}(y_{2t} - \mu_2))^2}{2(\sigma_{11} - \sigma_{12}^2/\sigma_{22})}\}. \tag{4.7}$$

After algebraic manipulations of the first order conditions for (4.6) and (4.7), we have 2SCMLEI (for α) as $(\hat{\alpha}_1, \hat{\alpha}_2) = (\bar{y}_1, \frac{1}{n}\sum_{t=1}^n (y_{1t} - \bar{y}_1)^2, \frac{1}{n}\sum_{t=1}^n (y_{1t} - \bar{y}_1)(y_{2t} - \bar{y}_2), \bar{y}_2, \frac{1}{n}\sum_{t=1}^n (y_{2t} - \bar{y}_2)^2)'$ where $\bar{y}_i = \frac{1}{n}\sum_{t=1}^n y_{it}$, $i = 1,2$.

As for 2SCMLEII, we maximize L_{1n}^β over $\beta_1 \in B_1 = \mathbb{R} \times (\mathbb{R} - \{0\}) \times \mathbb{R}_+$, and L_{2n}^β over $\beta_2 \in B_2(\hat{\beta}_1) = \mathbb{R} \times \mathbb{R}_+$, respectively, where $\hat{\beta}_1 = \underset{\beta_1 \in B_1}{\text{argmax}}\, L_{1n}^\beta$ and

$$L_{1n}^\beta = \sum_{t=1}^n \{- 1/2 \log 2\pi - 1/2 \log \beta_{13} - \frac{(y_{1t} - \beta_{11} - \beta_{12}y_{2t})^2}{2\beta_{13}}\}, \tag{4.8}$$

$$L^{\beta}_{2n} = \sum_{t=1}^{n} \{- 1/2 \log 2\pi - 1/2 \log \beta_{22}$$

$$- \left[y_{2t} - \frac{(\beta_{11} + \beta_{12}\beta_{21})(\beta_{13} + \beta^2_{12}\beta_{22}) - \beta_{11}\beta_{13}}{\beta_{12}\beta_{13}} \right]^2 /2\beta_{22}\}. \quad (4.9)$$

Again, after algebraic manipulations, from the first order conditions for (4.8) and (4.9) we obtain 2SCMLEII (for β), $\hat{\beta}$, as follows:

$$\hat{\beta}_{11} = \bar{y}_1 - \hat{\beta}_{12}\bar{y}_2,$$

$$\hat{\beta}_{12} = \left[\sum_{t=1}^{n} y_{1t}y_{2t} - \frac{1}{n}\sum_{t=1}^{n} y_{1t}\sum_{t=1}^{n} y_{2t}\right] / \left[\sum_{t=1}^{n} y^2_{2t} - \frac{1}{n}(\sum_{t=1}^{n} y_{2t})^2\right]$$

$$\hat{\beta}_{13} = \frac{1}{n}\sum_{t=1}^{n} (y_{1t} - \hat{\beta}_{11} - \hat{\beta}_{12}y_{2t})^2,$$

$$\hat{\beta}_{22} = \frac{1}{n}\sum_{t=1}^{n} (y_{2t} - \bar{y}_2)^2,$$

$$\hat{\beta}_{13}\bar{y}_1 = (\hat{\beta}_{11} + \hat{\beta}_{12}\hat{\beta}_{21})(\hat{\beta}_{13} + \hat{\beta}^2_{12}\hat{\beta}_{22}).$$

Solving the system in terms of θ, one can check that 2SCMLEI (for θ) = $\alpha^{-1}(\hat{\alpha}) = \beta^{-1}(\hat{\beta})$ = 2SCMLEII (for θ). In additions from the formula for the 2SCMLEI estimates of α, which is nothing else than θ, it can readily be seen that these estimates are also the CML/FIML estimates for θ. These results are expected since it can readily be checked that the assumption of Theorem 3 hold for this simple example.

5. EXAMPLES

As a more interesting application of Theorem 3, consider the following seemingly unrelated regression model:

$$Y_1 = Z_{11}\gamma_{11} + Z_{12}\gamma_{12} + U_1 \quad (5.1)$$

$$Y_2 = Z_{12}\gamma_{21} + U_2 \quad (5.2)$$

where Y_1 and Y_2 are both scalars, Z_{11} and Z_{12} are m-dimensional and n-dimensional vectors, respectively. Let us note that all the explanatory variables appearing in the second equation also appear in the first equation. In other words, there are no explanatory variables specific to the second equation. Assume that

$$\begin{bmatrix} U_1 \\ U_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right).^6 \quad (5.3)$$

Thus, we can characterize Y_2 and $Y_1|Y_2$ as:

$$Y_2 \sim N(Z_{12}\gamma_{21} + \mu_2, \sigma_{22}) \quad (5.4)$$

$$Y_1|Y_2 \sim N(Z_{11}\gamma_{11} + Z_{12}\gamma_{12} + \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(Y_2 - Z_{12}\gamma_{21} - \mu_2), \sigma_{11} - \frac{\sigma^2_{12}}{\sigma_{22}}). \quad (5.5)$$

Hence, let $\theta = (\gamma_{11}, \gamma_{12}, \gamma_{21}, \mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})'$ and define $\alpha(\theta) = (\alpha_1'(\theta), \alpha_2'(\theta))'$ such that $\alpha_1(\theta) = (\gamma_{11}, \gamma_{12}, \mu_1, \sigma_{12}, \sigma_{22})'$, $\alpha_2(\theta) = (\gamma_{21}, \mu_2, \sigma_{22})'$, then we have the appropriate framework for 2SCMLEI with $k_1 = m + n + 3$, $k_2 = n + 2$.

Alternatively, define $\beta(\theta) = (\beta_1'(\theta), \beta_2'(\theta))'$ such that

$$\beta_1(\theta) = (\gamma_{11}, \gamma_{12} - \frac{\sigma_{12}}{\sigma_{22}}\gamma_{21}, \mu_1 - \frac{\sigma_{12}}{\sigma_{22}}\mu_2, \frac{\sigma_{12}}{\sigma_{22}}, \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}})',$$

$$\beta_2(\theta) = (\gamma_{21} - \frac{\sigma_{12}}{\sigma_{11}}\gamma_{12}, \mu_2 - \frac{\sigma_{12}}{\sigma_{11}}\mu_1, \sigma_{22})'$$

then $\beta(\cdot)$ constructs the framework for 2SCMLEII with $\ell_1 = m + n + 3$, $\ell_2 = n + 2$.

To see that $\beta(\cdot)$ is a proper parameterization (see Definition 2.1) note that given $\beta(\theta) = \bar{\beta}$, σ_{22} is uniquely determined from $\bar{\beta}_{23}$. Thus, from $\bar{\beta}_{14}$, σ_{12} is uniquely determined which implies σ_{11} is also uniquely determined by $\bar{\beta}_{15}$. Now, $\bar{\beta}_{13}$ and $\bar{\beta}_{22}$ form a two-equation system for μ_1 and μ_2, as long as the determinant $(1 - \frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}})$ does not vanish, then μ_1 and μ_2 are both uniquely determined. Similarly, $\bar{\beta}_{12}$ and $\bar{\beta}_{21}$ construct a simultaneous equation system to solve for γ_{21} and γ_{21} which are also uniquely determined as long as $\sigma_{11}\sigma_{22} \neq \sigma_{12}^2$. Finally, σ_{11} is uniquely determined by $\bar{\beta}_{11}$. Therefore, $\beta(\cdot)$ is injective. Other properties, such as surjectiveness and continuous differentiability can also be checked.

In summary, for the model (5.1)-(5.3), $k_1 = m + n + 3 = \ell_1$, $k_2 = n + 2 = \ell_2$. From Theorem 3, we have therefore the following interesting property.

<u>Corollary 1</u>: For the seemingly unrelated regression model defined by Equations (5.1)-(5.3), the following two-stage procedure produces estimates that are numerically equal to the FIML estimates:

(i) Apply OLS to Equation (5.2) to derive $\hat{\gamma}_{21}, \hat{\mu}_2$, and $\hat{\sigma}_{22}$,

(ii) Apply OLS to Equation (5.1) <u>expanded</u> by the estimated residuals $Y_2 - Z_{12}\hat{\gamma}_{21} - \hat{\mu}_2$ to obtain $\hat{\gamma}_{11}, \hat{\gamma}_{12}, \hat{\mu}_1, \hat{\sigma}_{12}$, and $\hat{\sigma}_{11}$.

Since, $\ell_1 = k_1$, as mentioned above, the result follows from Theorem 3 by noticing that the simple procedure described in Corollary 1 actually generates the 2SCMLEI estimates of the initial parameters. By contrast the well-known GLS procedure proposed by Zellner (1962) on the "stacked" regression model, though asymptotically efficient, requires two OLS estimation (to estimate the covariance matrix), and one more burdensome GLS estimation on the stacked observations. In addition , for the model (5.1)-(5.3), the 2SCMLEI procedure gives exactly the FIML estimates of all the parameters including the variances and the covariance which are hence efficiently estimated. Finally, a Wald-type stastic can be readily constructed to test the hypothesis $\sigma_{12} = 0$ using the asymptotic covariance matrix given by Equation (2.8a).[7]

All the example considered up to now were linear models for which FIML estimates or asymptotically efficient estimates are not really difficult to obtain. Our results also apply to non-linear models for which FIML estimates are in general much more difficult to compute. As an illustration of the possible simplification, consider again the model (5.1)-(5.3), but suppose that Y_1 is observed only discretely. Then we have:

$$Y_1^* = Z_{11}\gamma_{11} + Z_{12}\gamma_{12} + U_1 \qquad (5.6)$$

$$Y_2 = Z_{12}\gamma_{21} + U_2$$
$$Y_1 = 1 \text{ if } Y_1^* > 0;\ 0, \text{ otherwise}, \tag{5.7}$$

where Y_1^* is an unobservable scalar and U_1, U_2, Z_{11}, Z_{12} all have the same structure as in the previous example except that we normalize σ_{11} to be equal to 1 for identification purpose. Now, we can characterize Y_2 and $Y_1|Y_2$ as:

$$Y_2 \sim N(Z_{12}\gamma_{21} + \mu_2, \sigma_{22})$$
$$Y_1|Y_2 \sim [\Phi(x)]^{1-Y_1}[1 - \Phi(x)]^{Y_1} \tag{5.8}$$

$$\text{where } x = \{-Z_{11}\gamma_{11} - Z_{12}\gamma_{12} - (\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(Y_2 - Z_{12}\gamma_{21} - \mu_2))\}/(1 - \frac{\sigma_{12}^2}{\sigma_{22}}), \tag{5.9}$$

and $\Phi(\cdot)$ is the cumulative density function for the standard normal distribution. Applying the reparameterizations $\alpha(\cdot)$ and $\beta(\cdot)$ defined in the previous example, then we obtain the two frameworks necessary for 2SCMLEI and 2SCMLEII. In addition, $k_1 = \ell_1$, $k_2 = \ell_2$. Therefore, 2SCMLEI is actually numerical equal to FIML. Yet, in this case, 2SCMLEI is much easier to compute. To be specific, the 2SCMLEI estimates are obtained by first applying OLS to Equation (5.2) to derive $\hat{\mu}_2$, $\hat{\gamma}_{21}$, and $\hat{\sigma}_{22}$. Then, one estimates Equation (5.9) by Probit analysis with μ_2, γ_{21}, σ_{22} replaced by $\hat{\mu}_2$, $\hat{\gamma}_{21}$, $\hat{\sigma}_{22}$. Or equivalently, the second step consists in doing a simple Probit analysis on the first equation with the estimated residuals $Y_2 - Z_{12}\hat{\gamma}_2 - \hat{\mu}_2$ as an additional regressor. In this case, it can readily be seen that this two-stage procedure, which generates the FIML estimates, is

computationally much easier than the direct maximization of the joint likelihood function for the discrete/continuous model (5.6)-(5.7).

6. CONCLUSION

In this paper, we considered the case when, after appropriate reparameterizations, both two-stage estimation procedures can be applied. In particular, the relationship between the number of parameters in the marginal [or conditional] density functions under two different kinds of parameterizations were characterized under some identification assumptions. Moreover, conditions for asymptotic equivalence and numerical equivalence between the two two-stage estimators were obtained. Finally, examples were provided to illustrate our results.

APPENDIX

<u>Proof of Lemma 3.1</u>: Let α be any point in A. Let $\theta = \alpha^{-1}(\alpha)$. From Assumption A4, there exists a (relative) neighborhood N_θ which is homeomorphic to an open set U of $\mathbb{R}^k$. Since $\alpha(\cdot)$ is continuous with respect to the relative topologies, then $\alpha(N_\theta)$ is a neighborhood of α since $\alpha(\cdot)$ are C^0, if follows that $\alpha(N_\theta)$ is homeomorphic to U, which establishes that dim A = k. Similarly dim B = k.

To prove the second part of the lemma, note that dim $A_1 \leq k_1$ since $A_1 \subset \mathbb{R}^{k_1}$. In addition, note that dim A $\leq$ dim A_1 + dim $A_2 \leq k_1 + k_2 = k$. From above dim A = k. Then dim $A_1 = k_1$. Similarly dim $B_1 = \ell_1$.

<u>Proof of Lemma 3.2</u>: The first part follows directly from the fact that $\beta(\cdot)$ is bijective. To show that $\phi_{\beta_1^0}$ is injective under assumption A5-II, assume $\alpha_2 = \phi_{\beta_1^0}(\beta_2) = \alpha_2' = \phi_{\beta_1^0}(\beta_2')$, then by definition of $\phi_{\beta_1^0}(\cdot)$:

$$\alpha_2(\beta^{-1}(\beta_1^0,\beta_2)) = \alpha_2(\beta^{-1}(\beta_1^0,\beta_2'))$$

which implies:

$$f_2^\alpha(y_2|z;\alpha_2(\beta^{-1}(\beta_1^0,\beta_2))) = f_2^\alpha(y_2|z;\alpha_2(\beta^{-1}(\beta_1^0,\beta_2')))$$

which is equivalent to:

$$f_2^\theta(y_2|z;\beta^{-1}(\beta_1^0,\beta_2)) = f_2^\theta(y_2|z;\beta^{-1}(\beta_1^0,\beta_2'))$$

which is equivalent to:

$$f_2^\beta(y_2|z;\beta_1^0,\beta_2) = f_2^\beta(y_2|z;\beta_1^0,\beta_2')$$

which implies $\beta_2 = \beta_2'$ under assumption A5- II.

Moreover, for every $\bar{\alpha}_2 \in \alpha_2[\beta_1^{-1}(\beta_1^0)]$, there exists a $\theta \in \Theta$ such that $\alpha_2(\theta) = \bar{\alpha}_2$ and $\beta_1(\theta) = \beta_1^0$. Let $\bar{\beta}_2 = \beta_2(\theta)$, then $\bar{\beta}_2 \in B_2(\beta_1^0)$ and $\beta^{-1}[\beta_1^0,\bar{\beta}_2] = \theta$ which implies $\alpha_2[\beta^{-1}(\beta_1^0,\bar{\beta}_2)] = \bar{\alpha}_2$. Therefore, $\phi_{\beta_1^0}$ is a mapping from $B_2(\beta_1^0)$ onto $\alpha_2[\beta_1^{-1}(\beta_1^0)]$.

Combining the above results, we have shown that $\phi_{\beta_1^0}$ is a well-defined bijective function and the proof is completed.

<u>Proof of Lemma 3.3</u>: The mapping $\phi_{\beta_1^0}(\cdot)$ is clearly continuous since $\phi_{\beta_1^0}(\cdot) = \alpha_2(\beta^{-1}(\beta_1^0,\cdot))$.

To prove that $\phi^{-1}_{\beta_1^0}(\cdot)$ is continuous, first note that $B_2(\beta_1^0)$ is compact since $B = \beta(\Theta)$ is compact. Thus $B_2(\beta_1^0)$ is a compact metric space. Moreover $\alpha_2[\beta_1^{-1}(\beta_1^0)]$ is a metric space. Since $\phi_{\beta_1^0}(\cdot)$ is bijective and continuous, the desired result follows from Dieudonné (1969, p. 64).

<u>Proof of Lemma 3.4</u>: For any $\beta = (\beta_1,\beta_2) \in B$ such that $\beta_1 \in B_1^0$, since dim B = k, there exists a relative neighborhood of β such that dim $N_\beta = k$. Yet, k = dim $N_\beta \leq$ dim $(N_{\beta_1} \times N_{\beta_2})$ = dim N_{β_1} + dim $N_{\beta_2} \leq \ell_1$ + dim N_{β_2}, where N_{β_1} is the projection of N_β on the β_1-plane, N_{β_2}

is the projection of N_β on the β_2-plane. Now, if dim $N_{\beta_2} < \ell_2$, then $k < \ell_1 + \ell_2$, which leads to a contradiction. Therefore, we must have dim $N_{\beta_2} \geq \ell_2$. But since $\beta_2 \in \mathbb{R}^{\ell_2}$, hence dim $N_{\beta_2} = \ell_2$. Since $N_{\beta_2} = N_\beta \cap B_2(\beta_1)$, thus dim $(N_\beta \cap B_2(\beta_1)) = \ell_2$ which implies dim $B_2(\beta_1) = \ell_2$, $\forall \beta_1 \in B_1^0$.

<u>Proof of Theorem 1</u>: If A5-II holds, then the result follows directly from Lemma 3.1 - 3.4.

Now suppose that A5-I holds. Pick any $\alpha_2^0 \in A_2$, and for every $\alpha_1^0 \in A_1(\alpha_2^0)$, define $\beta_1^0 = \beta_1[\alpha^{-1}(\alpha_1^0,\alpha_2^0)] \in \beta_1[\alpha_2^{-1}(\alpha_2^0)]$, where $\alpha_2^{-1}(\alpha_2^0)$ is the pre-image of α_2^0. Therefore, we may establish a function $\phi_{\alpha_2^0}: A_1(\alpha_2^0) \to \beta_1[\alpha_2^{-1}(\alpha_2^0)]$ such that $\phi_{\alpha_2^0}(\alpha_1^0) = \beta_1^0$. Following the similar arguments, we have $\ell_1 \geq k_1 \Rightarrow \ell_2 \leq k_2$.

<u>Proof of Theorem 2</u>: Since $f_2^\beta(y_2|z;\beta) = f_2^\alpha(y_2|z;\alpha_2)$, therefore

$$\frac{\partial\beta'}{\partial\alpha} \cdot \frac{\partial \log f_2^\beta(y_2|z;\beta)}{\partial\beta} = \frac{\partial \log f_2^\alpha(y_2|z;\alpha_2)}{\partial\alpha} .$$

Thus,

$$\frac{\partial\beta'}{\partial\alpha} \cdot E^0\left[\frac{\partial \log f_2^\beta(y_2|z;\beta)}{\partial\beta} \cdot \frac{\partial \log f_2^\beta(y_2|z;\beta)}{\partial\beta'}\right] \cdot \frac{\partial\beta}{\partial\alpha'}$$

$$= E^0\left[\frac{\partial \log f_2^\alpha(y_2|z;\alpha_2)}{\partial\alpha} \cdot \frac{\partial \log f_2^\alpha(y_2|z;\alpha_2)}{\partial\alpha'}\right] ,$$

which implies:

$$\frac{\partial\beta'}{\partial\alpha} \cdot D^2_{\beta\beta} \cdot \frac{\partial\beta}{\partial\alpha'} = \begin{bmatrix} 0 & 0 \\ 0 & B^2_{\alpha_2\alpha_2} \end{bmatrix} \tag{A1}$$

where $D^2_{\beta\beta}$ is the extended matrix of $D^2_{\beta_2\beta_2}$ as defined in (2.11b), and $B^2_{\alpha_2\alpha_2}$ is defined in (2.11a).

Similarly, $f_1^\alpha(y_1|y_2,z;\alpha) = f_1^\beta(y_1|y_2,z;\beta_1)$ implies

$$\frac{\partial\alpha'}{\partial\beta} \cdot \frac{\partial \log f_1^\alpha(y_1|y_2,z;\alpha)}{\partial\alpha} = \frac{\partial \log f_1^\beta(y_1|y_2,z;\beta_1)}{\partial\beta} ,$$

and

$$\frac{\partial\alpha'}{\partial\beta} \cdot E^0\left[\frac{\partial \log f_1^\alpha(y_1|y_2,z;\alpha)}{\partial\alpha} \cdot \frac{\partial \log f_1^\alpha(y_1|y_2,z;\alpha)}{\partial\alpha'}\right] \cdot \frac{\partial\alpha}{\partial\beta'}$$

$$= E^0\left[\frac{\partial \log f_1^\beta(y_1|y_2,z;\beta_1)}{\partial\beta} \cdot \frac{\partial \log f_1^\beta(y_1|y_2,z;\beta_1)}{\partial\beta'}\right] .$$

Therefore,
$$\frac{\partial\alpha'}{\partial\beta} \cdot B^1_{\alpha\alpha} \cdot \frac{\partial\alpha}{\partial\beta'} = \begin{bmatrix} D^1_{\beta_1\beta_1} & 0 \\ 0 & 0 \end{bmatrix} , \tag{A2}$$

where $B^1_{\alpha\alpha}$ is the extended matrix of $B^1_{\alpha_1\alpha_1}$ as defined in (2.10a) and $D^1_{\beta_1\beta_1}$ is defined in (2.9b).

Now, 2SCMLEI and 2SCMLEII will be asymptotically equivalent if and only if $[\sum^{(\beta)}(\beta^0)]^{-1} = \frac{\partial\alpha'}{\partial\beta}[\sum^{(\alpha)}(\alpha^0)]^{-1}\frac{\partial\alpha}{\partial\beta'}$. Also, from Equations (2.8a) and (2.8b), we have:

$$[\sum^{\alpha}(\alpha^0)]^{-1} = B^1_{\alpha\alpha} + \begin{bmatrix} 0 & 0 \\ 0 & B^2_{\alpha_2\alpha_2} - B^1_{\alpha_2\alpha_2} + B^1_{\alpha_2\alpha_1}[B^1_{\alpha_1\alpha_1}]^{-1}B^1_{\alpha_1\alpha_2} \end{bmatrix} \qquad \text{(A3)}$$

$$[\sum^{\beta}(\beta^0)]^{-1} = D^2_{\beta\beta} + \begin{bmatrix} D^1_{\beta_1\beta_1} - D^2_{\beta_1\beta_1} + D^2_{\beta_1\beta_2}[D^2_{\beta_2\beta_2}]^{-1}D^2_{\beta_2\beta_1} & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{(A4)}$$

where we utilized the facts the $B^1_{\alpha\alpha} = -A^1_{\alpha\alpha}$, $D^2_{\beta\beta} = -C^2_{\beta\beta}$ under correct "conditional" model specification. (see Vuong (1984)). Therefore,

$$\frac{\partial\alpha'}{\partial\beta}[\sum^{\alpha}(\alpha^0)]^{-1}\frac{\partial\alpha}{\partial\beta'} = \begin{bmatrix} D^1_{\beta_1\beta_1} & 0 \\ 0 & 0 \end{bmatrix} + \frac{\partial\alpha'}{\partial\beta}\begin{bmatrix} 0 & 0 \\ 0 & B_{\alpha_2\alpha_2} \end{bmatrix}\frac{\partial\alpha}{\partial\beta'}$$

$$+ \begin{bmatrix} \frac{\partial\alpha_1'}{\partial\beta}, \frac{\partial\alpha_2'}{\partial\beta} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & -B^1_{\alpha_2\alpha_2} + B^1_{\alpha_2\alpha_1}[B^1_{\alpha_1\alpha_1}]^{-1}B^1_{\alpha_1\alpha_2} \end{bmatrix} \begin{bmatrix} \frac{\partial\alpha_1}{\partial\beta'}, \frac{\partial\alpha_2}{\partial\beta'} \end{bmatrix}$$

$$= \begin{bmatrix} D^1_{\beta_1\beta_1} & 0 \\ 0 & 0 \end{bmatrix} + D^2_{\beta\beta} + \frac{\partial\alpha_2'}{\partial\beta}[-B^1_{\alpha_2\alpha_2} + B^1_{\alpha_2\alpha_1}[B^1_{\alpha_1\alpha_1}]^{-1}B^1_{\alpha_1\alpha_2}]\frac{\partial\alpha_2}{\partial\beta'}, \qquad \text{(A5)}$$

where Equations (A1) and (A2) were applied to derive these equalities.

Comparing (A4) and (A5), 2SCMLEI and 2SCMLEII will be asymptotically equivalent if and only if the following holds:

$$\begin{bmatrix} -D^2_{\beta_1\beta_1} + D^2_{\beta_1\beta_2}[D^2_{\beta_2\beta_2}]^{-1}D^2_{\beta_2\beta_1} & 0 \\ 0 & 0 \end{bmatrix}$$

$$= \frac{\partial\alpha'}{\partial\beta}\begin{bmatrix} 0 & 0 \\ 0 & -B^1_{\alpha_2\alpha_2} + B^1_{\alpha_2\alpha_1}[B^1_{\alpha_1\alpha_1}]^{-1}B^1_{\alpha_1\alpha_2} \end{bmatrix}\frac{\partial\alpha}{\partial\beta'}$$

which implies

$$\frac{\partial\alpha_2'}{\partial\beta_2}[B^1_{\alpha_2\alpha_2} - B^1_{\alpha_2\alpha_1}[B^1_{\alpha_1\alpha_1}]^{-1}B^1_{\alpha_1\alpha_2}]\frac{\partial\alpha_2}{\partial\beta_2'} = 0$$

which implies $B^1_{\alpha_2\alpha_2} - B^1_{\alpha_2\alpha_1}[B^1_{\alpha_1\alpha_1}]^{-1}B^1_{\alpha_1\alpha_2} = 0$, since $\frac{\partial\alpha_2}{\partial\beta_2'}$ has full rank. Furthermore, we also have $D^2_{\beta_1\beta_1} - D^2_{\beta_1\beta_2}[D^2_{\beta_2\beta_2}]^{-1}D^2_{\beta_2\beta_1} = 0$. Now, from Vuong (1984) and similar arguments, these two conditions actually guarantee that the two-stage estimators are asymptotically efficient. Thus, we have shown that 2SCMLEI is asymptotically equivalent to 2SCMLEII only if both are asymptotically efficient. The converse is obviously true.

<u>Proof of Theorem 3</u>: Since $\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_1,\alpha_2))}{\partial\alpha_1}\right|_{(\hat{\alpha}_1,\hat{\alpha}_2)} = 0,$

then applying the chain-rule, we have

$$\left.\frac{\partial\beta_1'}{\partial\alpha_1}\right|_{(\hat{\alpha}_1,\hat{\alpha}_2)} \left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\beta}(y_{1t}|y_{2t},z_t;\beta_1))}{\partial\beta_1}\right|_{(\beta_1(\hat{\alpha}_1,\hat{\alpha}_2),\beta_2(\hat{\alpha}_1,\hat{\alpha}_2))} = 0$$

since $[\partial\beta_1'/\partial\alpha_1]$ is a $(k_1 \times \ell_1)$ matrix with rank ℓ_1, therefore

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\beta}(y_{1t}|y_{2t},z_t;\beta_1))}{\partial\beta_1}\right|_{(\beta_1(\hat{\alpha}_1,\hat{\alpha}_2),\beta_2(\hat{\alpha}_1,\hat{\alpha}_2))} = 0$$

Now since $\beta_1(\hat{\alpha}_1,\hat{\alpha}_2) \varepsilon B_1$ and $(\hat{\beta}_1,\hat{\beta}_2)$ is the unique 2SCMLEII, hence

$$\hat{\beta}_1 = \beta_1(\hat{\alpha}_1,\hat{\alpha}_2) .$$

Similarly,

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_2^{\beta}(y_{2t}|z_t;\beta_1,\beta_2))}{\partial\beta_2}\right|_{(\hat{\beta}_1,\hat{\beta}_2)} = 0$$

which implies:

$$\left.\frac{\partial\alpha_2'}{\partial\beta_2}\right|_{(\hat{\beta}_1,\hat{\beta}_2)} \left.\frac{\partial(\sum_{t=1}^{n}\log f_2^{\beta}(y_{2t}|z_t;\alpha_2))}{\partial\alpha_2}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\alpha_2(\hat{\beta}_1,\hat{\beta}_2))} = 0,$$

where $\frac{\partial\alpha_2'}{\partial\beta_2}$ is an $(\ell_2 \times k_2)$ matrix with rank k_2, hence

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_2^{\alpha}(y_{2t}|z_t;\alpha_2))}{\partial\alpha_2}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\alpha_2(\hat{\beta}_1,\hat{\beta}_2))} = 0.$$

Now since $\alpha_2(\hat{\beta}_1,\hat{\beta}_2) \varepsilon A_2$ and $(\hat{\alpha}_1,\hat{\alpha}_2)$ is the unique 2SCMLEI, hence

$$\alpha_2(\hat{\beta}_1,\hat{\beta}_2) = \hat{\alpha}_2 .$$

Furthermore, since

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\beta}(y_{1t}|y_{2t},z_t;\beta_1))}{\partial\beta_1}\right|_{(\hat{\beta}_1,\hat{\beta}_2)} = 0,$$

hence

$$\left.\frac{\partial\alpha'}{\partial\beta_1}\right|_{(\hat{\beta}_1,\hat{\beta}_2)} \left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha))}{\partial\alpha}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\alpha_2(\hat{\beta}_1,\hat{\beta}_2))} = 0 \quad \text{(A6)}$$

Also, since

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\beta}(y_{1t}|y_{2t},z_t;\beta_1))}{\partial\beta_2}\right|_{(\hat{\beta}_1,\hat{\beta}_2)} = 0$$

hence

$$\left.\frac{\partial\alpha'}{\partial\beta_2}\right|_{(\hat{\beta}_1,\hat{\beta}_2)} \cdot \left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha))}{\partial\alpha}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\alpha_2(\hat{\beta}_1,\hat{\beta}_2))} = 0 \quad \text{(A7)}$$

Equations (A.6)-(A7) imply

$$\left.\frac{\partial\alpha'}{\alpha\beta}\right|_{(\hat{\beta}_1,\hat{\beta}_2)} \cdot \left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha))}{\partial\alpha}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\alpha_2(\hat{\beta}_1,\hat{\beta}_2))} = 0$$

where $\frac{\partial\alpha'}{\partial\beta}$ is a (k X k) matrix with rank k, therefore

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha))}{\partial\alpha}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\alpha_2(\hat{\beta}_1,\hat{\beta}_2))} = 0 . \qquad (A8)$$

Particularly,

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha))}{\partial\alpha_1}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\alpha_2(\hat{\beta}_1,\hat{\beta}_2))} = 0 . \qquad (A9)$$

Note that $\alpha_2(\hat{\beta}_1,\hat{\beta}_2) = \hat{\alpha}_2$, Equation (A9) can be rewritten as

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_1,\hat{\alpha}_2))}{\partial\alpha_1}\right|_{(\alpha_1(\hat{\beta}_1,\hat{\beta}_2),\hat{\alpha}_2)} = 0$$

Again, since $\hat{\alpha}$ is the unique 2SCMLEI, thus

$$\alpha_1(\hat{\beta}_1,\hat{\beta}_2) = \hat{\alpha}_1 .$$

Similarly, we can show $\hat{\beta}_2 = \beta_2(\hat{\alpha}_1,\hat{\alpha}_2)$. Now, note that the CMLE satisfies:

$$\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_1,\alpha_2) + \sum_{t=1}^{n}\log f_2^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_2))}{\partial\alpha_2} = 0$$

and

$$\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_1,\alpha_2))}{\partial\alpha_1} = 0 .$$

Yet, from Equation (A.8) we have:

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_1,\alpha_2))}{\partial\alpha_2}\right|_{(\hat{\alpha}_1,\hat{\alpha}_2)} = 0 .$$

Also by definition:

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_2^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_2))}{\partial\alpha_2}\right|_{(\hat{\alpha}_1,\hat{\alpha}_2)} = 0$$

and

$$\left.\frac{\partial(\sum_{t=1}^{n}\log f_1^{\alpha}(y_{1t}|y_{2t},z_t;\alpha_1,\alpha_2))}{\partial\alpha_1}\right|_{(\hat{\alpha}_1,\hat{\alpha}_2)} = 0 .$$

Hence, $(\hat{\alpha}_1,\hat{\alpha}_2)$ is also the CMLE. This completes the proof.

FOOTNOTES.

* This research was partially supported by National Science Foundation Grant SES-8410593. We are greatly indebted to D. Rivers and J. Sobel for helpful discussion, and to Shoo Bee Doo for moral support. Remaining errors are ours.

1. That is, $\alpha(\cdot)$ is an homeomorphism (see, e.g. Dieudonné (1969)).

2. Amemiya (1978) considers a two-step estimation of the bivariate logit model. The natural parameterization of this model satisfies Assumption A3-II. This example is therefore an illustration of 2SCMLEII.

3. Assumption A4 is satisfied under the regularity conditions of MLE, since the true parameter θ^0, whatever it may be, is assumed to belong to the interior of Θ which is therefore not empty (see e.g. White (1982), Vuong (1983)).

4. In other words, checking efficiency amounts to counting the number of parameters, while a direct approach would be to check the necessary and sufficient conditions given in Vuong (1984, Theorem 3).

5. This example can be considered as the simplest seemingly related regression model with the constant term as the only regressor. It is also a special case of the more general model considered below

(Equations (5.1)-(5.2)).

6. The reason for having non-zero means is to allow for the presence of constant terms in Equations (5.1) and (5.2) and to still satisfy Assumption A1 on random sampling.

7. A similar idea was exploited in Holly and Sargan (1982), Holly (1983), and Rivers and Vuong (1984) for testing exogeneity in simultaneous models.

REFERENCES

Amemiya, T. "On a Two-Step Estimation of a Multivariate Logit Model." Journal of Econometrics 8 (1978):13-22.

Barankin, E. W. "Sufficient Parameters: Solution of the Minimal Dimensional Problem." Annals of Institute of Statistical Mathematics 12 (1961):91-118.

Dieudonné, J. Foundations of Modern Analysis. New York: Academic Press, 1969.

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

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

Holly, A. and Sargan, J. D. "Testing for Exogeneity Within a Limited Information Framework." Discussion Paper no. 8204. Universite de Lausanne, 1982.

Le Cam, L. "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates." University of California Publication in Statistics 1 (1953):277-330.

Lehman, E. Testing Statistical Hypothesis. New York: John Wiley

and Sons, 1959.

Matsushima, Y. Differentiable Manifolds. New York: Marcel Dekker Inc., 1972.

Monfort, A. Cours de Statistique Mathematique. Paris: Economica, 1982.

Rivers, D. and Vuong, Q. H. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." Socisl Science Working Paper no. 539. Pasadena: California Institute of Technology, 1984.

Vuong, Q. H. "Misspecification and Conditional Maximum Likelihood Estimation." Social Science Working Paper no. 503. Pasadena: California Institute of Technology, 1983.

__________. "Two-Stage Conditional Maximum Likelihood Estimation of Econometric Models." Social Science Working Paper no. 538. Pasadena: California Institute of Technology, 1984.

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

White, H. "Maximum Likelihood Estimation of Misspecified Models." Econometrica 50 (1982):1-25.

Zellner, A. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Test for Aggregation Bias." Journal of the American Statistical Association 57 (1962):348-358.