DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

MISSPECIFICATION AND CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION

Quang H. Vuong

SOCIAL SCIENCE WORKING PAPER 503

April 1983
Revised November 1983

ABSTRACT

Recently White (1982) studied the properties of Maximum Likelihood estimation of possibly misspecifed models. The present paper extends Andersen (1970) results on Conditional Maximum Likelihood estimators (CMLE) to such a situation. In particular, the asymptotic properties of CMLE's are derived under correct and incorrect specification of the conditional model. Robustness of conditional inferences and estimation with respect to misspecification of the model for the conditioning variables is emphasized. Conditions for asymptotic efficiency of CMLE's are obtained, and specification tests _a la_ Hausman (1978) and White (1982) are derived. Examples are also given to illustrate the use of CMLE's properties. These examples include the simple linear model, the multinomial logit model, the simple Tobit model, and the multivariate logit model.

MISSPECIFICATION AND CONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION*

Quang H. Vuong

1. Introduction

In most applied work, estimation and inference are conducted conditional upon the observed values of some explanatory variables even though most data in the social sciences are not outcomes of well-defined experiments. This results in part because social scientists are often interested in estimating so-called structural or behavioral relationships between endogenous variables on the one hand and exogenous variables on the other hand.

The present paper first provides a justification for conditional estimation and inference by studying the properties of Conditional Maximum-Likelihood Estimators (CMLE). Specifically, we generalize Andersen's (1970) results by deriving the asymptotic properties of CMLE's under correct and incorrect specification of the conditional model. This is done by following the lines of White's (1982) important paper. It is then observed that the properties of CMLE's and the inferences based on CMLE's are robust with respect to misspecification of the model for the conditioning variables.

Conditional maximum-likelihood estimation may, however, entail a loss of efficiency especially when the model for the conditioning variables contains information on the estimated parameters. We then characterize conditions under which CML estimators are asymptotically as efficient as FIML estimators. These conditions are actually weaker than the condition that the conditioning variables be weakly exogenous in the sense of Engle, Hendry, and Richard (1983).

Finally, since it is often of great interest to know whether the conditional model is correctly specified, we also discuss specification tests within the present framework. It is argued that CMLE's can readily be used to construct specification tests _a la_ Hausman (1978) and White (1982). The essential reason comes from the fact that one can often choose or construct variables to conditon upon so that the resulting conditional likelihood contains the parameters of interest.

The paper is organized as follows. Section 2 presents our assumptions on the structure generating the data and on the chosen conditional model. Section 3 studies the asymptotic properties of CMLE's under correct or incorrect specification of the conditional model. Section 4 derives necessary and sufficient conditions for asymptotic efficiency of CMLE's. Section 5 uses CMLE's to construct Hausman-White type tests for misspecification. Particular care is given to the formulation of the null and alternative hypotheses for each test. Section 6 presents some applications of the properties of CMLE's. The first three examples are the simple linear model, the multinomial logit model, and the simple Tobit model. The fourth example considers the case in which there exists a sufficient statistic for some parameters. This is then illustrated by the multivariate logit model. Section 7 summarizes our results, and an appendix collects the proofs.

2. Notations and Assumptions[1]

Let X_t be an $m \times 1$ observed real random vector defined on a Euclidean measurable space (X, σ_x, ν_x). For instance X, σ_x, and ν_x can respectively be $\mathbb{R}^m$, the Borel σ-algebra, and the usual Lebesgue measure. The process generating the observations X_t, t=1,2,... satisfies the following assumption.

ASSUMPTION A1: The random vectors X_t, t=1,2,... are independent and identically distributed with common (true) cumulative distribution function H^o on (X, σ_x, ν_x).

The vector X_t is partitioned into $X_t = (Y_t', Z_t')'$ where Y_t and Z_t are respectively p and q dimensional vectors with $m = p + q$. Let (Y, σ_y, ν_y) and (Z, σ_z, ν_z) be the Euclidean measurable spaces associated with Y_t and Z_t.

For any t, let $F^o_{Y|Z}(.|.)$ be the true but unknown conditional distribution of Y_t given Z_t.[2] We are interested in estimating $F^o_{Y|Z}(.|.)$. To do so, we choose a (parametric) family of conditional distribution functions $F_{Y|Z}(.|.;\alpha)$, where α belongs to a subset A of $\mathbb{R}^k$. Such a family may or may not contain the true conditional distribution $F^o_{Y|Z}(.|.)$. It is, however, chosen so as to satisfy the following regularity conditions.

ASSUMPTION A2: (a) For every α in A, a compact subset of $\mathbb{R}^k$, and for

(H^o-almost) all z, the conditional distribution $F(.|\ z;\alpha)$ has a σ_y-measurable density $f(.|\ z;\alpha) = dF_{Y|Z}(.|\ z;\alpha)/d\nu_y$. (b) For ($H^o$-almost) all (y,z), $f(y|z;.)$ is continuous and strictly positive on A.

Assumption A2 ensures that we can define (almost surely) the conditional log-likelihood function:

$$L_n^c(Y|Z;\alpha) = \sum_{t=1}^{n} \log f(Y_t|Z_t;\alpha). \tag{2.1}$$

A Conditional Maximum-Likelihood Estimator (CMLE) is a σ_x^n-measurable function $\hat{\alpha}_n$ of $(X_1,\dots,X_n)$ such that:

$$L_n^c(Y|Z;\hat{\alpha}_n) = \sup_{\alpha \in A} L_n^c(Y|Z;\alpha).\text{[3]} \tag{2.2}$$

As stated below, Assumptions A1-A2 ensure the existence of a CMLE, $\hat{\alpha}_n$, for every n. To establish the strong consistency of a sequence of CMLE's, the next assumption is made.

ASSUMPTION A3: (a) For (H^o-almost) all (y,z), $|\log f(y|z;\alpha)| \leq M_1(y,z)$ for all α in A where $M_1(.,.)$ is H^o-integrable. (b) The function $z^f(\alpha) \equiv \int \log f(y|z;\alpha)\ dH^o(y,z)$ has a unique maximum on A at α^* (say).

Part (a) of assumption A3 ensures that $z^f(\alpha)$ is well-defined for any α in A, while part (b) requires global asymptotic identifiability of α^* (see e.g. Bowden (1973), Rothenberg (1971), and White (1982)).

Finally, to derive the asymptotic distribution of a CMLE, additional assumptions are made on the conditional density $f(y|z;\alpha)$.

ASSUMPTION A4: (a) For (H^o-almost) all (y,z), $\log f(y|z;.)$ is twice continuously differentiable on A. (b) For (H^o-almost) all (y,z), $|\partial^2 \log f(y|z;\alpha)/\partial\alpha\partial\alpha'| \leq M_2(y,z)$ for all α in A where $M_2(.,.)$ is H^o-integrable. (c) For (H^o-almost) all (y,z), $|\partial \log f(y|z;\alpha)/\partial\alpha \,.\, \partial \log f(y|z;\alpha)/\partial\alpha'| \leq M_3(y,z)$ for all α in A where $M_3(.,.)$ is H^o-integrable.

Parts (a) and (b) imply that $z^f(.)$ is twice continuously differentiable on A and that we can reverse the order of differentiation and integration when computing the first and second partial derivatives of $z^f(.)$. Given Parts (b) and (c), Jennrich's uniform strong Law of Large Numbers (1969, Theorem 2, p. 636) applies to:

$$A_n^f(\alpha) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial^2 \log f(Y_t|Z_t;\alpha)}{\partial\alpha\partial\alpha'}, \qquad (2.3)$$

and

$$B_n^f(\alpha) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial \log f(Y_t|Z_t;\alpha)}{\partial\alpha} \,.\, \frac{\partial \log f(Y_t|Z_t;\alpha)}{\partial\alpha'}. \qquad (2.4)$$

This ensures that $A_n^f(\hat{\alpha}_n)$ and $B_n^f(\hat{\alpha}_n)$ are strongly consistent estimators of

$$A_o^f(\alpha^*) = E^o\left[\frac{\partial^2 \log f(Y_t|Z_t;\alpha^*)}{\partial\alpha\partial\alpha'}\right], \qquad (2.5)$$

and

$$B_o^f(\alpha^*) = E^o\left[\frac{\partial \log f(Y_t|Z_t:\alpha^*)}{\partial\alpha} \,.\, \frac{\partial \log f(Y_t|Z_t;\alpha^*)}{\partial\alpha'}\right], \qquad (2.6)$$

where $E^o[.]$ is the expectation with respect to the true c.d.f. $H^o(.)$.

ASSUMPTION A5: (a) α^* is an interior point of A. (b) α^* is a regular point of $A_o^f(\alpha)$.

As is well known, Part (a) ensures that $\partial z^f/\partial\alpha$ is null at α^*. As in White's Theorem 3.1 (1982, p. 6), Part (b) together with Assumption A3-b implies that $A_o^f(\alpha^*)$ is nonsingular.

3. Asymptotic Properties of Conditional Maximum-Likelihood Estimators

Given the previous assumptions, the asymptotic properties of a sequence of CMLE's can readily be derived by standard techniques based on lemmas given by LeCam (1953) and Jennrich (1969). Alternatively, these properties can be obtained from White's (1982) results by noting that a CMLE can be thought of as a Quasi Maximum-Likelihood estimator (QMLE).

Let G be a postulated distribution for Z_t, and let H^G be the family of joint distributions $H^G(.,.;\alpha)$ for $X_t = (Y_t',Z_t')'$ defined as:

$$H^G = \{H^G(.,.;\alpha) : H^G(.,.;\alpha) = F_{Y|Z}(.|.;\alpha)G(.),\ \alpha \in A\}.$$

Suppose that G has a σ_z-measurable density $g(.) = dG/d\nu_z$ which is (H^o-almost surely) strictly positive. Then given A2, for any α in A,

$H^G(.,.;\alpha)$ has a σ_x-measurable density, $h^G(.,.;\alpha)$, which is (H^o-almost surely) strictly positive. A QMLE, $\hat{\alpha}_n^G$, for the family H^G of joint distributions for X_t is a σ_x^n-measurable function of $(X_1,\ldots,X_n)$ that satisfies:

$$L_n^G(Y,Z;\hat{\alpha}_n^G) = \sup_{\alpha \in A} L_n^G(Y,Z;\alpha) \tag{3.1}$$

where

$$\begin{aligned} L_n^G(Y,Z;\alpha) &= \sum_{t=1}^{n} \log h^G(Y_t,Z_t;\alpha) \\ &= L_n^c(Y|Z;\alpha) + \sum_{t=1}^{n} \log g(Z_t). \end{aligned} \tag{3.2}$$

We obviously have:

LEMMA 1: Given Assumptions A1-A2, a CMLE $\hat{\alpha}_n$ is a QMLE $\hat{\alpha}_n^G$ for the family H^G, where G can be any distribution for Z_t that has a (H^o-almost surely) strictly positive density.

The next result can be proved directly. Alternatively, since G can be arbitrarily chosen so as to satisfy the conditions of Lemma 1, then it is easy to check that the previous assumptions imply that White's assumptions hold for the family H^G so that the properties of QMLE's can be invoked (see White (1982, Theorems 2-1, 2-2, 3-2)).

THEOREM 1 (Asymptotic Properties of CMLE's): Let $\{\hat{\alpha}_n\}$ be a sequence of CMLE's.

(a) Given Assumptions A1-A2, for any n, there exists almost surely a CMLE $\hat{\alpha}_n$,

(b) Given Assumptions A1-A3, $\hat{\alpha}_n \xrightarrow{a.s.} \alpha^*$,

(c) Given Assumptions A1-A4,

$$A_n^f(\hat{\alpha}_n) \xrightarrow{a.s.} A_o^f(\alpha^*), \; B_n^f(\hat{\alpha}_n) \xrightarrow{a.s.} B_o^f(\alpha^*),$$

(d) Given Assumptions A1-A6, $\sqrt{n}(\hat{\alpha}_n - \alpha^*) \xrightarrow{D} N(0, C(\alpha^*))$ where

$$C(\alpha^*) = \left[A_o^f(\alpha^*)\right]^{-1} \left[B_o^f(\alpha^*)\right] \left[A_o^f(\alpha^*)\right]^{-1}.$$

Since Theorem 1 derives the properties of CMLE's under general conditions, it follows that we can make inference (on α^*) even when the conditional model for Y_t given Z_t is misspecified, i.e., even when $F^o_{Y|Z}(.|.)$ does not belong to the family of conditional distributions $\{F_{Y|Z}(.|.;\alpha);\ \alpha \in A\}$. From Lemma 1 and White (1982)'s Theorems 3.4 and 3.5, such inferences should be based on appropriate Wald statistics or Lagrange Multiplier statistics. It is also noteworthy that we are in a case in which $A_n^f(\hat{\alpha}_n)$ and $B_n^f(\hat{\alpha}_n)$ both consistently estimate $A_o^f(\alpha^*)$ and $B_o^f(\alpha^*)$ even though, conditional upon the observed $z_1,\ldots,z_n$, the random variables $X_1,\ldots,X_n$ are independent but not identically distributed (see White's corrigendum (1983)).

Suppose now that the conditional model for Y_t given Z_t is correctly specified, i.e., that $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;\alpha^o)$ for some α^o

in A. The next result follows from Jensen's inequality (Rao (1973, p. 58)) applied to the conditional densities $f(y|z;\alpha)$ and $f(y|z;\alpha^o)$.

LEMMA 2: Given Assumptions A1-A3, if $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;\alpha^o)$, then $\alpha^* = \alpha^o$.

Equality between $A^f_o(\alpha^o)$ and $-B^f_o(\alpha^o)$ is obtained under the next weak additional assumption which is similar to that used by e.g., Silvey (1959, Assumption 13, p. 394).

ASSUMPTION A6: For (H^o-almost) all z, $\int \partial^2 f(y|z;\alpha^*)/\partial\alpha\partial\alpha' \, d\nu_y = 0$.

It is then straightforward to prove the next result which is similar to the usual information matrix equivalence (see e.g., White (1982, Theorem 3.3)).

LEMMA 3: Given Assumptions A1-A4 and A6, if $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;\alpha^o)$, then for H^o-almost all z:

$$E^o_{Y|Z}\left[\frac{\partial^2 \log f(Y_t|z;\alpha^o)}{\partial\alpha\partial\alpha'}\right] = - E^o_{Y|Z}\left[\frac{\partial \log f(Y_t|z;\alpha^o)}{\partial\alpha} \cdot \frac{\partial \log f(Y_t|z;\alpha^o)}{\partial\alpha'}\right]$$

where $E^o_{Y|Z}[.]$ is the expectation with respect to the true conditional distribution of Y_t given Z=z.

By taking the total expectation of both sides of the previous equation with respect to the true distribution of Z, it follows that under the assumptions of Lemma 3:

$$A^f_o(\alpha^*) = -B^f_o(\alpha^*), \tag{3.3}$$

i.e., the usual information matrix equivalence.

The asymptotic properties of a sequence of CMLE's, when the conditional model for Y_t given Z_t is correctly specified, simply follow from Theorem 1 and Lemmas 2 and 3.

THEOREM 2 (Asymptotic Properties of CMLE's under correct specification of the conditional model): Let $\{\hat{\alpha}_n\}$ be a sequence of CMLE's. If $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.,\alpha^o)$, then:

(a) Given Assumptions A1-A3, $\hat{\alpha}_n \xrightarrow{a.s.} \alpha^o$,

(b) Given Assumptions A1-A4,

$$A^f_n(\hat{\alpha}_n) \xrightarrow{a.s.} A^f_o(\alpha^o), \quad B^f_n(\hat{\alpha}_n) \xrightarrow{a.s.} B^f_o(\alpha^o),$$

(c) Given Assumptions A1-A6, $\sqrt{n}(\hat{\alpha}_n - \alpha^o) \xrightarrow{D} N(0, C(\alpha^o))$ where

$$C(\alpha^o) = -\left[A^f_o(\alpha^o)\right]^{-1} = \left[B^f_o(\alpha^o)\right]^{-1}.$$

Theorem 2 is basically Andersen's (1970) result in a different framework (see Example 4 below, and footnote 14). Theorem 2 emphasizes, however, the robustness of a CMLE with respect to possible misspecification of the distribution of the conditioning variables Z_t. Specifically, suppose that one specifies a joint parametric model for (Y_t, Z_t), i.e., chooses some parametric family of joint distributions $\{H(.,.;\theta);\ \theta \in \Theta\}$ where Θ is some parameter space. Then, the associated family of conditional distributions for Y_t given Z_t is necessarily parameterized by some parameter α in some parameter space A so that it can be written as $\{F_{Y|Z}(.|.,\alpha);\ \alpha \in A\}$. For instance, α may be θ itself, a subvector of θ, or more generally a function of θ. If the conditional model for Y_t given Z_t is correctly specified and if Assumptions A1-A6 are satisfied, then the aforementioned properties of a CMLE and the classical inferences based on Wald, Lagrange Multiplier, and Log-Likelihood Ratio statistics are valid even though the induced marginal model for Z_t may be incorrectly specified., i.e., even though the true marginal distribution $G^o(.)$ of Z_t may not belong to the family $\{G(.;\theta);\ \theta \in \Theta\}$ where $G(.;\theta)$ is the marginal distribution of Z_t derived from $H(.,.;\theta)$.[4] In particular, strong consistency of a CMLE to α^o is robust with respect to misspecification of the marginal model for Z_t.

4. Asymptotic Efficiency of Conditional Maximum-Likelihood Estimators

Up to now, nothing has been said about asymptotic efficiency of CMLE's. This is so because Assumptions A2-A6 do not require that a probability model for the conditioning variables Z_t be specified. If, however, one specifies a joint probability model for (Y_t,Z_t) parameterized by θ in Θ, as above, then under suitable regularity conditions, one can define the information matrix (for one observation) as usual by:

$$I(\theta) = \mathrm{Var}^o\left[\frac{\partial \log h(Y_t,Z_t;\theta)}{\partial\theta}\right] \tag{4.1}$$

where "var ^{o}" means that the variance-covariance matrix is computed

with respect to the true distribution H^o of (Y_t,Z_t). Given the previous assumptions and some usual regularity conditions on the joint density $h(.,.;.)$, it follows from the asymptotic efficiency of FIML when $H^o(Y,Z) = H(Y,Z;\theta^o)$ for some θ^o (see e.g., Rao (1963)) that CMLE's are not in general asymptotically efficient estimators of α^o in the sense that:

$$C(\alpha^o) \geq \frac{\partial\alpha}{\partial\theta'}\bigg|_{\theta^o}\left[I(\theta^o)\right]^{-1}\frac{\partial\alpha'}{\partial\theta}\bigg|_{\theta^o}. \tag{4.2}$$

This is so because the marginal probability model for the conditioning variables Z_t may contain unused information on α^o. It follows that CMLE's are in general asymptotically inefficient estimators of α^o even when the conditional model for Y_t given Z_t is correctly specified. In some sense, we have traded off efficiency for some robustness by using CML estimation instead of FIML estimation. In this section, we shall recover the aforementioned result by embedding the issue of efficiency of CMLE's in a more general framework. In addition we shall characterize the conditions under which CMLE's are efficient.

Let $\{(Y_{1t}),(Y_{2t})\}$ be a partition of the set of variables (Y_t). Let p_i be the number of variables in Y_{it} where $p_i \geq 1$ for $i=1,2$.[5] Thus $p_1 + p_2 = p$. We shall consider the conditional model for Y_{1t} given (Y_{2t},Z_t) induced by the conditional model for Y_t given Z_t. Given Assumption A2, the conditional density of Y_{1t} given (Y_{2t},Z_t) is parameterized by some parameters α_1 in $\mathbb{R}^{k_1}$ that are functions of α, i.e., $\alpha_1 = \alpha_1(\alpha)$. Let $J(\alpha)$ be the Jacobian at α, if it exists, of the transformation $\alpha_1(.)$, i.e. $J(\alpha) = [\partial\alpha_1/\partial\alpha']$.

ASSUMPTION A7: (a) The function $\alpha_1(.)$ is continuously differentiable. (b) For any α in A, the Jacobian $J(\alpha)$ has full row rank.

Let $A_1 = \alpha_1(A)$. Given Assumption A7-(a), A_1 is a compact subset of $\mathbb{R}^{k_1}$. Assumption A7-(b) implies in particular that $k_1 \leq k$.

Let Assumptions A2'-A6' be the assumptions on the conditional model for Y_{1t} given (Y_{2t},Z_t) that correspond to the previous Assumptions A1-A6. For instance, Assumption A3'-(b) states that the function $z^{f_1}(\alpha_1)$ defined as $\int \log f_1(y_1|y_2,z;\alpha_1)dH^o(y,z)$, where $f_1(.|.,.;\alpha_1)$ denotes the conditional density of Y_{1t} given (Y_{2t},Z_t) for α_1 in A_1, has a unique maximum α_1^* on A_1.

It is important to note that in general α_1^* is not equal to $\alpha_1(\alpha^*)$ where α^* maximizes the function $z^f(\alpha)$ (see Assumption A3). This remark will be used in the following section. On the other hand, if the conditional model for Y_t given Z_t is correctly specified, i.e. $F^o_{Y|Z}(.|.) = F^o_{Y|Z}(.|.;\alpha^o)$ for some α^o in A, then the conditional model for Y_{1t} given (Y_{2t},Z_t) is necessarily correctly specified, i.e., $F^o_{Y_1|Y_2Z}(.|.,.) = F^o_{Y_1|Y_2Z}(.|.,.;\alpha_1^o)$ for some α_1^o in A_1^o. Moreover, we must have:

$$\alpha_1^o = \alpha_1(\alpha^o). \qquad (4.3)$$

Let $\hat{\alpha}_{1n}$ be the estimator defined by $\hat{\alpha}_{1n} = \alpha_1(\hat{\alpha}_n)$ where $\hat{\alpha}_n$ is the CMLE obtained by estimating the conditional model for (Y_{1t},Y_{2t}) given Z_t. Then, let $\tilde{\alpha}_{1n}$ be the CMLE obtained by estimating the conditional model for Y_{1t} given (Y_{2t},Z_t). We shall first study the properties of these two estimators under general conditions.

Let $A_o^{f_1}(.)$, $B_o^{f_1}(.)$, $A_n^{f_1}(.)$, and $B_n^{f_1}(.)$ be analogous to the corresponding matrices for $f(y|z;\alpha)$ defined in Section 2. Let

$$B_o^{ff_1}(\alpha^*,\alpha_1^*) = E^o\left[\frac{\partial \log f(Y_t|Z_t;\alpha^*)}{\partial \alpha} \; . \; \frac{\partial \log f_1(Y_{1t}|Y_{2t},Z_t;\alpha_1^*)}{\partial \alpha_1'}\right], \qquad (4.4)$$

and

$$B_n^{ff_1}(\alpha,\alpha_1) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial \log f(Y_t|Z_t;\alpha)}{\partial \alpha} \; . \; \frac{\partial \log f_1(Y_{1t}|Y_{2t},Z_t,\alpha_1)}{\partial \alpha_1'}. \qquad (4.5)$$

The existence of $B_o^{ff_1}(\alpha^*,\alpha_1^*)$ and the strong convergence of $B_n^{ff_1}(\hat{\alpha}_n,\tilde{\alpha}_{1n})$ to $B_o^{ff_1}(\alpha^*,\alpha_1^*)$ are ensured by the following assumption.

ASSUMPTION A8: For (H^o-almost) all (y,z), $|\partial \log f(y|z;\alpha)/\partial\alpha \; . \; \partial \log f_1(y_1|y_2,z;\alpha_1)/\partial\alpha_1'| \leq M_4(y,z)$ for all (α,α_1) in $A \times A_1$ where $M_4(.,.)$ is H^o integrable.

The following theorem gives the joint asymptotic distribution of $(\hat{\alpha}_{1n},\tilde{\alpha}_{1n})$ even when the conditional model for (Y_{1t},Y_{2t}) given Z_t is misspecified. Let

$$C(\alpha^*,\alpha_1^*) = \left[\begin{array}{ccc} C_{11}(\alpha^*) & & C_{12}(\alpha^*,\alpha_1^*) \\ C_{21}(\alpha^*,\alpha_1^*) & ; & C_{22}(\alpha_1^*) \end{array}\right] \qquad (4.6)$$

where

$$C_{11}(\alpha^*) = J(\alpha^*)\left[A_o^f(\alpha^*)\right]^{-1}B_o^f(\alpha^*)\left[A_o^f(\alpha^*)\right]^{-1}J'(\alpha^*),$$

$$C_{12}(\alpha^*,\alpha_1^*) = C_{21}'(\alpha^*,\alpha_1^*) = J(\alpha^*)\left[A_o^{f}(\alpha^*)\right]^{-1} B_o^{ff_1}(\alpha^*,\alpha_1^*)\left[A_o^{f_1}(\alpha_1^*)\right]^{-1},$$

$$C_{22}(\alpha_1^*) = \left[A_o^{f_1}(\alpha_1^*)\right]^{-1} B_o^{f_1}(\alpha_1^*)\left[A_o^{f_1}(\alpha_1^*)\right]^{-1}.$$

Let $C_n(\hat{\alpha}_n,\tilde{\alpha}_{1n})$ be the sample analog of $C(\alpha^*,\alpha_1^*)$ evaluated at $(\hat{\alpha}_n,\tilde{\alpha}_{1n})$.

THEOREM 3 (Joint Asymptotic Distribution of CMLE's): Given Assumptions A1-A5, A2'-A5', A7, and A8:

(a) For any n, the estimators $(\hat{\alpha}_{1n},\tilde{\alpha}_{1n})$ almost surely exist,

(b) $(\hat{\alpha}_{1n},\tilde{\alpha}_{1n}) \overset{a.s.}{\rightarrow} (\alpha_1(\alpha^*),\alpha_1^*)$,

(c) $\sqrt{n}\begin{bmatrix}\hat{\alpha}_{1n} - \alpha_1(\alpha^*) \\ \\ \tilde{\alpha}_{1n} - \alpha_1^*\end{bmatrix} \overset{D}{\rightarrow} N(0,C(\alpha^*,\alpha_1^*))$,

(d) $C_n(\hat{\alpha}_n,\tilde{\alpha}_{1n}) \overset{a.s.}{\rightarrow} C(\alpha^*,\alpha_1^*)$.

We now study the properties of the estimators $\hat{\alpha}_{1n}$ and $\tilde{\alpha}_{1n}$ under correct specification of the conditional model for (Y_{1t},Y_{2t}) given Z_t. As noted earlier, when the conditional model for (Y_{1t},Y_{2t}) given Z_t is correctly specified, Equation (4.3) holds. It follows that $\hat{\alpha}_{1n}$ and $\tilde{\alpha}_{1n}$ are both consistent estimators of α_1^o. The next theorem states that the estimator $\hat{\alpha}_{1n}$ defined by $\alpha_1(\hat{\alpha}_n)$ where $\hat{\alpha}_n$ is the CMLE obtained by estimating the conditional model for (Y_{1t},Y_{2t}) given Z_t is at least as efficient as the CMLE $\tilde{\alpha}_{1n}$ obtained by estimating the conditional model for Y_{1t} given (Y_{2t},Z_t). This is expected since the CMLE $\hat{\alpha}_n$ and hence the estimator $\hat{\alpha}_{1n}$ are

asymptotically efficient estimators of α^o and α_1^o respectively when the conditional model for (Y_{1t},Y_{2t}) given Z_t is correctly specified. The import of the theorem is that it also characterizes the cases for which the CMLE $\tilde{\alpha}_{1n}$ is as efficient as $\hat{\alpha}_{1n}$ and therefore asymptotically efficient.

We need some additional definitions and a lemma. Let, when it exists,

$$B_o^{f_2}(\alpha^o) = E^o\left[\frac{\partial \log\, f_2(Y_{2t}|Z_t;\alpha^o)}{\partial\alpha}\cdot\frac{\partial \log\, f_2(Y_{2t}|Z_t;\alpha^o)}{\partial\alpha'}\right] \qquad (4.7)$$

where $f_2(.|.;\alpha)$ is the conditional density of Y_{2t} given Z_t derived from the conditional density $f(.,.|.;\alpha)$ of (Y_{1t},Y_{2t}) given Z_t.

LEMMA 4: Given Assumptions A1-A5, A2'-A5', and A7, if $F_{Y|Z}^o(.|.) = F_{Y|Z}(.|.;\alpha^o)$, then:

$$B_o^f(\alpha^o) = J'(\alpha^o)B_o^{f_1}(\alpha_1^o)J(\alpha^o) + B_o^{f_2}(\alpha^o).$$

Where each of the above matrices is finite.

On the other hand, from the rank factorization of the $k_1 \times k$ matrix $J(\alpha^o)$ (see Rao (1973, p. 19)) we have:

$$J(\alpha^o) = MLN \qquad (4.8)$$

where M is a $k_1 \times k_1$ non-singular matrix, N is a $k \times k$ orthogonal matrix, and L is a $k_1 \times k_1$ matrix of which all the elements are null except the first r diagonal elements which are all equal to one, where $r = \text{rank } J(\alpha^o)$. From Assumption A7, it follows that $r = k_1 \leq k$. Therefore:

$$L = \begin{bmatrix} I_{k_1} & 0 \end{bmatrix} \qquad (4.9)$$

Then, when $k_1 < k$, we partition the $k \times k$ matrix $NB_o^{f_2}(\alpha^o)N'$ as follows:

$$NB_o^{f_2}(\alpha^o)N' = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \qquad (4.10)$$

where Z_{11} is a $k_1 \times k_1$ matrix.

THEOREM 4 (Asymptotic Efficiency of CMLE's): Given Assumptions A1-A6, A2'-A6', and A7, if $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;\alpha^o)$, then:

$$C_{11}(\alpha^o) \leq C_{22}(\alpha^o_1)$$

where the equality holds if and only if:

(i) $NB_o^{f_2}(\alpha^o)N' = 0$ when $k_1 = k$,

(ii) $Z_{11} - Z_{12} Z_{22}^{-1} Z_{21} = 0$ when $k_1 < k$.

It is easy to see that the inequality (4.2) discussed at the outset of this section is a special case of the above general result. Indeed, it suffices to let in Theorem 4, Z be the empty set, and the variables (Y_1,Y_2) and the parameters (α_1,α) be respectively the variables (Y,Z) and the parameters (α,θ) in the inequality (4.2).

As another special case of Theorem 4, let us consider the case in which the variables (Y_{2t},Z_t) are weakly exogenous for α_1 in the sense of Engle, Hendry, and Richard (1983). Let $\alpha = (\alpha_1,\alpha_2)$, and suppose that $A = A_1 \times A_2$, and

$$f_2(y_2|z;\alpha) = f_2(y_2|z;\alpha_2), \qquad (4.11)$$

i.e., that the (conditional) density of Y_{2t} depends only on α_2 Since, by definition of α_1, the conditional density of Y_{1t} given (Y_{2t},Z_t) depends only on α_1, then (α_1,α_2) operates a sequential cut. Let k_2 be the number of parameters in α_2, where $k_2 \geq 1$.

It readily follows that the $k \times k$ matrix $B_o^{f_2}$ is of the form:

$$B_o^{f_2}(\alpha^o) = \begin{bmatrix} 0 & 0 \\ 0 & \tilde{B}_o^{f_2}(\alpha_2^o) \end{bmatrix} \qquad (4.12)$$

where $\tilde{B}_o^{f_2}(\alpha_2^o)$ is the $k_2 \times k_2$ matrix defined as:

$$\tilde{B}_o^{f_2}(\alpha_2^o) = E^o \left[\frac{\partial \log f_2(Y_{2t}|Z_t;\alpha_2^o)}{\partial \alpha_2} \cdot \frac{\partial \log f_2(Y_{2t}|Z_t;\alpha_2^o)}{\partial \alpha_2'} \right] .$$

Since

$$J(\alpha^o) = \begin{bmatrix} I_{k_1} & 0 \end{bmatrix} \qquad (4.13)$$

it follows that $M = I_{k_1}$ and $N = I_k$. Therefore $NB_o^{f_2}N' = B_o^{f_2}$. It is now easy to see that $Z_{11}-Z_{12}\ Z_{22}^{-1}\ Z_{21} = 0$. Thus from Theorem 4, the CMLE $\tilde{\alpha}_{1n}$ is asymptotically efficient. As expected, when (Y_{2t}, Z_t) are strictly exogenous for α_1, no loss in efficiency is achieved by maximizing the conditional likelihood function for Y_1 given (Y_2, Z). As indicated by Theorem 4, the efficiency of CMLE's can arise, however, in other situations.

5. Specification Tests

Given the previous properties of CMLE's, it is of interest to know whether the chosen (or induced) conditional model for Y_t given Z_t is correctly specified, i.e., whether $F^o_{Y|Z}(.|.) = F_{Y|Z}(.,.;\alpha^o)$ for some α^o in A. From Lemma 1, it follows that to test such a hypothesis, we can apply the Information Matrix test proposed by White (1982, Theorems 4.1). This test is based on the nullity of $A^f_o(\alpha^*) + B^f_o(\alpha^*)$ which holds when the conditional model is correctly specified (see Equation (3.3) and Lemma 2). Then the appropriate assumptions and the appropriate statistic are obtained by replacing the joint density $h(.,.;\theta)$ by the conditional density $f(.|.;\alpha)$ in White's Assumptions A.8-A.10 and in White's statistic (4.1).

In this section, we shall argue that CML estimation is a convenient tool for carrying out the tests for parameter estimator consistency proposed by Hausman (1978) and White (1982, Section 5). The essential reason comes from the fact that we can choose or construct some appropriate variables to condition upon so that the parameters of interest appear in the resulting conditional likelihood.

We shall use the general framework introduced in Section 4.

The first specification test is based on the equation:

$$\alpha_1^* = \alpha_1(\alpha^*) \tag{5.1}$$

which holds when the conditional model for (Y_{1t}, Y_{2t}) given Z_t is correctly specified (see Equation (4.3)). From Section 4, $\hat{\alpha}_{1n}$ as defined by $\alpha_1(\hat{\alpha}_n)$, and $\tilde{\alpha}_{1n}$ are both consistent estimators of $\alpha_1^* = \alpha_1^o$ under correct specification. Moreover $\hat{\alpha}_{1n}$ is asymptotically efficient. Thus, following Hausman (1978), the difference $\tilde{\alpha}_{1n} - \hat{\alpha}_{1n}$ can be used to construct a test of equation (5.1).[6]

Let

$$V = C_{22}(\alpha_1^*) + C_{11}(\alpha^*) - C_{12}(\alpha^*, \alpha_1^*) - C_{21}(\alpha^*, \alpha_1^*) \tag{5.2}$$

where the matrices on the right-hand side are defined in Equation (4.6). Let V_n be the sample analog of V evaluated at $(\hat{\alpha}_n, \hat{\alpha}_{1n})$. Contrary to White's Assumption A.12, the $k_1 \times k_1$ matrices V and V_n may turn out to be singular. Let r and r_n be their respective rank.

When $1 \leq r < k_1$ and $1 \leq r_n < k_1$, we use generalized inverses of maximum rank, i.e., of rank k_1 of V and V_n. Specifically, let R be a $k_1 \times (k_1 - r)$ matrix such that the $k_1 \times (2k_1 - r)$ partitioned matrix [V, R] is of rank k_1. The $k_1 \times (k_1 - r_n)$ matrix R_n is similarly defined with respect to V_n. Then, from Rao and Mitra (1971, Section 2.7) it follows that the matrices $V + RR'$ and $V_n + R_nR_n'$ are non-singular and that $[V + RR']^{-1}$ and $[V_n + R_nR_n']^{-1}$ are generalized inverses of rank k_1 of V and V_n respectively. The matrices R and R_n

are not, however, unique. From now on, it is assumed that R and R_n are uniquely defined by the same continuous selection rule. This implies in particular that $V_n + R_nR_n'$ is a continuous function of V_n. This ensures that $V_n + R_nR_n'$ is a strongly consistent estimator of $V + RR'$ whenever V_n is a strongly consistent estimator of V.

Let

$$H_n = n(\tilde{\alpha}_{1n} - \hat{\alpha}_{1n})'[V_n + R_nR_n']^{-1} (\tilde{\alpha}_{1n} - \hat{\alpha}_{1n}). \qquad (5.3)$$

Let H_o be the hypothesis that $\alpha_1^* = \alpha_1(\alpha^*)$, and let H_1 be its complement.

THEOREM 5 (Hausman Test): Suppose that Assumptions A1-A5, A2'-A5', A7 and A8 hold. Suppose that $V \neq 0$. Then:

(a) under H_o, $H_n \overset{D}{\rightarrow} \chi_r^2$,

(b) under H_1, $H_n \overset{a.s.}{\rightarrow} \infty$.

Thus, if H_n exceeds the critical value for the χ_r^2 distribution at a given significance level, one must reject the hypothesis that $\alpha_1^*=\alpha_1(\alpha^*)$ and hence that the conditional model for (Y_{1t},Y_{2t}) given Z_t is correctly specified. Part (b) of Theorem 5 states that this test is actually consistent.

It is also important to note that the test is valid only when $V \neq 0$. Second, the asymptotic coveriance matrix of $\sqrt{n}(\tilde{\alpha}_{1n} - \hat{\alpha}_{1n})$ is not simply the difference between the asymptotic covariance matrices $C_{22}(\alpha_1^*)$ and $C_{11}(\alpha^*)$ of $\tilde{\alpha}_{1n}$ and $\hat{\alpha}_{1n}$. This latter convenient property

is, however, obtained under additional assymptions.

LEMMA 5: If in addition to the assumptions of Theorem 5, the following holds:

(i) $A_o^f = -B_o^f(\alpha^*)$, $A_o^{f_2}(\alpha_1^*) = -B_o^{f_1}(\alpha_1^*)$, and

(ii) for H^o-almost all (y_2,z)

$$E^o_{Y_1|Y_2Z}[\partial \log f_1(Y_{1t}|y_2,z;\alpha_1^*)/\partial\alpha_1] = 0,$$

then under H_o:

$$V = \left[B_o^{f_1}(\alpha_1^*)\right]^{-1} - J(\alpha^*)\left[B_o^f(\alpha^*)\right]^{-1}J'(\alpha^*). \qquad (5.4)$$

Condition (i) is just the information matrix equivalence at α^* and α_1^* for the conditional densities f and f_1. Condition (ii) is stronger than the requirement that α_1^* maximize $z^{f_1}(.)$ which is $E^o[\log f_1(Y_{1t}|Y_{2t},Z_t;.)]$. It is, however, worth noting that condition (ii) is automatically satisfied when the conditional model for Y_{1t} given (Y_{2t},Z_t) is correctly specified. (This follows from Lemma A2 in the Appendix.) Moreover, as emphasized by Lemma 5, Equation (5.4) holds only under H_o. In other words, Hausman's well-known formula (5.4) esentially holds under H_o and under correct specification of the conditional model for Y_{1t} given (Y_{2t},Z_t).

Finally, let us note from Theorem 4 that, under correct spedification of the conditional model for (Y_{1t},Y_{2t}) given Z_t the condition that V be non-zero is simply equivalent to the condition that $\tilde{\alpha}_{1n}$ be an inefficient estimator of α_1^*.

Our second specification test is based on the equation:

$$E^o[\partial \log f_1(Y_{1t}|Y_{2t},Z_t;\alpha_1(\alpha^*))/\partial\alpha_1] = 0. \qquad (5.5)$$

Indeed this must hold when the conditional model for (Y_{1t},Y_{2t}) given Z_t is correctly specified since in such a case $\alpha_1^* = \alpha_1(\alpha^*)$ (see Equation (4.3)), and since by definition α_1^* maximizes $z^{f_1}(.)$. Thus following White (1982), the appropriate test statistic is based on $(1/n)\ \partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial\alpha_1$ where $\hat{\alpha}_{1n} = \alpha_1(\hat{\alpha}_n)$. Let:

$$\begin{aligned}\Omega = {} & A_o^{f_1}(\alpha_1(\alpha^*))C_{11}(\alpha^*)A_o^{f_1}(\alpha_2(\alpha^*)) + B_o^{f_1}(\alpha_1(\alpha^*)) \\ & - A_o^{f_1}(\alpha_1(\alpha^*))J(\alpha^*)[A_o^f(\alpha^*)]^{-1}\ B_o^{ff_1}(\alpha^*,\alpha_1(\alpha^*)) \\ & - B_o^{f_1f}(\alpha^*,\alpha_1(\alpha^*))[A_o^f(\alpha^*)]^{-1}\ J'(\alpha^*)A_o^{f_1}(\alpha_1(\alpha^*)) \\ & - \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial\alpha_1} \cdot \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial\alpha_1'}. \end{aligned} \qquad (5.6)$$

It turns out that Ω is the asymptotic covariance matrix of $(1/\sqrt{n})\ \partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial\alpha_1$. Note that Ω depends only on α^*. Let Ω_n be the sample analog of Ω evaluated at $\hat{\alpha}_n$, i.e., where (say) the last term in (5.6) is replaced by:

$$\frac{1}{n}\ \frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial\alpha_1} \cdot \frac{1}{n}\ \frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial\alpha_1'}.$$

Let s and s_n be the respective rank of Ω and Ω_n. We consider again generalized inverses of maximum rank. Moreover, as before, we use a continuous selection rule for choosing the $k_1 \times (k_1 - s_n)$ matrix Q_n where Q_n is such that the $k_1 \times (2k_1 - s_n)$ matrix $[\Omega_n,Q_n]$ is of rank k_1.

Let

$$G_n = \frac{1}{n}\ \frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial\alpha_1'}\ [\Omega_n + Q_nQ_n']^{-1}\ \frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial\alpha_1}. \qquad (5.7)$$

Let H_o' be the hypothesis that Equation (5.5) holds, and let H_1' be its complement.

THEOREM 6 (Gradient Test): Suppose that Assumptions A1-A5, A2'-A5', A7 and A8 hold. Suppose that $\Omega \neq 0$. Then:

(a) under H_o', $G_n \overset{D}{\rightarrow} \chi_s^2$,

(b) under H_1', $G_n \overset{a.s.}{\rightarrow} \infty$.

Thus if G_n exceeds the critical value for the χ_s^2 distribution, one must reject the null hypothesis H_o' and hence that the conditional model for (Y_{1t},Y_{2t}) given Z_t is correctly specified. Moreover the gradient test is consistent.

Though the statistic (5.7) is similar in spirit to White's gradient test, it differs from it in the choice of the covariance matrix estimator. To simplify the discussion, let us restrict ourselves to the case studied by White in which the matrices Ω and Ω_n

are non-singular. Suppose in addition that $A_o^{f_1}(\alpha_1(\alpha^*))$ and $A_n^{f_1}(\hat{\alpha}_{1n})$ are non-singular. Then the covariance matrix estimator used by White (1982, Equation (5.2)) is:

$$\tilde{\Omega}_n = A_n^{f_1}(\hat{\alpha}_{1n})\ V_n(\hat{\alpha}_n,\hat{\alpha}_{1n})\ A_n^{f_1}(\hat{\alpha}_{1n}) \tag{5.8}$$

where

$$V_n(\hat{\alpha}_{n,}\hat{\alpha}_{1n}) = C_{22n}(\hat{\alpha}_{1n}) + C_{11}(\hat{\alpha}_n) - C_{12n}(\hat{\alpha}_n,\hat{\alpha}_{1n}) - C_{21n}(\hat{\alpha}_n,\hat{\alpha}_{1n}),$$

and $\hat{\alpha}_{1n} = \alpha_1(\hat{\alpha}_n)$.[7] Given the assumptions of Theorem 3, it follows from part b of that theorem that:

$$V_n(\hat{\alpha}_n,\hat{\alpha}_{1n}) \overset{a.s.}{\rightarrow} C_{22}(\alpha_1(\alpha^*)) + C_{11}(\alpha^*) - C_{12}(\alpha^*,\alpha_1(\alpha^*)) - C_{23}(\alpha^*,\alpha_1(\alpha^*))$$

$$A_n^{f_1}(\hat{\alpha}_{1n}) \overset{a.s.}{\rightarrow} A_o^{f_1}(\alpha_1(\alpha^*)).$$

Thus, from Equations (4.6) and (5.6), we get:

$$\tilde{\Omega}_n \overset{a.s.}{\rightarrow} \Omega + \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1} \cdot \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1'}. \tag{5.9}$$

Hence, only under the null hypothesis H_o' will $\tilde{\Omega}_n$ be a consistent estimator of Ω. Let $\tilde{G}_n$ be White's gradient statistic based on $\tilde{\Omega}_n$. It is easy to see that White's gradient test is consistent since $\tilde{\Omega}_n$ converges to a positive definite matrix. Let us, however, compare the asymptotic power of the two tests. We have:

$$\frac{1}{n}(G_n - \tilde{G}_n) = \frac{1}{n}\frac{\partial L_n^c(Y_1|Y_2,Z,\hat{\alpha}_{1n})}{\partial \alpha_1'}\left[\Omega_n^{-1} - \tilde{\Omega}_n^{-1}\right]\frac{1}{n}\frac{\partial L_n^c(Y_1|Y_2,Z,\hat{\alpha}_{1n})}{\partial \alpha_1} \tag{5.10}$$

Since $\Omega_n^{-1} - \tilde{\Omega}_n^{-1}$ converges almost surely to a positive semi-definite matrix, it follows that our test based on G_n is asymptotically as powerful as White's test based on $\tilde{G}_n$ for any alternatives, and strictly more powerful for some alternatives.[8]

Returning to the general case, it is worth noting that the numbers of degrees of freedom of the asymptotic distributions of H_n and G_n are not equal. This actually results from the fact that the two statistics H_n and G_n are not designed to test the same hypothesis. However, since H_o implies H_o', one may use the gradient statistic to test H_o.

LEMMA 6: Given the assumptions of Theorem 5, we have under H_o:

(a) $\Omega = A_o^{f_1}(\alpha_1^*)\ V\ A_o^{f_1}(\alpha_1^*)$,

(b) $r = s$,

(c) $H_n - G_n \overset{a.s.}{\rightarrow} 0$.

Lemma 6 generalizes White's (1982) Theorem 5.2 to the singular case $r = s < k_1$: Under H_o, the Hausman test and the gradient test have the same number of degrees of freedom and are asymptotically equivalent. It is noteworthy that our result holds irrespective of the choice of the generalized inverses V_n^- and Ω_n^-. However, while the Hausman test is consistent for any alternative H_1: $\alpha_1^* - \alpha_1(\alpha^*) = a \neq 0$

(see part b of Theorem 5), the gradient test may not have any power against some alternatives $a \neq 0$.

6. Examples

This section presents some applications of CMLE's and their properties. The first example deals with the simple linear model. The other three examples, which are the multinomial logit model, the Tobit model, and the bivariate logit model, illustrate different partitions of the parameter vector α. Specifically, these partitions are respectively:

(i) $f(Y_{1t},Y_{2t}|Z_t;\alpha) = f_1(Y_{1t}|Y_{2t},Z_t;\alpha) \ . \ f_2(Y_{2t}|Z_t;\alpha)$,

(ii) $f(Y_{1t},Y_{2t}|Z_t;\alpha_1,\alpha_2) = f_1(Y_{1t}|Y_{2t},Z_t;\alpha_1,\alpha_2) \ . \ f_2(Y_{2t}|Z_t;\alpha_2)$,

(iii) $f(Y_{1t},Y_{2t}|Z_t;\alpha_1,\alpha_2) = f_1(Y_{1t}|Y_{2t},Z_t;\alpha_1) \ . \ f_2(Y_{2t}|Z_t;\alpha_1,\alpha_2)$.

EXAMPLE 1: Suppose that one specifies the following simple linear model for (Y_t, Z_t), $t = 1,\dots,n$:

$$Y_t = \beta_1 + \beta_2 Z_t + u_t$$

where $E(u_t) = 0$ and $\text{var}(u_t) = \sigma^2$ for every t. We shall study the asymptotic properties of the OLS estimators $\hat\beta_1$, $\hat\beta_2$, and $\hat\sigma^2$ where $\hat\sigma^2$ is defined as the sum of squared residuals divided by n.

The OLS estimators, $\hat\alpha = (\hat\beta_1, \hat\beta_2, \hat\sigma^2)$, can be interpreted as CMLE's. Indeed they clearly maximize the conditional log-likelihood function $L_n^c(Y_1,\dots, Y_n \mid Z_1,\dots,Z_n;\alpha)$ where $\alpha = (\beta_1, \beta_2, \sigma^2)$ and

$$\log f(Y_t \mid Z_t;\alpha) = -.5 \log 2\pi -.5 \log \sigma^2 -.5(Y_t - \beta_1 - \beta_2 Z_t)^2/\sigma^2.$$

That is, the OLS estimators are identical to the CMLE's associated with the family of conditional normal distributions for Y_t given Z_t: $\{N(\beta_1 + \beta_2 Z_t;\sigma^2);\ (\beta_1,\beta_2)\in\mathbb{R}^2,\ \sigma^2 > 0\}$. Hence, the asymptotic properties of OLS follow from Section 3.

Specifically, let $(\mu_y^o, \mu_z^o, \sigma_{zy}^o, \sigma_{zz}^o, \sigma_{yz}^o)$ be the true means, variances and covariances of (Y_t, Z_t). Let $u_t^*=Y_t-\beta_1^*-\beta_2^*Z_t$ where $\alpha^*=(\beta_1^*,\beta_2^*,\sigma^{*2})$ is defined in Assumption A3-(b). Then it can be shown that α^* solves:

$$E^o(u_t^*) = 0 \quad , \quad E^o(u_t^*.Z_t) = 0 \quad , \quad E^o(u_t^{*2}) = \sigma^{*2}.$$

Hence:

$$\beta_1^* = \mu_y^o - \mu_z^o\,\sigma_{yz}^o \,/\, \sigma_{zz}^o \ ; \ \beta_2^* = \sigma_{yz}^o \,/\, \sigma_{zz}^o \ ; \ \sigma^{*2} = \sigma_{yy}^o - \sigma_{yz}^{o2} \,/\, \sigma_{zz}^o.$$

Then, the OLS estimator $\hat\alpha$ almost surely converges to α^*, whether or not the true conditional distribution of Y_t given Z_t is normal with a mean linear in Z_t and a variance independent of Z_t. Moreover, its asymptotic distribution is given by Theorem 1 so that one can conduct inferences on α^* through appropriately modified Wald or Lagrange Multiplier statistics. On the other hand, if the true conditional distribution of Y_t given Z_t is normal with a mean linear in Z_t and a variance independent of Z_t, i.e., $N(\beta_1^o + \beta_2^o Z_t;\sigma^{o2})$ for some $\alpha^o = (\beta_1^o, \beta_2^o, \sigma^{o2})$, then the OLS estimator $\hat\alpha$ consistently estimates α^o, as expected.

As mentioned in Section 3, these properties depend neither on the nature of the true distribution of Z_t nor on the choice of the marginal probability model for Z_t. Moreover, these properties are obtained whether or not Z_t is strictly or weakly exogenous for α (see Engle, Hendry, and Richard (1983) for definitions).[9,10]

Finally, one can test whether the true conditional distribution of Y_t given Z_t is normal with a mean linear in Z_t and a variance independent of Z_t. This is carried out by using White (1982) Information Matrix test as indicated above. It can be shown that $A_o^f(\alpha*) + B_o^f(\alpha*) = [d_{ij}]$ where:

$$
\begin{aligned}
d^*_{11} &= 0 &&, \quad d^*_{12} = \mathrm{cov}^o(u_t^{*2}, Z_t)/\sigma^{*4},\\
d^*_{22} &= \mathrm{cov}^o(u_t^{*2}, Z_t^2)/\sigma^{*4} &&, \quad d^*_{13} = E^o(u_t^{*3})/2\sigma^{*6} \quad ,\\
d^*_{23} &= E^o(u_t^{*3}.Z_t)/2\sigma^{*6} &&, \quad d^*_{33} = [E^o(u_t^{*4})-3(E^o(u_t^{*2}))^2]/4\sigma^{*8}.
\end{aligned}
$$

Thus the Information Matrix test is equivalent to testing $d^*_{12}= d^*_{22}= d^*_{13}= d^*_{23}= d^*_{33}= 0$. To carry out the test, consistent sample analogs $\hat{d}_{ijn}$ are used. For instance:

$$\hat{d}_{12n} = [\frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2 Z_t - (\frac{1}{n}\sum_{t=1}^{n}\hat{u}_t^2).(\frac{1}{n}\sum_{t=1}^{n}Z_t)] \;/\; \hat{\sigma}^4$$

where the $\hat{u}_t$ are the OLS residuals. It is worth noting that testing $d^*_{12}= d^*_{22}= 0$ is equivalent to testing that the squared OLS residuals are asymptotically uncorrelated with the cross products of the explanatory variables (see also White (1980)). On the other hand, testing $d^*_{13}= d^*_{23}= 0$ is equivalent to testing that the cubed OLS residuals are uncorrelated with the explanatory variables. Finally, while $d^*_{13}= 0$ means that the distribution of u^*_t is unskewed, the restriction $d^*_{33}= 0$ corresponds to the condition that the kurtosis of u^*_t be equal to 3 as required by the normal distribution (see White (1982)).

EXAMPLE 2: Let us consider the Multinomial Logit (MNL) model (see e.g., McFadden (1974), Nerlove and Press (1973)) for the random sample (Y_t,Z_t), t=1,...,n. Let Y_t be discrete with I categories. Then:

$$\log \Pr(Y_t=i) = \mu_t + v'_{it}\alpha$$

where $\mu_t = -\log\sum_{j=1}^{I}\exp(v'_{jt}\alpha)$, and where v_{it} combines characteristics of the alternative i and the individual characteristics Z_t.

Let B be a proper subset of the initial choice set. Define the statistic $S_t= 1_B(Y_t)$ where $1_B(.)$ is the indicator function of B. Let:

$$
\begin{aligned}
Y_{it} &= 1 \quad \text{if} \quad Y_t = i && ; \quad B_t = B \quad \text{if} \quad S_t = 1,\\
&= 0 \quad \text{otherwise} && ; \quad = B^c \quad \text{if} \quad S_t = 0,
\end{aligned}
$$

where B^c is the complement of B.

It is easy to show that the conditional log-likelihood of $(Y_1,\ldots,Y_n)$ given $(S_1,\ldots,S_n,Z_1,\ldots,Z_n)$ is:

$$L_n^c(Y|S,Z;\alpha) = \sum_{t=1}^{n}\sum_{i\in B_t} Y_{it}[v'_{it}\alpha - \log \sum_{j\in B_t}(\exp v'_{jt}\alpha)]$$

while the (conditional) log-likelihood of $(S_1,\ldots,S_n)$ given $(Z_1,\ldots,Z_n)$ is:

$$L_n^c(S|Z;\alpha) = \sum_{i=1}^{n} S_t \log p_t + (1-S_t) \log (1-p_t)$$

where

$$p_t = e^{b_t} / (e^{b_t} + e^{b_t^c})$$

and $b_t = \log(\sum_{j \in B} \exp v'_{jt}\alpha)$, $b_t^c = \log(\sum_{j \in B^c} \exp v'_{jt}\alpha)$.

From Section 3, it follows that the maximization of $L_n^c(Y|S;\alpha)$ with respect to the identified parameters α_1 of α gives consistent estimates $\tilde{\alpha}_{1n}$ of these parameters.[11] It is noteworthy that this conditional log-likelihood function can be obtained by acting as if the individuals who chose an alternative in the restricted choice set B (or B^c) had only B (or B^c respectively) as their initial choice set. Thus $\tilde{\alpha}_{1n}$ can readily be obtained from MNL package (e.g., QUAIL) that allows different choice sets for the individuals.

Moreover, as indicated in Section 5 we can construct specification tests for the MNL model based on the indicators $(\tilde{\alpha}_{1n} - \hat{\alpha}_{1n})$ or $\partial L_n^c(Y|S,Z;\hat{\alpha}_{1n})/\partial\alpha_1$, where $\hat{\alpha}_{1n} = \alpha_1(\hat{\alpha}_n)$ and $\hat{\alpha}_n$ is the ML estimator on the complete choice set. Though our test based on $(\tilde{\alpha}_{1n} - \hat{\alpha}_{1n})$ is similar in spirit to the specification test for the MNL model proposed by Hausman and McFadden (1981), it differs from it for the reason that the statistic used by these authors is $(\alpha_{1n}^B - \hat{\alpha}_{1n})$ where α_{1n}^B is obtained by maximizing:

$$L_n(\alpha) = \sum_{\{t;S_t=1\}} \sum_{j \in B} Y_{it}[v'_{it}\alpha - \log \sum_{j \in B} (\exp v'_{jt}\alpha)].$$

Hence $L_n(\alpha)$ neglects the information on the individuals who have chosen an alternative outside the restricted choice set B.[12] Since our tests use all the information, they are likely to be more powerful than the Hausman-McFadden specification test.

EXAMPLE 3: Suppose that one considers the simple Tobit model (Tobin (1958), Amemiya (1973)) for the random sample (Y_t, Z_t), t=1,...,n, i.e.:

$$\begin{aligned} Y_t &= Z_t'\beta + u_t \quad && \text{if } Z_t'\beta + u_t > 0 , \\ &= 0 && \text{otherwise} , \end{aligned}$$

where the u_t's are $N(0, \sigma^2)$ and independent given the Z_t's.

Let $S_t = 1$ if $Y_t > 0$, and 0 otherwise. The (conditional) likelihood function of $(Y_1, S_1, \ldots, Y_n, S_n)$ given $(Z_1, \ldots, Z_n)$ can be written as:

$$\begin{aligned} L_n^c(Y,S \mid Z;\beta,\sigma) = &\prod_{t=1}^{n} \left[1 - \Phi(Z_t'\gamma)\right]^{1-S_t} \left[\Phi(Z_t'\gamma)\right]^{S_t} \\ &\times \prod_{t=1}^{n} \left[\phi(Y_t/\sigma - Z_t'\gamma)/\sigma\Phi(Z_t'\gamma)\right]^{S_t}, \end{aligned}$$

where $\gamma = \beta/\sigma$, and $\phi(.)$ and $\Phi(.)$ are respectively the density and c.d.f. of the standard Normal.

Since the model for S_t given Z_t is a dichotomous Probit model, it follows that the first product in L_n^c is the conditional likelihood function associated with $(S_1,\dots,S_n)$ given $(Z_1,\dots,Z_n)$. Hence, the second product in L_n^c is simply the conditional likelihood function associated with $(Y_1,\dots,Y_n)$ given $(S_1,\dots,S_n,Z_1,\dots,Z_n)$. It is worth noting that this latter likelihood function is the one associated with a random sample drawn from a truncated normal distribution (for the distinction between censored and truncated, see Maddala (1983)).[13] Hence maximizing this second product with respect to σ and γ gives estimates $\tilde{\sigma}$ and $\tilde{\gamma}$ that have the properties of CMLE's.

Thus from Section 5 specification tests for the Tobit model can be constructed from e.g., the indicator $(\tilde{\gamma}-\hat{\gamma},\ \tilde{\sigma}-\hat{\sigma})$ where $(\tilde{\sigma},\tilde{\gamma})$ and $(\hat{\sigma},\hat{\gamma})$ are respectively the ML estimators associated with the truncated Tobit model and the simple Tobit model.

Incidently, let us note that the CMLE's $(\tilde{\sigma},\tilde{\gamma})$ must satisfy the normal equation associated with the partial derivative with respect to σ, i.e.:

$$\tilde{\sigma}^2 N_1 + \tilde{\sigma}\, Y_1'Z_1\tilde{\gamma} - Y_1'Y_1 = 0$$

where N_1 is the number of observations such that $Y_t > 0$, Y_1 is the $N_1 \times 1$ vector of such observations on Y, and Z_1 is the corresponding matrix of observations on the explanatory variables Z. Solving for $\tilde{\sigma}$, the positive root of this normal equation, gives:

$$\tilde{\sigma} = \left[\frac{Y_1'Y_1}{N_1} + \frac{1}{4}\left(\frac{Y_1'Z_1\tilde{\gamma}}{N_1}\right)^2\right]^{1/2} - \frac{1}{2}\,\frac{Y_1'Z_1\tilde{\gamma}}{N_1}.$$

From Section 2, $\tilde{\sigma}$ and $\tilde{\gamma}$ are strongly consistent estimators of σ^o and γ^o (under correct specification of the conditional model for Y_t given (S_t,Z_t)). If one has available another strongly consistent estimator of γ^o such as the ML estimator γ^P obtained by estimating the Probit model for S_t given Z_t (the first stage in Heckman's (1976) procedure), then it follows that the estimator σ^P obtained from the previous equation where $\tilde{\gamma}$ is replaced by γ^P is a strongly consistent estimator of σ^o. Then, one readily obtains strongly consistent estimates of β^o by using $\sigma^P\gamma^P$. This procedure has the following advantages: (i) it ensures that the estimate σ^P is always strictly positive, (ii) it is easy to carry out since it requires only one estimation, and (iii) it does not require the expensive computation of $\phi(Z_t'\gamma^P)$ and $\Phi(Z_t'\gamma^P)$ as in Heckman's second stage.

EXAMPLE 4: As in Andersen (1970), conditional ML estimation is particularly useful when there exist sufficient statistics for some parameters.[14] Specifically, suppose that $Y_1,\dots,Y_n$ are independent with a common distribution that is assumed to belong to the family $\{F_Y(.;\alpha);\ \alpha \in A\}$. Suppose that there exist (i) a partition (α_1,α_2) of the parameter vector α such that $A = A_1 \times A_2$, and (ii) a sufficient statistic $S_t = S(Y_t)$ for α_2, for any α_1 in A_1.[15]. Let $h(.,.;\alpha)$ be the joint density of (Y_t,S_t). Then we have:

$$h(Y_t,S_t;\alpha) = f(Y_t|S_t;\alpha_1)\ .\ g(S_t;\alpha_1,\alpha_2)$$

where $f(.|.;\alpha_1)$ and $g(.;\alpha_1,\alpha_2)$ are respectively the conditional density of Y_t given S_t and the marginal density of S_t. It follows that by maximizing the conditional log-likelihood function for $(Y_1,\ldots,Y_n)$ given $(S_1,\ldots,S_n)$, one obtains an estimator $\hat{\alpha}_1$ that has the properties mentioned in Section 3. In particular, these properties are robust with respect to misspecification of the marginal model for S_t.

An example of such a situation is the estimation of a multivariate logit model (see Nerlove and Press (1973), Amemiya (1978)).[16] Let Y_{1t} and Y_{2t} be qualitative variables with I and J categories respectively. For any $t = 1,\ldots,n$, $i = 1,\ldots,I$, and $j = 1,\ldots,J$, one assumes that:

$$\log \Pr(Y_{1t} = i,\ Y_{2t} = j) = \mu_t + v'_{ijt}\alpha$$

where $\mu_t = -\log\left[\sum_{i=1}^{I}\sum_{j=1}^{J}\exp(v'_{ijt}\alpha)\right]$ and v_{ijt} is a vector of explanatory variables. Let us partition the vector v_{ijt} into explanatory variables z_{ijt} that vary across i, and explanatory variables z_{jt} that do not. Let α be partitioned accordingly into α_1 and α_2. Then, the conditional probability that $Y_{1t} = i$ given that $Y_{2t} = j$ satisfies:

$$\log \Pr(Y_{1t} = i|Y_{2t} = j) = \mu_{jt} + z'_{ijt}\alpha_1$$

where $\mu_t = -\log\left[\sum_{i=1}^{I}\exp(z'_{ijt}\alpha_1)\right]$. Hence Y_{2t} is a sufficient statistic for α_2, and maximizing the conditional log-likelihood function for $(Y_{11},\ldots,Y_{1n})$ given $(Y_{21},\ldots,Y_{2n})$ gives a CMLE $\hat{\alpha}_{1n}$. This

estimator is not in general efficient since the marginal probability model for Y_{2t} depends in general on α_1.[17]

As in the previous examples, one can construct specification tests for the bivariate logit model based on $(\tilde{\alpha}_{1n} - \hat{\alpha}_{1n})$ or $\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial\alpha_1$ where $\tilde{\alpha}_{1n}$ is the CMLE for the conditional model for Y_1 given Y_2, and $\hat{\alpha}_{1n}$ is the ML estimator for the bivariate logit model.

Finally let us note that one case has not been covered: the case in which the variable Y_{2t} is, in addition to being sufficient for α_2, ancillary for α_1, i.e. such that the marginal density $f_2(.;\alpha_1,\alpha_2)$ is independent of α_1 (see Fisher (1956)). From Part (c) of Theorem 2 and the fact that in this case the information matrix becomes block diagonal with first block equal to $B_o^{f_1}(\alpha_1)$, it follows that CML estimation of α_1 is efficient, given of course correct specification of the conditional model for Y_{1t} given Y_{2t}. This is not surprising since Y_{2t} is then weakly exogenous for α_1 (see Engle, Hendry, and Richard (1983)).

7. Conclusion

In this paper we derived the asymptotic properties of CMLE's under correct or incorrect specification of the conditional model. CMLE's were found to be robust with respect to misspecification of the model for the conditioning variables. Efficiency of CMLE's as well as tests for misspecification of the conditional model were also discussed. It was argued that CML estimation provides a convenient way to test for parameter estimator inconsistency. Some examples were given to illustrate the range of application of the CMLE technique. Our results should prove to be useful to social scientists who conduct estimation and inferences conditional upon the observed values of some explanatory variables.

APPENDIX

PROOF OF LEMMA 1: Obvious.

PROOF OF THEOREM 1: The theorem follows from Lemma 1 and White (1982) Theorems 2.1, 2.2, and 3.2. Indeed it suffices to choose an arbitrary distribution for Z_t with a strictly positive density. It is then easy to check that Assumptions A1-A5 imply White's Assumptions A1-A6 on the resulting family of joint distribution H^G.

Q.E.D.

PROOF OF LEMMA 2: Let

$$w(z;\alpha) = \int \log f(y|z;\alpha)dF^o_{Y|Z}(y|z)$$

where the right-hand side exists by Assumptions A1-A3 for H^o-almost all z. Since $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;\alpha^o)$ by assumption, it follows from Jensen's inequality (see, e.g., Rao (1973, p. 58)) that for H^o-almost all z:

$$w(z;\alpha^o) \geq w(z;\alpha) \quad \text{for all } \alpha \text{ in A.}$$

Since $z^f(\alpha) = \int w(z;\alpha)dG^o(z)$ where $G^o(.)$ is the true distribution ot Z_t, it follows by integration that:

$$z^f(\alpha^o) \geq z^f(\alpha) \quad \text{for all } \alpha \text{ in A.}$$

From the uniqueness of α^* (Assumption A3-b), it follows that $\alpha^* = \alpha^o$.

Q.E.D.

PROOF OF LEMMA 3: We have:

$$\frac{\partial^2 \log f(y|z;\alpha)}{\partial\alpha\partial\alpha'} = \frac{\partial}{\partial\alpha'}\left[\frac{\partial f(y|z;\alpha)}{\partial\alpha} \cdot \frac{1}{f(y|z;\alpha)}\right]$$
$$= \frac{1}{f(y|z;\alpha)} \frac{\partial^2 f(y|z;\alpha)}{\partial\alpha\partial\alpha'} - \frac{\partial \log f(y|z;\alpha)}{\partial\alpha} \cdot \frac{\partial \log f(y|z;\alpha)}{\partial\alpha'}.$$

Taking expectations of both sides evaluated at α^o with respect to the true c.d.f. $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;\alpha^o)$, it follows from Lemma 2 and Assumption A6 that:

$$\int \frac{\partial^2 \log f(y|z;\alpha^o)}{\partial\alpha\partial\alpha'} f(y|z;\alpha^o)dy$$
$$= - \int \frac{\partial \log f(y|z;\alpha^o)}{\partial\alpha} \cdot \frac{\partial \log f(y|z;\alpha^o)}{\partial\alpha'} f(y|z;\alpha^o)dy$$

where both sides exist for H^o-almost all z because of Assumption A.4.

Q.E.D.

PROOF OF THEOREM 2: Straightforward from Theorem 1, Lemma 1, and Equation (3.3).

To prove Theorem 3, we use the following lemma.

LEMMA A1: Given Assumptions A1-A5, A2'-A5', and A8:

$$\begin{bmatrix} \frac{1}{\sqrt{n}} \frac{\partial L^c_n(Y|Z;\alpha*)}{\partial\alpha} \\ \frac{1}{\sqrt{n}} \frac{\partial L^c_n(Y_1|Y_2,Z;\alpha^*_1)}{\partial\alpha_1} \end{bmatrix} \overset{D}{\rightarrow} N(0, \begin{bmatrix} B^f_o(\alpha*) & B^{ff_1}_o(\alpha*,\alpha^*_1) \\ B^{f_1f}_o(\alpha*,\alpha^*_1) & B^{f_1}_o(\alpha^*_1) \end{bmatrix}).$$

Proof: The result follows from the multivariate version of the standard Central Limit Theorem applied to:

$$\frac{1}{\sqrt{n}}\begin{bmatrix} \frac{\partial L^c_n(Y|Z;\alpha*)}{\partial\alpha} \\ \frac{\partial L^c_n(Y_1|Y_2,Z;\alpha^*_1)}{\partial\alpha_1} \end{bmatrix} = \sqrt{n}\begin{bmatrix} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \log f(Y_t|Z_t;\alpha*)}{\partial\alpha} \\ \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \log f_1(Y_{1t}|Y_{2t},Z_t;\alpha^*_1)}{\partial\alpha_1} \end{bmatrix}.$$

Indeed, from Assumptions A3-A5, we have:

$$E^o\left[\frac{\partial \log f(Y_t|Z_t;\alpha*)}{\partial\alpha}\right] = \frac{\partial}{\partial\alpha} E^o\left[\log f(Y_t|Z_t;\alpha*)\right] = 0.$$

Thus, from Equation (2.6) and Assumption A5:

$$\text{var}^o\left[\frac{\partial \log f(Y_t|Z_t;\alpha*)}{\partial\alpha}\right] = B^f_o(\alpha*) < \infty.$$

Similarly, from Assumptions A3'-A5', we have:

$$E^o\left[\frac{\partial \log f(Y_{1t}|Y_{2t};Z_t;\alpha^*_1)}{\partial\alpha_1}\right] = 0,$$
$$\text{var}^o\left[\partial \log f(Y_{1t}|Y_{2t};Z_t;\alpha^*_1)\right] = B^{f_1}_o(\alpha^*_1) < \infty.$$

Moreover, from Equation (4.4) and Assumption A8, we have:

$$\text{cov}^o\left[\frac{\partial \log f(Y_t|Z_t;\alpha*)}{\partial\alpha}, \frac{\partial \log f_1(Y_{1t}|Y_{2t},Z_t;\alpha^*_1)}{\partial\alpha'_1}\right] = B^{ff_1}_o(\alpha*,\alpha^*_1) < \infty.$$

Q.E.D.

PROOF OF THEOREM 3: Part (a) simply follows from Theorem 1-a. Part (b) follows from Theorem 1-b and Assumption A7-a. To prove Part (d), we note that from Assumptions A4, A4', A8, and Jennerich's uniform strong Law of Large Numbers (1969, Theorem 2, p. 636) we have uniformly on A X A_1:

$$A_n^f(\alpha) \overset{a.s.}{\rightarrow} A_o^f(\alpha) \quad ; \quad A_n^{f_1}(\alpha_1) \overset{a.s.}{\rightarrow} A_o^{f_1}(\alpha_1),$$
$$B_n^f(\alpha) \overset{a.s.}{\rightarrow} B_o^f(\alpha) \quad ; \quad B_n^{f_1}(\alpha_1) \overset{a.s.}{\rightarrow} B_o^{f_1}(\alpha_1),$$
$$B_n^{ff_1}(\alpha,\alpha_1) \overset{a.s.}{\rightarrow} B_o^{ff_1}(\alpha,\alpha_1).$$

Given Equation (4.6), Assumptions A5-b, A5'-b, and A7, it follows from the uniform convergence that:

$$C_n(\bar{\alpha}_n,\bar{\alpha}_{1n}) \overset{a.s.}{\rightarrow} C(\alpha^*,\alpha_1^*)$$

for any estimator $(\bar{\alpha}_n,\bar{\alpha}_{1n})$ that converges almost surely to (α^*,α_1^*). Part (d) follows.

Moreover, by taking a Taylor expansion around α^* and α_1^* respectively, and using the definitions of $\hat{\alpha}_n$ and $\tilde{\alpha}_{1n}$ it follows that:

$$0 = \frac{1}{\sqrt{n}} \frac{L_n^c(Y|Z;\alpha^*)}{\partial\alpha} + A_o^f(\alpha^*)\sqrt{n}(\hat{\alpha}_n - \alpha^*) + o_p(1),$$

$$0 = \frac{1}{\sqrt{n}} \frac{\partial L_n^c(Y_1|Y_2,Z;\alpha_1^*)}{\partial\alpha_1} + A_o^{f_1}(\alpha_1^*)\sqrt{n}(\hat{\alpha}_n - \alpha_1^*) + o_p(1).$$

Thus, given Assumption A5 and A5':

$$\sqrt{n}\begin{bmatrix} \hat{\alpha}_n - \alpha^* \\ \\ \tilde{\alpha}_{1n} - \alpha_1^* \end{bmatrix} = - \begin{bmatrix} [A_o^f(\alpha^*)]^{-1} & 0 \\ \\ 0 & [A_o^{f_1}(\alpha_1^*)]^{-1} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{n}} \frac{\partial L_n^c(Y|Z;\alpha^*)}{\partial\alpha} \\ \\ \frac{1}{\sqrt{n}} \frac{\partial L_n^c(Y_1|Y_2,Z;\alpha_1^*)}{\partial\alpha_1} \end{bmatrix} + o_p(1)$$

Using Lemma 1, we obtain:

$$\sqrt{n}\begin{bmatrix} \hat{\alpha}_n - \alpha^* \\ \\ \tilde{\alpha}_{1n} - \alpha_1^* \end{bmatrix} \overset{D}{\rightarrow} N(0,\textstyle\sum)$$

where

$$\textstyle\sum = \begin{bmatrix} [A_o^f(\alpha^*)]^{-1}B_o^f(\alpha^*)[A_o^f(\alpha^*)]^{-1} & ; & [A_o^f(\alpha^*)]^{-1}B_o^{ff_1}(\alpha^*,\alpha_1^*)[A_o^f(\alpha_1^*)]^{-1} \\ [A_o^{f_1}(\alpha_1^*)]^{-1}B_o^{ff_1}(\alpha^*,\alpha_1^*)[A_o^f(\alpha^*)]^{-1} & ; & [A_o^{f_2}(\alpha_1^*)]^{-1}B_o^{f_1}(\alpha_1^*)[A_o^{f_1}(\alpha_1^*)]^{-1} \end{bmatrix}$$

On the other hand, we have $(\hat{\alpha}_{1n},\tilde{\alpha}_{1n}) = (\alpha_1(\hat{\alpha}_n),\tilde{\alpha}_{1n}) = h(\hat{\alpha}_n,\tilde{\alpha}_{1n})$ so that from Assumption A7, the Jacobian of the transformation h(.,.) at (α^*,α_1^*) is:

$$\tilde{J}(\alpha^*) = \begin{bmatrix} J(\alpha^*) & 0 \\ 0 & I_{k_1} \end{bmatrix}$$

From a well-known property of convergence in distribution, it follows that:

$$\sqrt{n}\begin{bmatrix} \hat{\alpha}_{1n} - \alpha_1(\alpha^*) \\ \\ \tilde{\alpha}_{1n} - \alpha_1^* \end{bmatrix} \overset{D}{\rightarrow} N(0, \tilde{J}(\alpha^*) \textstyle\sum \tilde{J}'(\alpha^*)).$$

Part (c) straightforwardly follows.

Q.E.D.

To prove Lemma 4, we use the following lemma.

LEMMA A2: Given Assumptions A1, A2'-A5', if $F^o_{Y_1|Y_2Z}(.|.,.) = F_{Y_1|Y_2Z}(.|.,.;\alpha^o_1)$ then for H^o-almost all (y_2,z):

$$E^o_{Y_1|Y_2Z}\left[\frac{\partial \log f_1(Y_{1t}|y_2,z;\alpha^o_1)}{\partial\alpha_1}\right] = 0$$

where $E_{Y_1|Y_2Z}$ is the expectation with respect to the true conditional distribution of Y_{1t} given $Y_2 = y_2$ and $Z = z$.

Proof: From Jensen's Inequality, we have for all α_1 in A_1:

$$\int \log f_1(y_1|y_2,z;\alpha^o_1) f_1(y_1|y_2,z;\alpha^o_1)dy_1 \geq \int \log f_1(y_1|y_2,z;\alpha_1) f_1(y_1|y_2,z;\alpha^o_1)dy_1$$

Since $\alpha^o_1 = \alpha^*_1$ (Lemma 2), and since α^*_1 belongs to the interior of A_1 (Assumption A5), it follows that, at $\alpha_1 = \alpha^o_1$:

$$\frac{\partial}{\partial\alpha_1}\int \log f_1(y_1|y_2,z;\alpha_1)\ f_1(y_1|y_2,z;\alpha^o_1)dy_1 = 0 \qquad (*)$$

when the left-hand side exists.

On the other hand, $\partial\log f_1(y_1|y_2,z;\alpha_1)/\partial\alpha_1$ is, from Assumption A5'-c, dominated by a function $M(y_1,y_2,z)$ which is H^o-integrable. But

$$\int M(y_1,y_2,z)dH^o(y_1,y_2,z) = \int\left[\int M(y_1,y_2,z)dF^o_{Y_1|Y_2Z}(y_1|y_2,z)\right]dH^o(y_2,z)$$

Hence, for H^o-almost all (y_2,z), $M(y_1,y_2,z)$ is integrable with respect to $F^o_{Y_1|Y_2Z}(.|.,.) = F_{Y_1|Y_2Z}(.|.,.;\alpha^o_1)$. It follows that we can reverse the order of the derivation sign and the integration sign in (*). The result now follows from the definition of $E^o_{Y_1|Y_2Z}$.

Q.E.D.

It is worth noting, by taking the total expectation of the above equation with respect to the true distribution of (Y_{2t},Z_t) that this equation implies (but is not implied by):

$$E^o\left[\frac{\partial\log f_1(Y_{1t}|Y_{2t},Z_t;\alpha^o_1)}{\partial\alpha_1}\right] = 0$$

which is a standard property of α^o_2.

PROOF OF LEMMA 4: Since

$$\frac{\partial\log f(y_1,y_2|z;\alpha)}{\partial\alpha} = J'(\alpha)\ \frac{\partial\log f_1(y_1|y_2,z;\alpha_1)}{\partial\alpha_1} + \frac{\partial\log f_2(y_2|z;\alpha)}{\partial\alpha}$$

it follows from the definitions of $B^f_o(\alpha)$ and $B^{f_2}_o(\alpha_1)$ that:

$$\begin{aligned} B^f_o(\alpha) = {} & J'(\alpha)B^{f_1}_o(\alpha_1)J(\alpha) + B^{f_2}_o(\alpha) \\ & + J'(\alpha)E^o\left[\frac{\partial\log f_1(Y_{1t}|Y_{2t},Z_t;\alpha_1)}{\partial\alpha_1}\ .\ \frac{\partial\log f_2(Y_{2t}|Z_t;\alpha)}{\partial\alpha'}\right] \\ & + E^o\left[\frac{\partial\log f_2(Y_{2t}|Z_t;\alpha)}{\partial\alpha}\ .\ \frac{\partial\log f_1(Y_{1t}|Y_{2t},Z_t;\alpha_1)}{\partial\alpha_1'}\right]J(\alpha). \end{aligned}$$

Since $\alpha^o_1 = \alpha_1(\alpha^o)$ when $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;\alpha^o)$, it suffices to show that the last two matrices on the right hand side of the previous equation are identically null when evaluated at (α^o,α^o_1). This property follows by noting that:

$$E^o\left[\frac{\partial \log f_1(Y_{1t}|Y_{2t},Z_t;a_1^o)}{\partial a_1}\cdot\frac{\partial \log f_2(Y_{2t}|Z_t;a^o)}{\partial a'}\right]$$

$$= E^o_{Y_2Z}\left[E^o_{Y_1|Y_2,Z}\left[\frac{\partial \log f_1(Y_{1t}|y_2,z;a_1^o)}{\partial a_1}\right]\cdot\frac{\partial \log f_2(Y_{2t}|Z_t;a^o)}{\partial a'}\right]$$

where $E^o_{Y_2Z}$ denotes the expectation with respect to the true distribution of (Y_{2t},Z_t). The desired property then follows from Lemma A2.

Finally, since $B_o^f(a^o)$ and $B_o^{f_1}(a_1^o)$ are finite (from Assumptions A5 and A5'), then $B_o^{f_2}(a)$ is also finite.

Q.E.D.

PROOF OF THEOREM 4: To show that $\hat{a}_{1n}$ is at least as efficient as $\tilde{a}_{1n}$, we can use some general properties on FIML estimates (see e.g., Rao (1963)). Indeed $\hat{a}_{1n}$ and $\tilde{a}_{1n}$ are in fact jointly consistent and uniformly asymptotically normal (JCUAN) estimates of a_1^o. Moreover, by picking an arbitrary but fixed distribution for Z_t, it is easy to see that $\hat{a}_n$ is then the FIML estimator so that $\hat{a}_{1n}$ is an asymptotically efficient estimator of a_1^o.

Parts (i) and (ii) of Theorem 4 requires, however, a direct proof. When $F^o_{Y|Z}(.|.) = F_{Y|Z}(.|.;a^o)$, it follows from Theorem 3, Assumptions A6, A6', and Lemmas 2 and 3 that:

$$\text{Asym.Var}^o(\hat{a}_{1n}) = C_{11}(a^o) = J(a^o)\left[B_o^f(a^o)\right]^{-1} J'(a^o),$$

$$\text{Asym.Var}^o(\tilde{a}_{1n}) = C_{22}(a_1^o) = \left[B_o^{f_1}(a_1^o)\right]^{-1}.$$

We want to show that:

$$\left[B_o^{f_1}\right]^{-1} \geq J\left[B_o^f\right]^{-1} J'$$

where we have dropped a^o and a_1^o to simplify the notations. From Lemma 4, this is equivalent to showing that:

$$\left[B_o^{f_1}\right]^{-1} = J\left[J'B_o^{f_1} J + B_o^{f_2}\right]^{-1} J'$$

which is, from Equation (4.8), equivalent to:

$$\left[M'B_o^{f_1}M\right]^{-1} \geq L\left[L'M'B_o^{f_1}ML + NB_o^{f_2}N'\right]^{-1} L' \qquad (*)$$

after having used the non-singularity of M and the orthogonality of N.

Suppose first that $k_1 = k$. Then, from Equation (4.9), L is the identity matrix. Hence (*) is equivalent to:

$$M'B_o^{f_1}M \leq M'B_o^{f_1}M + NB_o^{f_2}N'$$

which is true since $B_o^{f_2}$ and hence $NB_o^{f_2}N'$ are positive semi-definite. Moreover the equality holds if and only if $NB_o^{f_2}N' = 0$.

Suppose now that $k_1 < k$. From Equations (4.9) and (4.10) it follows that (*) is equivalent to:

$$\left[M'B_o^{f_1}M\right]^{-1} \geq \left[I_{k_1}\ 0\right]\begin{bmatrix} M'B_o^{f_1}M + Z_{11} & ; & Z_{12} \\ Z_{21} & ; & Z_{22}\end{bmatrix}^{-1}\begin{bmatrix} I_{k_1} \\ 0\end{bmatrix}$$

Note that the square matrices $M'B_o^{f_1}M + Z_{11}$ and Z_{22} must be non-singular since B_o^f is non- singular (from Assumptions A5, A6, and Lemma 3). Using the formula for the inverse of a partitioned matrix, we obtain:

$$\left[M'B_o^{f_1}M\right]^{-1} \geq \left[M'B_o^{f_1}M + Z_{11} - Z_{12}\, Z_{22}^{-1}\, Z_{21}\right]^{-1}$$

or equivalently, after simplification:

$$Z_{11} - Z_{12}Z_{22}^{-1}Z_{21} \geq 0$$

which is true since the left-hand side is equal to

$$\left[I_{k_1} ; -Z_{12}Z_{22}^{-1}\right] NB_o^{f_2}N' \begin{bmatrix} I_{k_1} \\ -Z_{22}^{-1}Z_{21}\end{bmatrix}$$

where $B_o^{f_2}$ is positive semi-definite.

Q.E.D.

PROOF OF THEOREM 5: From Theorem 3, it follows that:

$$\sqrt{n}[\tilde{\alpha}_{1n} - \hat{\alpha}_{1n} - (\alpha_1^* - \alpha_1(\alpha^*))] \xrightarrow{D} N(0,V)$$

where V is defined by (5.2). Since by assumption V ≠ 0, it follows from Rao and Mitra (1971, Theorem 9.2.2) that:

$$n[\tilde{\alpha}_{1n} - \hat{\alpha}_{1n} - (\alpha_1^* - \alpha_1(\alpha^*))]'V^-[\tilde{\alpha}_{1n} - \hat{\alpha}_{1n} - (\alpha_1^* - \alpha_1(\alpha^*))] \xrightarrow{D} \chi_r^2$$

where V^- is any generalized inverse of V, and r = rank V. Thus under the null hypothesis H_o:

$$n[\tilde{\alpha}_{1n} - \hat{\alpha}_{1n}]'\ V^-[\tilde{\alpha}_{1n} - \hat{\alpha}_{1n}] \xrightarrow{D} \chi_r^2.$$

Moreover, from part (c) of Theorem 3, V_n converges almost surely to V. Thus, by construction, $[V_n + R_nR_n']^{-1}$ converges almost surely to $[V + RR']^{-1}$ which is a generalized inverse of V. Therefore, under H_o, H_n converges in distribution to a chi-square with r degrees of freedom.

Under H_1, it follows from part (b) of Theorem 3 that:

$$\tilde{\alpha}_{1n} - \hat{\alpha}_{1n} \xrightarrow{a.s.} \alpha_1^* - \alpha_1(\alpha^*) = a \neq 0.$$

Since $[V_n + R_nR_n']^{-1}$ converges almost surely to $[V + RR']^{-1}$, we have:

$$(\tilde{\alpha}_{1n} - \hat{\alpha}_{1n})'[V_n + R_nR_n']^{-1}(\tilde{\alpha}_{1n} - \hat{\alpha}_{1n}) \xrightarrow{a.s.} a'[V + RR']^{-1}a.$$

Since, by construction, V + RR' is non-singular, it follows that V + RR' and hence $[V + RR']^{-1}$ are positive definite. Thus $a'[V + RR']^{-1}a \neq 0$ for any a ≠ 0. Hence under H_1, H_n converges almost surely to ∞.

Q.E.D.

PROOF OF LEMMA 5: Given condition (a), it follows from Equations (5.2) and (4.6) that:

$$
\begin{aligned}
V = {} & [B_o^{f_1}(\alpha_1^*)]^{-1} + J(\alpha^*)B_o^f(\alpha^*)J'(\alpha^*) \\
& - J(\alpha^*)[B_o^f(\alpha^*)]^{-1}B_o^{ff_1}(\alpha^*,\alpha_1^*)[B_o^{f_1}(\alpha_1^o)]^{-1} \\
& - [B_o^{f_1}(\alpha_1^*)]^{-1}B_o^{f_1 f}(\alpha*,\alpha_1^*)[B_o^f(\alpha^*)]^{-1}J'(\alpha^*).
\end{aligned}
$$

On the other hand, we have:

$$
\frac{\partial \log f(y_1,y_2|z;\alpha^*)}{\partial \alpha} = J'(\alpha^*)\frac{\partial \log f_1(y_1|y_2,z;\alpha_1(\alpha^*))}{\partial \alpha_1} + \frac{\partial \log f_2(y_2|z,\alpha^*)}{\partial \alpha}.
$$

Thus from Equation (4.4) and the definition of $B_o^{f_1}(.)$, we have under H_o:

$$
\begin{aligned}
B_o^{ff_1}(\alpha^*,\alpha_1^*) = {} & J'(\alpha^*)B_o^{f_1}(\alpha_1^*) \\
& + E^o\left[\frac{\partial \log f_2(Y_{2t}|Z_t;\alpha^*)}{\partial \alpha} \cdot \frac{\partial \log f_1(Y_{1t}|Y_{2t},Z_t;\alpha_1^*)}{\partial \alpha_1'}\right]
\end{aligned}
$$

But the second term is null since it is equal to:

$$
E^o_{Y_2Z}[\frac{\partial \log f_2(Y_{2t}|Z_t;\alpha^*)}{\partial \alpha} \cdot E^o_{Y_1|Y_2Z}[\frac{\partial \log f_1(Y_{1t}|y_2,z;\alpha_1^*)}{\partial \alpha_1'}]]
$$

where the conditional expectation is null because of condition (b). Therefore:

$$
V = [B_o^f(\alpha_1^*)]^{-1} - J(\alpha^*)[B_o^f(\alpha^*)]^{-1}J'(\alpha^*).
$$

Q.E.D.

To prove Theorem 5, we use the following Lemma which is similar to Lemma A2.

LEMMA A3: Given Assumptions A1-A5, A2'-A4', and A8:

$$
\begin{bmatrix}
\frac{1}{\sqrt{n}} & \frac{\partial L_n^c(Y|Z;\alpha^*)}{\partial \alpha} \\
\frac{1}{\sqrt{n}} & \frac{\partial L_n^c(Y_1|Y_2,Z;\alpha_1(\alpha^*))}{\partial \alpha_1}
\end{bmatrix}
\overset{D}{\rightarrow} N(
\begin{bmatrix}
0 \\
\frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1}
\end{bmatrix}, W)
$$

where

$$
W = \begin{bmatrix}
B_o^f(\alpha^*) & ; & B_o^{ff_1}(\alpha^*,\alpha_1(\alpha^*)) \\
B_o^{f_1 f}(\alpha^*,\alpha_1(\alpha^*)) & ; & B_o^{f_1}(\alpha_1(\alpha^*)) - \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1} \cdot \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1'}
\end{bmatrix}.
$$

Proof: The proof is similar to the proof of Lemma A2, and is based on the multivariate version of the standard Central Limit Theorem. The only difference is that:

$$
E^o\left[\frac{\partial \log f(Y_{1t}|Y_{2t},Z_t;\ \alpha^*))}{\partial \alpha_1}\right] = \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1}
$$

which may not be zero so that:

$$
\mathrm{var}^o\left[\frac{\partial \log f(Y_{1t}|Y_{2t},Z_t;\alpha_1(\alpha^*))}{\partial \alpha_1}\right] = B_o^{f_1}(\alpha_1(\alpha^*)) - \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1} \cdot \frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1'}
$$

Q.E.D.

PROOF OF THEOREM 6: Since $\sqrt{n}(\hat{\alpha}_{1n} - \alpha_1(\alpha^*)) = O_p(1)$, we obtain from a Taylor expansion around $\alpha_1(\alpha^*)$:

$$\frac{1}{\sqrt{n}}\frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial \alpha_1} = \frac{1}{\sqrt{n}}\frac{\partial L_n^c(Y_1|Y_2,Z;\alpha_1(\alpha*))}{\partial \alpha_1} + \frac{1}{n}\frac{\partial^2 L_n^c(Y_1|Y_2,Z;\alpha_1(\alpha*))}{\partial \alpha_1 \partial \alpha_1'}\sqrt{n}(\hat{\alpha}_{1n} - \alpha_1(\alpha^*)) + o_p(1).$$

On the other hand, since $\sqrt{n}(\hat{\alpha}_n - \alpha^*) = O_p(1)$, we have from a Taylor expansion of $\alpha_1 = \alpha_1(\alpha)$ around α^*:

$$\sqrt{n}\hat{\alpha}_{1n} = \sqrt{n}\alpha_1(\alpha^*) + J(\alpha^*)\sqrt{n}(\hat{\alpha}_n - \alpha^*) + o_p(1).$$

Moreover, from the proof of Theorem 3, we have:

$$\sqrt{n}(\hat{\alpha}_n - \alpha^*) = - [A_o^f(\alpha^*)]^{-1}(1/\sqrt{n})\partial L_n^c(Y|Z;\alpha^*)/\partial \alpha + o_p(1),$$

and

$$(1/n)\partial^2 L_n^c(Y_1|Y_2,Z;\alpha_1(\alpha^*))/\partial \alpha_1 \partial \alpha_1' = A_o^{f_1}(\alpha_1(\alpha^*)) + o_p(1).$$

Collecting these results, the first equation becomes:

$$\frac{1}{\sqrt{n}}\frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial \alpha_1} = \frac{1}{\sqrt{n}}\frac{\partial L_n^c(Y_1|Y_2,Z;\alpha_1(\alpha^*))}{\partial \alpha_1} - A_o^{f_1}(\alpha_1(\alpha^*))J(\alpha^*)[A_o^f(\alpha^*)]^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_n^c(Y|Z;\alpha^*)}{\partial \alpha} + o_p(1)$$

From Lemma A3, it follows that:

$$\frac{1}{\sqrt{n}}\frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial \alpha_1} \xrightarrow{D} N(\frac{\partial z^{f_1}(\alpha_1(\alpha^*))}{\partial \alpha_1},\Omega)$$

where Ω is given by Equation (5.6) using the definition (4.6) of $C_{11}(\alpha^*)$.

Under H_o', $\partial z^{f_1}(\alpha_1(\alpha^*))/\partial \alpha_1 = 0$. Thus:

$$\frac{1}{\sqrt{n}}\frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial \alpha_1'}\Omega^-\frac{\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})}{\partial \alpha_1'} \xrightarrow{D} \chi^2_s$$

for any choice of the generalized inverse Ω^- of Ω when $\Omega \neq 0$ (see Rao and Mitra (1971, Theorem 9.2.2)). Since by construction $[\Omega_n + Q_nQ_n']^{-1}$ converges almost surely to $[\Omega + QQ']^{-1}$ which is a generalized inverse of Ω, it follows that under H_o', G_n converges in distribution to a chi-square with s degrees of freedom.

Under H_1', $\partial z^{f_1}(\alpha_1(\alpha^*))/\partial \alpha_1 = a \neq 0$. Since $(1/n)\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial \alpha_1$ converges almost surely to $a \neq 0$, and since $[\Omega_n + Q_nQ_n']^{-1}$ converges almost surely to a positive definite matrix, it follows that under H_1', G_n converges almost surely to ∞.

Q.E.D.

PROOF OF LEMMA 6: We use the first equation of the proof of Theorem 6. Since under H_0, $\alpha_1^* = \alpha_1(\alpha^*)$, we get:

$$(1/\sqrt{n})\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial \alpha_1 = (1/\sqrt{n})\partial L_n^c(Y_1|Y_2,Z;\alpha_1^*)/\partial \alpha_1 + A_o^{f_1}(\alpha_1^*)\sqrt{n}(\hat{\alpha}_{1n} - \alpha_1^*) + o_p(1).$$

On the other hand from the proof of Theorem 3, we have:

$$\sqrt{n}(\tilde{\alpha}_{1n} - \alpha_1^*) = - [A_o^{f_1}(\alpha_1^*)]^{-1}(1/\sqrt{n})\partial L_n^c(Y_1|Y_2,Z;\alpha_1^*)/\partial\alpha_1 + o_p(1).$$

Hence, under H_0:

$$(1/\sqrt{n})\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial\alpha_1 = A_o^{f_1}(\alpha_1^*)\sqrt{n}(\hat{\alpha}_{1n} - \tilde{\alpha}_{1n}) + o_p(1). \qquad (*)$$

Since Ω is the asymptotic covariance matrix of the left-hand side, and since V is the asymptotic covariance matrix of $\sqrt{n}(\hat{\alpha}_{1n} - \tilde{\alpha}_{1n})$, it follows that:

$$\Omega = A_o^{f_1}(\alpha_1^*)\ V\ A_o^{f_1}(\alpha_1^*). \qquad (**)$$

Since by Assumption A5', the matrix $A_o^{f_1}(\alpha_1^*)$ is non-singular, it follows that $r = s$.

To prove (c), we use the fact that:

$$H_n = n(\hat{\alpha}_{1n} - \tilde{\alpha}_{1n})'[V + RR']^{-1}(\hat{\alpha}_{1n} - \tilde{\alpha}_{1n})' + o_p(1),$$

$$G_n = (1/n)\partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial\alpha_1'\ [\Omega + QQ']^{-1}\ \partial L_n^c(Y_1|Y_2,Z;\hat{\alpha}_{1n})/\partial\alpha_1 + o_p(1).$$

These equations follow from the fact that $V_n + R_nR_n'$ and $\Omega_n + Q_nQ_n'$ are consistent estimators of $V + RR'$ and $\Omega + QQ'$ respectively. Hence, using (*) we get:

$$H_n - G_n = n(\hat{\alpha}_{1n} - \tilde{\alpha}_{1n})'[(V + RR')^{-1} - A(\Omega + QQ')^{-1}A](\hat{\alpha}_{1n} - \tilde{\alpha}_{1n}) + o_p(1)$$

where $A = A_o^{f_1}(\alpha_1^*)$. Since A is non-singular by Assumption A5'-b, it is clear from (**) that, in the non-singular case ($r = s = k_1$), the first term on the right-hand side is identically null so that $H_n - G_n \rightarrow 0$ in probability.

In the singular case the derived result is, however, more difficult to establish. To see that, let $V^- = (V + RR')^{-1}$ and $\Omega^- = (\Omega + QQ')^{-1}$. Though for any generalized inverse (of maximum rank) V^- of V the matrix $A^{-1}\ V^-\ A^{-1}$ is a generalized inverse (of maximum rank) of Ω, nothing ensures that $V^- = \Omega^-$ since this depends on the choice of R_n and Q_n, i.e., on the choice of generalized inverses (of maximum rank) of V_n and Ω_n. We shall nevertheless show that the first term on the right-hand side converges in distribution and hence in probability to 0.

From the proof of Theorem 5 we have:

$$\sqrt{n}(\hat{\alpha}_{1n} - \tilde{\alpha}_{1n}) \xrightarrow{D} N(0,V).$$

Moreover

$$\begin{aligned} V(V^- - A\Omega^-A)V &= V - VA\Omega^-AV \\ &= V - (A^{-1}\Omega A^{-1})A\Omega^-A(A^{-1}\Omega A^{-1}) \\ &= 0 \end{aligned}$$

where we have used that V^- and Ω^- are generalized inverses of V and Ω, and that $V = A^{-1}\Omega A^{-1}$ which follows from part (a). Hence:

$$[V(V^- - A\Omega^-A)]^3 = [V(V^- - A\Omega^-A)]^2.$$

Therefore from Theorem 9.2.1 in Rao and Mitra (1971) it follows that:

$$n(\hat{\alpha}_{1n} - \tilde{\alpha}_{1n})'[V^- - A\Omega^- A](\hat{\alpha}_{1n} - \tilde{\alpha}_{1n}) \overset{D}{\rightarrow} \chi^2_m$$

where

$$\begin{aligned} m &= \text{trace } [V^- - A\Omega^- A]V \\ &= \text{trace } (V^- V) - \text{trace } (\Omega^- AVA) \\ &= \text{trace } (V^- V) - \text{trace } (\Omega^- \Omega) \\ &= \text{rank } V - \text{rank } \Omega \\ &= 0 , \end{aligned}$$

where the fourth equality follows from a property of a generalized inverse (see Rao and Mitra (1971, Definition 3, p. 21). Thus $H_n - G_n \rightarrow 0$ in probability.

Q.E.D.

FOOTNOTES

[*] I am much indebted to J. Dubin, R. Engle, D. Grether, J. Link, D. Rivers, and H. White for helpful comments and criticism. Remaining errors are of course mine.

[1] The following assumptions, with the exception of Assumption 6 in Section 3, are similar to those of White (1982). The basic difference is that our assumptions bear on the conditional density instead of on the joint density.

[2] The existence of a conditional distribution is ensured by Jirina's Theorem (see e.g., Loeve (1955), Monfort (1980)).

[3] Nothing is said about uniqueness of a CMLE. Our definition corresponds to Wald's (1949) approach to ML estimation. On the other hand, Andersen (1970) takes Cramer's (1946) approach so that his assumptions are somewhat different from ours.

[4] Note that this holds even if $G^o(.)$ does not have a density.

[5] As a matter of fact Y_{2t} may simply be a function of Y_{1t} (see Examples 2 and 3 below).

[6] From Theorem 3, Equation (5.1) can also be written as $\text{plim } \tilde{\alpha}_{1n} = \text{plim } \hat{\alpha}_{1n}$.

[7] Note, however, that in our case the efficient estimator $\hat{\alpha}_{1n}$ is used in evaluating the gradient of $L^c_n(Y_1|Y_2,Z;.)$ and $\tilde{\Omega}_n$. This

contrasts with White's statistic (5.2) where the inefficient estimator is used.

8. I owe this point to a discussion with D. Rivers.

9. Consider the simultaneous system $Y_t = \beta_1 + \beta_2 Z_t + u_t$ and $Z_t = \gamma + v_t$ where u_t and v_t may be correlated. [The first equation is not identified and, Z_t is neither weakly nor strictly exogenous for $\alpha = (\beta_1, \beta_2, \sigma_{uu})$.] OLS on the first equation consistently estimates the conditional distribution of Y_t given Z_t when the true conditional distribution of u_t given v_t is normal, a condition that is satisfied when u_t and v_t are jointly normally distributed.

10. Since no assumption is made on Z_t, one can also consider the "reverse" regression of Z_t on Y_t. The resulting parameter estimates and the direct OLS estimates must then satisfy some compatibility conditions in order to define a proper estimated joint distribution for (Y_t, Z_t) (see Gourieroux and Monfort (1979)). See also footnote 17.

11. Whether or not all the parameters in α are identified clearly depends on the choice of B.

12. If B is the complete choice set minus one alternative, then clearly $\tilde{\alpha}_{1n} = \alpha^B_{1n}$. Our test becomes identical to the one proposed by Hausman and McFadden (1981).

13. Though the log-likelihood function associated with the (censored) Tobit model is globally concave in γ and $1/\sigma$ (see Olsen (1978)), the log-likelihood function associated with the truncated Tobit model is only partially concave in γ and $1/\sigma$ in the sense that given γ it is concave in $1/\sigma$, and given $1/\sigma$ it is concave in γ.

14. Andersen (1970) suggests CML estimation instead of ML estimation when there are incidental parameters. The appropriate conditioning variable to use is a sufficient statistic for the incidental parameters. Assumption A1 rules out such a situation since the Z_t must be identically distributed.

15. In fact S_t is a marginal sufficient statistic since it depends only on Y_t (see Rao (1973, p. 132)). Sudakov (1971), however, shows that when the Y_t's are i.i.d., then a marginal sufficient statistic is also sufficient in the usual sense.

16. For more complex examples, see Vuong (1982b).

17. See Amemiya (1978) and Vuong (1982c). In this latter paper, CML estimation and ML estimation of the marginal model for Y_{2t} are used iteratively in order to produce efficient estimators of all the parameters. Note that, instead of considering the estimation of the marginal model for Y_{2t}, one can consider the CML estimation of the conditional model for Y_{2t} given Y_{1t}. This second approach was suggested by Nerlove and Press (1973) and studied by Guilkey and Schmidt (1979) and Vuong (1982a).

REFERENCES

Amemiya, T.: "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica*, 41 (1973), 997-1016.

_________: "On a Two-Step Estimation of a Multivariate Logit Model," *Journal of Econometrics*, 8 (1978), 13-21.

Andersen, E. B.: "Asymptotic Properties of Conditional Maximum-Likelihood Estimators," *Journal of the Royal Statistical Society*, Series B, 32 (1970), 283-301.

Bowden, R.: "The Theory of Parametric Identification," *Econometrica*, 41 (1973), 1069-174.

Cramer, H.: *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

Engle, R. F., D. Hendry, and J. F. Richard: "Exogeneity," *Econometrica*, 51 (1983), 277-304.

Fisher, R. A.: *Statistical Methods and Scientific Inference*. London: Oliver and Boyd, 1956.

Gourieroux, C., and A. Monfort: "On the Characterization of a Joint Probability Distribution by Conditional Distributions, *Journal of Econometrics*, 9 (1979), 115-118.

Guilkey, D., and P. Schmidt: "Some Small Sample Properties of Estimators and Test Statistics in the Multivariate Logit Model," *Journal of Econometrics*, 10 (1979), 33-42.

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

Hausman, J., and D. McFadden: "A Specification Test for the Multinomial Logit Model," MIT Working Paper,No 292, 1981.

Heckman, J.: "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables, and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5 (1976), 475-492.

Jennrich, R. I.: "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40 (1969), 633-643.

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

Loeve, M.: *Probability Theory*. New York: D. Van Nostrand Company, 1955.

McFadden. D.: "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press, 1974.

Maddala, G. S.: *Limited-Dependent and Qualitative Variables in*

Econometrics. New York: Cambridge University Press, 1983.

Monfort, A.: *Cours de Probabilites*. Paris: Economica, 1980.

Nerlove, M. and S. J. Presa: *Univariate and Multivariate Log-Linear and Logistic Models*. Santa Monica: Rand Corporation, 1973.

Olsen, R. J.: "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," *Econometrica*, 46 (1978), 1211-1215.

Rao, C. R.: "Criteria of Estimation in Large Samples," *Sankkya*, 25 (1963), 189-206.

__________: *Linear Statistical Inference and Its Applications*. New York: John Wiley and Sons, 1973.

Rao, C. R., and S. K. Mitra: *Generalized Inverse of Matrices and its Applications*. John Wiley and Sons, 1971.

Rothenberg, T.: "Identification in Parametric Models," *Econometrica*, 39 (1971), 577-591.

Silvey, S. D.: "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics*, 30 (1959), 389-407.

Sudakov, V. N.: "A Remark on Marginal Sufficiency," *Doklady*, 198 (1971), 796-798.

Tobin, J.: "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26 (1958), 24-36.

Vuong, Q. H.: "Conditional Log-Linear Probability Models: A Theoretical Development with an Empirical Application," Ph.D Thesis, Northwestern University, 1982.

__________: "Probability Feedback in a Recursive System of Probability Models," Social Science Working Paper No. 443, California Institute of Technology, 1982.

__________: "Probability Feedback in a Recursive System of Logit Models: Estimation," Social Science Working Paper No. 444, California Institute of Technology, 1982.

Wald, A.: "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics*, 60 (1949), 595-603.

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

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

__________: "Corrigendum," *Econometrica*, 51 (1983), 513.