CentER
for
Economic Research

Discussion paper

[redacted]

TILBURG UNIVERSITY

No. 8944

IDENTIFICATION AND ESTIMATION OF DICHOTOMOUS LATENT VARIABLES MODELS USING PANEL DATA

by Cheng Hsiao

October, 1989

Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data*

Cheng Hsiao

University of Southern California
Los Angeles, California 90089
U.S.A.

Revised September 1989

Abstract

Identification conditions for binary choice errors-in-variables models are explored. Conditions for the consistency and asymptotic normality of the maximum likelihood estimators of binary choice models with unbounded explanatory variables are given. Two or three step estimators to simplify computation are also suggested. Their consistency are proved and asymptotic variance-covariance matrices are derived. Conditions for the two or three step estimator to achieve asymptotic efficiency are also given.

JEL Code 212

1. Introduction

When variables enter into an equation nonlinearly, and these variables are subject to measurement errors, complicated identification and estimation issues arise. Realizing that the usual large sample theory seems to fail to provide useful results for nonlinear errors-in-variables models and consistent estimators in the usual sense are not readily available, Y. Amemiya (1985), Y. Amemiya and Fuller (1985, 88), Stefanski and Carroll (1985), Wolter and Fuller (1982a,b), etc. have developed alternative asymptotic theories in terms of the index $n = a_n b_n$ with $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ being sequences of positive real numbers representing the magnitudes of reciprocals of error variances and the number of data points, respectively, for $n = 1, 2, \cdots$. Consistency and asymptotic normality of nonlinear least squares and instrumental variable estimators are demonstrated when $n \to \infty$ and $a_n^{-1} = o(n^{-\frac{1}{2}})$ or $a_n^{-1} = o(n^{-\frac{1}{3}})$.

While this alternative approach is ingenious and yields useful approximations to the properties of estimators when error variances are small and when sample size is large, not all economic data possess the property of shrinking error variances when the number of observations increases. Unfortunately, without imposing more structural information, there appears no alternative to the assumption of shrinking error variances for deriving consistent estimators for general nonlinear errors-in-variables models. However, in many samples individual observations may be viewed as random draws from a common population. In this paper we wish to explore the type of data which would allow us to identify a nonlinear errors-in-variables model under this assumption and provide consistent estimators for the unknown structural parameters when measurement error variances stay constant. We shall focus our issues on the binary choice models.

We set up the model in section 2. Issues of identification are explored in section 3. Conditions for the consistency and asymptotic normality of maximum likelihood estimators are established in section 4. A computationally simpler two-step conditional maximum likelihood estimator conditioning on a subset of the estimated parameters is suggested and its asymptotic variance-covariance matrix is derived in section 5. Simple consistent estimators for probit model and logit model are suggested in section 6. Conclusions are in section 7.

2. The Model

Let $(y_i, \underset{\sim}{x}_i')$ be $(K+1)$ dimensional independently distributed random variables with finite second order moments. Let the expected value of y conditional on $\underset{\sim}{x}$ be

$$E(y_i \mid \underset{\sim}{x}_i) = g(\underset{\sim}{x}_i, \underset{\sim}{\theta}^0). \tag{2.1}$$

where $\underset{\sim}{\theta}^0$ is a $p \times 1$ vector of unknown parameters, assumed to lie in the interior of a convex compact set $\Theta \subset R^p$, with R^p denoting a p-dimensional Euclidean space. Model (2.1) is nonlinear if $g(\underset{\sim}{x}_i; \underset{\sim}{\theta}^0)$ is nonlinear in either $\underset{\sim}{x}$ or $\underset{\sim}{\theta}^0$.

The class of nonlinear models we will concentrate on is that of qualitative choice models, originally developed by psychologists and later adapted and extended by economists for describing consumers' choices (e.g. see Amemiya (1981), McFadden (1976, 81, 84), and Train (1986) for general reviews). This class of models assumes that y takes a binary outcome and relates the choice probabilities to observed attributes of the alternatives (such as the price or cost associated with each alternative) and of the individuals (such as income) in the form

$$\text{Prob}(y = 1 \mid \underset{\sim}{x}) = E(y \mid \underset{\sim}{x}) = g(\underset{\sim}{\theta}^{0\prime} \underset{\sim}{x}). \tag{2.2}$$

When $g(\underset{\sim}{\theta}^{0\prime} \underset{\sim}{x})$ takes the form of the integrated standard normal, $\Phi(\underset{\sim}{\theta}^{0\prime} \underset{\sim}{x})$, we have the probit model. When $g(\underset{\sim}{\theta}^{0\prime} \underset{\sim}{x})$ takes the form of the logistic distribution, $\exp(\underset{\sim}{\theta}^{0\prime} \underset{\sim}{x})/[1 + \exp(\underset{\sim}{\theta}^{0\prime} \underset{\sim}{x})]$, we have the logit model.

Suppose $\underset{\sim}{x}$ are unobservable.[1] Instead, we observe

$$\underset{\sim}{z} = \Lambda \underset{\sim}{x} + \underset{\sim}{\epsilon}, \tag{2.3}$$

where $\underset{\sim}{\epsilon}$ are assumed to be independent of $\underset{\sim}{x}$ and u and are assumed to be independently normally distributed with mean zero and variance-covariance matrix Ω. When $\Lambda = I_K$, and

$\underset{\sim}{\epsilon}_i$ is a vector of K errors, (2.3) corresponds to the standard measurement errors situation. When Λ is an $M \times K$ matrix with rank $K(M > K)$ and $\underset{\sim}{\epsilon}_i$ is an $M \times 1$ vector of errors, (2.4) corresponds to the factor analysis model with $\underset{\sim}{z}_i$ being the indicator (or manifest) variables of the K latent variables $\underset{\sim}{x}_i$ (e.g. Anderson (1985), Anderson and Rubin (1956), Lawley and Maxwell (1971, 73)).

Let the conditional distribution of $\underset{\sim}{x}$ given $\underset{\sim}{z}$ be $f(\underset{\sim}{x} \mid \underset{\sim}{z}; \underset{\sim}{\delta}^0)$, where $\underset{\sim}{\delta}^0$ is an $s \times 1$ vector of unknown parameters. For instance, under the assumption that $\underset{\sim}{x}$ is also normally distributed with mean $\underset{\sim}{\mu}$ and variance-covariance matrix Σ_x, the conditional distribution of $\underset{\sim}{x}$ given $\underset{\sim}{z}$ also has a multivariate normal distribution. This distribution is characterized by the (conditional) mean

$$E(\underset{\sim}{x} \mid \underset{\sim}{z}) = \underset{\sim}{\mu} + \Sigma_x \Lambda (\Omega + \Lambda \Sigma_x \Lambda')^{-1} (\underset{\sim}{z} - \Lambda \underset{\sim}{\mu}), \tag{2.4}$$

and (conditional) variance-covariance matrix,

$$\mathrm{Var}(\underset{\sim}{x} \mid \underset{\sim}{z}) = \Sigma_x - \Sigma_x \Lambda' (\Omega + \Lambda \Sigma_x \Lambda')^{-1} \Lambda \Sigma_x = A. \tag{2.5}$$

The joint density of $(y, \underset{\sim}{z})$ can be written in any one of the following equivalent forms:

$$\begin{aligned} f(y, \underset{\sim}{z}) &= f(y \mid \underset{\sim}{z}) f(\underset{\sim}{z}) \\ &= \int f(y, \underset{\sim}{x}, \underset{\sim}{z}) d\underset{\sim}{x} \\ &= \int f(y \mid \underset{\sim}{x}, \underset{\sim}{z}) f(\underset{\sim}{x} \mid \underset{\sim}{z}) d\underset{\sim}{x} \cdot f(\underset{\sim}{z}) \\ &= \int f(y \mid \underset{\sim}{x}) f(\underset{\sim}{x} \mid \underset{\sim}{z}) d\underset{\sim}{x} \cdot f(\underset{\sim}{z}), \end{aligned} \tag{2.6}$$

where the last equality follows from the assumption that under (2.1) and (2.3), we have

$$f(y \mid \underset{\sim}{x}, \underset{\sim}{z}) = f(y \mid \underset{\sim}{x}). \tag{2.7}$$

Thus, for the model (2.1), we have

$$\mathrm{Prob}(y = 1 \mid \underset{\sim}{z}) = \int g(\underset{\sim}{\theta}^{0\prime} \underset{\sim}{x}) \cdot f(\underset{\sim}{x} \mid \underset{\sim}{z}; \underset{\sim}{\delta}^0) d\underset{\sim}{x}. \tag{2.8}$$

Under the assumption of (2.3), we have

$$\underset{\sim}{x} = (\Lambda'\Lambda)^{-1}\Lambda'(\underset{\sim}{z} - \underset{\sim}{\epsilon}). \tag{2.9}$$

Substituting (2.9) into (2.8), the specific type of model we are considering has the form

$$\begin{aligned} \text{Prob } (y = 1 \mid \underset{\sim}{z}) &= \int g\left\{\underset{\sim}{\theta}^{0\prime}\left[(\Lambda'\Lambda)^{-1}(\underset{\sim}{z} - \underset{\sim}{\epsilon})\right]\right\} f(\underset{\sim}{\epsilon} \mid \underset{\sim}{z}) d\underset{\sim}{\epsilon} \\ &= G(\underset{\sim}{b}^{0\prime}\underset{\sim}{z}). \end{aligned} \tag{2.10}$$

where G is some transformation of $\underset{\sim}{b}^{0\prime}z$ and $\underset{\sim}{b}^0$ is a nonlinear continuous function of $\underset{\sim}{\theta}^0$ and $\underset{\sim}{\delta}^0$, $\underset{\sim}{b}^0 = h(\underset{\sim}{\theta}^0 , \underset{\sim}{\delta}^0)$. When $\underset{\sim}{x}$ and $\underset{\sim}{\epsilon}$ are normally distributed, from (2.8) we know that the probability of $y = 1$, conditional on $\underset{\sim}{z}$, for the probit model, is (e.g. Lien (1986))[2]

$$\text{Prob}(y = 1 \mid \underset{\sim}{z}) = G(\underset{\sim}{b}^{0\prime}\underset{\sim}{z}) = \Phi\left\{\frac{\underset{\sim}{\theta}^{0\prime}[\underset{\sim}{\mu} + \Sigma_x\Lambda(\Omega + \Lambda\Sigma_x\Lambda')^{-1}(\underset{\sim}{z} - \Lambda\underset{\sim}{\mu})]}{(1 + \underset{\sim}{\theta}^{0\prime}A\underset{\sim}{\theta}^0)^{\frac{1}{2}}}\right\} \tag{2.11}$$

and, for the logit model, is

$$\begin{aligned} &\text{Prob}(y = 1 \mid \underset{\sim}{z}) = G(\underset{\sim}{z}) \\ &= \int_{-\infty}^{\infty} g\left\{(\underset{\sim}{\theta}^{0\prime}A\underset{\sim}{\theta}^0)^{\frac{1}{2}}v + \underset{\sim}{\theta}'[\mu + \Sigma_x\Lambda(\Omega + \Lambda\Sigma_x\Lambda')^{-1}(\underset{\sim}{z} - \Lambda\underset{\sim}{\mu})]\right\} \cdot \phi(v)dv, \end{aligned} \tag{2.12}$$

where $g(a) = \exp(a)[1 + \exp(a)]^{-1}$, and $\phi(\cdot)$ is the standard normal density function.

3. Identification

A structure S is a complete specification of the probability distribution function of the random variable $y, F(y)$. The set of all a priori possible structure J is called a model. The identification problem consists in making judgements about structures, given the model J and the observations y (e.g. Hsiao (1983)). In most applications, conditional on the $m \times 1$ parameter vector $\underset{\sim}{\gamma}^0$, y is assumed to be generated by a known probability distribution function $F(y \mid \underset{\sim}{\gamma})$, but $\underset{\sim}{\gamma}^0$ is unknown. Thus, the problem of distinguishing structures is reduced to the problem of distinguishing between parameter points. In this framework, we have

DEF. 3.1: *For $\underset{\sim}{\gamma}^0 \epsilon N$, when N is a convex compact subset of R^m, the structure $F(y \mid \underset{\sim}{\gamma}^0)$ is said to be identified if there is no other $\underset{\sim}{\gamma}^1 \epsilon N$ such that $F(y \mid \underset{\sim}{\gamma}^0) = F(y \mid \underset{\sim}{\gamma}^1)$ for all y.*

Given Definition 3.1, the general condition for the local identification of a structure as worked out by Rothenberg (1971), etc. is that:

THEOREM 3.1: *Let $\underset{\sim}{\gamma}^0 \epsilon N$ be a regular point of the information matrix, i.e. the information matrix has constant rank for $\underset{\sim}{\gamma}$ in an open neighborhood of $\underset{\sim}{\gamma}^0$. Then $\underset{\sim}{\gamma}^0$ is locally identifiable if and only if the information matrix evaluated at $\underset{\sim}{\gamma}^0$ is nonsingular.*

From Theorem 3.1 we can conclude that

THEOREM 3.2: *The binary choice model of (2.2)-(2.3) is (locally) identified only if $\underset{\sim}{\delta}^0$ can be identified from the marginal distribution of $\underset{\sim}{z}$. When $\underset{\sim}{\theta}^0$ is identifiable given $\underset{\sim}{y}$ and $\underset{\sim}{x}$, this condiiton is also sufficient provided $E\underset{\sim}{z}\underset{\sim}{z}' = M^*$ being an $M \times M$ nonsingular matrix with $M \geq K$.*

For proof, see the appendix.

Theorem 3.2 states that in order to identify $\underset{\sim}{\theta}^0$ from $(\underset{\sim}{y}, \underset{\sim}{z})$ it is essential that $\underset{\sim}{\delta}^0$ can be identifiable from the sample information of $\underset{\sim}{z}$. This immediately imposes restrictions on the type of data for which a nonlinear errors-in-variables model is identifiable. For instance, under the assumption that $\underset{\sim}{x}$ and $\underset{\sim}{\epsilon}$ in (2.3) are mutually independent multivariate normally distributed random variables, the conditional distribution of $\underset{\sim}{x}$ given $\underset{\sim}{z}$ is characterized by the conditional mean (2.4) and conditional variance-covariance matrix (2.5). In other words, in order to identify $\underset{\sim}{\delta}^0$, we have to know $\Lambda, \underset{\sim}{\mu}$ and the measurement error covariance matrix Ω. The mere existence of instruments $\underset{\sim}{w}$ which are correlated with $\underset{\sim}{x}$ and uncorrelated with $\underset{\sim}{\epsilon}$ is not sufficient to ensure the identifiability of $\underset{\sim}{\delta}^0$, hence $\underset{\sim}{\theta}^0$. Stronger conditions on the probability distribution of $\underset{\sim}{z}$ are needed in order to identify $\underset{\sim}{\theta}^0$.

There are many different ways one can identify Λ, Ω and Σ_x from observed $\underset{\sim}{z}$. For instance, there could be consumer's responses (the indicator variables $\underset{\sim}{z}$) to a series of attitudinal questions (e.g. Train, McFadden and Goett (1987)). In such a factor analysis framework, one set of conditions for the identification of a factor analysis model (2.3) is that (i) Ω is diagonal, (ii) each column of Λ has at least $K-1$ specified 0's in a certain column and the matrix composed of the rows of Λ corresponding to the 0's in a certain column has rank $K-1$, (iii) a normalization rule such as an element in each column of Λ is 1, (iv) the number of components in the variance-covariance matrix of $\underset{\sim}{z}, \Sigma_z$, and the number of conditions (sum of (ii) and (iii)), $\frac{1}{2}M(M+1)+K^2$ exceeds the number of parameters in Λ, Σ_x, and Ω, i.e., $\frac{1}{2}[(M-K)^2-M-K] \geq 0$. (Anderson and Rubin (1956)).

In the standard measurement error framework, we have $\Lambda = I_K$. One way to identify Ω, hence Σ_x, from $\underset{\sim}{z}$, is to obtain replicated observations for $\underset{\sim}{x}_i$,

$$\underset{\sim}{z}_{it} = \underset{\sim}{x}_i + \underset{\sim}{\epsilon}_{it}, \qquad t = 1, \ldots, n_i, \tag{3.1}$$

where $n_i \geq 2$. Under the assumption that $\underset{\sim}{x}_i$ and $\underset{\sim}{\epsilon}_{it}$ are multivariate normal, the joint density of $(\underset{\sim}{z}_{i1}, \ldots, \underset{\sim}{z}_{in_i})$ is

$$\begin{aligned} f(\underset{\sim}{z}_{i1}, \ldots, \underset{\sim}{z}_{in_i}) &= f(\underset{\sim}{z}_{i1}, \ldots, \underset{\sim}{z}_{in_i} \mid \bar{\underset{\sim}{z}}_i) f(\bar{\underset{\sim}{z}}_i) \\ &= (2\pi)^{-\frac{n_i-1}{2}} \mid \Omega \mid^{-\frac{(n_i-1)K}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{n_i}(\underset{\sim}{z}_{it} - \bar{\underset{\sim}{z}}_i)'\Omega^{-1}(\underset{\sim}{z}_{it} - \bar{\underset{\sim}{z}}_i)\right\} \\ &\cdot (2\pi)^{-\frac{K}{2}} \mid \Sigma_x + \frac{1}{n_i}\Omega \mid^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\bar{\underset{\sim}{z}}_i - \underset{\sim}{\mu})'(\Sigma_x + \frac{1}{n_i}\Omega)^{-1}(\bar{\underset{\sim}{z}}_i - \underset{\sim}{\mu})\right\} \end{aligned} \tag{3.2}$$

where $\bar{\underset{\sim}{z}}_i = \frac{1}{n_i}\sum_{t=1}^{n_i} \underset{\sim}{z}_{it}$. Taking the second partial derivatives of the logarithm of $\Pi_{i=1}^n f(\underset{\sim}{z}_i)$ with respect to $\underset{\sim}{\mu}$ and Ω, the limiting matrix has the form[3]

$$-\begin{bmatrix} E(\Sigma_x + \frac{1}{n_i}\Omega) & \underset{\sim}{0} \\ \underset{\sim}{0} & \frac{E(n_i-1)}{2}\Omega \otimes \Omega \end{bmatrix} \tag{3.3}$$

It is clear from (3.3) that a necessary and sufficient condition for (3.3) to be nonsingular is that $E(n_i - 1) \neq 0$. This is so if and only if there are repeated observations. That is $\underset{\sim}{\mu}$ and Ω can be identified if we have replicated observations.

4. Maximum Likelihood Estimation

In this section we will focus on the issues of estimation when the identification of (2.2)-(2.3) is achieved through replicated observations (3.1). Conventional proofs of the consistency and asymptotic normality of the maximum likelihood estimator (MLE) for the binary choice models typically assume that the explanatory variables are bounded (e.g. Amemiya (1985), Giourieroux and Monfort (1981), McFadden (1974)). But in our formulation, in particular, when $\underset{\sim}{x}$ and $\underset{\sim}{\epsilon}$ are assumed to be normally distributed, the $\underset{\sim}{z}$ are clearly unbounded. Therefore, we shall give a set of regularity conditions which will ensure the desirable properties of the MLE. However, these conditions are not necessarily the weekest. They are chosen for simplicity and ease in verification.

Consider the likelihood function

$$\begin{aligned} f(y_{i1},\cdots,y_{in_i},\underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i}) &= f(y_{i1},\cdots,y_{in_i} \mid \underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i})f(\underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i}) \\ &= \left[\int f(y_{i1},\cdots,y_{in_i} \mid \underset{\sim}{x}_i)f(\underset{\sim}{x}_i \mid \underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i})d\underset{\sim}{x}_i\right] f(\underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i} \mid \bar{\underset{\sim}{z}}_i)f(\bar{\underset{\sim}{z}}_i) \end{aligned} \tag{4.1}$$

where $\bar{\underset{\sim}{z}}_i = \frac{1}{n_i}\sum_{t=1}^{n_i} \underset{\sim}{z}_{it}$. Taking the logarithm of $\Pi_{i=1}^{n} f(\underset{\sim}{y}_i, \underset{\sim}{z}_i)$, we have

$$L_{n^*} = \log L = \sum_{i=1}^{n} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i) + \log L_z, \tag{4.2}$$

where $n^* = \sum_{i=1}^{n} n_i$,

$$H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i) = \int \Pi_{t=1}^{n_i} f(y_{it} \mid \underset{\sim}{x}_i)f(\underset{\sim}{x}_i \mid \underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i})d\underset{\sim}{x}_i \tag{4.3}$$

and

$$\begin{aligned} \log L_z &= \sum_{i=1}^{n} \log f(\underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i} \mid \bar{\underset{\sim}{z}}_i) + \sum_{i=1}^{n} \log f(\bar{\underset{\sim}{z}}_i) = -\frac{K}{2}\left[\sum_{i=1}^{n}(n_i-1)\right]\log 2\pi \\ &-\frac{1}{2}\left[\sum_{i=1}^{n}(n_i-1)\right]\log|\Omega| - \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{n_i}(\underset{\sim}{z}_{it}-\bar{\underset{\sim}{z}}_i)'\Omega^{-1}(\underset{\sim}{z}_{it}-\bar{\underset{\sim}{z}}_i) \\ &-\frac{nK}{2}\log 2\pi - \frac{n}{2}\log|\Sigma_x + \frac{1}{n_i}\Omega| - \frac{1}{2}\sum_{i=1}^{n}(\bar{\underset{\sim}{z}}_i - \underset{\sim}{\mu})'(\Sigma_x + \frac{1}{n_i}\Omega)^{-1}(\bar{\underset{\sim}{z}}_i - \underset{\sim}{\mu}). \end{aligned} \tag{4.4}$$

Using a similar argument as that of Gourieroux and Monfort (1981), Hoadley (1971) and White (1980), we can show that

THEOREM 4.1: *Let $\hat{\underset{\sim}{\gamma}}_{n^*}$ be the estimator that $L_{n^*}(\hat{\underset{\sim}{\gamma}}_n) = \max_{\underset{\sim}{\gamma}\epsilon N} L_{n^*}(\underset{\sim}{\gamma})$. Assume that*

(A1): $\log \int \Pi_{t=1}^{n_i} f(y_{it} \mid \underset{\sim}{x}_i;\underset{\sim}{\theta})f(\underset{\sim}{x}_i \mid \underset{\sim}{z}_{i1},\cdots,\underset{\sim}{z}_{in_i},\underset{\sim}{\delta})d\underset{\sim}{x}_i = \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\gamma})$ is a measureable function on $D_y \times D_z$, where D_y and D_z denote the support of y and $\underset{\sim}{z}$, respectively.

(A2): $\log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma})$ is a continuous function of $\underset{\sim}{\gamma}$ on N, uniformly in i, a.s. $[P]$; where P denotes the probability law governing the data generating process of y and $\underset{\sim}{z}$.

(A3): There exists measurable function $m(\underset{\sim}{y}_i, \underset{\sim}{z}_i)$ such that $\mid \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) \mid < m(\underset{\sim}{y}_i, \underset{\sim}{z}_i)$ for all $\underset{\sim}{\gamma} \in N$ and for all i, $E \mid m(\underset{\sim}{y}_i, \underset{\sim}{z}_i) \mid^{1+\nu} \leq \Delta < \infty$ for some $\nu > 0$, and $0 < \Delta < \infty$.

(A4): The model is identifiable. In other words, there exists n_0 such that

$$\inf_{n^* \geq n_0} \left[\bar{L}(\underset{\sim}{\gamma}^0) - \max_{\underset{\sim}{\gamma} \in \bar{N}} \bar{L}(\underset{\sim}{\gamma}) \right] > 0.$$

where $\bar{N}$ denotes the complement of a neighborhood of $\underset{\sim}{\gamma}^0$ in N, and

$$\bar{L} = \frac{1}{n^*} \sum_{i=1}^{n} E \left[\log f(\underset{\sim}{y}_i, \underset{\sim}{z}_i \mid \underset{\sim}{\gamma}) \right].$$

hold, then $\hat{\underset{\sim}{\gamma}}_{n^*} \xrightarrow{a.s.} \underset{\sim}{\gamma}^0$.

Let $\tau_{\min}(\underset{\sim}{\gamma})$ and $\tau_{\max}(\underset{\sim}{\gamma})$ be the smallest and largest eigenvalues of $\frac{\partial^2 L_{n^*}(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma} \partial \underset{\sim}{\gamma}'} = I_{n^*}(\underset{\sim}{\gamma})$, the following conditions will imply A4:

(A5): $\tau_{\min}(\underset{\sim}{\gamma}^0) \longrightarrow \infty$ as $n^* \longrightarrow \infty$

(A6): for $\underset{\sim}{\gamma}$ in the neighborhood of $\underset{\sim}{\gamma}^0$, there exists a c such that $\frac{\tau_{\max}(\underset{\sim}{\gamma})}{\tau_{\min}(\underset{\sim}{\gamma})} < c$, $\forall n^*$.

THEOREM 4.2.: *Under the assumptions of Theorem 4.1 and*

(A7): $\frac{\partial}{\partial \underset{\sim}{\gamma}} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma})$ and $\frac{\partial^2}{\partial \underset{\sim}{\gamma} \partial \underset{\sim}{\gamma}'} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma})$ exist and are continuous function of $\underset{\sim}{\gamma}$ for $\underset{\sim}{\gamma} \in N$, uniformly in i, a.s. $[P]$, and is a measurable function of $\underset{\sim}{y}$ and $\underset{\sim}{z}$.

(A8): $E[\frac{\partial}{\partial \underset{\sim}{\theta}} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) \mid \underset{\sim}{\gamma}] = 0$

(A9): $E[\frac{\partial}{\partial \underset{\sim}{\theta}} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) \frac{\partial}{\partial \underset{\sim}{\theta}'} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) \mid \underset{\sim}{\gamma}] = -E(\frac{\partial^2}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) \mid \underset{\sim}{\gamma}] = I^*_{\underset{\sim}{y}|z}$, and $I^*_{\underset{\sim}{y}|z}$ is positive definite.

(A10): For some $\nu > 0, \sum_{i=1}^{n} E \mid \underset{\sim}{\lambda}' \frac{\partial}{\partial \underset{\sim}{\theta}} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) \mid^{2+\nu} / \, n^{*(2+\nu)/2} \to 0$ for all $\underset{\sim}{\lambda}$ in R^P.

(A11): For some $\nu > 0, E \sup_{\underset{\sim}{\gamma} \epsilon N} \mid \frac{\partial^2}{\partial \underset{\sim}{\gamma} \partial \underset{\sim}{\gamma}'} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) \mid^{1+\nu} < \infty$.

Then

$$n^{*\frac{1}{2}}(\hat{\underset{\sim}{\gamma}}_n - \underset{\sim}{\gamma}^0) \xrightarrow{d} N(0, \Gamma^{-1}),$$

where

$$\Gamma = -E\left(\frac{\partial^2 \log L}{\partial \underset{\sim}{\gamma} \partial \underset{\sim}{\gamma}'}\right)_{\underset{\sim}{\gamma} = \underset{\sim}{\gamma}^0}.$$

As pointed out by Andrews (1987) and Pötscher and Prucha (1986), assumption A.2 (of theorem 4.1) is a fairly restrictive one. It often implies that $\underset{\sim}{z}$ has to be bounded (e.g. Amemiya (1985, p. 270)). But under our formulation $\underset{\sim}{z}$ clearly is not bounded. However, in the case that y_{it} can only take value of zero and 1, $H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i)$ takes the form

$$H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\delta}) = \int P(\underset{\sim}{x}; \underset{\sim}{\theta})^{\sum_{t=1}^{n_i} y_{it}} [1 - P(\underset{\sim}{x}; \underset{\sim}{\theta})]^{n_i - \sum_{t=1}^{n_i} y_{it}} f(\underset{\sim}{x} \mid \underset{\sim}{z}_i; \underset{\sim}{\delta}) d\underset{\sim}{x} \tag{4.5}$$

where $P(\underset{\sim}{x}; \underset{\sim}{\theta}) = \text{Prob}(y = 1 \mid \underset{\sim}{x}; \underset{\sim}{\theta})$. We only need relatively mild assumption to ensure the consistency and asymptotic normality of the maximum likelihood estimator without having to impose the assumption that $\underset{\sim}{z}$ belongs to a bounded set.

Lemma 4.1 In addition to assumptions **A.1** and **A.4** (or **A1**, **A5** and **A.6**), we assume that

(A12): $P(\underset{\sim}{x};\underset{\sim}{\theta})$ is a continuous function of $\underset{\sim}{x}$ and $\underset{\sim}{\theta}$ on $D_x \times \Theta$, where D_x denotes the support of $\underset{\sim}{x}$ and Θ is a p-dimensional convex compact set.

(A13): There exists a $\Delta < \infty$ such that for each $\underset{\sim}{\theta}$ in Θ, $E \mid \frac{\partial P(\underset{\sim}{x};\theta)}{\partial \underset{\sim}{x}} \mid < \Delta$ and $E \mid \frac{\partial P(\underset{\sim}{x};\theta)}{\partial \underset{\sim}{\theta}} \mid < \Delta$.

(A14): For each $\underset{\sim}{\theta}$ in Θ, there exists an $\underset{\sim}{\bar{x}}$ such that $0 < P(\underset{\sim}{\bar{x}};\underset{\sim}{\theta}) < 1$.

(A15): $f(\underset{\sim}{x} \mid \underset{\sim}{z}) > 0$ and is continuous for $\underset{\sim}{x} \in D_x$.

(A16):[4] For some $\nu > 0$, there exists a $\Delta < \infty$ such that for $\| z \| > c > 0$ $\sup_{\underset{\sim}{\gamma} \epsilon N} \frac{f(\underset{\sim}{z};\gamma)}{H(\underset{\sim}{y}|\underset{\sim}{z};\underset{\sim}{\gamma})} \leq \Delta \|\underset{\sim}{z}\|^{-(1+\nu)}$.

hold. Then the MLE of $\underset{\sim}{\gamma}$ is consistent and asymptotically normally distributed.

Proof: Under A12, $P(\underset{\sim}{x};\underset{\sim}{\theta})$ is continous in $\underset{\sim}{x}$. Under A14, we know that there exists an $\underset{\sim}{\bar{x}}$ in D_x and $a > 0$ such that $P(\underset{\sim}{\bar{x}};\underset{\sim}{\theta}) \geq a > 0$. Furthermore, A.15 ensures that $f(\underset{\sim}{\bar{x}} \mid \underset{\sim}{z};\underset{\sim}{\delta}) > 0$. By the continuity argument, we have $H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\gamma}) \neq 0$ and $H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\gamma})$ is bounded because $\mid P(\underset{\sim}{x};\underset{\sim}{\theta}) \mid \leq 1$. Therefore, $\log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\gamma})$ is bounded and there exists a positive $\nu < 1$ such that $E \sup_{\underset{\sim}{\gamma} \epsilon N} \mid \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\gamma}) \mid^{1+\nu} < \infty$.

Furthermore, for $\underset{\sim}{z}$ in D_z under A12-A15,

$$
\begin{aligned}
&\frac{\partial}{\partial \underset{\sim}{\gamma}} \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\gamma}) \\
&= \frac{1}{H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i;\underset{\sim}{\gamma})} \Bigg[\int (\Sigma_t y_{it})\, P(\underset{\sim}{x};\underset{\sim}{\theta})^{\Sigma_t y_{it}-1} [1 - P(\underset{\sim}{x};\underset{\sim}{\theta})]^{n_i - \Sigma_t y_{it}} \frac{\partial P(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial \underset{\sim}{\gamma}} f(\underset{\sim}{x} \mid \underset{\sim}{z};\underset{\sim}{\delta}) d\underset{\sim}{x} \\
&- \int (n_i - \Sigma_t y_{it} - 1)\, P(\underset{\sim}{x};\underset{\sim}{\theta})^{\Sigma y_{it}} [1 - P(\underset{\sim}{x};\underset{\sim}{\theta})]^{n_i - \Sigma_t y_{it} - 1} \frac{\partial P(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial \underset{\sim}{\gamma}} f(\underset{\sim}{x} \mid \underset{\sim}{z};\underset{\sim}{\delta}) d\underset{\sim}{x} \\
&+ \int P(\underset{\sim}{x};\underset{\sim}{\theta})^{\Sigma_t y_{it}} [1 - P(\underset{\sim}{x};\underset{\sim}{\theta})]^{n_i - \Sigma_t y_{it}} \frac{\partial}{\partial \underset{\sim}{\gamma}} f(\underset{\sim}{x} \mid \underset{\sim}{z};\underset{\sim}{\delta}) d\underset{\sim}{x} \Bigg]
\end{aligned}
\tag{4.6}
$$

is bounded because there exists an $a > 0$ (a may depend on $\underset{\sim}{z}_i$ and $\underset{\sim}{\gamma}$) such that $H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma}) > a$ and $E\frac{\partial P(\underset{\sim}{x};\theta)}{\partial \underset{\sim}{\gamma}}$ and $\int \frac{\partial f(\underset{\sim}{x}|\underset{\sim}{z};\delta)}{\partial \underset{\sim}{\gamma}} d\underset{\sim}{x}$ are bounded by A.13 and A.15. Furthermore, the independence assumption of $\underset{\sim}{z}_i$ and A16 ensures that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{\underset{\sim}{\gamma} \epsilon N} \| \frac{\partial \log H(\underset{\sim}{y}_i \mid \underset{\sim}{z}_i; \underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}} \| < \infty \tag{4.7}$$

Therefore, by a theorem of Andrews (1987), $\frac{1}{n^*} \log L$ converges to $E \log L$. Hence the MLE is consistent.

Similarly, we can show that A12-A16 are sufficient for the conditions of theorem 4.2 to hold. Hence $\hat{\underset{\sim}{\gamma}}_{MLE}$ is asymptotically normally distributed.

In the case that $P(\underset{\sim}{x}; \underset{\sim}{\theta})$ is an integrated normal (probit model) or logistic distribution (logit model), $P(\underset{\sim}{x}; \underset{\sim}{\theta})$ is continous on $D_x \times \Theta$ with bounded derivatives. Furthermore, since $\underset{\sim}{\theta} \epsilon \Theta$, a compact set, there exists an $\bar{\underset{\sim}{x}}$ in D_x such that $0 < P(\bar{\underset{\sim}{x}}; \underset{\sim}{\theta}) < 1$. Moreover, for the probit model we have

$$\frac{\partial P(\underset{\sim}{x}; \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} = \phi(\underset{\sim}{x}'\underset{\sim}{\theta})\underset{\sim}{x}. \tag{4.8}$$

For the logit model, we have

$$\begin{aligned} \frac{\partial P(\underset{\sim}{x}; \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} &= \frac{e^{\underset{\sim}{\theta}'\underset{\sim}{x}}}{(1 + e^{\underset{\sim}{\theta}'\underset{\sim}{x}})} \times \frac{1}{(1 + e^{\underset{\sim}{\theta}'\underset{\sim}{x}})} \underset{\sim}{x} \\ &= P(\underset{\sim}{x}; \underset{\sim}{\theta}) \left[1 - P(\underset{\sim}{x}; \underset{\sim}{\theta})\right] \underset{\sim}{x}. \end{aligned} \tag{4.9}$$

Therefore A.12-A13 hold.

When $\underset{\sim}{x}$ and $\underset{\sim}{\epsilon}$ are normally distributed, A.15 holds. So is A.16 because $H(\underset{\sim}{y} \mid \underset{\sim}{z}; \underset{\sim}{\gamma})$ is bounded away from zero. Furthermore, using an argument similar to the proof of Corollary 3 of Fahrmeir and Kaufman (1985) we can show that the following sampling scheme is sufficient to ensure A.5 and A.6:

(i) Let the number of data points (n) tend to infinity. The number of replications at each data point (n_i) does not necessarily have to go to infinity as long as $\frac{\tilde{n}}{n}$ tends to a nonzero constant, where $\tilde{n}$ = the number of $\{i \mid n_i > 1\}$.

Then, the MLE of $\underset{\sim}{\theta}$ and $\underset{\sim}{\delta}$ are consistent and asymptotically normally distributed.

On the other hand, the following sampling scheme:

(ii) The number of data points (n) is finite and the number of replications (n_i) tends to infinity provided $\sum_{i=1}^{n} \bar{\underset{\sim}{z}}_i \bar{\underset{\sim}{z}}_i'$ has full rank,

is not sufficient to ensure A5 and A6. However, using an argument similar to that of Y. Amemiya and Fuller (1988) we can show that the MLE of $\underset{\sim}{\theta}$ is still consistent.

Hence, if one is only interested in the estimation of $\underset{\sim}{\theta}$ for probit and logit models under the normality assumption of $\underset{\sim}{x}$ and $\underset{\sim}{\epsilon}$, the MLE of $\underset{\sim}{\theta}$ is consistent and asymptotically normally distributed when the number of data points remains finite but the number of replication increases or when the number of replications remains finite as long as the number of data points increases or when both the replications and data points increase. Of course, under the latter sampling schemes other parameters can also be consistently estimated.

5. A Conditional Maximum Likelihood Estimation Procedure

The simultaneous estimation of $(\underset{\sim}{\theta}', \underset{\sim}{\delta}') = \underset{\sim}{\gamma}'$ can be very complicated. However, as discussed in section 3, under the assumption that $\underset{\sim}{\theta}$ is identifiable had $\underset{\sim}{x}$ been observable, $\underset{\sim}{\gamma}$ is identifiable from the probability distribution of $(\underset{\sim}{y}, \underset{\sim}{z})$ if and only if $\underset{\sim}{\delta}$ is identifiable from the marginal distribution of $\underset{\sim}{z}$ alone. This suggests that we may first estimate $\underset{\sim}{\delta}$ from the marginal distribution of $\underset{\sim}{z}$ alone, substitute the estimated $\hat{\underset{\sim}{\delta}}$ into the joint likelihood of $\underset{\sim}{y}$ and $\underset{\sim}{z}$, $f(\underset{\sim}{y}, \underset{\sim}{z} \mid \underset{\sim}{\theta}, \underset{\sim}{\delta} = \hat{\underset{\sim}{\delta}}_n)$ assuming that $\underset{\sim}{\delta} = \hat{\underset{\sim}{\delta}}_n$, then maximize the likelihood function with respect to $\underset{\sim}{\theta}$ alone. For instance, the conditional distribution, $f(\underset{\sim}{x} \mid \underset{\sim}{z}; \underset{\sim}{\delta})$, is characterized by $\Lambda, \underset{\sim}{\mu}, \Omega$ and Σ_x((2.4) and (2.5)). In the standard measurement errors framework $\Lambda = I_k$, if the identification of $\underset{\sim}{\mu}, \Omega$, hence Σ_x, from $\underset{\sim}{z}$ is achieved through replicated observations (3.1), we can maximize the conditional likelihood $\Pi_{i=1}^{n} f(\underset{\sim}{z}_{i1}, \cdots, \underset{\sim}{z}_{in_i} \mid \bar{\underset{\sim}{z}}_i)$ with respect to Ω (Andersen (1970)) and obtain

$$\hat{\Omega} = \left[\sum_{i=1}^{n}(n_i - 1)\right]^{-1} \sum_{i=1}^{n}\sum_{t=1}^{n_i}(\underset{\sim}{z}_{it} - \bar{\underset{\sim}{z}}_i)(\underset{\sim}{z}_{it} - \bar{\underset{\sim}{z}}_i)', \tag{5.1}$$

where

$$\bar{z}_i = \frac{1}{n_i} \sum_{t=1}^{n_i} z_{it} \tag{5.2}$$

We obtain an esimate of μ by maximizing the marginal likelihood of $\Pi_{i=1}^{n} f(\bar{z}_i)$,

$$\bar{z} = \left(\sum_{i=1}^{n} n_i \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{n_i} z_{it} \right). \tag{5.3}$$

Once $\hat{\Omega}$ is obtained, we can estimate Σ_x by

$$\hat{\Sigma}_x = \hat{\Sigma}_z - \hat{\Omega}, \tag{5.4}$$

where

$$\hat{\Sigma}_z = \left(\sum_{i=1}^{n_i} n_i \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{n_i} (z_{it} - \bar{z})(z_{it} - \bar{z})' \right). \tag{5.5}$$

The estimates of $\bar{z}, \hat{\Omega}$ and $\hat{\Sigma}_x$ converges to μ, Ω and $\sum_x$ at the rate $0(n^{-\frac{1}{2}})$. Thus, we can substitute $\hat{\delta}_n$ for δ^0 in (4.2) and estimate θ conditioning on $\delta^0 = \hat{\delta}_n$. The resulting conditional maximum likelihood estimator is consistent due to the following theorem:

THEOREM 5.1: *Let (y_i, z_i') be independent random variables from a distribution depending on $(\theta^{0\prime}, \delta^{0\prime}) = \gamma^{0\prime} \epsilon N$, where N is a convex compact subset of m-dimensional Euclidean space. Let $\hat{\delta}_n = \psi(z_1, \ldots, z_n)$ be such that $\hat{\delta}_n$ converges to δ^0 in probability. Let $\phi(y_i, z_i; \gamma)$ be a differentiable function of γ for $\gamma \epsilon N$ and $E \mid \phi(y, z; \gamma) \mid < \infty$. Suppose $\mid \frac{\partial \phi}{\partial \gamma} \mid \leq L(y, z)$ for all $\gamma \epsilon N$ and $E \mid L(y, z) \mid < \infty$. Then*

$$\frac{1}{n} \sum_{i=1}^{n} \phi(y_i, z_i; \theta, \hat{\delta}_n) \rightarrow \frac{1}{n} \sum_{i=1}^{n} \phi(y_i, z_i; \theta, \delta^0) \tag{5.1}$$

in probability.

The asymptotic variance-covariance matrix of the two-step estimator follows from

THEOREM 5.2: *Let the model (2.1) and (3.1) hold. Assume that the partial derivatives of* $\log L = \mathcal{L}_n(\gamma)$ *up to the third order exist on* $\gamma \epsilon N$ *and are bounded by integrable functions. Let* $\hat{\delta}_n = \psi(z_1, \dots, z_n)$ *be a consistent estimator of* δ^0. *Let* $\hat{\theta}_n$ *be a root of the equation*

$$\frac{\partial}{\partial \theta} \mathcal{L}_n(\theta, \hat{\delta}_n) = 0 \tag{5.2}$$

Then $\sqrt{n}(\hat{\theta}_n - \theta^0)$ will have asymptotic variance-covariance matrix

$$\begin{aligned} \mathcal{L}_{\theta\theta}^{-1}\{E(\frac{1}{n}\frac{\partial \mathcal{L}_n(\theta^0,\delta^0)}{\partial \theta} \cdot \frac{\partial \mathcal{L}_n(\theta^0,\delta^0)}{\partial \theta'}) + \mathcal{L}_{\theta\delta}\mathrm{Var}(\sqrt{n}\hat{\delta}_n)\mathcal{L}_{\delta\theta} \\ + \sum \mathcal{L}_{\delta\theta} + \mathcal{L}_{\theta\delta}\sum{}'\}\mathcal{L}_{\theta\theta}^{-1} \end{aligned} \tag{5.3}$$

where $\mathcal{L}_{\theta\delta} = E(\frac{1}{n}\frac{\partial \mathcal{L}_n^2(\theta^0,\delta^0)}{\partial \theta \partial \delta'})$ and $\sum = E[\frac{1}{\sqrt{n}}\frac{\partial \mathcal{L}_n(\theta^0,\delta^0)}{\partial \theta} \cdot \sqrt{n}(\hat{\delta}_n - \delta^0)]$. Moreover, if the conditions that ensure $\sqrt{n}(\hat{\gamma} - \gamma^0)$ being asymptotically normally distributed holds, $\sqrt{n}(\hat{\theta}_n - \theta^0)$ is also asymptotically normally distributed.

For proof of theorems 5.1 and 5.2, see Gourieroux and Monfort (1987) or Hsiao (1989).

As shown in (5.3), the asymptotic variance-covariance matrix of a two-step estimator of $\theta^0, \hat{\theta}_n$, is, in general, larger than the MLE of $\hat{\theta}_{MLE}$. The conditional MLE will have the same asymptotic efficiency as the MLE if and only if either of the following conditions hold:

Lemma 5.1. The MLE and the two-step estimator of θ^0 will have the same asymptotic variance-covariance matrix if and only if either

i. $E(\frac{1}{n}\frac{\partial^2 \log \mathcal{L}_n(\theta^0,\delta^0)}{\partial \theta \partial \delta'}) = 0$

or

ii. The asymptotic variance-covariance matrix of $\hat{\delta}_n$ is the same as the asymptotic variance-covariance matrix of the MLE of δ^0.

Proof: When (i) holds, then (5.4) becomes $E(-\frac{1}{n}\frac{\partial^2 \log \mathcal{L}}{\partial \theta \partial \theta'})_{\theta^0}^{-1}$, which is the asymptotic variance-covariance matrix of the MLE of θ^0 under the assumption that $E(\frac{\partial^2 \log \mathcal{L}(\theta^0,\delta^0)}{\partial \theta \partial \delta'}) = 0$.

When (ii) holds, we have $\Sigma = \underset{\sim}{0}$. To show this, we follow a similar argument as that of Rao (1973) by defining a new estimator

$$\tilde{\underset{\sim}{\delta}}_n = \hat{\underset{\sim}{\delta}}_n + \lambda\Sigma'\Big[\frac{1}{n}\frac{\partial\mathcal{L}(\underset{\sim}{\theta}^0,\underset{\sim}{\delta}^0)}{\partial\underset{\sim}{\theta}}\Big], \tag{5.4}$$

which is consistent and has the asymptotic variance covariance matrix of the form

$$\mathrm{Var}(\sqrt{n}\tilde{\underset{\sim}{\delta}}_n) = \mathrm{Var}(\sqrt{n}\hat{\underset{\sim}{\delta}}_n) + 2\lambda\Sigma'\Sigma + \lambda^2\Sigma'\,\mathrm{Var}(\frac{1}{\sqrt{n}}\frac{\partial\mathcal{L}(\underset{\sim}{\theta}^0,\underset{\sim}{\delta}^0)}{\partial\underset{\sim}{\theta}})\Sigma. \tag{5.5}$$

Let $\underset{\sim}{c}$ be an arbitrary $s \times 1$ vector of constants, then

$$\underset{\sim}{c}'\Big[\mathrm{Var}(\sqrt{n}\hat{\underset{\sim}{\delta}}_n) - \mathrm{Var}(\sqrt{n}\tilde{\underset{\sim}{\delta}}_n)\Big]\underset{\sim}{c} = -\lambda\Big[2\underset{\sim}{c}'\Sigma'\Sigma\underset{\sim}{c} + \lambda\underset{\sim}{c}\Sigma'\,\mathrm{Var}(\frac{1}{\sqrt{n}}\frac{\partial\mathcal{L}(\underset{\sim}{\theta}^0,\underset{\sim}{\delta}^0)}{\partial\theta})\Sigma\underset{\sim}{c}\Big] \tag{5.6}$$

It is obvious that for $\underset{\sim}{\lambda} \in (0, -2(\underset{\sim}{c}'\Sigma'\,\mathrm{Var}(\frac{1}{\sqrt{n}}\frac{\partial\mathcal{L}(\underset{\sim}{\theta}^0,\underset{\sim}{\delta}^0)}{\partial\underset{\sim}{\theta}})\Sigma\underset{\sim}{c})^{-1}\underset{\sim}{c}'\Sigma'\Sigma\underset{\sim}{c})$, (5.6) is non-negative. This is a direct contradiction to the statement that $\mathrm{Var}(\sqrt{n}\hat{\underset{\sim}{\delta}}_n)$ achieves the Cramer-Rao lower bound unless $\Sigma = \underset{\sim}{0}$.

When $\Sigma = \underset{\sim}{0}$, the asymptotic variance-covariance matrix of the two-step estimator of $\underset{\sim}{\theta}^0$, (5.3) becomes

$$\begin{aligned} &-\mathcal{L}_{\theta\theta}^{-1} + \mathcal{L}_{\theta\theta}^{-1}\mathcal{L}_{\theta\delta}\big[-\mathcal{L}_{\delta\delta} + \mathcal{L}_{\delta\theta}\mathcal{L}_{\theta\theta}^{-1}\mathcal{L}_{\theta\delta}\big]^{-1}\mathcal{L}_{\delta\theta}\mathcal{L}_{\theta\theta}^{-1} \\ &= -(\mathcal{L}_{\theta\theta} - \mathcal{L}_{\theta\delta}\mathcal{L}_{\delta\delta}^{-1}\mathcal{L}_{\delta\theta})^{-1}. \end{aligned} \tag{5.7}$$

which is identical to the asymptotic variance-covariance matrix of the MLE of $\underset{\sim}{\theta}^0$, where $\mathcal{L}_{\delta\delta} = E\frac{\partial^2 \log \mathcal{L}_n(\underset{\sim}{\theta}^0,\underset{\sim}{\delta}^0)}{\partial\underset{\sim}{\delta}\partial\underset{\sim}{\delta}'}$.

Although the conditional maximum likelihood estimator may be less efficient than the maximum likelihood estimator, it does simplify the computation substantially. Moreover, if the surface of $\mathcal{L}_n(\underset{\sim}{\gamma})$ is ill-behaved, an iterative procedure to solve for $\hat{\underset{\sim}{\theta}}$ and $\hat{\underset{\sim}{\delta}}$ will have difficulty converging. On the other hand, there could be cases where conditional on $\underset{\sim}{\delta} = \hat{\underset{\sim}{\delta}}_n$, the surface of $\mathcal{L}_n(\underset{\sim}{\theta}, \hat{\underset{\sim}{\delta}}_n)$ is well-behaved in $\underset{\sim}{\theta} \epsilon \Theta$, hence making the iterative procedure to solve for $\frac{\partial\mathcal{L}_n(\underset{\sim}{\theta}_n,\hat{\underset{\sim}{\delta}}_n)}{\partial\underset{\sim}{\theta}} = \underset{\sim}{0}$ more easily convergeable to $\hat{\underset{\sim}{\theta}}_n$.

6. A Simple Consistent Estimator for the Probit and Logit Models

In section 5 we suggested a conditional MLE of $\underset{\sim}{\theta}$ to simplify the computation. However, its implication remains complicated. For instance, consider the case of probit model where $g(\underset{\sim}{\theta}'\underset{\sim}{x}) = \Phi(\underset{\sim}{\theta}'\underset{\sim}{x})$. Conditioning on $\underset{\sim}{x}$, the likelihood function of $(y_{i1}, \cdots, y_{in_i})$ has a univariate probit form

$$\Pi_{t=1}^{n_i}\Phi(\underset{\sim}{\theta}'\underset{\sim}{x}_i)^{y_{it}}[1 - \Phi(\underset{\sim}{\theta}'\underset{\sim}{x}_i)]^{1-y_{it}}. \tag{6.1}$$

Conditioning on $(\underset{\sim}{z}_{i1}, \cdots, \underset{\sim}{z}_{in_i})$, Ω and Σ_x, the likelihood function of $(y_{i1}, \cdots, y_{in_i})$ becomes a multivariate probit model involving n_i-dimensional integration. Reformulating the likelihood of $(y_{i1}, \cdots, y_{in_i})$ in the form

$$\begin{aligned} f(y_{i1}, \cdots, y_{in_i} \mid \bar{\underset{\sim}{z}}_i) = \int \Pi_{t=1}^{n_i}\Phi\{c_i + \underset{\sim}{\theta}'[I - \Omega(n_i\Sigma_x + \Omega)^{-1}]\bar{\underset{\sim}{z}}_i + \underset{\sim}{\theta}'\underset{\sim}{v}_i\} \\ (2y_{it} - 1)f(\underset{\sim}{v}_i)d\underset{\sim}{v}_i, \end{aligned} \tag{6.2}$$

where

$$\begin{aligned} f(\underset{\sim}{v}_i) &\sim N[\underset{\sim}{O}, \sigma_u^2 I_{n_i} + \underset{\sim}{e}_i\underset{\sim}{e}_i' \cdot \underset{\sim}{\theta}' A_i \underset{\sim}{\theta}], \\ A_i &= \frac{1}{n_i}[\Omega - \Omega(n_i\Sigma_x + \Omega)^{-1}\Omega], \\ c_i &= \underset{\sim}{\theta}'[\Omega(\Omega + n_i\Sigma_x)^{-1}]\underset{\sim}{\mu} \end{aligned}$$

and $\underset{\sim}{e}_i$ is an $n_i \times 1$ vector of ones, I_{n_i} is an n_i-rowed identity matrix, we can reduce the n_i-dimensional integration into a K-dimensional integration, where K is the number of explanatory variables. However, if K is large, the K-dimensional integration will still be quite complicated.[5] To obtain simple consistent estimate of $\underset{\sim}{\theta}$, we note that conditional on $\bar{\underset{\sim}{z}}_i$, we have

$$\text{Prob}(y_{it} = 1 \mid \bar{\underset{\sim}{z}}_i) = \Phi\left[\frac{c + \underset{\sim}{\theta}'(I - \Omega\Sigma_i^{-1})(\bar{\underset{\sim}{z}}_i - \underset{\sim}{\mu})}{(1 + \underset{\sim}{\theta}'A_i\underset{\sim}{\theta})^{\frac{1}{2}}}\right] \tag{6.3}$$

where

$$\begin{aligned} c &= \underset{\sim}{\theta}'\underset{\sim}{\mu}, \\ \Sigma_i &= \Omega + n_i\Sigma_x. \end{aligned}$$

Suppose $n_i = n_j = T,\ \forall i,j$, then we have

$$\Sigma_i = \Omega + T\Sigma_x = \Sigma_j = \Sigma_{\bar{z}},$$

and

$$A_i = A_j = A^*.$$

If we ignore the correlation between y_{it} and y_{is}, we can write down the pseudo likelihood function of $\underset{\sim}{y}, i = 1, \cdots, n$, as[6]

$$Q_n = \sum_{i=1}^{n}\{\sum_{t=1}^{T}[y_{it}\log\Phi(a^* + \underset{\sim}{\theta}^{*\prime}\bar{z}_i) + (1 - y_{it})\log (1 - \Phi(a^* + \underset{\sim}{\theta}^{*\prime}\bar{\underset{\sim}{z}}_i))\}, \tag{6.4}$$

and use the standard Probit computer package to estimate a^* and $\underset{\sim}{\theta}^*$. Once an estimated $\underset{\sim}{\theta}^*$ is obtained, we can solve for $\underset{\sim}{\theta}$ by letting

$$\underset{\sim}{\theta} = (1 - \underset{\sim}{\theta}^{*\prime}C^{-1}A^*C^{-1}\underset{\sim}{\theta}^*)^{-\frac{1}{2}}C^{-1}\underset{\sim}{\theta}^*, \tag{6.5}$$

where

$$C = [I - \Omega(\Omega + T\Sigma_x)^{-1}].$$

Provided that the estimated Ω and Σ_x are $n^{\frac{1}{2}}$ - consistent, one can show that the two step pseudo maximum likelihood estimator of $\underset{\sim}{\theta}$ is consistent.

Similarly, we can obtain simplified estimators of $\underset{\sim}{\theta}$ for logit models. As shown by Efron (1975), Maddala (1983), and McFadden (1976), etc. if the conditional distribution of $\underset{\sim}{x}$ given y is multivariate normal with mean $\underset{\sim}{\mu}_y$ and common variance-covariance matrix D. Then through the use of Bayes' formula, we can express the coefficients of the logit model

$$\text{Prob}(y = 1 \mid \underset{\sim}{x}) = \exp(\theta_o + \underset{\sim}{\theta}_1'\underset{\sim}{x})/[1 + \exp(\theta_o + \underset{\sim}{\theta}_1'\underset{\sim}{x})], \tag{6.6}$$

in terms of $\underset{\sim}{\mu}_y$, D, and the marginal distribution of y, π_y:

$$\theta_o = \frac{1}{2}(\underset{\sim}{\mu}_o - \underset{\sim}{\mu}_1)'D^{-1}(\underset{\sim}{\mu}_o + \underset{\sim}{\mu}_1) - \ell n(\pi_o/\pi_1), \qquad (6.7))$$

$$\underset{\sim}{\theta}_1 = D^{-1}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_o). \tag{6.8}$$

If there are repeated observations as in (3.1), we can obtain simple consistent estimators of $\underset{\sim}{\mu}_y, \pi_y$, and D by

$$\begin{aligned}
&\hat{\underline{\mu}}_o = \frac{1}{T_o} \sum_{\substack{i \\ i,t\epsilon I_o}} \sum_t \underline{z}_{it}, \qquad \hat{\underline{\mu}}_1 = \frac{1}{T_1} \sum_{\substack{i \\ i,t\epsilon I_1}} \sum_t \underline{z}_{it}, \\
&\hat{\pi}_o = \frac{T_o}{T^*}, \qquad \hat{\pi}_i = \frac{T_1}{T^*}, \\
&\hat{D} = \frac{1}{T}\{ \sum_{i,t\epsilon I_o} (\underline{z}_{it} - \hat{\underline{\mu}}_o)(\underline{z}_{it} - \hat{\underline{\mu}})' + \sum_{i,t\epsilon I_1} (\underline{z}_{it} - \hat{\underline{\mu}}_1)(\underline{z}_{it} - \hat{\underline{\mu}}_1)' \} - \hat{\Omega},
\end{aligned} \tag{6.9}$$

where $T^* = \Sigma_{i=1}^n n_i, T_1 = \Sigma_i \Sigma_t y_{it}, \quad T_o = T^* - T_1, I_o = \{(i,t) \mid y_{it} = 0\}, I_1 = \{(i,t) \mid y_{it} = 1)$, and $\hat{\Omega}$ is estimated by (5.1). Thus, we can substitute (6.9) into (6.7) and (6.8) to obtain consistent estimators of $\underline{\theta}$.

The simple consistent estimator can be used for its own right. Or it can be used as an initial estimator to start the iterative process of obtaining maximum likelihood or conditional maximum likelihood estimator.

7. Conclusions:

In this paper we have explored conditions for identifying binary errors-in-variables models. It is shown that when measurement error variances do not decrease with the sample size, contrary to the linear errors-in-variables models, it is almost impossible to get a model identified unless there are replicated observations. We have also explored conditions for the maximum likelihood estimators to be consistent and asymptotically normally distributed. Some two or three step estimators which substantially simplifies the computation are also suggested and their loss of efficiency is also explored.

The discussion in the paper is based on the assumption that all components of $\underset{\sim}{x}$ are observed with errors. Similar conclusions can easily be drawn when only part of the components of $\underset{\sim}{x}$ are observed with errors.

APPENDIX

To prove Theorem 3.2, we note that the joint likelihood function of $f(\underset{\sim}{y}, \underset{\sim}{z}; \underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0)$ can be written as the product of the conditional likelihood, $f(\underset{\sim}{y} \mid \underset{\sim}{z}; \underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0)$ and the marginal likelihood $f(\underset{\sim}{z}; \underset{\sim}{\delta}^0)$, then we have

$$\begin{aligned}\log f(\underset{\sim}{y}, \underset{\sim}{z}; \underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0) &= \log f(\underset{\sim}{y} \mid \underset{\sim}{z}; \underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0) + \log f(\underset{\sim}{z}; \underset{\sim}{\delta}^0) \\ &= \log \int f(\underset{\sim}{y} \mid \underset{\sim}{x}; \underset{\sim}{\theta}^0) f(\underset{\sim}{x} \mid \underset{\sim}{z}; \underset{\sim}{\delta}^0) d\underset{\sim}{x} + \log f(\underset{\sim}{z}; \underset{\sim}{\delta}^0).\end{aligned} \tag{A.1}$$

Hence, the information matrix of $(\underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0)$ is of the form

$$I = I_{\underset{\sim}{y}|\underset{\sim}{z}}(\underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0) + I_z(\underset{\sim}{\delta}^0), \tag{A.2}$$

where $I_{\underset{\sim}{y}|\underset{\sim}{z}}(\underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0)$ and $I_z(\underset{\sim}{\delta}^0)$ are the information matrices of $\log f(\underset{\sim}{y} \mid \underset{\sim}{z}; \underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0)$ and $\log f(\underset{\sim}{z}; \underset{\sim}{\delta}^0)$, respectively. By (2.10), we have

$$\begin{aligned}L_{\underset{\sim}{y}|\underset{\sim}{z}} = \log f(\underset{\sim}{y} \mid \underset{\sim}{z}; \underset{\sim}{\theta}^0, \underset{\sim}{\delta}^0) = \sum_i \Big\{ & y_i \log G(\underset{\sim}{b}^{0\prime} z_i) \\ & + (1 - y_i) \log\left[1 - G(\underset{\sim}{b}^{0\prime} z_i)\right] \Big\}\end{aligned} \tag{A.3}$$

Using the chain rule of diffentiation, the score vector of (A.3) is

$$\frac{\partial L_{\underset{\sim}{y}|\underset{\sim}{z}}}{\partial \underset{\sim}{\theta}} = B \sum_i \left(\frac{y_i}{G_i} - \frac{1 - y_i}{1 - G_i} \right) G_i' z_i, \tag{A.4}$$

$$\frac{\partial L_{\underset{\sim}{y}|\underset{\sim}{z}}}{\partial \underset{\sim}{\delta}} = C \sum_i \left(\frac{y_i}{G_i} - \frac{1 - y_i}{1 - G_i} \right) G_i' z_i, \tag{A.5}$$

where $B = \frac{\partial \underset{\sim}{b}'}{\partial \underset{\sim}{\theta}}, C = \frac{\partial \underset{\sim}{b}'}{\partial \underset{\sim}{\delta}}$, and $G_i = G(\underset{\sim}{b}' z_i)$. Therefore, the information matrix of (A.3) is equal to

$$I_{\underset{\sim}{y}|\underset{\sim}{z}} = -E_z E_y \Big(\frac{\partial L_{y|z}}{\partial \underset{\sim}{\gamma}} \cdot \frac{\partial L_{y|z}}{\partial \underset{\sim}{\gamma}'} \Big) = E_z \binom{B}{C} \Big(\sum_i h_i \underset{\sim}{z}_i \underset{\sim}{z}_i' \Big) (B' \; C')$$
$$= \begin{pmatrix} B\tilde{M}B' & B\tilde{M}C' \\ C\tilde{M}B' & C\tilde{M}C' \end{pmatrix} \tag{A.6}$$

where $h_i = -[G_i(1-G_i)]^{-1}(G_i')^2$ and $\tilde{M} = E(\sum_i h_i \underset{\sim}{z}_i \underset{\sim}{z}_i')$. The information matrix of $\log f(\underset{\sim}{z}; \underset{\sim}{\delta}^0)$ has the form

$$\begin{pmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & D \end{pmatrix}, \tag{A.7}$$

because $f(\underset{\sim}{z}; \underset{\sim}{\delta}^0)$ is assumed not a function of $\underset{\sim}{\theta}^0$.

If we premultiply the first p rows of (A.6) by - CB^- and add them to the last s rows we have a matrix of the form

$$\begin{pmatrix} B\tilde{M}B' & B\tilde{M}C' \\ \underset{\sim}{0} & \underset{\sim}{0} \end{pmatrix}, \tag{A.8}$$

where B^- is the left inverse of B. Therefore, for I to have full rank, it is necessary that D is of full rank s. Furthermore, if $B\tilde{M}B'$ is of full rank p, then I is of full rank $m = p + s$.

The rank condition of $B\tilde{M}B'$ is assured if M^* is a full rank matrix and B is of rank p. The rank condition of B is asured by the assumption that $\underset{\sim}{\theta}^0$ is identifiable from $(\underset{\sim}{y}, \underset{\sim}{x})$. The identification of $\underset{\sim}{\theta}^0$ also implies that the information matrix of $(\underset{\sim}{y}, \underset{\sim}{x})$ is nonsingular. The nonsingularity of the information matrix of $(\underset{\sim}{y}, \underset{\sim}{x})$ together with the relation (2.3) implies that $\tilde{M}$ will be nonsingular if $E\underset{\sim}{z}\underset{\sim}{z}'$ is nonsingular.

FOOTNOTES

* This work was partially supported by National Science Foundation grant SES88-21205. An earlier version of this paper was presented at Deuxiemes Journees D'etude Sur L'tutilisation Des Donnees De Panel held at Universite Paris Val De Marne on June 6 and 7, 1988. The revision was carried out while the author was visiting the Center for Economic Research, Tilburg University. The author wishes to thank three referees for very helpful comments. He also wishes to thank H. Cheng, F. Kozin, W. Shafer and conference participants for helpful comments and discussions.

1. For examples of models when y are measured with errors, see Stapelton and Young (1984).

2. An alternative derivation of (2.11) is to rewrite the probit model in terms of the conventional latent variable formulation of

$$y_i^* = \underset{\sim}{\theta}^{0\prime}\underset{\sim}{x}_i + v_i \quad , \quad v_i \sim N(0,1),$$

and

$$y_i = \left\{ \begin{matrix} 1 & \text{if } y_i^* > 0, \\ 0 & \text{otherwise} \end{matrix} \right\}$$

Substituting $\underset{\sim}{x}_i$ by $E(\underset{\sim}{x}_i \mid \underset{\sim}{z}_i) + \eta_i$ where $\eta_i \sim N(0, A)$ we have $y_i^* = \underset{\sim}{\theta}^{0\prime}E(\underset{\sim}{x}_i \mid \underset{\sim}{z}_i) + v_i + \underset{\sim}{\theta}^{0\prime}\underset{\sim}{\eta}_i = \underset{\sim}{\theta}^{0\prime}E(\underset{\sim}{x}_i \mid \underset{\sim}{z}_i) + v_i^*, \underset{\sim}{v}_i^* \sim N(0, 1 + \underset{\sim}{\theta}^{0\prime}A\underset{\sim}{\theta}^0)$

3. The symmetry condition in Ω has been ignored.

4. Alternatively, one may replace A.16 by some smoothness condition (e.g. McFadden (1984, p. 1407)).

5. Hajivassiliou (1989) has provided an efficient smooth unibased simulator for the score corresponding to (6.2).

6. This formulation can be viewed as a special case of those suggested by Gourieroux et.al. (1985).

References

Amemiya, T. (1981) "Qualitative Response Models: A Survey", *Journal of Economic Literature,* 19, 1483-1536.

____________ (1985), *Advanced Econometrics*, Cambridge: Harvard University Press.

Amemiya, Y. (1985) "Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model", *Journal of Econometrics* 28, 273-290.

Amemiya, Y. and W.A. Fuller (1985) "Estimation for the Nonlinear Functional Relationship", *Proceedings of the 1983 Business Economics Section, ASA*, 45-56.

____________ and ____________ (1988) "Estimation for the Nonlinear Functional Relationship", *The Annals of Statistics*, 16, 147-160.

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

Anderson, T.W. (1985), *An Introduction to Multivariate Analysis,* 2nd ed. Wiley, New York.

Anderson, T.W. and H. Rubin (1956) "Statistical Inference in Factor Analysis", in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probabilities*, vol. V, Berkeley and Los Angeles: University of California Press, 111-156.

Andrews, D.W.K. (1987), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers", *Econometrica*, 55, 1465-1471.

Efron, B. (1975) "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis", *Journal of the American Statistical Association* 70, 892-898.

Fahrmeir, L. and H. Kaufmann (1985), "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models", *The Annuals of Statistics* 13, 342-368.

Gourieroux, G. and A. Monfort (1981), "Asymptotic Properties of the Maximum Likelihood Estimator in Dichotomous Logit Models", *Journal of Econometrics* 17, 83-97.

____________ (1987), "Statistique des Modéles Econométriques", Cours polycopié ENSAE.

Gourieroux, G., A. Monfort and A. Trognon (1985), "A General Approach to Serial Correlation", *Econometric Theory*, 1, 315-340.

Hajivassiliou, V.A. (1989), "Measurement Errors in Switching Regression Models: With Applications to Price-Fixing Behaviour", Mimeo.

Hsiao, C. (1983), "Identification", in *Handbook of Econometrics*, vol. I. ed. by Z. Griliches and M. Intriligator, North-Holland: Amsterdam.

___________ (1989), "Consistent Estimation for Some Nonlinear Errors-in-Variables Models", *Journal of Econometrics*, 41, 159-185.

Lawley, D.N. and A.E. Maxwell (1971), *Factor Analysis as a Statistical Method,* London: Butterworth.

___________ (1973), "Regression and Factor Analysis", *Biometrika* 60, 331-338.

Lien, D.H. (1986), "Moments of Ordered Bivariate Log-Normal Distributions", *Economic Letters* 20, 45-47.

Maddala, G.S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press: Cambridge.

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

McFadden, D.J. (1976), "A Comment on Discriminant Analysis versus Logit Analysis", *Annals of Economic and Social Measurement,* 5, 511-523.

___________ (1981) "Econometric Models of Probabilistic Choice", in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C.F. Manski and D. McFadden, Cambridge: MIT Press.

___________ (1984), "Qualitative Response Models", in *Handbook of Econometrics*, Vol. II, edited by Z. Giriliches and M. Intriligator, Amersterdam: North-Holland, 1395-1457.

Pötscher, B.M. and I.R. Prucha (1986), "Consistency in Nonlinear Econometrics: A Generic Uniform Law of Large Numbers and Some Comments on Recent Results", Department of Economics Working Paper #86-9, University of Maryland.

Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley: New York.

Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica* 39, 577-592.

Stapelton, D.C. and D.J. Young (1984), "Censored Normal Regression with Measurement Error on the Dependent Variable", *Econometrica* 52, 737-760.

Stefanski, L.A. and R.J. Carroll (1985), "Covariate Measurement Error in Logistic Regression", *The Annals of Statistics* 13, 1335-1351.

Train, K., D. McFadden and A. Goett (1987), "The Incorporation of Attitudes in Econometric Models of Consumer Choice", *Review of Economics and Statistics* 69, 383-391.

Train, K. (1985) *Qualitative Choice Analysis*, Cambridge: MIT Press.

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

Wolter, K.M. and W.A. Fuller (1982a), "Estimation of Nonlinear Errors-in-Variables Models", *Annuals of Statistics* 10, 539-548.

___________ (1982b) "Estimation of the Quadratic Errors-in-Variables Model", *Biometrika* 69, 175-182.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

No.	Author(s)	Title
8801	Th. van de Klundert and F. van der Ploeg	Fiscal Policy and Finite Lives in Interdependent Economies with Real and Nominal Wage Rigidity
8802	J.R. Magnus and B. Pesaran	The Bias of Forecasts from a First-order Autoregression
8803	A.A. Weber	The Credibility of Monetary Policies, Policy-makers' Reputation and the EMS-Hypothesis: Empirical Evidence from 13 Countries
8804	F. van der Ploeg and A.J. de Zeeuw	Perfect Equilibrium in a Model of Competitive Arms Accumulation
8805	M.F.J. Steel	Seemingly Unrelated Regression Equation Systems under Diffuse Stochastic Prior Information: A Recursive Analytical Approach
8806	Th. Ten Raa and E.N. Wolff	Secondary Products and the Measurement of Productivity Growth
8807	F. van der Ploeg	Monetary and Fiscal Policy in Interdependent Economies with Capital Accumulation, Death and Population Growth
8901	Th. Ten Raa and P. Kop Jansen	The Choice of Model in the Construction of Input-Output Coefficients Matrices
8902	Th. Nijman and F. Palm	Generalized Least Squares Estimation of Linear Models Containing Rational Future Expectations
8903	A. van Soest, I. Woittiez, A. Kapteyn	Labour Supply, Income Taxes and Hours Restrictions in The Netherlands
8904	F. van der Ploeg	Capital Accumulation, Inflation and Long-Run Conflict in International Objectives
8905	Th. van de Klundert and A. van Schaik	Unemployment Persistence and Loss of Productive Capacity: A Keynesian Approach
8906	A.J. Markink and F. van der Ploeg	Dynamic Policy Simulation of Linear Models with Rational Expectations of Future Events: A Computer Package
8907	J. Osiewalski	Posterior Densities for Nonlinear Regression with Equicorrelated Errors
8908	M.F.J. Steel	A Bayesian Analysis of Simultaneous Equation Models by Combining Recursive Analytical and Numerical Approaches

No.	Author(s)	Title
8909	F. van der Ploeg	Two Essays on Political Economy (i) The Political Economy of Overvaluation (ii) Election Outcomes and the Stockmarket
8910	R. Gradus and A. de Zeeuw	Corporate Tax Rate Policy and Public and Private Employment
8911	A.P. Barten	Allais Characterisation of Preference Structures and the Structure of Demand
8912	K. Kamiya and A.J.J. Talman	Simplicial Algorithm to Find Zero Points of a Function with Special Structure on a Simplotope
8913	G. van der Laan and A.J.J. Talman	Price Rigidities and Rationing
8914	J. Osiewalski and M.F.J. Steel	A Bayesian Analysis of Exogeneity in Models Pooling Time-Series and Cross-Section Data
8915	R.P. Gilles, P.H. Ruys and J. Shou	On the Existence of Networks in Relational Models
8916	A. Kapteyn, P. Kooreman and A. van Soest	Quantity Rationing and Concavity in a Flexible Household Labor Supply Model
8917	F. Canova	Seasonalities in Foreign Exchange Markets
8918	F. van der Ploeg	Monetary Disinflation, Fiscal Expansion and the Current Account in an Interdependent World
8919	W. Bossert and F. Stehling	On the Uniqueness of Cardinally Interpreted Utility Functions
8920	F. van der Ploeg	Monetary Interdependence under Alternative Exchange-Rate Regimes
8921	D. Canning	Bottlenecks and Persistent Unemployment: Why Do Booms End?
8922	C. Fershtman and A. Fishman	Price Cycles and Booms: Dynamic Search Equilibrium
8923	M.B. Canzoneri and C.A. Rogers	Is the European Community an Optimal Currency Area? Optimal Tax Smoothing versus the Cost of Multiple Currencies
8924	F. Groot, C. Withagen and A. de Zeeuw	Theory of Natural Exhaustible Resources: The Cartel-Versus-Fringe Model Reconsidered

No.	Author(s)	Title
8925	O.P. Attanasio and G. Weber	Consumption, Productivity Growth and the Interest Rate
8926	N. Rankin	Monetary and Fiscal Policy in a 'Hartian' Model of Imperfect Competition
8927	Th. van de Klundert	Reducing External Debt in a World with Imperfect Asset and Imperfect Commodity Substitution
8928	C. Dang	The D_1-Triangulation of R^n for Simplicial Algorithms for Computing Solutions of Nonlinear Equations
8929	M.F.J. Steel and J.F. Richard	Bayesian Multivariate Exogeneity Analysis: An Application to a UK Money Demand Equation
8930	F. van der Ploeg	Fiscal Aspects of Monetary Integration in Europe
8931	H.A. Keuzenkamp	The Prehistory of Rational Expectations
8932	E. van Damme, R. Selten and E. Winter	Alternating Bid Bargaining with a Smallest Money Unit
8933	H. Carlsson and E. van Damme	Global Payoff Uncertainty and Risk Dominance
8934	H. Huizinga	National Tax Policies towards Product-Innovating Multinational Enterprises
8935	C. Dang and D. Talman	A New Triangulation of the Unit Simplex for Computing Economic Equilibria
8936	Th. Nijman and M. Verbeek	The Nonresponse Bias in the Analysis of the Determinants of Total Annual Expenditures of Households Based on Panel Data
8937	A.P. Barten	The Estimation of Mixed Demand Systems
8938	G. Marini	Monetary Shocks and the Nominal Interest Rate
8939	W. Güth and E. van Damme	Equilibrium Selection in the Spence Signaling Game
8940	G. Marini and P. Scaramozzino	Monopolistic Competition, Expected Inflation and Contract Length
8941	J.K. Dagsvik	The Generalized Extreme Value Random Utility Model for Continuous Choice

No.	Author(s)	Title
8942	M.F.J. Steel	Weak Exogenity in Misspecified Sequential Models
8943	A. Roell	Dual Capacity Trading and the Quality of the Market
8944	C. Hsiao	Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLANDS