A SEMIPARAMETRIC ESTIMATION OF NONLINEAR ERRORS-IN-VARIABLES MODELS*

Liqun Wang, Institute for Statistics and Econometrics, University of Basel, Holbeinstrasse 12, CH-4051 Basel, Switzerland
Cheng Hsiao, Department of Economics, University of Southern California, University Park, Los Angeles, CA 90089-0253, USA

Key Words: Measurement errors, nonparametric estimation, method of simulated moments, identification, deconvolution, simulation estimation.

1 Introduction

Measurement error (errors in-variables) models have been applied in many fields of science (e.g., Aigner et al (1984), Fuller (1987), Hsiao (1992) and Carroll, Ruppert and Stefanski (1995)). If a model is linear in variables, the issues of random measurement errors can often be overcome through the use of the instrumental variable method. If a model is nonlinear in variables, the conventional instrumental variable method, in general, does not yield consistent estimator of the unknown parameters when variables are subject to random measurement errors (e.g., Y. Amemiya (1985) and Hsiao (1989)).

To obtain consistent estimators for nonlinear errors-in-variables models, two approaches have been adopted. One is to assume that the variances of the measurement errors shrink towards zero when sample size increases (e.g., Wolter and Fuller (1982a, b), Amemiya and Fuller (1988) and Amemiya (1990)). The other is to assume that sample observations are random draws from a common population (the so called structural models, see, e.g., Kendall and Stuart (1977)). The former approach may not be applicable to data sets often encountered by economists. The latter approach will yield consistent estimators of the unknown parameters through the use of the maximum likelihood or minimum distance principle only if the conditional distribution of the measurement errors given the observed covariates are known a priori. Unfortunately, the probability distribution of the measurement errors typically is unknown to investigators unless validation data are available (e.g. Carroll and Stefanski (1990), Sepanski and Carroll (1993)).

In this paper we propose an alternative approach to derive the consistent estimators for nonlinear errors-in-variables models. We combine the nonparametric estimation method with the method of Fourier deconvolution to separate the systematic part of the regression model and the probability distribution of the unobservables. We demonstrate that, contrary to the common belief, instrumental variables do yield useful information with regard to identification and estimation of the unknown parameters. To derive the estimators, we use a simulation based procedure. While the basic idea of simulation is similar to the method of simulated moments (MSM) of McFadden (1989) or Pakes and Pollard (1989), it is different in the sense that the knowledge of the true density function of the unobservables is not required. Essentially, simulation generated from any arbitrary distribution is capable of yielding consistent and asymptotically normally distributed estimators. The method is also easier to implement than the semiparametric method recently proposed by Newey (1993). To save space, we present here only conditions and main results. More detailed derivations and proofs may be found in Wang and Hsiao (1996).

2 The Model

Consider the regression model

$$y = g(x; \theta_0) + \eta, \tag{2.1}$$

where $y \in I\!R$, $x \in I\!R^k$ and $\theta_0 \in I\!R^p$ is a vector of unknown parameters. The function $g(x;\theta_0)$ is nonlinear in x. Suppose x is unobservable. Instead we observe

$$z = x + \zeta. \tag{2.2}$$

*This work was supported in part by National Science Foundation grant SBA 94-09540. The main results of the paper was derived while the first author was visiting USC. He wishes to thank the Department of Economics and especially the econometrics group for the hospitality.

In addition we assume that there exists a instrumental variable $w \in I\!R^l$ which is related to x through

$$x = \Gamma_0 w + u, \tag{2.3}$$

where Γ_0 is a $k \times l$ matrix of unknown parameters. The η, ζ and u are unobserved errors which we assume to satisfy $E(u \mid w) = 0$, $E(\eta \mid w, u) = 0$ and $E(\zeta \mid w, u, \eta) = 0$. There is no assumption about the functional form of the error distributions. Thus, the model is semiparametric. The primary interest is to estimate the parameters θ_0, Γ_0 and the distribution of $u, F_u(u)$.

Hausman et al (1991) and Hausman, Newey and Powell (1995) analyzed a special form of model (2.1) - (2.3) where x is a scalar variable and $g(x;\theta_0)$ is a polynomial of x. Newey (1993) considered more general model of form (2.1) - (2.3) with the assumption that the error u in (2.3) is independent of the instrumental variable w. He proposed consistent estimators of the model based on the following moment equations

$$E(y \mid w) = \int g(\Gamma_0 w + u; \theta_0)\, dF_u(u), \tag{2.4}$$

$$E(zy \mid w) = \int (\Gamma_0 w + u)\, g(\Gamma_0 w + u; \theta_0)\, dF_u(u), \tag{2.5}$$

$$E(z \mid w) = \Gamma_0 w, \tag{2.6}$$

under the assumption that the parameters θ_0, Γ_0 and the distribution $F_u(u)$ are simultaneously identified by (2.4) - (2.6). Hausman et al (1991) showed that the polynomial model is identifiable. Newey (1993) conjectured that the identifiability holds for more general models. Also, Newey (1993) derived a consistent simulated moment estimator for the model when $F_u(u)$ belongs to a parametric family and a consistent semiparametric estimator when $F_u(u)$ is nonparametric but may be approximated by a parametric family.

3 Identification

Following Newey (1993) we consider the question of identifiability of the parameters θ_0, Γ_0 and the distribution $F_u(u)$ based on moment equations (2.4) - (2.6). Obviously (2.6) is a usual linear regression equation and, therefore, Γ_0 is identified in general. In the following we show how θ_0 and $F_u(u)$ are identified by (2.4) and (2.5), given that Γ_0 is identified. We make the following assumptions.

A 1 *The distribution of w is absolutely continuous w.r.t. Lebesgue measure and has support $I\!R^l$.*

A 2 Γ_0 *has full rank k.*

A 3 *The functions $g(x;\theta_0)$, $xg(x;\theta_0) \in L^1(I\!R^k)$.*

Let $m_1(\Gamma_0 w) = E(y \mid w)$ and $m_2(\Gamma_0 w) = E(zy \mid w)$. Then, since the conditional expectations in (2.4) and (2.5) depend on w only through $v = \Gamma_0 w$, (2.4) and (2.5) can be respectively written as

$$m_1(v) = \int g(v+u;\theta_0) dF_u(u), \tag{3.1}$$

$$m_2(v) = \int (v+u) g(v+u;\theta_0) dF_u(u). \tag{3.2}$$

Unless otherwise indicated explicitly, all integrals in this paper are taken to be over the space $I\!R^k$. Condition A3 implies that $m_1(v), m_2(v) \in L^1(I\!R^k)$ and the Fourier transforms $\widetilde{g}(\lambda;\theta_0), \widetilde{m}_1(\lambda)$ and $\widetilde{m}_2(\lambda)$ of $g(x;\theta_0), m_1(v)$ and $m_2(v)$ respectively exist, where, e.g.,

$$\tilde{g}(\lambda;\theta_0) = \int e^{-i\lambda' x} g(x;\theta_0)\, dx.$$

Then taking Fourier transformation on both sides of (3.1) and applying the Fubini Theorem we have

$$\begin{aligned}\widetilde{m}_1(\lambda) &= \int e^{-i\lambda' v} \int g(v+u;\theta_0) dF_u(u) dv \\ &= \widetilde{g}(\lambda;\theta_0) \widetilde{f}_u(\lambda), \end{aligned} \tag{3.3}$$

where

$$\widetilde{f}_u(\lambda) = \int e^{i\lambda' u} dF_u(u)$$

is the characteristic function of $F_u(u)$. Likewise taking Fourier transformation on both sides of (3.2) yields

$$\widetilde{m}_2(\lambda) = \widetilde{g}_\lambda(\lambda;\theta_0) \widetilde{f}_u(\lambda), \tag{3.4}$$

where

$$\widetilde{g}_\lambda(\lambda;\theta_0) = \int e^{-i\lambda' x} x g(x;\theta_0)\, dx.$$

Now, if $\widetilde{g}(\lambda;\theta_0) \neq 0$, then (3.3) is equivalent to

$$\widetilde{f}_u(\lambda) = \widetilde{m}_1(\lambda) / \widetilde{g}(\lambda;\theta_0). \tag{3.5}$$

It is apparent now that $\widetilde{f}_u(\lambda)$, hence the distribution $Fu(u)$, is uniquely determined by θ_0 through (3.5). In order to derive the condition under which θ_0 is identified, we substitute (3.5) into (3.4) and obtain

$$\widetilde{g}(\lambda;\theta_0)\, \widetilde{m}_2(\lambda) = \widetilde{g}_\lambda(\lambda;\theta_0)\, \widetilde{m}_1(\lambda). \tag{3.6}$$

It follows from the Uniqueness Theorem of Fourier transformation, that

$$\int g(\xi - v; \theta_0)\, m_2(v)\, dv = \int (\xi - v)\, g(\xi - v; \theta_0)\, m_1(v)\, dv \qquad (3.7)$$

holds almost everywhere on $\mathbb{R}^k$ (with respect to Lebesgue measure). In fact, from equations (3.1) and (3.2) it is easy to verify directly that (3.7) holds for all $\xi \in \mathbb{R}^k$. As a result, we have

$$\begin{aligned} & G(\xi; \theta_0) \\ = & \int g(\xi - v; \theta_0)\left[(\xi - v)\, m_1(v) - m_2(v)\right] dv \\ \equiv & \; 0. \end{aligned} \qquad (3.8)$$

Let $\Theta \subseteq \mathbb{R}^p$ denote the parameter space. Then we have the following results.

Theorem 1 *Suppose A1 - A3 hold for model (2.1) - (2.3) and* $\widetilde{g}(\lambda; \theta_0) \neq 0, \forall \lambda \in \mathbb{R}^k$. *Then*

(1) if there exists a point $\xi_0 \in \mathbb{R}^k$, *such that* $G(\xi_0; \theta_0) = 0$ *has unique solution* $\theta_0 \in \Theta$, *then* (θ_0, F_u) *is identified (by (2.4) and (2.5));*

(2) (θ_0, F_u) *is identified if and only if* θ_0 *is the unique point in* Θ *satisfying (3.8).*

Since (3.8) contains k equations, from Theorem 3.1 we have immediately the following identification conditions.

Corollary 2 *Under the condition of Theorem 3.1,*

(1) a necessary condition for θ_0 *to be identified by (3.8) is that* $k \geq p$;

(2) if $k \geq p$ *and the function* $G(\xi; \theta)$ *in (3.8) is differentiable at* θ_0, *then a sufficient condition for identification is that there exists* $\xi_0 \in \mathbb{R}^k$, *such that* $\operatorname{rank}\left[\partial G(\xi_0; \theta_0) / \partial \theta'\right] = p$.

4 Estimation

Let the data (y_t, z_t, w_t), $t = 1, 2, ..., T$ be given with sample size T. First we note that, if we have a consistent estimator of θ_0, say $\widehat{\theta}$, then the distribution of u can be estimated through (3.5) as

$$\widetilde{\widehat{f}}_u(\lambda) = \widetilde{\widehat{m}}_1(\lambda) / \widetilde{g}\left(\lambda; \widehat{\theta}\right), \qquad (4.1)$$

where $\widetilde{\widehat{m}}_1(\lambda)$ is the Fourier transform of $\widehat{m}_1$ which is a consistent nonparametric estimator of m_1. The estimator (4.1) is point-wise consistent under certain regularity conditions. Therefore, our focus will be on deriving consistent estimator of θ_0.

The identification condition (3.8) also suggests a method to estimate θ_0. Indeed, equation (3.8) provides k orthogonality conditions which may be used to estimate the parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$ by a method similar to the generalized method of moments (GMM) of Hansen (1982) or the method of simulated moments (MSM) of McFadden (1989) or Pakes and Pollard (1989), i.e., an estimator of θ_0 can be constructed by making the sample analog of (3.8) as close to zero as possible. This estimation procedure however may not always yield unique estimate even when the θ_0 is identified. Furthermore, it is not known generally, at which point of $\xi \in \mathbb{R}^k$ is the θ_0 uniquely determined by (3.8). To make use of condition (3.8), we propose a stochastic version of Theorem 3.1. Let $f_\xi(\xi)$ be a positive function on $\mathbb{R}^k$. Then a necessary and sufficient condition for (3.8) is

$$\int \|G(\xi; \theta_0)\|^2 f_\xi(\xi)\, d\xi = 0, \qquad (4.2)$$

where $\| \quad \|$ denotes the Euclidean norm. Then an estimator of θ_0 may be obtained by minimizing the function

$$\widetilde{Q}(\theta) = \int \|G(\xi; \theta)\|^2 f_\xi(\xi)\, d\xi, \qquad (4.3)$$

where

$$G(\xi; \theta) = \int g(\xi - v; \theta)\left[(\xi - v)\, m_1(v) - m_2(v)\right] dv. \qquad (4.4)$$

However, $\widetilde{Q}(\theta)$ is a multiple integral which often causes complications and difficulties in numerical computation. To make the idea operational, we "discretize" the integral (4.3) by

$$Q(\theta) = \frac{1}{S} \sum_{s=1}^{S} \|G(\xi_s; \theta)\|^2, \qquad (4.5)$$

where $\xi_1, \xi_2, ..., \xi_S$ are randomly generated from an arbitrary density function $f_\xi(\xi)$ having support $\mathbb{R}^k$ and S is large enough such that $\partial^2 Q(\theta_0) / \partial\theta\partial\theta'$ is nonsingular (see assumption A18 below). It is clear that under some mild conditions $Q(\theta)$ converges in probability to $\widetilde{Q}(\theta)$ uniformly in a neighborhood of $\theta_0 \in \Theta$.

Thus we propose the following procedure of estimation:

Step 1. From (2.6) estimate Γ_0 by the LS estimator

$$\widehat{\Gamma} = \left(\sum_{t=1}^{T} z_t w_t'\right)\left(\sum_{t=1}^{T} w_t w_t'\right)^{-1}. \qquad (4.6)$$

Then let $v_t = \widehat{\Gamma} w_t, t = 1, 2, ..., T$ and estimate the density function $f_v(v)$ of $v = \Gamma_0 w$, the conditional mean functions $m_1(v) = E(y \mid v)$ and $m_2(v) = E(zy \mid v)$ by kernel method as

$$\widehat{f}_v(v) = \frac{1}{Th_T^k} \sum_{t=1}^{T} K\left(\frac{v - v_t}{h_T}\right), \tag{4.7}$$

$$\widehat{m}_1(v) = \frac{1}{Th_T^k} \sum_{t=1}^{T} y_t K\left(\frac{v - v_t}{h_T}\right) / \widehat{f}_v(v) \tag{4.8}$$

and

$$\widehat{m}_2(v) = \frac{1}{Th_T^k} \sum_{t=1}^{T} z_D y_t K\left(\frac{v - vOt}{h_T}\right) / \widehat{f}_v(v), \tag{4.9}$$

where $K(\cdot)$ is the kernel function and h_T is the bandwidth.

Step 2. Approximate the integral (4.4) by

$$\widehat{G}_T(\xi;\theta) = \tfrac{1}{T}\textstyle\sum_{t=1}^{T} I\left(\left|\widehat{f}_v(v_t)\right| \geq b_T\right) g(\xi - v_t;\theta) \times \left[(\xi - v_t)\widehat{m}_1(v_t) - \widehat{m}_2(v_t)\right] / \widehat{f}_v(v_t), \tag{4.10}$$

where $I(\cdot)$ is the indicator function and b_T are positive constants satisfying $b_T \to 0$ as $T \to \infty$.

Step 3. Construct the sample analog of (4.5) as

$$Q_T(\theta) = \frac{1}{S} \sum_{s=1}^{S} \left\| \widehat{G}_T(\xi_s;\theta) \right\|^2 \tag{4.11}$$

where each term $\widehat{G}_T(\xi_s;\theta)$ is computed according to (4.10).

Step 4. The simulation estimator (SE) $\widehat{\theta}_T$ is defined as the measurable function satisfying

$$Q_T\left(\widehat{\theta}_T\right) = \inf_{\theta \in \Theta} Q_T(\theta). \tag{4.12}$$

The consistency of the SE $\widehat{\theta}_T$ may be derived following the traditional fashion by establishing the uniform convergence of $Q_T(\theta)$ to $Q(\theta)$ which has unique minimizer $\theta_0 \in \Theta$ and Θ is compact. From (4.4) - (4.5) and (4.10) - (4.11) it is easily seen that the convergence of $Q_T(\theta)$ to $Q(\theta)$ requires the consistencies of the LS and nonparametric estimators (4.6) - (4.9) in the first step of the estimation procedure. In fact, even the uniform convergence of the first stage estimators are desired. In this paper we use the results of Andrews (1995) on the rate of uniform convergence of kernel estimators of density functions and conditional mean functions.

Definition 1 *Let $\mathcal{D}_q$, $q \geq 1$, be the class of all real functions $f(\cdot)$ on $\mathbb{R}^k$ such that all partial derivatives of order* 0 *through q are continuous and uniformly bounded.*

To use the results of Andrews (1995), we assume that

A 4 $(y_t, z_t, w_t), t = 1, 2, ..., T$ *are independent and identically distributed.*

A 5 $Ey^2 < \infty, E\|yz\|^2 < \infty, E\|w\|^4 < \infty$ *and* $M_w = Eww'$ *is nonsingular.*

A 6 *For some* $q \geq 1$*, the functions* $f_v(v), m_1(v), m_2(v) \in \mathcal{D}_q$.

A 7 *For the $q \geq 1$ in A6, the kernel function $K(v)$ is bounded on $\mathbb{R}^k$ and satisfies:*

(1) $\int K(v)\,dv = 1$ *and* $\int v_1^{q_1} v_2^{q_2} \cdots v_k^{q_k} K(v)\,dv = 0$, *for* $q_j \geq 0$ *and* $1 \leq \sum_{j=1}^{k} q_j \leq q - 1$;

(2) $\int \|v\|^j |K(v)|\,dv < \infty$, *for* $j = 0$ *or* q;

(3) $\sup_{v \in \mathbb{R}^k} \|\partial K(v)/\partial v\| (\|v\| + 1) < \infty$;

(4) $\int e^{i\lambda' v} K(v)\,dv \in L^1\left(\mathbb{R}^k\right)$.

A 8 *As* $T \to \infty, h_T \to 0, b_T \to 0, Th_T^{2k} b_T^6 \to \infty$ *and* $h_T^q b_T^{-3} \to 0$, *where $q \geq 1$ is as in A6.*

To derive the consistency of the SE $\widehat{\theta}_T$ defined by (4.12), we assume further that

A 9 *The function $g(x;\theta)$ satisfies*

(1) $\sup_{\theta \in \Theta} \|g(x;\theta)\|$, $\sup_{\theta \in \Theta} \|xg(x;\theta)\| \in L^1\left(\mathbb{R}^k\right)$;

(2) $\sup_{\theta \in \Theta} \|\partial g(x;\theta)/\partial x\|$ *and* $\sup_{\theta \in \Theta} \|\partial xg(x;\theta)/\partial x\|$ *are uniformly bounded.*

A 10 θ_0 *is the unique point in Θ for which* $Q(\theta_0) = 0$ *and* $\Theta \subset \mathbb{R}^p$ *is compact.*

Then the consistency of $\widehat{\theta}_T$ is given in the following theorem.

Theorem 3 *Under A1 – A10,* $\widehat{\theta}_T \xrightarrow{P} \theta_0$, *as* $T \to \infty$.

Similar to the consistency, the asymptotic normality of $\widehat{\theta}_T$ also may be obtained in the traditional way by first Taylor expanding the derivative of $Q_T(\theta)$ at θ_0 and then showing that the Hessian $\partial^2 Q_T/\partial\theta\partial\theta'$ converges to a nonsingular matrix and the gradient $\partial Q_T/\partial\theta$ times $\sqrt{T}$ has an asymptotic normal distribution. However, as in the case of Robinson (1188), the derivation becomes much more complicated because of the presence of nonparametric estimators in function $Q_T(\theta)$, which have the convergence rate lower than $\sqrt{T}$. To achieve the $\sqrt{T}$-consistency of his semiparametric estimator, Robinson (1988) used higher order kernels combined with certain smoothness conditions for the density and conditional mean functions. Essentially, he assumed the density and conditional mean functions belong to $\mathcal{G}_{\mu}^{\alpha}, \alpha > 0, \mu > 0$, which is defined as a class of functions $f : I\!R^k \to I\!R$ satisfying: (1) $f(\cdot)$ is $(q-1)$-times partially differentiable, $q - 1 < \mu \leq q$; (2) for some $\rho > 0$, $\sup_{\|u-v\|<\rho} |f(u) - f(v) - F(u,v)| / \|u - v\|^{\mu} \leq \gamma(v)$ for all v, where $F = 0$, when $q = 1$; and F is a $(q-1)$-th degree homogeneous polynomial in $u - v$ with coefficients the partial derivatives of f at v of orders 1 through $(q-1)$, when $q > 1$; and (3) the function $\gamma(\cdot)$, $f(\cdot)$ and all its partial derivatives of order $q - 1$ and less have finite α-th moments. It is easy to see that every function in $\mathcal{D}_q$ belongs to $\mathcal{G}_q^{\alpha}$ and, thus, $\mathcal{D}_q \subseteq \mathcal{G}_q^{\alpha}$.

Following Robinson (1988), to obtain the $\sqrt{T}$-consistency, we use the product kernel $K(v) = \prod_{j=1}^{k} \kappa(v_j)$ in the nonparametric estimators (4.8) - (4.10), where $\kappa(\cdot)$ is a univariate kernel and v_j is the j-th component of $v \in I\!R^k$. However, to adapt to our consistency assumptions A7, we use a modification of his definition for the class of kernel functions.

Definition 2 *Let $\mathcal{K}_q, q \geq 1$, be the class of all even functions $\kappa(\cdot) : I\!R \to I\!R$ satisfying*

(1) $\int_{I\!R} r^j \kappa(r)\, dr = \delta_{0b}, j = 0, 1, ..., q-1$, *where δ_{ij} is Kronecker's delta;*

(2) $\kappa(r) = O\left(\left(1 + |r|^{q+1+\epsilon}\right)^{-1}\right)$, *for some* $\epsilon > 0$;

(3) $\sup_{r \in I\!R} |\partial \kappa(r) / \partial r| (|r| + 1) < \infty$ *and* $\sup_{r \in I\!R} \left|\partial^2 \kappa(r) / \partial r^2\right| < \infty$;

(4) $\int e^{i\mu r} \kappa(r)\, dr \in L^1(I\!R)$.

Thus, we make the following assumption.

A 11 *For the $q \geq 1$ in A6, the kernel function $K(v) = \prod_{j=1}^{k} \kappa(v_j)$ with $\kappa(\cdot) \in \mathcal{K}_{2q-1}$.*

A 12 *As $T \to \infty, Th_T^{4k+2} b_T^6 \to \infty$ and $Th_T^{4q} b_T^{-4} \to 0$, where $q \geq 1$ is as in A6.*

It is easily seen that every kernel function $K(v)$ satisfying A11 satisfies A7 too. The following discussion and result apply not only to the estimator defined by (4.12) but also to those satisfying the score equation $\partial Q_T(\theta) / \partial \theta = 0$,though we will continue to use the notation $\widehat{\theta}_T$. The estimators defined as the roots of the score equation are local optima. As far as the local optima are concerned, only the local analogs of A9 and A10 are needed.

A 13 *In addition to A9, $g(x; \theta_0), xg(x; \theta_0) \in \mathcal{D}_2$.*

A 14 *For each $\xi \in I\!R^k$, the following moments are finite:*

$$A(\pi) = E\left[\frac{m_1(v)}{f_v(v)} \frac{\partial g(\xi - v; \theta_0)(\xi - v)}{\partial(\text{vec}\,\Gamma)'} - \frac{m_2(v)}{f_v(v)} \frac{\partial g(\xi - v; \theta_0)}{\partial(\text{vec}\,\Gamma)'}\right],$$

$$B(\xi) = Eg(\xi - v; \theta_0) \times \left[(\xi - v)\frac{\partial[m_1(v)/f_v(v)]}{\partial(\text{vec}\,\Gamma)'} - \frac{\partial[m_2(v)/f_v(v)]}{\partial(\text{vec}\,\Gamma)'}\right],$$

where $v = \Gamma_0 w$ and vec *is the column vector operator.*

A 15 *For each $\xi \in I\!R^k$,*

$$E\left\|I\left(\left|\widehat{f}_v\left(\widehat{\Gamma}w\right)\right| < b_T\right) F(\xi)\right\| = o\left(T^{-1/2}\right),$$

where

$$F(\xi) = g(\xi - v; \theta_0)\left[(\xi - v) m_1(v) - m_2(v)\right] / f_v(v).$$

A 16 *For each $\xi \in I\!R^k$, the covariance matrix $\Sigma(\xi) = EF(\xi) F(\xi)'$ exists, where $F(\xi)$ is defined in A15. The covariance matrix $\Sigma_\Gamma = E\left[(z - \Gamma_0 w)(z - \Gamma_0 w)' \mid w\right]$ exists and does not depend on w.*

A 17 *Θ contains an open neighborhood of θ_0 in which $\partial g(x; \theta) / \partial \theta$ and $\partial^2 g(x; \theta) / \partial \theta \partial \theta'$ exist and have the same property as A9 for the function $g(x; \theta)$.*

A 18 *The matrix $H = \frac{1}{S} \sum_{s=1}^{S} D(\xi_s) D(\xi_s)'$ is nonsingular, where*

$$D(\xi) = \int \frac{\partial g(\xi - v; \theta_0)}{\partial \theta} \left[(\xi - v) m_1(v) - m_2(v)\right]' dv.$$

Then we have the following result.

Theorem 4 *Under A1 - A18, for any estimator $\widehat{\theta}_T$ satisfying (4.12),*

$$\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \xrightarrow{L} N\left(0, H^{-1} \Omega H^{-1}\right),$$

where

$$\Omega = \frac{1}{S} \sum_{s=1}^{S} D(\xi_s) \{\Sigma(\xi_s) + [A(\xi_s) + B(\xi_s)] \left(M_w^{-1} \otimes \Sigma_\Gamma\right) [A(\xi_s) + B(\xi_s)]'\} D(\xi_s)'.$$

References

[1] Aigner, D.J., C. Hsiao, A. Kapteyn and T. Wansbeek (1984). Latent Variable Models in Econometrics. In Z. Griliches and M.D. Intriligator eds. *Handbook of Econometrics,* Vol. II, North-Holland, Amsterdam.

[2] Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press.

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

[4] Amemiya, Y. (1990). Two-Stage Instrumental Variable Estimators for the Nonlinear Errors-in-Variables Model. *Journal of Econometrics*, 44, 311-322.

[5] Amemiya, Y. and W.A. Fuller (1988). Estimation for the Nonlinear Functional Relationship. *Annals of Statistics*, 16, 147-160.

[6] Andrews, D.W.K. (1995). Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, 11, 560-596.

[7] Carroll, R.J., D. Ruppert and L.A. Stefanski (1995). *Measurement Error in Nonlinear Models.* Chapman & Hall, London.

[8] Carroll, R.J. and L.A. Stefanski (1990). Approximate Quasi-likelihood Estimation in Models with Surrogate Predictors. *Journal of American Statistical Association,* 85, 652-663.

[9] Fuller, A.W. (1987). *Measurement Error Models.* Wiley, New York.

[10] Hansen, L.P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica,* 50, 1029-1054.

[11] Hausman, J.A., W.K. Newey, H. Ichimura and J.L. Powell (1991). Estimation of Polynomial Errors in Variables Models. *Journal of Econometrics*, 50, 273-295.

[12] Hausman, J.A., W.K. Newey and J.L. Powell (1995). Nonlinear Errors in Variables: Estimation of Some Engel Curves. *Journal of Econometrics*, 65, 205-233.

[13] Hsiao, C. (1989). Consistent Estimation for Some Nonlinear Erosion-Variables Models. *Journal of Econometrics*, 41, 159-185.

[14] Hsiao, C. (1992). Nonlinear Latent Variable Models. In L. Matyas and P. Sevestre eds., The Econometrics of Panel Data, 242-261.

[15] Kendall, M.G. and A. Stuart (1977). *The Advanced Theory of Statistics.* 4th ed. Hafner, New York.

[16] McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica,* 57, 995-1026.

[17] Newey, W.K. (1993). Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models. Manuscript.

[18] Pakes, A. and D. Pollard (1989). Simulation and the Asymptotics of Optimization Estimators. *Econometrica,* 57, 1027-1057.

[19] Robinson, P.M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica*, 56, 931-954.

[20] Sepanski J.H. and R.J. Carroll (1993). Semiparametric Quasilikelihood and Variance Function Estimation in Measurement Error Models. *Journal of Econometrics*, 58, 223-256.

[21] Wang, Liqun and Cheng Hsiao (1996). A Simulated Semiparametric Estimation of Nonlinear Errors-in-Variables Models. *WWZ-Discussion Papers, No. 9602,* Center of Economics and Business Administration, University of Basel.

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

[23] Wolter, K.M. and W.A. Fuller (1982b). Estimation of the Quadratic Errors-in-Variables Models. *Biometrika*, 69, 175-182.