A Simulated Semiparametric Estimation of Nonlinear Errors-in-Variables Models*

Liqun Wang[1,2] and Cheng Hsiao[1]

[1]Department of Economics, University of Southern California, Los Angeles, USA
[2]Institute for Statistics and Econometrics, University of Basel, Switzerland

October 1995

Abstract

This paper deals with a nonlinear errors-in-variables model where the error distributions are nonparametric and a prediction equation for unobserved covariates is available. Consistent and asymptotically normally distributed estimator is constructed using the combined approach of the method of nonparametric estimation and the method of simulated moments (MSM) of McFadden (1989) or Pakes and Pollard (1989). Necessary and sufficient conditions of identifiability of the model are derived by using the Fourier deconvolution method.

Keywords: Measurement error models, nonparametric estimation, identification, Fourier transform, deconvolution, method of simulated moments, simulation estimation.

1 Introduction

Measurement errors occur frequently (e.g. Aigner et al (1984), Fuller (1987) and Hsiao (1992)). 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. Amemiya and Fuller (1988) and Wolter and Fuller (1982a, b)). The other is to assume that sample observations are random draws from a common population (the

*This work was supported in part by the Swiss National Science Foundation and US National Science Foundation grant SBA 94-09540.

so called structural errors-in-variables 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 unobservables 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).

In section 2 we set up the basic nonlinear errors-in-variables model under the assumption that the instruments exist. Section 3 considers the identification condition. Section 4 proposes a simulated semiparametric estimator which is consistent and asymptotically normally distributed. Conclusions are in section 5. The proofs of the theorems and lemmas are in appendices A and B.

2 The Model

Consider the regression model

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

where y is the dependent variable, x is a k dimensional explanatory variable and θ_0 is a p dimensional vector of unknown parameters. The function $g(x; \theta_0)$ is nonlinear in x. The explanatory variable x is assumed to be related to the l dimensional instrumental variable w by

$$x = \Gamma_0 w + u, \tag{2.2}$$

where Γ_0 is a $k \times l$ matrix of unknown parameters and the error u is assumed to be independent of w. Suppose x is unobservable. Instead we observe z, where

$$z = x + \zeta. \tag{2.3}$$

In other words, we observe y, w and z, but not x. The η, ζ and u are unobserved errors which we assume to satisfy $E(\eta \mid w, u) = 0$ 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) analyzed a special form of model (2.1) - (2.3) where $g(x;\theta_0)$ is a polynomial of x. Hausman, Newey and Powell (1995) considered the general regression model via polynomial approximations. Newey (1993) proposed consistent estimators of model (2.1) - (2.3) based on the following moment equations

$$\begin{aligned} E(y \mid w) &= \int g(\Gamma_0 w + u;\theta_0) dF_u(u), \\ E(zy \mid w) &= \int (\Gamma_0 w + u) g(\Gamma_0 w + u;\theta_0) dF_u(u), \\ E(z \mid w) &= \Gamma_0 w, \end{aligned} \tag{2.4}$$

under the assumption that the parameters θ_0, Γ_0 and the distribution $F_u(u)$ are simultaneously identified by (2.4). Hausman et al (1991) showed that the polynomial model is identifiable. Newey (1993) conjectured that the identifiability holds for more general models. Assuming the model to be identifiable, Newey (1993) derived a consistent simulated moment estimator for the model where $F_u(u)$ 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). That is, given the observed information (the left-hand side of (2.4)), are θ_0, Γ_0 and $F_u(u)$ uniquely determined by (2.4)? Thus, in this paper, that $(\theta_0, \Gamma_0, F_u)$ are identified means that they are uniquely determined by (2.4).

Obviously the last equation of (2.4) 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 the first two equations of (2.4), given that Γ_0 is identified. We assume that:

A 1. *The distribution of $w, F_w(\cdot)$, is absolutely continuous with respect to Lebesgue measure and has support $\mathbb{R}^l$.*

A 2. Γ_0 *has full rank k.*

A 3. *The functions $g(x;\theta_0)$, $xg(x;\theta_0) \in L^1(\mathbb{R}^k)$, the space of all absolutely integrable functions on $\mathbb{R}^k$.*

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) depend on w only through $v = \Gamma_0 w$, the first two equations of (2.4) can be respectively written as

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

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

In this paper, unless otherwise indicated explicitly, all integrals are taken to be over the space $I\!R^k$. Considered as functions of $v \in I\!R^k$, $m_1(v)$ and $m_2(v)$ are well-defined by (3.1) and (3.2). Furthermore, condition A3 implies that $m_1(v), m_2(v) \in L^1(I\!R^k)$ and the Fourier transforms $\tilde{g}(\lambda;\theta_0), \tilde{m}_1(\lambda)$ and $\tilde{m}_2(\lambda)$ 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

$$\begin{aligned} \tilde{m}_1(\lambda) &= \int e^{-i\lambda' v} \int g(v+u;\theta_0)dF_u(u)dv && (3.3) \\ &= \int [\int e^{-i\lambda'(v+u)} g(v+u;\theta_0)dv] e^{i\lambda' u} dF_u(u) && (3.4) \\ &= \int e^{-i\lambda' x} g(x;\theta_0)dx \cdot \int e^{i\lambda' u} dF_u(u) && (3.5) \\ &= \tilde{g}(\lambda;\theta_0)\tilde{f}_u(\lambda), && (3.6) \end{aligned}$$

where $\tilde{f}_u(\lambda) = \int e^{i\lambda' u} dF_u(u)$ is the characteristic function of $F_u(u)$. Likewise taking Fourier transformation on both sides of (3.2) yields

$$\tilde{m}_2(\lambda) = \tilde{g}_\lambda(\lambda;\theta_0)\tilde{f}_u(\lambda), \tag{3.7}$$

where

$$\begin{aligned} \tilde{g}_\lambda(\lambda;\theta_0) &= \int e^{-i\lambda' x} x g(x;\theta_0)dx \\ &= i\frac{\partial \tilde{g}(\lambda;\theta_0)}{\partial \lambda}. \end{aligned}$$

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

$$\tilde{f}_u(\lambda) = \frac{\tilde{m}_1(\lambda)}{\tilde{g}(\lambda;\theta_0)}. \tag{3.8}$$

It is apparent now that $\tilde{f}_u(\lambda)$, hence the distribution $F_u(u)$, 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

$$\tilde{g}(\lambda;\theta_0)\tilde{m}_2(\lambda) = \tilde{g}_\lambda(\lambda;\theta_0)\tilde{m}_1(\lambda). \tag{3.9}$$

It follows from the Uniqueness Theorem of Fourier transformation, that

$$\int g(\xi - v;\theta_0)m_2(v)dv = \int (\xi - v)g(\xi - v;\theta_0)m_1(v)dv \tag{3.10}$$

holds almost everywhere on $I\!R^k$ (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 $\xi \in I\!R^k$. As a result, we have

$$G(\xi;\theta_0) = \int g(\xi - v;\theta_0)\left[(\xi - v)m_1(v) - m_2(v)\right]dv \equiv 0. \tag{3.11}$$

Let $\Theta \subseteq I\!R^p$ denote the parameter space. Then we have the following results.

Theorem 3.1. *(Pointwise Identification). Suppose A1 - A3 hold for model (2.1) - (2.3) and* $\tilde{g}(\lambda;\theta_0) \neq 0, \forall\lambda \in I\!R^k$. *Then*

1. *if there exists a point* $\xi_0 \in I\!R^k$, *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));*

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 3.1. *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* $\xi_0 \in I\!R^k$, *such that*

$$rank(\frac{\partial G(\xi_0;\theta_0)}{\partial\theta'}) = p.$$

Remark 3.1. *Assumptions A1 and A2 are made to ensure that the conditional moments* $m_1(v)$ *and* $m_2(v)$ *are fully observed. It is easily seen that technically the derivation of the results in this section still goes through even without assumptions A1 and A2, because* $m_1(v)$ *and* $m_2(v)$ *are well-defined functions by the right-hand sides of (3.1) and (3.2). In this sense* $m_1(v)$ *and* $m_2(v)$ *may be viewed as extensions of the conditional expectations* $E(y \mid w)$ *and* $E(zy \mid w)$ *respectively. However,* $m_1(v)$ *and* $m_2(v)$ *may not be observed at every point* $v \in I\!R^k$ *without A1 or A2. In applications point of view neither is the assumption A3 as restrictive as it appears, because the relation* $g(x;\theta_0)$ *can be considered as holding in a sufficiently large compact subset of* $I\!R^k$ *and outside this compact subset it can be redefined such that A3 is satisfied.*

Remark 3.2. *The condition in Theorem 3.1 that* $\tilde{g}(\lambda;\theta_0) \neq 0$, *for all* $\lambda \in I\!R^k$ *may be replaced by the condition that* $\tilde{g}(\lambda;\theta_0) \neq 0$ *at* $\lambda = 0$ *and the characteristic function* $\tilde{f}Wu(\lambda)$ *is analytic. This is easy to see because* $\tilde{g}(\lambda;\theta_0)$ *is a continuous function of* λ *and, hence,* $\tilde{g}(0;\theta_0) \neq 0$ *implies* $\tilde{g}(\lambda;\theta_0) \neq 0$ *in a neighborhood of zero, which in turn implies that (3.5) holds in a neighborhood of zero. Since now* $\tilde{f}_u(\lambda)$ *is analytic, (3.5) must hold for all* $\lambda \in I\!R^k$.

The Fourier transformation turns out to be a useful tool for solving the problem of identifiability, because the convolution of two functions can be transformed into the product of the corresponding Fourier counterparts. In order to get more insight about the structure of model (2.1) - (2.3), we proceed to find the solution $\tilde{g}(\lambda;\theta_0)$ of the differential equation (3.6). As is shown in Appendix A, the solution is given by

$$\tilde{g}(\lambda;\theta_0) = \tilde{g}(0;\theta_0)e^{h(\lambda)}, \tag{3.12}$$

where

$$h(\lambda) = \int_0^{\lambda_j} \frac{\tilde{m}_{2j}(\lambda)}{i\tilde{m}_1(\lambda)} d\lambda_j$$

and λ_j and $\tilde{m}_{2j}(\lambda)$ are the j-th component of λ and $\tilde{m}_2(\lambda)$ respectively. By (3.5)

$$\tilde{f}_u(\lambda) = \frac{\tilde{m}_1(\lambda)e^{-h(\lambda)}}{\tilde{g}(0;\theta_0)}. \tag{3.13}$$

Equations (3.9) and (3.10) suggest that the Fourier transform $\tilde{g}(\lambda;\theta_0)$ and the characteristic function $\tilde{f}_u(\lambda)$ can be respectively decomposed as a known function of θ_0 multiplied by a known function of observed information. Furthermore, if the left-hand sides of (3.9) and (3.10) are absolutely integrable on $I\!R^k$, then applying the Fourier inversion formula to them yields

$$g(x;\theta_0) = \frac{\tilde{g}(0;\theta_0)}{(2\pi)^k}\int e^{i\lambda' x + h(\lambda)}d\lambda \tag{3.14}$$

and

$$f_u(u) = \frac{1}{\tilde{g}(0;\theta_0)(2\pi)^k}\int e^{-i\lambda' u - h(\lambda)}\tilde{m}_1(\lambda)d\lambda. \tag{3.15}$$

To summarize, we have the following results with regard to the structure of the model.

Theorem 3.2. *Under the conditions of Theorem 3.1,*

1. *the Fourier transform $\tilde{g}(\lambda;\theta_0)$ has representation (3.9);*
2. *the characteristic function $\tilde{f}_u(\lambda)$ has representation (3.10);*
3. *if $\tilde{g}(\lambda;\theta_0) \in L^1(I\!R^k)$, then the regression function $g(x;\theta_0)$ has representation (3.11);*
4. *the density function of u, if exists, has representation (3.12).*

To illustrate the results of this section, we consider a simple example.

Example 3.1. *Consider a simple case of model (2.1) - (2.3) where all variables are scalars, the regression function $g(x;\theta_0) = e^{-\theta_0 x^2}$ with $\theta_0 > 0$ and $\Gamma_0 = 1$. Suppose the true distribution $F_u(u)$ of u is the standard normal distribution $N(0,1)$. Then by (3.1) and (3.2) it is straightforward to calculate*

$$m_1(v) = (1+2\theta_0)^{-1/2}\exp(-\frac{\theta_0 v^2}{1+2\theta_0}),$$

$$m_2(v) = (1+2\theta_0)^{-3/2}v\exp(-\frac{\theta_0 v^2}{1+2\theta_0}).$$

To see that (3.8) identifies θ_0, we calculate, for any $\theta > 0$,

$$G(\xi;\theta) = \frac{\sqrt{\pi}(\theta-\theta_0)}{(1+2\theta_0)(\theta+\theta_0+2\theta\theta_0)^{3/2}}\exp(-\frac{\theta\theta_0\xi^2}{\theta+\theta_0+2\theta\theta_0})$$

which equals zero if and only if $\theta = \theta_0$. To see the decompositions (3.9) - (3.12), we calculate further the Fourier transforms

$$\tilde{m}_1(\lambda) = \sqrt{\frac{\pi}{\theta_0}}\exp(-\frac{(1+2\theta_0)\lambda^2}{4\theta_0}),$$

$$\tilde{m}_2(\lambda) = -\frac{i\lambda}{2\theta_0}\sqrt{\frac{\pi}{\theta_0}}\exp(-\frac{(1+2\theta_0)\lambda^2}{4\theta_0}).$$

It follows that

$$\frac{\tilde{m}_2(\lambda)}{i\tilde{m}_1(\lambda)} = -\frac{\lambda}{2\theta_0}$$

and hence

$$h(\lambda) = -\int_0^\lambda \frac{\lambda}{2\theta_0}d\lambda = -\frac{\lambda^2}{4\theta_0}.$$

Therefore, the Fourier transform of $g(x;\theta)$,

$$\tilde{g}(\lambda;\theta_0) = \sqrt{\frac{\pi}{\theta_0}}\exp(-\frac{\lambda^2}{4\theta_0}) = \tilde{g}(0;\theta_0)e^{h(\lambda)}$$

and , by (3.5), the characteristic function is given by

$$\tilde{f}_u(\lambda) = \frac{\tilde{m}_1(\lambda)}{\tilde{g}(\lambda;\theta_0)} = e^{-\lambda^2/2}$$

which is easily seen to satisfy (3.10). The decompositions (3.11) and (2.12) can be verified analogously.

4 Estimation

In this section we consider estimation of model (2.1) - (2.3). 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 $\hat{\theta}$, then the distribution of u can be estimated through (3.5) as

$$\hat{\tilde{f}}_u(\lambda) = \frac{\tilde{\hat{m}}_1(\lambda)}{\tilde{g}(\lambda;\hat{\theta})}, \tag{4.1}$$

where $\tilde{\hat{m}}_1(\lambda)$ is the Fourier transform of $\hat{m}_1$ which is a consistent nonparametric estimator of m_1. The estimator (4.1) is pointwise consistent under certain regularity conditions. Therefore, our focus will be on deriving consistent estimator of θ_0. For ease of reading, we will only state the conditions and results. The proofs are given in Appendix B.

The identification condition (3.8) also suggests a method to estimate θ_0. Since it is not known generally, at which point of $\xi \in I\!R^k$ 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 function on $I\!R^k$. Then a necessary condition for (3.8) is

$$\int G(\xi;\theta_0)f_\xi(\xi)d\xi = 0. \tag{4.2}$$

However, (4.2) is not sufficient for (3.8). A necessary and sufficient condition for (3.8) is

$$\int \|G(\xi;\theta_0)\|^2 f_\xi(\xi)d\xi = 0, \tag{4.3}$$

where $\|\cdot\|$ denotes the Euclidean norm and $f_\xi(\xi)$ is positive on $I\!R^k$.

Theorem 4.1. *(Integrated Identification). Let $f_\xi(\xi)$ be a function on $I\!R^k$. Then under the conditions of Theorem 3.1,*

1. *if (4.2) has a unique solution $\theta_0 \in \Theta$, then (θ_0, F_u) is identified;*
2. *if $f_\xi(\xi) > 0, \forall \xi \in I\!R^k$, then (θ_0, F_u) is identified if and only if $\theta_0 \in \Theta$ is the unique solution of (4.3).*

4.1 A Simulation Estimator (SE)

Equation (4.2) provides k orthogonality conditions which may be used to estimate the parameter $\theta_0 \in \Theta \subset I\!R^p$ 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 (4.2) as close to zero as possible. This estimation procedure however may not always yield unique estimate even when the θ_0 is identified.

To derive a more general estimation procedure, we use the condition (4.3) directly. Specifically, define

$$\tilde{Q}(\theta) = \int \|G(\xi;\theta)\|^2 f_\xi(\xi) d\xi, \tag{4.4}$$

where

$$G(\xi;\theta) = \int g(\xi - v;\theta)\left[(\xi - v)m_1(v) - m_2(v)\right] dv. \tag{4.5}$$

Then an estimator of θ_0 may be obtained by minimizing the function $\tilde{Q}(\theta)$. However, $\tilde{Q}(\theta)$ is a multiple integral which often causes complications and difficulties in numerical computation. To make the idea operational, we "discretize" the integral (4.4) by

$$Q(\theta) = \frac{1}{S}\sum_{s=1}^{S} \|G(\xi_s;\theta)\|^2, \tag{4.6}$$

where $\xi_1, \xi_2, ..., \xi_S$ are randomly generated from an arbitrary density function $f_\xi(\xi)$ having support $I\!R^k$ and S is large enough such that $\partial^2 Q(\theta_0)/\partial\theta\partial\theta'$ is nonsingular (see assumption A16 below). It is clear that under some mild conditions $Q(\theta)$ converges in probability to $\tilde{Q}(\theta)$ uniformly in a neighborhood of $\theta_0 \in \Theta$.

Thus we propose the following procedure of estimation:

Step 1. From the third equation of (2.4) estimate Γ_0 by the LS estimator

$$\hat{\Gamma} = (\sum_{t=1}^{T} z_t w_t')(\sum_{t=1}^{T} w_t w_t')^{-1}. \tag{4.7}$$

Then let $v_t = \hat{\Gamma} w_t, t = 1, 2, ..., T$ 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

$$\hat{f}_v(v) = \frac{1}{Th_T^k}\sum_{t=1}^{T} K(\frac{v - v_t}{h_T}), \tag{4.8}$$

$$\hat{m}_1(v) = \frac{1}{Th_T^k} \sum_{t=1}^{T} y_t K(\frac{v - v_t}{h_T}) / \hat{f}_v(v) \tag{4.9}$$

and

$$\hat{m}_2(v) = \frac{1}{Th_T^k} \sum_{t=1}^{T} z_t y_t K(\frac{v - v_t}{h_T}) / \hat{f}_v(v), \tag{4.10}$$

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

Step 2. Approximate the integral (4.5) by

$$\hat{G}_T(\xi;\theta) = \frac{1}{T} \sum_{t=1}^{T} \frac{I(|\hat{f}_v(v_t)| \geq b_T) g(\xi - v_t;\theta)}{\hat{f}_v(v_t)} \left[(\xi - v_t)\hat{m}_1(v_t) - \hat{m}_2(v_t) \right], \tag{4.11}$$

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 (4.6) as

$$Q_T(\theta) = \frac{1}{S} \sum_{s=1}^{S} \|\hat{G}_T(\xi_s;\theta)\|^2 \tag{4.12}$$

where each term $\hat{G}_T(\xi_s;\theta)$ is computed according to (4.11).

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

$$Q_T(\hat{\theta}_T) = \inf_{\theta \in \Theta} Q_T(\theta). \tag{4.13}$$

The asymptotic properties of the estimator $\hat{\theta}_T$ are derived in the next subsection.

Remark 4.1. *It should be noted here that there are several ways to proceed to construct consistent estimators for θ_0, depending on the definition of the optimality criterion. McFadden (1989) defines the MSM estimators $\hat{\theta}$ which satisfy*

$$Q_T(\hat{\theta}) \leq \inf_{\theta \in \Theta} Q_T(\theta) + O_p(1). \tag{4.14}$$

This definition is similar to that of the estimators considered by Pakes and Pollard (1989) in which $O_p(1)$ is replaced by $o_p(1)$. One advantage of criterion (4.14) is that the estimators thus defined always exist, even if the infimum in (4.13) is not attained. The later can be avoided when the parameter space Θ is subject to certain restrictions, e.g., when Θ is compact. Clearly any estimator satisfying (4.13) satisfies (4.14) too. The estimators defined by (4.13) and (4.14) are global optima. In general, to compute the global minimum can be a burdensome task. In such case it is more convenient to define the estimators as the roots of the score equation

$$\frac{\partial Q_T(\theta)}{\partial \theta} = \frac{2}{S} \sum_{s=1}^{S} \frac{\partial \hat{G}_T'(\xi_s;\theta)}{\partial \theta} \hat{G}_T(\xi_s;\theta) = 0. \tag{4.15}$$

The asymptotic properties of the estimators under these criteria are discussed in the next two subsections.

4.2 Consistency

The consistency of the SE $\hat{\theta}_T$ 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 (4.5) - (4.6) and (4.11) - (4.12) it is easily seen that the convergence of $Q_T(\theta)$ to $Q(\theta)$ requires the consistencies of the LS and nonparametric estimators (4.7) - (4.10) in the first step of the estimation procedure. In fact, even the uniform convergence of the first stage estimators are desired. There is large amount of papers in the literature dealing with convergence of nonparametric estimators. One of the most recent one is Andrews (1995), which gives results on the rate of uniform convergence of nonparametric estimators of density functions and conditional mean functions under general conditions.

Definition 4.1. Let $\mathcal{D}_q$, $q \geq 1$, be the class of all real functions $f(\cdot)$ on $I\!R^k$ 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. $Ey^2 < \infty, E\|yz\|^2 < \infty, E\|w\|^4 < \infty$, *$Eww'$ is nonsingular and $l = k$.*

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

(4) $\int e^{i\lambda' v} K(v)dv \in L^1(I\!R^k)$.

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 $\hat{\theta}_T$ defined by (4.13), we assume further that

A 9. *The function $g(x;\theta)$ satisfies that, for each $\xi \in I\!R^k$,*

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

(2) $E \sup_{\theta \in \Theta} \|g(\xi - v;\theta)\| < \infty$ *and* $E \sup_{\theta \in \Theta} \|(\xi - v) g(\xi - v;\theta)\| < \infty$;

(3) $E \sup_{\theta \in \Theta} \|g(\xi - v;\theta) [(\xi - v)m_1(v) - m_2(v)] / f_v(v)\|^2 < \infty$.

A 10. $\Theta \subset I\!R^p$ *is compact.*

A 11. θ_0 *is the unique point in Θ for which $Q(\theta_0) = 0$.*

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

Theorem 4.2. *Under A1 – A11,* $\hat{\theta}_T \xrightarrow{P} \theta_0$, *as* $T \to \infty$.

4.3 Asymptotic Normality

Similar to the consistency, the asymptotic normality of $\hat{\theta}_T$ 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 gradian $\partial Q_T/\partial\theta$ times $\sqrt{T}$ has an asymptotic normal distribution. However, as in the case of Robinson (1988), 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 $\mathcal{G}^\alpha_\mu, \alpha > 0, \mu > 0$, 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 homogenous 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 $\mathcal{G}^\alpha_q$ and, thus, $\mathcal{D}_q \subseteq \mathcal{G}^\alpha_q$.

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

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

(2) $\kappa(r) = O((1 + |r|^{q+1+\epsilon})^{-1})$, 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) $Tdinte^{i\mu r}\kappa(r)dr \in L^1(I\!R)$.

Thus, we make the following assumption.

A 12. *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 13. *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 A12 satisfies A7 too. The following discussion and result apply not only to the estimator defined by (4.13) but also to those satisfying (4.15), though we will continue to use the notation $\hat{\theta}_T$. The estimators defined as the roots of the score equation (4.15) are local optima. As far as the local optima are concerned, only the local analogs of A9 - A11 are needed.

A 14. *θ_0 is an interior point of Θ.*

A 15. *In addition to A9, the function $g(x;\theta)$ satisfies*

(1! $g(x;\theta_0), xg(x;\theta_0) \in \mathcal{D}_2$;

(2) There is 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 16. *For each $\xi \in I\!R^k$,*

$$E\|\frac{m_1(v)}{f_v(v)}\frac{\partial g(\xi - v;\theta_0)(\xi - v)}{\partial(vec\Gamma)'} - \frac{m_2(v)}{f_v(v)}\frac{\partial g(\xi - v;\theta_0)}{\partial(vec\Gamma)'}\|^2 < \infty,$$

$$E\|g(\xi - v;\theta_0)[(\xi - v)\frac{\partial\,[m_1(v)/f_v(v)]}{\partial(vec\Gamma)'} - \frac{\partial\,[m_2(v)/f_v(v)]}{\partial(vec\Gamma)'}]\|^2 < \infty.$$

where vec is the column vector operator.

A 17. *The matrix $H = \frac{1}{S}\sum_{s=1}^{S} D(\xi_s;\theta_0)D(\xi_s;\theta_0)'$ is nonsingular, where*

$$D(\xi;\theta_0) = \int \partial g(\xi - v;\theta_0)/\partial\theta\,[(\xi - v)m_1(v) - m_2(v)]'\,dv. \tag{4.16}$$

A 18. *For each $\xi \in I\!R^k$, $V(\xi;\theta_0) = \lim_{T\to\infty} ET\hat{G}_T(\xi;\theta_0)\hat{G}_T(\xi;\theta_0)'$ exists.*

Then we have the following result.

Theorem 4.3. *Under A1 - A18, for any estimator $\hat\theta_T$ satisfying (4.15), $\sqrt{T}(\hat\theta_T - \theta_0)\xrightarrow{L} N(0, H^{-1}VH^{-1})$, where $V = \frac{1}{S}\sum_{s=1}^{S} D(\xi_s;\theta_0)V(\xi_s;\theta_0)D(\xi_s;\theta_0)'$.*

5 Conclusion

In this paper we combine the nonparametric estimation of conditional moments and the Fourier deconvolution method to separate the systematic part of the regression model from the errors. We demonstrate that, contrary to the common belief, instrumental variables do yield useful information with regard to the identification and estimation of the unknown parameters and the probability distribution of the errors in nonlinear errors-in-variables models. We propose a simulated method of moment estimator. However, this simulation based estimator is different from the conventional simulated moments estimators in the sense that there is no need to perform the simulation based on the probability distribution of the unobservables. In fact, simulation generated by any arbitrary distribution is capable of yielding consistent and asymptotically normally distributed estimators and the rate of convergence is $\sqrt{T}$. This remarkable result is achieved through the combined use of nonparametric estimation and the Fourier deconvolution method. Although in this paper we have confined the application of this approach to analyze the nonlinear errors-in-variables models, it appears that this novel approach should have wider applicability.

Appendices

A Proof of (3.9)

The result (3.9) follows from the following lemma.

Lemma A.1. *Suppose $p(x,y),\ q(x,y) \in C^1(\mathbb{R}^2)$ and $f(x,y)$ satisfies the differential equations*

$$\frac{\partial f(x,y)}{\partial x} = f(x,y)p(x,y), \tag{A.1}$$

$$\frac{\partial f(x,y)}{\partial y} = f(x,y)q(x,y). \tag{A.2}$$

Then the solution to (A.1)-(A.2) is given by

$$f(x,y) = e^{\int p(x,y)dx + c_1} = e^{\int q(x,y)dy + c_2},$$

where c_1 and c_2 are constants.

Proof: Considering the solution to each equation of (A.1)-(A.2) we have

$$f(x,y) = e^{\int p(x,y)dx + p_1(y)} = e^{\int q(x,y)dy + q_1(x)}$$

which implies

$$\int p(x,y)dx + p_1(y) = \int q(x,y)dy + q_1(x). \tag{A.3}$$

Furthermore, (A.1)-(A.2) and equality $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial x$ imply that $\partial p/\partial y = \partial q/\partial x$. Then differentiating both sides of equation (A.3) about x yields

$$\begin{aligned} p(x,y) &= \int \frac{\partial q(x,y)}{\partial x} dy + \frac{\partial q_1(x)}{\partial x} \\ &= p(x,y) + \frac{\partial q_1(x)}{\partial x}. \end{aligned}$$

It follows that $\partial q_1(x)/\partial x \equiv 0$ and, hence, $q_1(x)$ is a constant. Similarly $p_1(y)$ can be shown to be a constant too. □

B Proofs of Results in Section 4

Throughout the Appendix B we will maintain the following convention and notations.

1. For the sake of notational simplicity, when it does not cause confusion we will omit the arguments of some functions and denote, e.g., $\hat{g}_t = g(\xi - \hat{\Gamma} w_t; \theta)$, $\hat{f}_t = \hat{f}_v(\hat{\Gamma} w_t)$, $\hat{m}_{it} = \hat{m}_i(\hat{\Gamma} w_t)$, $\hat{r}_{it} = \hat{m}_{it}\hat{f}_t$, $J_t = I(|\hat{f}_t| \ge b_T)$ and

$$\begin{aligned} \hat{G}_T &= \frac{1}{T}\sum_{t=1}^{T} J_t \hat{g}_t (\xi - \hat{\Gamma} w_T)\hat{m}_{1t}/\hat{f}_t - \frac{1}{T}\sum_{t=1}^{T} J_t \hat{g}_t \hat{m}_{2t}/\hat{f}_t \\ &= \hat{G}_{1T} - \hat{G}_{2T}. \end{aligned}$$

2. Define $\bar{f}_v(v)$ and $\bar{m}_i(v)$, $i = 1, 2$ as the analog of $\hat{f}_v(v)$ and $\hat{m}_i(v)$ respectively where $\hat{\Gamma}$ is replaced by Γ_0 (see (4.8) – (4.10)). Further, denote $g_t = g(\xi - \Gamma_0 w_t; \theta)$, $\bar{f}_t = \bar{f}_v(\Gamma_0 w_t)$, $\bar{m}_{it} = \bar{m}_i(\Gamma_0 w_t)$, $\bar{r}_{it} = \bar{m}_{it}\bar{f}_t$ and

$$\begin{aligned}\bar{G}_T &= \frac{1}{T}\sum_{t=1}^{T} J_t g_t(\xi - \Gamma_0 w_t)\bar{m}_{1t}/\bar{f}_t - \frac{1}{T}\sum_{t=1}^{T} J_t g_t \bar{m}_{2t}/\bar{f}_t \\ &= \bar{G}_{1T} - \bar{G}_{2T}.\end{aligned}$$

3. Similarly, let $f_t = f_v(\Gamma_0 w_t)$, $m_{it} = m_i(\Gamma_0 w_t)$, $r_{it} = m_{it} f_t$ and

$$\begin{aligned}\tilde{G}_T &= \frac{1}{T}\sum_{t=1} NT J_t g_t(\xi - \Gamma_0 w_t) m_{1t}/f_t - \frac{1}{T}\sum_{t=1}^{T} J_t g_t m_{2t}/f_t \\ &= \tilde{G}_{1T} - \tilde{G}_{2T}.\end{aligned}$$

Further, let

$$\begin{aligned}G_T &= \frac{1}{T}\sum_{t=1}^{T} g_T(\xi - \Gamma_0 w_t) m_{1t}/f_t - \frac{1}{T}\sum_{t=1}^{T} g_t m_{2t}/f_t \\ &= G_{1T} - G_{2T}.\end{aligned}$$

4. If $F(A)$ is a matrix function of a matrix argument A, then the partial derivative is denoted as $\partial F/\partial A = \partial \mathrm{vec} F/\partial(\mathrm{vec} A)'$ (Magnus and Neudecker (1988)).

5. The phrase "with probability converging to one" is denoted as $w.p. \to 1$.

6. C always denotes a generic constant.

The proofs of Theorem 4.2 and Theorem 4.3 are based on the following Lemmas.

Lemma B.1. *Under A1 – A8,*

$$\sup_{v\in \mathbb{R}^k} \left|\hat{f}_v(v) - f_v(v)\right| = O_p(T^{-1/2}h_T^{-k} + h_T^q), \tag{B.1}$$

$$\sup_{|\hat{f}_v(v)|\geq b_T} \|\hat{m}_i(v) - m_i(v)\| = O_p(T^{-1/2}h_T^{-k}b_T^{-2} + h_T^q b_T^{-2}), \tag{B.2}$$

$$\sup_{v\in \mathbb{R}^k} \left|\bar{f}_v(v) - f_v(v)\right| = O_p(T^{-1/2}h_T^{-k} + h_T^q), \tag{B.3}$$

$$\sup_{|\bar{f}_v(v)|\geq b_T} \|\bar{m}_i(v) - m_i(v)\| = O_p(T^{-1/2}h_T^{-k}b_T^{-2} + h_T^q b_T^{-2}). \tag{B.4}$$

Proof: It is easy to check that all conditions of Theorem 1 of Andrews (1995) are implied by A1 – A8 and, therefore, the results follow immediately from that theorem. □

Lemma B.2. *Under A1 – A8,*

$$\sup_{w\in \mathbb{R}^k} \left|\hat{f}_v(\hat{\Gamma}w) - \bar{f}_v(\Gamma_0 w)\right| = O_p(T^{-1/2}h_T^{-k}), \tag{B.5}$$

$$\sup_{|\hat{f}_v(\hat{\Gamma}w)|\geq b_T} \left\|\hat{m}_i(\hat{\Gamma}w) - \bar{m}_i(\Gamma_0 w)\right\| = O_p(T^{-1/2}h_T^{-k}b_T^{-1}). \tag{B.6}$$

Proof: For any $w \in I\!R^k$, by the Mean-value Theorem we have

$$\hat{f}_v(\hat{\Gamma}w) - \bar{f}_v(\Gamma_0 w) 5 \frac{1}{Th_T^k} \sum_{j=1}^{T} \frac{\partial K(\tilde{v}_j)}{\partial v\prime} \frac{(\hat{\Gamma} - \Gamma_0)(w - w_j)}{h},$$

where $\tilde{v}_j$ lies between $\hat{\Gamma}(w - w_j)/h$ and $\Gamma_0(w - w_j)/h$. It follows from A7, $\sqrt{T}\|\hat{\Gamma} - \Gamma_0\| = O_p(1)$ and $\sup_w \|(w - w_j)/h\| / (\|\tilde{v}_j\| + 1) = O_p(1)$, that

$$\left|\hat{f}_v(\hat{\Gamma}w) - \bar{f}_v(\Gamma_0 w)\right| \leq \frac{\|\hat{\Gamma} - \Gamma_0\|}{Th_T^k} \sum_{j=1}^{T} \|\frac{\partial K(\tilde{v}_j)}{\partial v}\| \|\frac{w - w_j}{h}\| = O_p(T^{-1/2} h_T^{-k}),$$

which implies (B.5). Analogous it can be shown that, for $i = 1, 2$,

$$\sup_{w \in I\!R^k} \left\|\hat{r}_i(\hat{\Gamma}w) - \bar{r}_i(\Gamma_0 w)\right\| = O_p(T^{-1/2} h_T^{-k}). \tag{B.7}$$

Further, because of (B.5) and A8,

$$\lim_{T \to \infty} P\big(\inf_{\left|\hat{f}_v(\hat{\Gamma}w)\right| \geq b_T} \left|\bar{f}_v(\Gamma_0 w)\right| \geq \frac{b_T}{2}\big) = 1. \tag{B.8}$$

Hence, by (B.4) and A6, we have, $w.p. \to 1$,

$$\sup_{|\hat{f}_v| \geq b_T} \|\bar{m}_i(\Gamma_0 w)\| \leq \sup_{|\hat{f}_v| \geq b_T} \|\bar{m}_i(\Gamma_0 w) - m_i(\Gamma_0 w)\| + \|m_i(\Gamma_0 w)\| = O_p(1). \tag{B.9}$$

(B.6) follows then from

$$\|\hat{m}_i(\hat{\Gamma}w) - \bar{m}_i(\Gamma_0 w)\| \leq \frac{\|\hat{r}_i(\hat{\Gamma}w) - \bar{r}_i(\Gamma_0 w)\|}{\left|\hat{f}_v(\hat{\Gamma}w)\right|} + \frac{\|\bar{m}_i(\Gamma_0 w)\| \|\hat{f}_v(\hat{\Gamma}w) - \bar{f}_v(\Gamma_0 w)\|}{\left|\hat{f}_v(\hat{\Gamma}w)\right|},$$

(B.1), (B.7) and (B.9). □

Lemma B.3. *Under A1 – A9, for any* $\xi \in I\!R^k$, $\sup_{\theta \in \Theta} \left\|\hat{G}_T(\xi; \theta) - G(\xi; \theta)\right\| = o_p(1)$.

Proof: For any $\xi \in I\!R^k$, first we show that

$$\sup_{\theta \in \Theta} \|\hat{G}_T - \tilde{G}_T\| \leq \sup_{\theta \in \Theta} \|\hat{G}_{1T} - \tilde{G}_{1T}\| + \sup_{\theta \in \Theta} \|\hat{G}_{2T} - \tilde{G}_{2T}\| = o_p(1). \tag{B.10}$$

For any $\theta \in \Theta$, we have

$$\begin{aligned} \|\hat{G}_{2T} - \tilde{G}_{2T}\| &\leq \frac{1}{T} \sum_{t=1}^{T} J_t \|\hat{g}_t - g_t\| \|\frac{\hat{m}_{2t}}{\hat{f}_t}\| + \frac{1}{T} \sum_{t=1}^{T} J_t \, \|g_t\| \, \|\frac{\hat{m}_{2t}}{\hat{f}_t} - \frac{m_{2t}}{f_t}\| \\ &= A_1 + A_2. \end{aligned} \tag{B.11}$$

Now since the first derivative $\partial g(x; \theta)/\partial x$ is bounded by A9(1) and

$$J_t \, \|\hat{m}_{2t}\| \leq J_t \, \|\hat{m}_{2t} - m_{2t}\| + \|m_{2t}\| = O_p(1)$$

by (B.2) and A6, it follows from the mean-value theorem that

$$A_1 \leq \frac{C\|\hat{\Gamma}-\Gamma_0\|}{Tb_T}\sum_{t=1}^{T} J_t\,\|w_t\|\,\|\hat{m}_{2t}\| = O_p(T^{-1/2}b_T^{-1}). \tag{B.12}$$

Further, because of (B.1)

$$\lim_{T\to\infty} P(\inf_{|\hat{f}_v(\hat{\Gamma}w)|\geq b_T} |f_v(\Gamma_0 w)| \geq \frac{b_T}{2}) = 1. \tag{B.13}$$

It follows from Lemma B.1 and A9(2) that, $w.p. \to 1$,

$$\begin{aligned} A_2 &\leq \frac{1}{T}\sum_{t=1}^{T} J_t\,\|g_t\|\,(\|\frac{\hat{m}_{2t}-m_{2t}}{\hat{f}_t}\| + \|\frac{m_{2t}(\hat{f}_t - f_t)}{\hat{f}_t f_t}\|) \\ &\leq O_p(T^{-1/2}h_T^{-k}b_T^{-3} + h_T^q b_T^{-3}). \end{aligned} \tag{B.14}$$

Thus, $\|\hat{G}_{2T} - \tilde{G}_{2T}\| = o_p(1)$ by (B.11) - (B.14) and A8. Furthermore, since only the functions $g(x;\theta)$ and $\partial g(x;\theta)/\partial x$ involve the parameter θ, it is easily seen by A9 that (B12) and (B.14) hold uniformly in $\theta \in \Theta$, which together with (B.11) and A8 implies $\sup_{\theta\in\Theta}\|\hat{G}_{2T} - \tilde{G}_{2T}\| = o_p(1)$. Completely analogous it can be shown that $\sup_{\theta\in\Theta}\|\hat{G}_{1T} - \tilde{G}_{1T}\| = o_p(1)$, which implies (B.10).

Next we show that $\sup_{\theta\in\Theta}\|\tilde{G}_T - G\| = o_p(1)$. First by Cauchy-Schwarz inequality

$$\begin{aligned} E\sup_{\theta\in\Theta}\|\tilde{G}_T - G_T\| &\leq E\sup_{\theta\in\Theta}\|(1-J_t)g_t\left[(\xi-\Gamma_0 w_t)m_{1t} - m_{2t}\right]/f_t\| \\ &\leq [P(\left|\hat{f}_t\right| < b_T)E\sup_{\theta\in\Theta}\|g_t\left[(\xi-\Gamma_0 w_t)m_{1t} - m_{2t}\right]/f_t\|^2]^{1/2} \\ &= o(1), \end{aligned}$$

where the last equation follows from A9(3) and the fact

$$\begin{aligned} P(\left|\hat{f}_t\right| < b_T) &\leq P(|f_t| < b_T + \left|\hat{f}_t - f_t\right|) \\ &= o(1). \end{aligned}$$

Finally, by A9(3) and Theorem 4.2.1 of Amemiya (1985), $\sup_{\theta\in\Theta}\|G_T - G\| = o_p(1)$, which completes the proof. □

Proof of Theorem 4.2*:* First, since S is finite and fixed and the Euclidean norm $\|\cdot\|$ is a continuous function, Lemma B.3 implies that $Q_T(\theta)$ converges to $Q(\theta)$ in probability uniformly in $\theta \in \Theta$. Then the consistency of the estimator $\hat{\theta}_T$ follows from A10, A11 and Theorem 4.1.1 of Amemiya (1985). □

Lemma B.4. *For any* $\xi \in \mathbb{R}^k$,

$$\hat{G}_T(\xi;\theta_0) = \bar{G}_T(\xi;\theta_0) + (A_T + B_T)vec(\hat{\Gamma} - \Gamma_0) + o_p(T^{-1/2}), \tag{B.15}$$

where

$$A_T = \frac{1}{T}\sum_{t=1}^{T}\frac{J_t}{\bar{f}_t}[\bar{m}_{1t}\frac{\partial g(\xi-\Gamma_0 w_t;\theta_0)(\xi-\Gamma_0 w_t)}{\partial\Gamma} - \bar{m}_{2t}\frac{\partial g(\xi-\Gamma_0 w_t;\theta_0)}{\partial\Gamma}],$$

$$B_T = \frac{1}{T}\sum_{t=1}^{T} J_t g(\xi-\Gamma_0 w_t;\theta_0)[(\xi-\Gamma_0 w_t)\frac{\partial(\bar{m}_{1t}/\bar{f}_t)}{\partial\Gamma} - \frac{\partial(\bar{m}_{2t}/\bar{f}_t)}{\partial\Gamma}],$$

$$\frac{\partial g(\xi-\Gamma_0 w_t;\theta_0)}{\partial\Gamma} = -\frac{\partial g(\xi-\Gamma_0 w_t;\theta_0)}{\partial x'}(w_t\otimes I_k)',$$

$$\frac{\partial(\bar{m}_{it}/\bar{F}_t)}{\partial\Gamma} = \frac{1}{\bar{f}_t^2 T h^{k+1}}\sum_{j=1}^{T}(z_j y_j - 2\bar{m}_{at})\frac{\partial}{\partial v'}K(\frac{\Gamma_0 w_t - \Gamma_0 w_j}{h})(w_t - w_j)'\otimes I_k,$$

and where $\otimes$ is the Kronecker product and I_k is the $k\times k$ identity matrix.

Proof: First we consider the decomposition of $\hat{G}_{2T}$. By A15(1) the function $\hat{g}_t$ has the second order Taylor expansion at Γ_0:

$$g(\xi-\hat{\Gamma}w_t;\theta_0) = g(\xi-\Gamma_0 w_t;\theta_0) + \frac{\partial g(\xi-\Gamma_0 w_t;\theta_0)}{\partial\Gamma}\text{vec}(\hat{\Gamma}-\Gamma_0) + r_t, \tag{B.16}$$

where

$$r_t = \text{vec}(\hat{\Gamma}-\Gamma_0)'(wWt\otimes I_k)\frac{\partial^2 g(\xi-\tilde{\Gamma}w_t;\theta_0)}{\partial x\partial x'}(w_t\otimes I_k)'\text{vec}(\hat{\Gamma}-\Gamma_0)$$

and where $\tilde{\Gamma}$ lies on the line segment connecting $\hat{\Gamma}$ and Γ_0. Similarly, by A12 the second order Taylor expansion of $\hat{m}_{2t}/\hat{f}_t$ at Γ_0 is

$$\frac{\hat{m}_{2t}}{\hat{f}_t} = \frac{\bar{m}_{2t}}{\bar{f}_t} + \frac{\partial(\bar{m}_{2t}/\bar{f}_t)}{\partial\Gamma}\text{vec}(\hat{\Gamma}-\Gamma_0) + R_t, \tag{B.17}$$

where

$$R_t = \text{vec}(\hat{\Gamma}-\Gamma_0)'\frac{\partial^2(\tilde{m}_{it}/\tilde{f}_t)}{\partial\Gamma^2}\text{vec}(\hat{\Gamma}-\Gamma_0),$$

$\tilde{f}_t$ and $\tilde{m}_{it}$ are defined as $\bar{f}_t$ and $\bar{m}_{it}$ with Γ_0 replaced by $\tilde{\Gamma}$, and $\partial^2(\tilde{m}_{it}/\tilde{f}_t)/\partial\Gamma^2$ is given by

$$\begin{aligned}
&\frac{1}{\tilde{f}_t^2 T h^{k+2}}\sum_{j=1}^{T}[(w_t-w_j)\otimes I_k\otimes(z_j y_j - 2\tilde{m}_{it})]\frac{\partial^2 K(\tilde{v}_j)}{\partial v\partial v'}(w_t-w_j)'\otimes I_k\\
&-\frac{2}{\tilde{f}_t^3 T^2 h^{2(k+1)}}\sum_{j=1}^{T}[(w_t-w_j)\otimes I_k\otimes(z_j y_j - 3\tilde{m}_{it})]\frac{\partial K(\tilde{v}_j)}{\partial v}\cdot\sum_{j=1}^{T}\frac{\partial K(\tilde{v}_j)}{\partial v'}(w_t-w_j)'\otimes I_k\\
&-\frac{2}{\tilde{f}_t^3 T^2 h^{2(k+1)}}\sum_{j=1}^{T}[(w_t-w_j)\otimes I_k]\frac{\partial K(\tilde{v}_j)}{\partial v}\otimes I_k\cdot\sum_{j=1}^{T}z_j y_j\frac{\partial K(\tilde{v}_j)}{\partial v'}(w_t-w_j)'\otimes I_k
\end{aligned}$$

and where $\tilde{v}_j = \tilde{\Gamma}(w_t - w_j)/h$. It is easy to verify by A15(1) that

$$r_t = \|w_t\|^2 O_p(T^{-1})$$

and by A5–A8 and A12 that

$$R_t = (\|w_t\|+1)^2 O_p(T^{-1}h_T^{-k-2}b_T^{-2}) + O_p(T^{-1}h_T^{-2k}b_T^{-3}).$$

Substituting the right-hand side of (B.16) and (B.17) into $\hat{G}_{2T}$ yields

$$\hat{G}_{2T} = \frac{1}{T}\sum_{t=1}^{T} J_t g_t \bar{m}_{2t}/\bar{f}_t + \frac{1}{T}\sum_{t=1}^{T}(J_t \bar{m}_{2t}/\bar{f}_t)\frac{\partial g_t}{\partial \Gamma}\text{vec}(\hat{\Gamma}-\Gamma_0)$$
$$+ \frac{1}{T}\sum_{t=1}^{T} J_t g_t \frac{\partial(\bar{m}_{2t}/\bar{f}_t)}{\partial \Gamma}\text{vec}(\hat{\Gamma}-\Gamma_0) + o_p(T^{-1/2}), \tag{B.18}$$

where the last term is $o_p(T^{-1/2})$ because of A13. The analog of (B.18) for $\hat{G}_{1T}$ can be derived in the same way, which, combined with (B.18), leads to (B.15). □

Lemma B.5. *For any* $\xi \in I\!R^k$,

$$A_T \xrightarrow{P} A = E[\frac{m_1(v)}{f_v(v)}\frac{\partial g(\Xi - v;\theta_0)(\xi - v)}{\partial \Gamma} - \frac{m_2(v)}{f_v(v)}\frac{\partial g(\xi - v;\theta_0)}{\partial \Gamma}], \tag{B.19}$$

$$B_T \xrightarrow{P} B = Eg(\xi - v;\theta_0)[(\xi - v)\frac{\partial\left[m_1(v)/f_v(v)\right]}{\partial \Gamma} - \frac{\partial\left[m_2(v)/f_v(v)\right]}{\partial \Gamma}]. \tag{B.20}$$

Proof: Analogous to the decomposition of $\hat{G}_T$, we denote

$$A_T = \frac{1}{T}\sum_{t=1}^{T}\frac{J_t \bar{m}_{1t}}{\bar{f}_t}\frac{\partial(\xi - \Gamma_0 w_t)g_t}{\partial \Gamma} - \frac{1}{T}\sum_{t=1}^{T}\frac{J_T \bar{m}_{2t}}{\bar{f}_t}\frac{\partial g_t}{\partial \Gamma}$$
$$= A_{1T} - A_{2T}$$

and

$$\tilde{A}_T = \frac{1}{T}\sum_{t=1}^{T}\frac{J_t m_{1t}}{f_t}\frac{\partial(\xi - \Gamma_0 w_t)g_t}{\partial \Gamma} - \frac{1}{T}\sum_{t=1}^{T}\frac{J_t m_{2t}}{f_t}\frac{\partial g_t}{\partial \Gamma}$$
$$= \tilde{A}_{1T} - \tilde{A}_{2T}.$$

Then by A9(1)

$$\|A_{2T} - \tilde{A}_{2T}\| \le \frac{1}{T}\sum_{t=1}^{T} J_t \|(\frac{\bar{m}_{2t}}{\bar{f}_t} - \frac{m_{2t}}{f_t})\frac{\partial g_t}{\partial x'}(w_t \otimes I_k)'\|$$
$$\le \frac{C}{T}\sum_{t=1}^{T} J_t\|w_t\|(\|\frac{\bar{m}_{2t} - m_{2t}}{\bar{f}_t}\| + \|\frac{m_{2t}(\bar{f}_t - f_t)}{\bar{f}_t f_t}\|). \tag{B.21}$$

It follows from (B.8), (B.13), Lemma B.1 and A8 that, $w.p. \to 1$,

$$\left\|A_{2T} - \tilde{A}_{2T}\right\| \le O_p(T^{-1/2}h_T^{-k}b_T^{-3} + h_T^q b_T^{-3}) = o_p(1).$$

Analogous we have $A_{1T} = \tilde{A}_{1T} + o_p(1)$ and, hence, $A_T = \tilde{A}_T + o_p(1)$. Further, by Cauchy-Schwarz inequality and A16

$$E\|\tilde{A}_T - \frac{1}{T}\sum_{t=1}^{T}[\frac{m_{1t}}{f_t}\frac{\partial g_t(\xi - \Gamma_0 w_t)}{\partial \Gamma} - \frac{m_{2t}}{f_t}\frac{\partial g_t}{\partial \Gamma}]\|$$
$$\le \{P(\left|\hat{f}_t\right| < b_T)E\|\frac{m_{1t}}{f_t}\frac{\partial(\xi - \Gamma_0 w_t)g_t}{\partial \Gamma} - \frac{m_{2t}}{f_t}\frac{\partial g_t}{\partial \Gamma}\|^2\}^{1/2}$$
$$= o(1).$$

It follows that

$$A_T = \frac{1}{T}\sum_{t=1}^{T}[\frac{m_{1t}}{f_t}\frac{\partial g_t(\xi - \Gamma_0 w_t)}{\partial \Gamma} - \frac{m_{2t}}{f_t}\frac{\partial g_t}{\partial \Gamma}] + o_p(1),$$

which implies (B.19) by the Markov law of large numbers (LLN). (B.20) may be proved analogously. □

Lemma B.6. *For any* $\xi \in I\!R^k$, $\bar{G}_T(\xi;\theta_0) = G_T(\xi;\theta_0) + o_p(T^{-1/2})$.

Proof: As the proof of Lemma B.5, we first show that $\bar{G}_T = \tilde{G}_T + o_p(T^{-1/2})$. Write

$$\begin{aligned} \bar{G}_{2T} - \tilde{G}_{2T} &= \frac{1}{T}\sum_{t=1}^{T} J_t g_t \frac{\bar{m}_{2t} - m_{2t}}{\bar{f}_t} - \frac{1}{T}\sum_{t=1}^{T} J_t g_t \frac{m_{2t}(\bar{f}_t - f_t)}{\bar{f}_t f_t} \\ &= A_1 + A_2. \end{aligned}$$

Then

$$\begin{aligned} E\|A_1\|^2 &= \frac{1}{T^2} E \sum_{t,s} \frac{J_t g_t J_s g_s}{\bar{f}_t \bar{f}_s}(\bar{m}_{2t} - m_{2t})'(\bar{m}_{2s} - m_{2s}) \\ &= \frac{1}{T} E \frac{J_1 g_1^2}{\bar{f}_1^2}\|\bar{m}_{21} - m_{21}\|^2 + \frac{2(T-1)}{T} E \frac{J_1 g_{1t} J_2 g_2}{\bar{f}_1 \bar{f}_2}(\bar{m}_{21} - m_{21})'(\bar{m}_{22} - m_{22}) \\ &= B_1 + B_2. \end{aligned}$$

Now

$$\begin{aligned} B_1 &= \frac{1}{T} E \frac{J_1 g_1^2}{\bar{f}_1^2}\| \frac{1}{\bar{F}_1 T h_T^k}\sum_{j=1}^{T} y_j z_j K_{1j} - m_{21}\|^2 \\ &\leq \frac{1}{T^3 h_T^{2k} b_T^4} E g_1^2 \|\sum_{j=1}^{T}(y_j z_j - m_{21})K_{1j}\|^2 \\ &\leq \frac{2}{T^3 h_T^{2k} b_T^4} E g_1^2 \|(y_1 z_1 - m_{21})K(0)\|^2 + \frac{2}{T^3 h_T^{2k} b_T^4} E g_1^2 \|\sum_{j=2}^{T}(y_j z_j - m_{21})K_{1j}\|^2, \end{aligned}$$

where

$$K_{tj} = K(\frac{v_t - v_j}{h_T}).$$

By A5 and A15(1),

$$\frac{2}{T^3 h_T^{2k} b_T^4} E g_1^2 \|(y_1 z_1 - m_{21})K(0)\|^2 = O(T^{-3} h_T^{-2k} b_T^{-4}).$$

It follows that

$$\begin{aligned} B_1 &\leq O(T^{-3} h_T^{-2k} b_T^{-4}) + \frac{2(T-1)}{T^3 h_T^{2k} b_T^4} E g_1^2 \|(y_2 z_2 - m_{21})K_{12}\|^2 \\ &\quad + \frac{2(T-1)(T-2)}{T^3 h_T^{2k} b_T^4} E g_1^2 K_{12}^2 K_{13}^2 (y_2 z_2 - m_{21})'(y_3 z_3 - m_{21}) \\ &\leq O(T^{-2} h_T^{-2k} b_T^{-4}) + \frac{2}{T h_T^{2k} b_T^4} E g_1^2 K_{12}^2 K_{13}^2 (y_2 z_2 - m_{21})'(y_3 z_3 - m_{21}) \\ &\leq O(T^{-2} h_T^{-2k} b_T^{-4}) + \frac{2}{T h_T^{2k} b_T^4}\left[E\|g_1 E_1 K_{12}(m_{22} - m_{21})\|^2 E\|g_1 E_1 K_{13}(m_{23} - m_{21})\|^2\right]^{1/2}, \end{aligned}$$

where the second inequality follows from A5 and A15(1) and the last follows from Cauchy-Schwarz inequality and $E_1 = E(\cdot \mid v_1)$. By Lemma 5 of Robinson (1988)

$$E \left\| g_1 E_1 K_{12}(m_{2j} - m_{21}) \right\|^2 = h_T^{2(k+q)}, j = 2, 3.$$

It follows that

$$B_1 \leq O_p(T^{-2} h_T^{-2k} b_T^{-4}) + O_p(T^{-1} h_T^{4q} b_T^{-4}) = o_p(T^{-1})$$

by A13. Similarly it can be shown that $B_2 = o_p(T^{-1})$, which implies that $A_1 = o_p(T^{-1/2})$. Analogous we have $A_2 = o_p(T^{-1/2})$. The Lemma follows then from

$$\begin{aligned} \tilde{G}_T - G_T &= \frac{1}{T} \sum_{t=1}^{T} (1 - J_t) g_t \left[(\xi - \Gamma_0 w_t) m_{1t} - m_{2t} \right] / f_t \\ &= o_p(T^{-1/2}). \end{aligned}$$

$\square$

Lemma B.7. *For any $\xi \in I\!R^k$, the random vectors $\sqrt{T} G_T(\xi; \theta_0)$ and $\sqrt{T} vec(\hat{\Gamma} - \Gamma_0)$ jointly have the asymptotic normal distribution with Zero mean.*

Proof: By definition (4.7)

$$\hat{\Gamma} = \Gamma_0 + (\sum_{t=1}^{T} (z_t - \Gamma_0 w_t) w_t')(\sum_{t=1}^{T} w_t w_t')^{-1}$$

and hence

$$\text{vec}(\hat{\Gamma} - \Gamma_0) = [(\sum_{t=1}^{T} w_t w_t')^{-1} \otimes I_k](\sum_{t=1}^{T} (w_t \otimes I_k)(z_t - \Gamma_0 w_t)).$$

Let

$$D_T = \begin{pmatrix} I_k & 0 \\ 0 & (\frac{1}{T} \sum_{t=1}^{T} w_t w_t')^{-1} \otimes I_k \end{pmatrix}.$$

Then

$$\sqrt{T} \begin{pmatrix} G_T(\xi; \theta_0) \\ \text{vec}(\hat{\Gamma} - \Gamma_0) \end{pmatrix} = \frac{D_T}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} g_t \left[(\xi - \Gamma_0 w_t) m_{1t} - m_{2t} \right] / f_t \\ (w_t \otimes I_k)(z_t - \Gamma_0 w_t) \end{pmatrix}$$

which converges to normal distribution with zero mean by the Lindeberg-Levy central limit theorem (CLT) and the Slutsky's theorem (Amemiya (1985)). $\square$

Proof of Theorem 4.3*:* By A12 the first derivative $\partial Q_T(\theta)/\partial\theta$ exists and hAs the first order Taylor expansion in A neighborhood of θ_0. Since $\partial Q_T(\hat{\theta}_T)/\partial\theta = 0$ and $\hat{\theta}_T \xrightarrow{P} \theta_0$, for sufficiently large T, we have

$$0 = \frac{\partial Q_T(\theta_0)}{\partial\theta} + \frac{\partial^2 Q_T(\tilde{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}_T - \theta_0), \tag{B.22}$$

where $\tilde{\theta}$ lies between $\hat{\theta}_T$ and θ_0. The first and second derivatives in (B.22) are given by

$$\frac{\partial Q_T(\theta_0)}{\partial\theta} = \frac{2}{S} \sum_{s=1}^{S} \frac{\partial \hat{G}_T'(\xi_s; \theta_0)}{\partial\theta} \hat{G}_T(\xi_s; \theta_0) \tag{B.23}$$

and

$$\frac{2}{S}\sum_{s=1}^{S}[\frac{\partial \hat{G}_T'(\xi_s;\tilde{\theta})}{\partial\theta}\frac{\partial \hat{G}_T(\xi_s;\tilde{\theta})}{\partial\theta'} + (\hat{G}_T(\xi_s;\tilde{\theta})\otimes I_p)\frac{\partial \mathrm{vec}(\partial \hat{G}_T'(\xi_s;\tilde{\theta})/\partial\theta)}{\partial\theta'}]$$

respectively. Since in $\hat{G}_T(\xi;\theta)$ only $g(\xi;\theta)$ involves θ, analogous to the proof of Lemma B.3 it can be shown by A1 that the second derivative $\partial^2 Q_T(\theta)/\partial\theta\partial\theta'$ converges in probability to $\partial^2 Q(\theta)/\partial\theta\partial\theta'$ uniformly in the neighborhood of θ_0. It follows that

$$\frac{\partial^2 Q(\tilde{\theta})}{\partial\theta\partial\theta'} \xrightarrow{P} \frac{\partial^2 Q(\theta_0)}{\partial\theta\partial\theta'} = \frac{2}{S}\sum_{s=1}^{S} D(\xi_s;\theta_0)D(\xi_s;\theta_0)', \tag{B.24}$$

where the last equation follows from $G(\xi_s;\theta_0)=0$. Furthermore, analogous to the proof of Lemma B.3 it is easy to show

$$\frac{\partial \hat{G}_T(\xi_s;\theta_0)}{\partial\theta'} \xrightarrow{P} D(\xi_s;\theta_0). \tag{B.25}$$

By Lemma B.4 - B.7,

$$\begin{aligned}\sqrt{T}\hat{G}_T(\xi_s;\theta_0) &= \sqrt{T}G_T(\xi_s;\theta_0) + (A+B)\sqrt{T}\mathrm{vec}(\hat{\Gamma}-\Gamma_0) + o_p(1)\\ &\xrightarrow{L} N(0, V(\xi;\theta_0)),\end{aligned} \tag{B.26}$$

where $V(\xi;\theta_0)$ is given as in A16. The theorem follows then from (B.22) – (B.26). $\square$

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

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

[6] 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.

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

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

[9] 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.

[10] 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.

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

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

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

[14] Magnus, J.R. and H. Neudecker (1988). Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York.

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

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

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

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

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

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

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